RICE UNIVERSITY

The Distractor Frequency Effect in Stroop and Picture-word Interference Paradigms

by

Jingyi Geng

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE Master of Arts

APPROVED, THESIS COMMITTEE:

Tatiana Schnur, Assistant Professor of Psychology

Randi Martin, Elma Schneider
Professor of Psychology

Jessica Logan, Assistant Professor of Psychology

HOUSTON, TEXAS
MAY 2010
ABSTRACT

The Distractor Frequency Effect in Stroop and Picture-word Interference Paradigms

by

Jingyi Geng

The color naming (Stroop) and picture-word interference (PWI) paradigms play a pivotal role in theorizing about cognitive processes in general, and language production in particular. Despite their assumed similarities, there exist discrepancies. In this study, I compared the effect of distractor word frequency between Stroop and PWI paradigms (in PWI picture naming is faster for high frequency than low frequency word distractors; Miozzo & Caramazza, 2003). In five experiments, I confirm the presence of DFE in both Stroop and PWI paradigms when the naming latencies are longer but the absence of DFE in both paradigms when the naming latencies are shorter and found that when naming latencies are fast, the distractor words are most likely processed to the phonological level by the time targets are named in both paradigms in the last experiment. Thus, my results are consistent with the assumption that the two paradigms operate in the same manner.
ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chair, Dr. Tatiana Schnur. Her guidance and persistent help were crucial to the completion of this master thesis. In addition, special thanks go to Dr. Niels Janssen. Without his guidance and great suggestions this thesis would not have been possible. I thank my committee members, Dr. Randi Martin and Dr. Jessica Logan.

I would like to thank my parents for their financial and never ending support, for the help in my study and for its success. I am grateful the undergraduate research assistants and Sujuan Lin for assisting and helping me in my research.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables and Figures</td>
<td>V</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>6</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>17</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>19</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>24</td>
</tr>
<tr>
<td>Experiment 5</td>
<td>29</td>
</tr>
<tr>
<td>General Discussion</td>
<td>35</td>
</tr>
<tr>
<td>Conclusions</td>
<td>41</td>
</tr>
<tr>
<td>Reference</td>
<td>42</td>
</tr>
<tr>
<td>Appendix</td>
<td>47</td>
</tr>
</tbody>
</table>
LIST OF TABLES AND FIGURES

Table 1  Naming latencies (ms) and error rate (in parentheses) for the HF and LF words shown in Experiment 1a & b  10

Table 2  Mean Measurements of colors and pictures in Experiment 2  18

Table 3  Naming latencies (ms) and error rate (in the parentheses) for the HF and LF words in Experiment 2  19

Table 4  Naming latencies (ms) and error rate (in the parentheses) for the four conditions in Experiment 3  22

Table 5  Naming latencies (ms) and error rate (in parentheses) for PWI and Stroop tasks in Experiment 4.  25

Table 6  Naming latencies (ms) and error rate (in parentheses) for PWI and Stroop tasks in Experiment 5.  33

Figure 1  Illustration about how the difference in speed of picture and color naming impacts the Dfe in PWI and Stroop along the process chain of word production  15

Figure 2  Illustration of loci of the PFE and SIE  27

Figure 3  Illustration of loci of the PFE, SIE and DFE  34

Figure 4  Levels of processing in PWI and Stroop paradigms  39
The Stroop and picture-word interference (PWI) paradigms play a pivotal role in theorizing about cognitive processes in general, and language production in particular. Recently, the comparison of Stroop and PWI paradigms has received much attention, because experimental findings suggest that the paradigms, may not in fact, be comparable for how word production works. In this master’s thesis, I conducted five experiments to directly compare Stroop and PWI paradigms. Results suggest that the paradigms are comparable, but the speed of naming colors or pictures affects the appearance of interference effects.

There are at least two ways to test whether the two paradigms are comparable. One way is to test whether similar effects observed in both paradigms have the same locus along the process chain of language production which is commonly assumed to have three stages in language production: conceptual preparation (meaning of a word), lexical selection (retrieving a word) and phonological encoding (sound of a word). (e.g., Levelt, Roelofs, & Meyer, 1999; Roelofs, 2003). Dell’ Acqua, Job, Peressortti, and Pascali (2007) tested the PWI paradigm in this way and claimed that the PWI effect (i.e., more interference produced by semantically related picture-word pairs (e.g., picture: dog; word: cat) than semantically unrelated picture-word pairs (e.g., picture: dog; word: table)) is located prior to lexical selection, while the Stroop effect is located at the level of lexical selection. However, different loci do not necessarily indicate different mechanisms. Van Maanen, Hedderik, and Borst (2009) presented a single computational model to account for both PWI and Stroop effects and argued that the mechanism underlying the two effects remains the same, though they have different loci (I discuss it in the general discussion). Another way to test paradigm similarity is to do the same manipulation in
both paradigms to see whether comparable effects are observed. I adopted the latter method in this study by manipulating the frequency of distractor words in the Stroop and PWI paradigms and comparing the size of the distractor frequency effects (i.e., high frequency word distractors create faster picture naming response times than low frequency distractors, Miozzo & Caramazza, 2003).

For both PWI and Stroop paradigms, a similar relationship between distractor word and the color/picture produces similar behavioral effects. In the standard Stroop paradigm, there are a set of words written in different color inks. Participants name the ink color and ignore the written word. A well-established finding with this paradigm is that the naming times are slower for the colors in the incongruent condition (e.g., red is written in green) than in the neutral condition (e.g., red is written in black) (e.g., Stroop, 1935). Subsequently, Hentschel (1973) modified the Stroop paradigm by embedding words inside line drawings of familiar objects creating the picture-word interference paradigm (PWI). Similar to the Stroop paradigm in PWI, naming is slower for pictures (e.g., dog) paired with semantically related words (e.g., cat) than semantically unrelated words (e.g., table). The slowing in naming in both paradigms in these situations is referred to as a semantic interference effect. Beyond the semantic interference effect, the PWI and Stroop paradigms produce other similar behavioral effects. First, in both paradigms, naming latencies are faster for the color/picture paired with a distractor word that describes the same color/picture (congruent/identity condition) than when a distractor word is unrelated to the color/picture (neutral condition) (PWI: Rayner & Posnansky, 1978; Stroop: Glaser & Glaser, 1982). Secondly, in both paradigms, phonologically related distractor words create faster naming latencies than unrelated word distractors (e.g. PWI: picture: dog,
word: doll; e.g., Damian & Martin, 1999; Stroop: color: green, word: great; e.g., Spinks, Liu, Perfetti, & Tan, 2000). This is often referred to as the PFE. Finally, when the task is word reading instead of picture/color naming, incongruent colors/pictures cannot interfere with word reading (e.g., Glaser & Glaser, 1989). In other words, there is no reverse Stroop/PWI effect. This evidence supports the assumption that these two paradigms operate using similar cognitive mechanisms.

However, the Stroop and PWI paradigms do not produce similar behavioral effects under all circumstances. Using a psychological refractory period (PRP) paradigm, Dell'Acqua et al. (2007) demonstrated that the classic semantic interference effect in the PWI does not produce effects similar to the semantic interference effect in Stroop (Fagot & Pashler, 1992). In Dell'Acqua et al.'s version of the PRP task, subjects identify one of three tones with a key press (task 1), and at a variable stimulus onset asynchrony (SOA) later, name a picture while ignoring a distractor word (PWI paradigm) (task 2). Distractor words were semantically related or unrelated to the pictures. Processing of the two tasks can be broken down into three stages: perceptual encoding, response selection, and response execution (Pashler, 1991). Different stages can occur in parallel; however, response selection is assumed to proceed only one task at one time. Therefore, at short SOAs, the limitation in performing response selection for two stimuli postpones the response selection for task 2 until the response selection for the first stimulus is completed. Thus, if the effect for the task 2 is located in the response selection stage, it should be observed across all short and long SOAs. However, if the effect is located in perceptual encoding, it can be observed for long SOAs but not short SOAs. The rationale here is that the perceptual encoding effect for the second stimulus occurs simultaneously
with response selection for the first stimulus for short SOAs, and thus any interference is
resolved before response selection occurs for the second stimulus. Dell’ Acqua et al.
(2007) found that the semantic interference effect was observed at a long SOA (1000ms)
but not at short SOAs (100ms and 350ms). In contrast, Fagot and Pashler (1992) found
that the Stroop effect was constantly observed at all SOAs (i.e., -50, 50, 150, and 450ms).
Dell’ Acqua et al. (2007) concluded that the locus of the PWI effect is different from the
Stroop effect where the PWI effect occurs in the perceptual encoding stage but the Stroop
effect occurs during response selection. Therefore they claimed that the two paradigms
are different.

However, in Dell’ Acqua et al. (2007)’s experiment, they used only PWI paradigm and
compared their results with those of the Stroop paradigm in Fagot and Pashler (1992).
Due to different subjects, SOAs, and materials (e.g., distractor words) in the two studies,
the conclusion about the difference between the two paradigms is not convincing. Given
the similarities between the two paradigms (e.g., Glaser & Glaser, 1989), the conclusion
that they have different mechanisms is too strong if it is only based on one PWI study.

Nevertheless, the Stroop and PWI paradigms may also produce different results for
other manipulations, namely the distractor frequency effect. Miozzo and Caramazza
(2003) repeatedly observed that high frequency (HF) word distractors create faster
picture naming response times than low frequency (LF) distractors (PWI Experiment 1-7;
of the DFE—active blocking of distractor selection. The central claim of this hypothesis
is that the target picture is not available for production until the distractor word is actively
blocked. The response time for the target picture depends on how fast the distractor word
is blocked. The faster the distractor word is blocked, the faster the picture name is produced. Due to the higher activation level of the HF distractor word, it is processed sooner than the LF distractor word and thus is blocked more quickly. That is why the DFE is observed. This hypothesis predicts that there is a DFE in the Stroop paradigm.

However, in Stroop, the results are much less robust than the PWI paradigm: the DFE was not significant in Monsell, Taylor, and Murphy (2001) though they observed a DFE trend (Experiment 1-2), while Burt (2002) found the DFE (Experiment 4) but the HF and LF distractors were not well-matched on linguistic factors known to affect word reading and degree of word distractor interference (e.g., grammatical class, number of syllables). Exploring the role of emotional valence and frequency in the Stroop paradigm, Kahan and Hely (2008) found the DFE only for negative emotion-laden distractor words (e.g., *death*) but not for emotionally neutral H/LF (e.g., *news*) or positive emotion-laden H/LF words (e.g., *free*).

However, these studies (Stroop: Monsell et al., 2001; Burt, 2002; Kahan & Hely, 2008; PWI: Miozzo & Caramazza, 2003) which demonstrated a difference in the DFE for the two paradigms are not convincing, as they used different materials and subjects. Thus, firstly, I need to establish the reliability of the difference between the Stroop and PWI paradigms for the DFE. Once I establish the difference, I can explore why the difference in the DFE occurs.

The purpose of this study is to establish the reliability of the difference of the DFE in the Stroop and PWI paradigms and then ascertain why the difference in DFE occurs. My work contributes to interpreting why the difference in DFE was observed in previous Stroop and PWI experiments (e.g., Stroop: Monsell et al., 2001; PWI: Miozzo &
Caramazza, 2003) and understanding whether interference effects in both paradigms arise from similar processes.

I compared the DFE in Stroop and PWI paradigms using a within-subjects and between-materials design. Across all experiments, subjects named the color ink of words in a Stroop paradigm while ignoring the printed word, and/or named pictures and ignored word distractors in PWI. If differences between materials accounted for the previously observed contrast, I expected to find the same DFE in the Stroop and PWI tasks in Experiment 1. By contrast, if previous contrasting results reflect genuine differences between PWI and Stroop, I expected a DFE in PWI but not in Stroop.

**EXPERIMENT 1**

The objective of Experiment 1a is to test whether there is the DFE in the Stroop paradigm and compare the DFEs in Stroop and PWI paradigms. Participants did both Stroop and PWI paradigms in the context of two types of word distractors: high frequency (HF) versus low frequency (LF) (e.g., Stroop: color: red, distractor word: *city* vs. *lava*; PWI: picture: bell, distractor word: *city* vs. *lava*). The same distractors were used for both tasks. According to Miozzo and Caramazza (2003), I expected participants to take longer to name pictures in the presence of LF compared to HF words (the standard DFE) in the PWI paradigm. Based on the data from the previous Stroop studies, I expected no DFE or a small DFE in the Stroop paradigm. Thus, I predicted an interaction between the two paradigms and the DFE. In Experiment 1b, in order to exclude the possibility that there was not enough power to detect the DFE in the Stroop task in Experiment 1a, I replicated the null effect in the Stroop task with another 24 participants.
METHOD

Participants

In this and the following experiments, all participants were native English speakers at Rice University and received credit or were paid for their participation. All subjects gave informed consent in accordance with the protocol approved by the Institutional Review Board of Rice University. Twenty-four native English speakers (10 male) took part in Experiment la and a different twenty-four (9 male) in Experiment 1b.

Materials

I used 80 distractor words, 40 high frequency and 40 low frequency words from Miozzo and Caramazza (2003; Experiments 3-4) in both PWI and Stroop tasks. As Miozzo and Caramazza (2003) reported, the high frequency (HF) and low frequency (LF) words (all nouns) were matched for word length, concreteness, the number of neighbors, word mean bigram frequency, and the measure of grapheme-to-phoneme. For the Stroop task, four colors (red, blue, green and yellow) were randomly assigned to all 80 words, such that each word was seen in one of the four colors. For the PWI task, I used the same 20 target picture names from Miozzo and Caramazza (2003; Experiment 3-4), selecting pictures from the Snodgras and Vanderwart (1980) set. I paired each picture with two HF and two LF words. The picture-word pairs used in Experiment 1a are reported in Appendix A.

Design

For both tasks, the two conditions (40 HF and 40 LF words) were combined into four blocks. For Stroop paradigm, each block contained 10 HF and 10 LF words paired with
four colors. For PWI paradigm, each block contained 10 HF and 10 LF words paired with 20 target pictures. Trial presentation within each block was pseudo-randomized such that no print color, picture or word was repeated on consecutive trials. In Experiment 1a, each subject saw all four blocks for both tasks. For each task, the total number of trails is 138. The order of the tasks was balanced among participants and the order of block presentation was counterbalanced across subjects. In Experiment 1b, each subject saw all four blocks only for the Stroop task. Three warm-up trials using practice items were included at the beginning of each block. The total number of trail is 138.

**Procedure**

The following procedure was used in all of the Stroop and PWI tasks reported in this article. The experiment was operated by the software DmDX (Foster & Foster, 2003). In Stroop task, words were presented in capital letters in Times New Roman 16-point font on a color monitor with a dark background. The participants were instructed to name the colors and ignore the distractor words. The procedure for Stroop task is identical as that in Burt (2002)'s Experiment 4. On each trial, the fixation point(+) was shown in the center of the screen for 500 ms, the screen was dark for 250 ms, and the word appeared in red, green, blue, or yellow color ink until a vocal response was detected. An interval of 3 seconds then elapsed before the next fixation display. A 12-trial practice session which was never used in the experiment session preceded the experiment session. In the PWI task, words were presented in capital letters in Times New Roman 12-point font. The participants were instructed to name the pictures and ignore the distractor words. There were three sessions: learning session, practice session and experimental session. In the learning session, the participants familiarized themselves with the pictures and their
names. Pictures were shown with their names underneath but without superimposed
distractor words and they would appear until participants named it aloud. In the practice
session, the participants were presented with all the pictures with distractor words that did
not appear in any of the experimental trails. A trial was structured as follows: a fixation
point (+) was displayed in the center of the screen for 700 ms and then was replaced by
the stimulus. Stimuli disappeared once participants responded. If the voice key was not
triggered, the picture was shown for 3 seconds. The subjects were asked to respond as
quickly and as accurately as possible and an experimenter coded errors. Experiment 1a
lasted for 20 minutes. Experiment 1b lasted for 10 minutes.

RESULTS

Experiment 1a

Naming latencies were discarded from the analyses whenever any of the following
occurred: (a) A picture was named incorrectly; (b) subjects made a noise (e.g. cough); (c)
the voice-key failed to trigger; and (d) RTs deviated from a participant’s mean by more
than three standard deviations. For the Stroop task, 3.4% of the data points were removed.
For the PWI task, 4.7% of the data points were removed. Two ANOVAs were computed
with participants and items as random variables. Fixed variables were distractor
frequency condition (HF vs. LF), and task type (Stroop vs. PWI). The distractor
frequency is a within-subject and between-item variable and task type is a within-subject
and within-item variable. For the item analysis in all the experiments, I treated items as
distractors (e.g., H/LF words: df=78), as more degrees of freedom means more reliable
interpretation. An overview of the mean RTs and error rate for the two tasks is presented in Table 1.

Table 1. Naming latencies (ms) and error rate (in parentheses) for the HF and LF words shown in Experiment 1a & b

<table>
<thead>
<tr>
<th></th>
<th>LF words</th>
<th>HF words</th>
<th>DFE (LF-HF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1a</td>
<td>Stroop</td>
<td>629 (.03)</td>
<td>626 (.04)</td>
</tr>
<tr>
<td></td>
<td>PWI</td>
<td>705 (.05)</td>
<td>680 (.05)</td>
</tr>
<tr>
<td>Experiment 1b</td>
<td>Stroop</td>
<td>639 (.04)</td>
<td>633 (.04)</td>
</tr>
</tbody>
</table>

* p < .05

In the error analysis, there were no significant effects ($F$s < 1). In the analysis of naming latencies, color naming in the Stroop task was significantly faster than picture naming in PWI (64 ms) ($F_1(1,23) = 16.98, p < .001$, $MSE = 5935.15$; $F_2(1,78) = 153.67, p < .001$, $MSE = 1015.13$). There was a main effect of the distractor frequency ($F_1(1,23)=11.91, p < .01, MSE=377.49$; $F_2(1,78)=8.27, p < .01, MSE = 1137.42$) indicating slower naming latencies for low frequency words than high frequency words. The size of the DFE was significantly larger in the PWI compared to Stroop paradigm as indicated by a significant interaction between frequency condition and paradigms ($F_1(1,23)=5.07, p = .03$, $MSE=618.84$; $F_2(1,78) = 5.44, p < .05, MSE = 1015.13$). In the PWI task, a planned t-test (two-tailed) showed that naming latencies were 26ms slower for LF than HF distractors ($t_1(23) = 3.40, p < .01; t_2(78) = 3.28, p < .01$). However, in the Stroop task, there was no significant difference between HF and LF conditions ($t < 1$).
Experiment 1b (replication of the null effect in Stroop paradigm)

Latency data were preprocessed in the same way as in Experiment 1a. 4.1% of the data points were removed. An overview of the mean RTs and error rate for the two tasks is presented in Table 1. For the subject analysis, a paired-samples t-test was computed comparing color naming with HF and LF distractor words in the Stroop paradigm. For the item analysis, an unpaired-samples t-test was computed. In the error analysis, there was no difference between HF and LF conditions ($t_1 < 1$). In the analysis of naming latencies, there was no difference between HF and LF conditions ($t_1(23) = 1.52, p = .14; t_2(78) = 1.24, p = .22$). In order to see whether the DFE in the Stroop paradigm in Experiment 1b was different from that in the PWI of Experiment 1a, I computed an ANOVA with participants and items as random variables. Fixed variables were distractor frequency condition (HF vs. LF), and paradigm (Stroop vs. PWI).

In the analysis of naming latencies, I found again that color naming in the Stroop task was significantly faster than picture naming in the PWI paradigm ($F_1(1,46) = 5.19, p = .03, MSE = 12970.99; F_2(1,78) = 115.60, p < .001, MSE = 1043.77$), and overall naming latencies were significantly slower for low frequency than high frequency words ($F_1(1,46) = 13.76, p = .001, MSE = 460.49; F_2(1,78) = 12.25, p < .001, MSE = 943.53$). However the significant DFE was driven solely by the PWI paradigm, as demonstrated by the significant interaction between frequency condition and paradigm, indicating that the size of the DFE was again smaller (and non-significant) in the Stroop compared to the PWI paradigm ($F_1(1,46) = 5.03, p = .03, MSE = 460.49; F_2(1,78) = 3.91, p = .05, MSE = 1043.77$).
DISCUSSION

The purpose of Experiment 1 was to test the DFE in Stroop and PWI paradigms. In Experiment 1a, I found that LF distractors interfered with the production of target responses more than HF distractors in PWI paradigm. This is consistent with the results for the PWI paradigm in Miozzo and Caramazza (2003). In the Stroop paradigm, I did not observe the DFE in Experiment 1a nor in its replication in Experiment 1b. This is also consistent with the previous Stroop studies (e.g., Monsell et al., 2001). Also, as expected, there was a significant interaction between paradigm and distractor frequency conditions for Experiment 1. Having established the pattern of DFE in the two paradigms observed in the literature, I explore why the DFE may produce different results in each of the two paradigms.

As mentioned in the introduction, according to Miozzo and Caramazza (2003), during the DFE, the target picture will not be available for production until the distractor word is actively blocked and HF distractors are blocked sooner than LF distractors, producing faster picture naming latencies. This "blocking" hypothesis is the early version of the response exclusion hypothesis proposed by Mahon, Costa, Peterson, and Vargas (2007). Both hypotheses have the same interpretation for the DFE and assume that the DFE is a post-lexical effect (i.e., it occurs at a level which is later than the lexical level), though they have slightly different assumptions about the locus of DFE (i.e., Miozzo & Caramazza, 2003) assume that the DFE occurs at the phonological level. Mahon et al. (2007) assume that the DFE occurs at the output buffer which is later than the phonological level. I discuss it in the discussion of Experiment 5.). Now, I only use the new version—response exclusion hypothesis to represent the two hypotheses. With the
assumption that both PWI and Stroop are similar, the response exclusion hypothesis predicts the DFE in Stroop paradigm. Thus, it cannot provide an explanation for the null effect in Stroop of Experiment 1.

Since Monsell et al. (2001) found no DFE in the Stroop paradigm, they generated an account to interpret it. Does their account explain the contrasting results in Experiment 1? In their Stroop experiments, they included not only HF and LF distractors but also pronounceable nonwords (e.g., bennel) and nonalphanumeric symbols (e.g., @#$%&). They found that the color naming was slower for HF, LF and pronounceable nonwords than nonalphanumeric symbols but there was no difference among the first three item types. Thus, they argued that if the letter string is pronounceable, no matter whether it is real word or nonword, it produced similar interference in Stroop task. They referred this interference as task set competition between word reading and color naming. Since word reading is automatic, it causes a longer RT when subjects need to name the color and inhibit word reading. Hence, for every trial, if the distractor is a word or a word-like letter string, it elicits word reading which competes with color naming. That is why color naming was slower for HF, LF and pronounceable nonwords than nonalphanumeric symbols. Furthermore, they argued that the noncolor words in Stroop task do not access their lexical representations, as there was no difference in response times between HF and LF words. In contrast, due to the fact that color words are either in the response set or category coordinators for the targets, they get primed, access their lexical representations, and then compete with the selection of targets. That is why the standard Stroop effect is observed in the literature. However, if Monsell et al. (2001) assume that the Stroop and PWI paradigms are comparable, the aforementioned account cannot explain the DFE
observed in PWI. The HF and LF distractors are not related to the target pictures in PWI and so should not access their lexical representations. Thus, their hypothesis cannot account for the results of Experiment 1.

Is it possible that the DFE difference across paradigms is due to the paradigms operating in a different way via the nature of one task requiring picture naming and the other requiring more repetitive ink color naming? As seen in Experiment 1, color naming in Stroop was significantly faster than picture naming in PWI (Stroop: 628ms; PWI: 693ms). There are four colors repeated twenty times in the Stroop paradigm but twenty pictures repeated only four times in the PWI paradigm. It is well-known that repetition results in faster naming latencies in both Stroop and PWI paradigms (e.g., Macleod, 1998; La Heij & van den Hof, 1995). La Heij and van den Hof (1995) found that the semantic interference effect in PWI decreased as the target-set size decreased. Interestingly, when there were only four pictures repeated eighteen times, no semantic interference effect was observed. They argued that high repetition results in a reduction of the threshold of the target name or an increase in the strength of the association between the target’s concept and its lexical representation. In my experiment, there was no DFE in Stroop with four colors but a DFE in PWI with twenty pictures. Thus, according to La Heij and van den Hof (1995), it is possible that high repetition in Stroop eliminated the DFE. However, if it is true, why is the standard Stroop effect always observed when four distractor color words and four target colors are included in Stroop studies (i.e., naming times are slower for colors in the incongruent than in the neutral condition.)? La Heij and van den Hof (1995) did not provide any explanation for the Stroop effect. Thus, this explanation of interference cannot accommodate the range of results seen here and elsewhere.
Figure 1. Illustration about how the difference in speed of picture and color naming impacts the DFE in PWI (Panel A) and Stroop (Panel B) along the process chain of word production. The dashed box refers to post-lexical levels (the locus of DFE, M&C, 2003; Mahon et al., 2007). In Stroop (Panel B), I am agnostic to what level the distractor word is processed by the time the color is named. I only know it should be earlier than the locus of DFE. For illustration purposes only, I put it at the lexical selection level.

There is another way to explain the results based on the different naming speeds in the two paradigms. Although the contrasting results in PWI and Stroop are problematic for the response exclusion hypothesis proposed by Mahon et al. (2007), it is still plausible when the different naming speeds in the two paradigms are taken into account. According to response exclusion hypothesis, there are two premises for the interpretation of the DFE: a) words have a privileged relationship to the articulation in a way that pictures do not and b) in PWI, all distractor words get to the post-lexical level faster than target pictures. Thus, those words have to be excluded before the targets are named. The naming
latencies reflect how fast the HF and LF words are excluded. That is why there is a DFE in PWI (see Figure 1, Panel A). In contrast, there was no DFE in Stroop. In order to interpret it, the response exclusion hypothesis needs to drop the aforementioned two assumptions and include a new speed hypothesis. The speed hypothesis proposes that since there are faster naming latencies for Stroop than PWI, it is possible that in Stroop, distractor words are still being processed at earlier levels when target names are processed to the post-lexical level (where the DFE occurs, see Figure 1). Thus, no distractor words compete with the target at the post-lexical level. That is why there is no DFE in Stroop (see Figure 1, Panel B). Once the response exclusion hypothesis takes the speed hypothesis (the relative speed of processing pictures/colors vs. words) into account, it explains why there is the DFE in PWI but not in Stroop. But how does this hypothesis explain the Stroop effect? In the typical Stroop studies, the naming latencies for the incongruent color words are more than 750ms (e.g., Dunbar & Macleod, 1984), which is much slower than the naming latencies (e.g., 628ms) for HF and LF words in our Experiment 1. When incongruent color words are included, response times become slower, as color words are more salient and cause more attention from participants as negative emotional words do in the emotional Stroop task (i.e., color naming for negative emotional words is slower than neutral words). Some studies also showed that attention modulates the Stroop effect (e.g., Lowe & Mitterer, 1982). Thus, according to our speed hypothesis, when color naming gets slower, distractors are most likely fully processed, and thus, the Stroop effect is observed. Thus, this hypothesis predicts how the speed of naming impacts the DFE.

According to the speed hypothesis, I predicted that the DFE would be observed when
naming was slow but not observed when naming is fast in both Stroop and PWI paradigms. In the following experiments, we tested this hypothesis in two ways. In Experiment 2, we diminished the DFE in PWI by speeding up picture naming (such that the frequency of the word distractors no longer impacts response times). In Experiment 3, I obtained the DFE in Stroop by slowing down color naming.

**EXPERIMENT 2**

The aim of Experiment 2 is to test whether the DFE can be eliminated by speeding up picture naming in the PWI paradigm. For the PWI paradigm, there are at least two ways to speed up picture naming. One is to use high frequency pictures names. Miozzo and Caramazza (2003, Experiment 1) manipulated the frequency of both pictures and distractor words and found the DFE for both low and high frequency pictures. The DFE appeared for high frequency pictures most likely because the picture frequency manipulation did not greatly speed up picture naming as high frequency picture naming was still well over 700ms. Another way to speed up naming is to increase naming repetition by reducing the number of pictures named. La Heij and van den Hof (1995) found that semantic interference effect decreased as the size of the target set decreased. They observed no semantic interference effect when there were only four pictures. Thus, I chose to use fewer pictures (four pictures, akin to the four colors in Stroop) to speed up naming.

The four pictures were matched on several variables known to affect picture naming with the four colors used in Stroop task. According to the speed hypothesis, if the
distractor words are not processed to the postlexical level before the pictures are produced, I should not observe the DFE in PWI task.

METHOD

Participants

Twenty native English speakers (9 male) took part in Experiment 2.

Materials and Procedure

I used the same materials and design of the Stroop task in Experiment 1 (only HF and LF conditions). The only difference was that I used four pictures instead of four colors. The four pictures were CAR, PIG, TREE and HAT. The color and picture names were matched on frequency (Kučera & Francis, 1967), word length, imageability (Gilhooly & Logie, 1980) and number of syllables (see Table 2, no significant difference between four colors and four pictures in all the variables) so that picture names might be named as quickly as color names. Four pictures were randomly assigned to all 80 HF and LF words, such that each word was seen in one of the four pictures. Procedure was exactly the same as that for PWI task in Experiment 1a. The whole experiment lasted for 10 minutes.

Table 2 Mean Measurements of colors and pictures in Experiment 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Color</th>
<th>Picture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>128</td>
<td>99</td>
</tr>
<tr>
<td>Imageability rating</td>
<td>590</td>
<td>614</td>
</tr>
<tr>
<td>Word Length</td>
<td>4.5</td>
<td>3.3</td>
</tr>
<tr>
<td>Number of syllables</td>
<td>1.3</td>
<td>1</td>
</tr>
</tbody>
</table>

*p < .05
RESULTS

Latency data were preprocessed in the same way as in Experiment 1. 4.2% of the data points were removed. An overview of the mean RTs and error rate for the two tasks is presented in Table 3.

Table 3. Naming latencies (ms) and error rate (in the parentheses) for the HF and LF words in Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>LF words</th>
<th>HF words</th>
<th>Diff (LF-HF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PWI</td>
<td>592 (.04)</td>
<td>592 (.05)</td>
<td>0</td>
</tr>
</tbody>
</table>

* p < .05

A paired-samples t-test was computed comparing HF and LF words for subject analysis. For item analysis, an unpaired-samples t-test was computed. In the error analysis, there was no difference between HF and LF conditions ($t_{19} = 1.05, p = .31; t_{78} = 0.94, p = .35$). In the analysis of naming latencies, there was no difference between HF and LF conditions ($ts < 1$).

DISCUSSION

As predicted, in Experiment 2 I did not observe the DFE in the PWI paradigm when I sped up response naming by reducing the number of pictures named. The picture naming latencies in this experiment were almost 100ms faster than that in PWI task of Experiment 1a ($t(42) = 4.43, p < .001$). So I succeeded in speeding up picture naming by using four pictures. My results are consistent with the speed hypothesis. According to this hypothesis, if the distractor word is not processed to the post-lexical level before the target response is available for production, the frequency of the distractor words will not
impact picture/color naming. Thus, the DFE disappeared in the PWI paradigm with four pictures.

**EXPERIMENT 3**

In Experiment 2, I successfully sped up picture naming in the PWI paradigm which eliminated the DFE. In Experiment 3, I attempted to slow down color naming in the Stroop paradigm in order to allow an impact of the word distractors at the post-lexical level for the DFE to emerge. For the Stroop paradigm, one way to slow down color naming is to increase the overall number of target colors to make it comparable to the PWI task (normally 20 to 40 pictures are named). However, it is impossible to include 20 colors in the Stroop paradigm, as 20 colors are hard to identify. Another way is to include a condition that slows down naming (e.g., incongruent color names; standard Stroop effect). By including a distractor condition which slows down naming for that condition, overall naming in the block can be slowed down due to participants setting a criterion for initiating articulation based on the average between “slow” and “fast” trials within a block (e.g., Lupker, Brown, & Colombo, 1997). Thus, the response times for the “fast” trials in a mixed block (i.e., “slow” and “fast” trials) are slower than in a pure block with only “fast” trials. Given that the average response time for the incongruent condition is more than 780ms (e.g., Dunbar & Macleod, 1984) (which is much slower than the response time for colors paired with HF and LF words in Experiment 1), in Experiment 3 I included incongruent color words to slow down overall color naming across all conditions. Additionally, I included the congruent color condition as an added control to verify that participants were in fact reading words when naming colors in the Stroop task.
Although the congruent condition could facilitate the color naming, the facilitation effect is much less than the corresponding interference effects produced by incongruent condition (see review: MacLeod, 1991). Thus, I included both congruent and incongruent words to slow down color naming for the high and low frequency conditions predicting as a result that I will observe the DFE in the Stroop paradigm.

**METHOD**

**Participants**

Twenty-four native English speakers (7 male) took part in Experiment 3.

**Materials**

I used the same HF/LF word distractors from Experiments 1-3. A control condition was included four color words (red, blue, green and yellow) divided into two conditions: a) congruent condition: the color and the word matched; b) incongruent condition: the color and the word did not match (e.g., the word “red” printed in blue ink). The incongruent condition was included to slow down the color naming in the frequency conditions and the congruent condition was to verify that the standard Stroop effect was obtained in this experiment. There were 80 fillers so that the color congruency conditions (congruent and incongruent) only occurred 1/3 of the time. This manipulation was introduced to rule out the possibility that the participants have the expectation for the color words when they do the naming task. No color name shared the same phonological onset as a written word (except in congruent color condition). The 40 HF and LF words were each seen once, while the color words were seen 20 times each (10 times in the congruent condition and 10 times in the incongruent condition). I repeated the control condition 20 times in order
to get the enough power for the standard Stroop effect (Kieley & Hartley, 1997; Klopfer, 1996).

**Design & Procedure**

The four conditions (40 HF and 40 LF words, 4 incongruent and congruent color words) and 80 fillers were combined into six blocks, each containing approximately six experimental trials for each distractor condition (HF/LF and incongruent/congruent) and filler. Each block was 40 trials long. Trial presentation within each block was pseudo-randomized such that no print color or word was repeated on consecutive trials and that the color denoted by a word and the color of print of the word were different on consecutive trials. Each subject saw all six blocks and the order of block presentation was counterbalanced across subjects. Three warm-up trials using practice items were included at the beginning of each block. The procedure is identical to the Stroop task in Exp 1. The experiment lasted for 15 minutes.

**RESULTS**

Latency data were preprocessed in the same way as Experiment 1. 4.6% of the data points were removed. An overview of the mean RTs and error rate for each of the four conditions is presented in Table 4.

Two paired-samples t-tests were computed comparing HF and LF words (only for subject analysis) and congruent and incongruent words. An unpaired-samples t-test was computed comparing HF and LF words for item analysis. In the error analysis, there was
Table 4. Naming latencies (ms) and error rate (in the parentheses) for the four conditions in Experiment 3.

<table>
<thead>
<tr>
<th>Distractor words conditions</th>
<th>Incongruent color words</th>
<th>Congruent color words</th>
<th>Stroop effect (Incong-Cong)</th>
<th>LF words (Incong)</th>
<th>HF words (Incong)</th>
<th>DFE (LF-HF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroop</td>
<td>794 (.06)</td>
<td>667 (.02)</td>
<td>127*</td>
<td>705 (.02)</td>
<td>689 (.03)</td>
<td>16*</td>
</tr>
</tbody>
</table>

* p < .05

no difference between HF and LF conditions (ts < 1) but there was a significantly higher error rate for incongruent color words compared to congruent color words (t(23) = 5.09, p < .001) In the analysis of naming latencies, naming latencies were 16ms slower for LF distractors than HF distractors (t(23) = 2.16, p < .05). Naming latencies were 127ms slower for incongruent color words than congruent color words (t(23) = 14.05, p < .001). In sum, I observed both the distractor frequency effect and standard Stroop effect in the Stroop paradigm.

DISCUSSION

In this Stroop experiment, we included both congruent and incongruent color words to slow down color naming for the two frequency conditions in order to allow an impact of the word distractors at the post-lexical level for the DFE to emerge. As predicted, the DFE was observed for the Stroop paradigm. The response times for HF and LF conditions in this experiment were significantly slower than that for the Stroop task in Experiment 1a (t(46) = 3.90, p < .001).
Thus, the inclusion of color words in the Stroop paradigm slowed down color naming for the frequency conditions which I assume made the distractor words available before or currently with the color names, resulting in a significant DFE. In Experiment 2, I sped up picture naming in the PWI paradigm which I assume made the distractor words available only after the picture names resulting in a null DFE. Combining the results of the two experiments, I conclude that the reason why there was a DFE in the PWI paradigm and no DFE in the Stroop paradigm in Experiment 1 is most likely because in the PWI paradigm the distractor words were fully processed before or at the same time as the picture names were available for production but in the Stroop task not processed to the post-lexical level before the color names were available.

As I stated in the introduction, if the two paradigms operate in a similar way, the same manipulation should yield similar results, a pattern demonstrated in Experiments 2 and 3. However, although the DFE was observed in both paradigms, the DFE may differ in magnitude. A difference in magnitude may suggest that in fact, the two paradigms do not depend on the same mechanism creating the DFE. Thus, in next experiment, I tested whether the magnitude for the DFE in Stroop and PWI paradigms is comparable.

**EXPERIMENT 4**

In this experiment, I used the same HF and LF distractor words in both paradigms within the same subject group in order to see whether the magnitude of DFE in both paradigms was similar.
METHOD

Participants

Twenty-four native English speakers (14 male) took part in Experiment 4.

Materials and Procedure

I used the same materials and design in Experiment 3 for the Stroop task and the same materials and design of the PWI task in Experiment 1a for the PWI task. So, there were four conditions (40 HF and 40 LF words, 4 incongruent and congruent color words) in Stroop task and two conditions (40 HF and 40 LF words) in the PWI task. Procedure was exactly the same as that in Experiment 1a. The whole experiment lasted for 30 minutes.

RESULTS

Latency data were preprocessed in the same way as Experiment 1. For the Stroop task, 4.0% of the data points were removed. For the PWI task, 3.3% of the data points were removed. An overview of the mean RTs and error rate for the two tasks is presented in Table 5.

Two ANOVAs were computed with participants and items as random variables. Fixed variables were distractor frequency condition (HF vs. LF) which was a within-subject and between-item variable and task type (Stroop vs. PWI) which was a within-subject and within-item variable. In the error analysis, there were no significant effects (Fs<1). In the analysis of naming latencies, there was a main effect of the distractor frequency ($F_1$(1,23) = 41.37, $p < .001$, $MSE = 8856.98$; $F_2$(1,78) = 12.30, $p = .001$, $MSE = 1418.99$), indicating slower naming latencies for low frequency words than high frequency words.
Table 5. Naming latencies (ms) and error rate (in parentheses) for PWI and Stroop tasks in Experiment 4.

<table>
<thead>
<tr>
<th>Distractor words conditions</th>
<th>Incongruent color words</th>
<th>Congruent color words</th>
<th>Stroop effect</th>
<th>LF words</th>
<th>HF words</th>
<th>DFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroop</td>
<td>830 (.07)</td>
<td>717 (.04)</td>
<td>113*</td>
<td>725 (.03)</td>
<td>705 (.03)</td>
<td>20*</td>
</tr>
<tr>
<td>PWI</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>704 (.04)</td>
<td>684 (.03)</td>
<td>19*</td>
</tr>
</tbody>
</table>

*p < .05

Color naming in the Stroop paradigm was marginally slower than picture naming in the PWI paradigm ($F_1(1,23) = 1.04, p = .32, MSE = 10540.73; F_2(1,78) = 12.30, p = .001, MSE = 1418.99$). The interaction between frequency condition and task was not significant ($F_s < 1$). Although the interaction was not significant, based on a priori hypothesis, I performed individual t-tests in order to examine whether the DFE was significant within each task. In the PWI task, naming latencies were significantly slower for LF distractors than HF distractors (19 ms, $t_1(23) = 4.13, p < .001; t_2(78) = 2.02, p = .05$). In the Stroop task, naming latencies were significantly slower for LF distractors than HF distractors (16 ms, $t_1(23) = 4.32, p < .001; t_2(78) = 2.67, p < .01$). Naming latencies were significantly slower (113ms) for incongruent color words than congruent color words ($t_1(23) = 9.25, p < .001; t_2(78) = 12.03, p < .001$) and there was a significantly higher error rate for incongruent color words compared to congruent color words ($t_1(23) = 2.99, p < .01; t_2(78) = 2.77, p < .01$).
INTERIM DISCUSSION

The DFE was observed in both Stroop and PWI paradigms and there was no interaction between frequency effect and paradigm, indicating that the magnitude of the DFE in these two paradigms is comparable. Based on the results of the above four experiments, I concluded that the different processing speeds of the two paradigms account for the observation of the DFE in all my experiments and the two paradigms are comparable at least for the DFE.

Given that the DFE disappears if naming latency is sped up (Experiments 1-2), the question that follows is why the H/LF word distractors do not impact naming when naming is fast, but do impact naming when naming is slow. Specifically, the next experiment tests to what level (conceptual, lexical, phonological) the distractor words are processed in the Stroop and PWI paradigms with faster naming latencies. I address this issue by exploring the semantic interference (SIE) and phonological facilitation effects (PFE) in the PWI and Stroop paradigms. The SIE refers to the greater interference (slower naming latencies) produced by a distractor semantically related (vs. unrelated) to the picture (e.g., picture: dog; distractor: cat). The PFE refers to reduced interference (faster naming latencies) produced by distractors phonologically related (vs. unrelated) to...
pictures (e.g., picture: dog; distractor: doll).

Empirical evidence indicates that the loci of the semantic interference and PFE are different (e.g., Schriefers, Meyer, & Levelt, 1990). It is generally assumed that the PFE occurs at the phonological level (e.g., Roelofs, 1992; Levelt, et al., 1999, See Figure 2). However, the locus of the SIE is controversial. An influential theory in language production --- lexical selection by competition, assumes that the SIE occurs at the lexical selection level earlier than the phonological level (e.g., Roelofs, 1992; Levelt, et al., 1999; See figure 2). In contrast, the response exclusion hypothesis proposed by Mahon et al. (2007) assumes that the SIE occurs at the output buffer which is later than the phonological level (See Figure 2).

All the predictions for the last two experiments are based on the response exclusion hypothesis, as it is to my knowledge the only theory that can explain both types of distractor interference, semantic and distractor frequency. The lexical selection by competition hypothesis cannot account for the DFE observed in the literature. It assumes that the response times for selection of targets increase as the activation level of words increases and so predicts that the response times for picture/color paired with HF distractors are slower than LF distractors, as HF distractors have higher activation levels than LF distractors (McClelland & Rumelhart, 1981). Thus, it predicts the opposite direction of DFE observed in the literature. I return to a discussion of the lexical selection by competition hypothesis in the general discussion. Now, I only consider the response exclusion hypothesis.

According to the response exclusion hypothesis, it assumes that the SIE occurs via a similar mechanism and locus as the DFE. The central assumption is that the decision
mechanism for excluding distractor words from the output buffer involves general semantic information and the provenance (e.g., picture or word) of the representations. For example, for the Stroop paradigm, all the targets are colors and so the criteria would be naming colors. The semantically related words (e.g., red) meet the criteria but unrelated words (e.g., table) do not meet the criteria. So subjects can use the criteria to exclude unrelated words sooner than related words. That is why the naming latencies for semantically related words are slower than unrelated words.

Since this hypothesis assumes that both the SIE and DFE occur at the output buffer and I did not observe the DFE in Experiment 1 (Stroop) and 2 (when naming is fast), I predict that no SIE would be observed either for both Stroop and PWI when naming is fast. Moreover, it is generally assumed that the PFE occurs at an earlier level—the phonological level. Thus, I predicted that I may observe PFE in both paradigms when naming is fast. I reasoned that by exploring the presence of the semantic interference and PFEs in the two paradigms, I can ascertain to what level the distractor words are processed.

**EXPERIMENT 5**

The purposes of this experiment were to figure out when naming latencies are fast, to what level in the processing stream of word production distractor words were processed in the Stroop and PWI paradigms and to make sure that the two paradigms are comparable---similar effects observed in Experiment 5a and b. In order to speed up naming latencies, I only included four pictures to be named (see Experiment 2) in PWI in Experiment 5a and non-color words in Stroop (see Experiment 1b) In Experiment 5b. I
manipulated the frequency of distractors, the semantic and phonological relationship between the pictures/colors and distractors. Thus, there were three distractor manipulations: Distractor Frequency (HF, LF), Semantic (semantically related, semantically unrelated), and phonological (phonologically related and unrelated). However, for Stroop, as Experiment 3 showed that the inclusion of color words slowed down color naming, which results in a significant DFE, I did not use color words as semantically related distractors. Instead, I used color associates as semantically related distractors (e.g., color: red; word: grass). Studies in Stroop showed that incongruent color associates produce more interference than neutral words (e.g., Klein, 1964; Glaser & Glaser, 1989, Experiment 5). Hence, I also refer to this effect as the SIE.

METHOD

Participants

Twenty-eight (13 male) native English speakers took part in Experiment 5a and a different twenty-eight (12 male) in Experiment 5b.

Materials and Procedure

I used the same four pictures from Experiment 2 for the PWI paradigm (Experiment 5a), the same four colors in Experiment 1&3 for the Stroop paradigm (Experiment 5b), and the same HF and LF distractors from Experiment 1-4. The four pictures were CAR, PIG, TREE and HAT and the four colors are GREEN, RED, BLUE and YELLOW. Four semantic and phonological conditions were included: a) semantically related/associated condition: category coordinates of target pictures (e.g., picture: pig; word: cat) in PWI and color associates (e.g., color: red; word: sky) in Stroop; b) semantically unrelated
condition: words from different category of the targets (e.g., picture: pig; word: deck) in PWI and neutral words (e.g., color: red; word: hat) in Stroop; c) phonologically related condition: the picture/color and the word share the same phonological onset (e.g., PWI: picture: pig; word: picture. Stroop: color: red; word: rat); d) phonologically unrelated condition: the picture/color and the word do not share the same phonological onset (e.g., PWI: picture: pig; word: subject. Stroop: color: red; word: web). I also included word distractors phonologically related to the picture (the picture and word share the same phonological onset (e.g., PWI: picture: pig; word: picture. Stroop: color: red; word: rat) and phonologically unrelated to the picture (the picture and the word do not share the same phonological onset (e.g., PWI: picture: pig; word: subject. Stroop: color: red; word: web). For both paradigms, semantically related/associated words and unrelated words were matched on frequency (Kučera & Francis, 1967), word length, imageability (Gilhooly & Logie, 1980) and number of syllables. Phonologically related words and unrelated words were matched on the same variables. There were 40 fillers so that the semantically associated/related and phonologically related conditions occur less than 1/3 of the time. This manipulation was introduced to rule out the possibility that the participants develop expectations for the semantically associated/related or phonologically related words when they did the naming paradigm. The 40 HF and LF words were each seen once, while the words in other four conditions were seen 10 times each.

**Design & Procedure**

For both paradigms, the six conditions (40 HF and 40 LF words, 4 semantically associated/related words and 4 semantically unrelated words, 4 phonologically related
words and 4 phonological unrelated words) and 40 fillers were combined into seven blocks, each containing approximately six experimental trials for each distractor condition (HF/LF and semantically related/unrelated and phonological related/unrelated) and filler. Each block was 40 trials long. Trial presentation within each block was pseudo-randomized such that no picture/color or word was repeated on consecutive trials and no word from the same distractor condition on consecutive trials. For each paradigm, each subject saw all seven blocks and the order of block presentation was counterbalanced across subjects. Three warm-up trials using practice items were included at the beginning of each block. The procedure is identical to the PWI and Stroop paradigms in Experiment 1. Each experiment lasted for 20 minutes.

RESULTS AND DISCUSSION

Experiment 5a

Latency data were preprocessed in the same way as in Experiment 1a. 4.5% of the data points were removed. An overview of the mean RTs and error rate for the PWI paradigm is presented in Table 6. Two paired-samples t-tests were computed comparing semantically related and unrelated words and phonologically related and unrelated words. For HF and LF distractors, a paired-samples t-test was computed for the subject analysis and an unpaired-samples t-test for the item analysis. I found no significant differences in naming errors for the error analysis ($t_{s} < 1$). In the analysis of naming latencies, there was no difference for HF and LF words and semantically related and unrelated words ($t_{s} < 1$). But naming latencies were 12ms faster for phonologically related words than unrelated
words \( t_1(27) = 3.60, p < .01; t_2(3) = 6.36, p < .01 \). In sum, I observed the PFE but not the DFE and SIE in the PWI paradigm.

**Experiment 5b**

Latency data were preprocessed in the same way as in Experiment 1a. 4.7% of the data points were removed. An overview of the mean RTs and error rate for the Stroop paradigm is presented in Table 6. Two paired-samples t-tests were computed comparing semantically associated and unrelated words and phonologically related and unrelated words. For HF and LF distractors, a paired-samples t-test was computed for the subject analysis and an unpaired-samples t-test for the item analysis.

Table 6. Naming latencies (ms) and error rate (in parentheses) for PWI and Stroop tasks in Experiment 5.

<table>
<thead>
<tr>
<th>Distractor words conditions</th>
<th>LF</th>
<th>HF</th>
<th>DFE</th>
<th>sr</th>
<th>s ur</th>
<th>SIE</th>
<th>pho r</th>
<th>pho ur</th>
<th>PFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PWI (exp 5a)</td>
<td>594(.02)</td>
<td>591(.03)</td>
<td>3</td>
<td>590(.02)</td>
<td>587(.02)</td>
<td>3</td>
<td>564(.02)</td>
<td>578(.01)</td>
<td>-14*</td>
</tr>
<tr>
<td>Stroop (exp 5b)</td>
<td>637(.02)</td>
<td>627(.01)</td>
<td>10</td>
<td>639(.03)</td>
<td>610(.01)</td>
<td>29</td>
<td>587(.02)</td>
<td>606(.02)</td>
<td>-19*</td>
</tr>
</tbody>
</table>

* \( p < .05 \). sr: semantically related condition; s ur: semantically unrelated condition; pho r: phonologically related condition; pho ur: phonologically unrelated condition.

DFE = LF-HF; SIE = sr - s ur; PFE = pho r - pho ur.

In the error analysis, there was no difference in naming errors between HF and LF words and phonologically related and unrelated words \( ts < 1 \) but a significant difference between semantically related and unrelated words (see Table 6, \( t_1(27) = 4.18, p < .001; t_2(3) = 3.54, p < .05 \)). In the analysis of naming latencies, naming latencies were faster for the phonologically related than unrelated words \( t_1(27) = 4.22, p < .001; t_2(3) = 9.97, p \)
<.01). For the semantic condition, the naming latencies were marginally slower for semantically related words than unrelated words ($t_1(27) = 8.818, p < .001; t_2(3) = 2.57, p = .08$). For the frequency condition, naming latencies were marginally slower for LF distractors than HF distractors ($t_1(27) = 2.30, p < .05; t_2(78) = 1.66, p = .10$).

According my speed hypothesis, fast naming can eliminate the effects which occur at late levels in language production, such as the DFE or the SIE. The reason why the DFE and SIE were marginally significant in Stroop paradigm could be that the color naming was slower so that some distractors were processed to the output buffer by the time colors were named. Thus, I did a post hoc analysis. First, I compared the overall naming latencies between Experiment 5a and 5b and the naming latencies for the frequency and semantic conditions in Experiment 5a and 5b. For the overall naming latencies, the color naming in Stroop (Experiment 5b) was marginally significant slower than the picture naming in PWI (Experiment 5) (PWI: 581ms; Stroop: 617ms; $t(52) = 1.87, p = .07$).

For the frequency condition, the color naming in Stroop was significantly slower than the picture naming in PWI (PWI: 592ms; Stroop: 632ms; $t(52) = 2.03, p < .05$). For the semantic condition, the color naming in Stroop was significantly slower than the picture naming in PWI (PWI: 588ms; Stroop: 624ms; $t(52) = 2.03, p < .06$). The slower naming in Stroop

![Diagram of the loci of the phonological facilitation (PFE), semantic interference (SIE), distractor frequency effects (DFE). The locus of PFE is phonological encoding level (earlier level). Both SIE and DFE occur at output buffer (later level).](image-url)
might impact the processing of distractor words and some distractor words in frequency and semantic conditions are processed to the output buffer where DFE and SIE occur. Thus, there is a trend of DFE and SIE in Stroop.

For both paradigms, I observed the PFE but not the DFE and the SIE. Thus, the results of Experiment 5 are consistent with the response exclusion hypothesis which assumes that the PFE occurs at earlier level (phonological level, see Figure 3) and the DFE and the SIE occur at later level (output buffer, see Figure 3).

As mentioned in the discussion of Experiment 1, Miozzo and Caramazza (2003) and Mahon et al. (2007) have different assumptions about the locus of the DFE. Miozzo and Caramazza (2003) assumed that the DFE occurs at the phonological level, while Mahon et al. (2007) proposed that the DFE occurs at the output buffer later than the phonological level. My results of last two experiments are consistent with the response exclusion hypothesis (Mahon et al., 2007), as I found the PFE but no DFE in both paradigms.

GENERAL DISCUSSION

I explored the issue of whether the Stroop and PWI paradigms operate in the same way by testing the DFE in both paradigms. I also tested to what level the distractor words are processed along the process chain of language production in both paradigms. For the comparison of the two paradigms, the five experiments produced four major findings. First, the DFE (i.e., picture naming latencies are slower in the presence of LF distractor words compared to HF distractor words) was observed in the standard PWI paradigm (Experiment 1a&4) and there was no DFE in Stroop paradigm with non-color word distractors (Experiment 1&5b). Second, the DFE can be eliminated in the PWI paradigm
by speeding up picture naming (Experiment 2&5a) and be observed in the Stroop paradigm by slowing down color naming (Experiment 3&4). From this, I conclude that the DFE only appears when the target responses are available for production after or currently with distractor words processed to the post-lexical level. Third, the magnitude of the DFE in Stroop with color words and PWI with twenty pictures is comparable. This indicates that the two paradigms operate in a similar way, with regards to how naming latency is affected in the presence of high and low frequency word distractors. Finally, There were similar result patterns (PFE but no DFE and SIE) for the Stroop and PWI paradigms when naming is fast (Experiment 5). Furthermore, Experiment 5 tested to what level the distractor words are processed when naming is fast in both paradigms. The presence of the PFE suggests that the words are processed to the phonological level by the time targets are named, when naming is fast in both paradigms. In addition, the last experiment suggests that the locus of DFE is at output buffer which is later than the phonological level.

Although the two paradigms are comparable for the DFE, the different processing speeds for naming pictures and colors in the two paradigms play the key role in observing the DFE. According to the response exclusion hypothesis, the DFE occurs at the output buffer. In the Stroop paradigm with non-color word distractors, the words are not processed to the output buffer by the time color names are produced, which results in no DFE (see Figure 1). In contrast, in standard PWI paradigm, the distractor words are processed to the output buffer by the time the target responses are available, allowing the frequency of the distractor word to impact naming (see Figure 1).
This speed hypothesis can also explain the presence and absence of the DFE in the Stroop task in previous studies. Monsell et al. (2001) did not find any difference between HF and LF distractors in the Stroop task without color words. I hypothesize that the naming latencies (570ms) were too fast to observe the DFE. Kahan and Hely (2008) found the DFE only for negative emotion words but not for neutral and positive emotion words in an emotional Stroop task. The naming latencies for negative emotion words were significantly longer than neutral and positive emotion words. Thus, the DFE appeared when response times were slower but not for the quicker response time, a pattern consistent with my hypothesis.

Why is naming generally faster in the Stroop vs. PWI paradigms? There are at least two possible factors. First, pictures are processed slower than colors in general (e.g., Vukovic, Wilson, & Nash, 2004). Second, there are usually more pictures than colors in experimental settings, which leads to more repetitions of colors than pictures. Experiment 2 showed that when there were only four pictures in PWI matched with four colors in Stroop and the number of repetition in both paradigms was the same, the picture naming in PWI was even faster than color naming in Stroop (picture naming: 592; color naming: 627, \( p < .05 \)). It may suggest that the number of repetitions is the main factor to impact observing interference effects in the two paradigms (e.g., DFE).

Importantly, the response exclusion hypothesis needs to be modified in order to account for the results of all the experiments. As mentioned in the discussion of Experiment 1, the original response exclusion hypothesis proposed by Mahon et al. (2007) gives an interpretation about the mechanism of DFE and predicts the DFE in both PWI and Stroop paradigms, regardless of whether naming is slow or fast. However, this is not
the case. In my experiments, there was no DFE observed when naming was fast but DFE when naming was slow in both paradigms. The speed hypothesis gives a good account for the results by assuming that when naming is fast, distractor words are not processed to the output buffer by the time targets are named. Thus, in order to interpret the results, the response exclusion hypothesis has to take the speed hypothesis into account. Furthermore, if the response exclusion hypothesis includes the speed hypothesis, two original assumptions need to be dropped: a) words have a privileged relationship to articulation in a way that pictures/colors do not and b) in PWI and Stroop, all distractor words get to the post-lexical level faster than target pictures/colors. When naming is fast in both Stroop and PWI, no DFE was observed, which suggests that sometimes words get to the post-lexical level later than target pictures/colors.

Is the modified response exclusion hypothesis the only one that accounts for my data? As mentioned in the discussion of Experiment 4, there is a more influential hypothesis—lexical selection by competition. It assumes that the SIE occurs at the lexical level. According to this hypothesis, it proposes that the time for target selection depends not only on the activation level of target but also on the activation levels of non-target words. For example, when the target concept corresponding to the target picture/color (e.g., blue) gets activated, it spreads activation not only to the target word but also the semantically related distractor words (e.g., red) at the lexical level. For semantically related distractor words, they get activation from both the target and presentation of the word. In contrast, for unrelated distractor words (e.g., table), they get activation from only the presentation of the word. Thus, semantically related words have higher activation levels than unrelated words. The hypothesis assumes that the response times for targets increase as the
activation levels of competitors increase. That is why the semantically related words interfere with the picture/color naming more than unrelated words. However, for the DFE, as mentioned in the discussion of Experiment 4, the competition hypothesis predicts that HF distractors should interfere more than LF distractors and the DFE should be observed in both paradigms, with the assumptions that HF words have higher activation levels than LF words and that the two paradigms operate in a similar way. Thus, this competition hypothesis makes incorrect predictions about the presence of the DFE in the two paradigms and the direction of the DFE.

How does it account for the presence of the PFE and the absence of the SIE in the last experiment? According to the competition hypothesis, the SIE occurs at the lexical level. Previous studies showed that the PFE occurs at the phonological level (e.g., Schriefers et al., 1990). It is generally assumed that there are two routes for word processing in the word production model (see Figure 4; e.g., Rolofes, 1992). One route runs from lexical level to articulation and it is called “the lexical route”. The other one runs from the phonological level to articulation and it is called “the sub-lexical route”. All words go through both routes in word reading task. All pronounceable non-words only go through the sub-lexical route, due to the fact that they do not have corresponding lexical representations in the lexical level. Previous studies showed that in the PWI paradigm, phonologically related non-words produced facilitation effect, compared to unrelated non-words (e.g., Lupker, 1982; Jerger, Martin, Damian, 2002). This indicates that the PFE can be observed when words or non-words are processed only via the sub-lexical route. The SIE can be observed only when words are processed to the lexical level
through the lexical route. In the last experiment, there was no SIE but PFE in both paradigms. According to the model, it is possible that when naming is fast, all words have

![Diagram of processing levels](image)

Figure 4. Levels of processing in PWI and Stroop paradigms. Route a is a lexical route. Route b is a sub-lexical route. Adapted from *Psychological Review, 110*, A. Roelofs, "Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task." p. 98, Figure 7, copyright 2003, with permission from the author.

not yet been processed to the lexical level via the lexical route by the time targets are named but they might be processed to the phonological level via the sub-lexical route. That is why there was no SIE but PFE in the last experiment. Therefore, although the lexical selection by competition hypothesis cannot interpret the DFE, it could still account for the presence of the PFE and the absence of the SIE in the last experiment.

As mentioned in the introduction, Van Maanen et al. (2009) provided a computational model of SIEs in both Stroop and PWI paradigms to account for Dell’ Acqua et al. (2007)’s and Fagot et al. (1992)’s results and argued that although the SIEs in both paradigms have different loci, the mechanisms of the effects are still the same. Does their model account for my results? They assume there is a competition at the conceptual
(perceptual) and lexical (response selection) levels for both PWI and Stroop paradigms and the SIE in PWI primarily arises during the perceptual stage and very little at the response selection stage, while the effects in Stroop arise at the response selection and not at the perceptual stages. This model does not say anything about DFE. But if assuming frequency is modeled in terms of activation levels at the response stage and the activation spreads proportionally between stages, this model makes incorrect predictions: 1) the presence of the DFE in PWI and Stroop, and 2) the direction of the DFE. For the PWI paradigm, the model assumes that the interference effects in PWI arise primarily at the perceptual level. Given the competition between target and non-target words and higher activation levels of HF words than LF words, the direction of this effect should be that naming latencies should be longer for high than for low frequency distractors.

For the Stroop task, the model assumes that the interference effects arise primarily at the response selection level. Thus, the model predicts a strong interference effect in the Stroop task for unrelated low and high frequency distractors. The direction of this effect should be longer naming latencies for high than for low frequency distractors.

The results in Experiment 1 are inconsistent with these predictions. Firstly, the DFE appears reliably in PWI but not in Stroop. Secondly, the direction of the effects is opposite from the predictions of the model. So, this model cannot explain my results.

CONCLUSIONS

The results reported in this master thesis firmly established two findings: 1) the Stroop and PWI paradigms operate in a similar way but the speed of naming colors or pictures affects the appearance of interference effects. The response exclusion hypothesis could
account for the results if it takes the speed hypothesis into account; 2) when naming is fast, the distractor words are most likely processed to the phonological level by the time targets are named. This could be interpreted by either the modified response exclusion hypothesis or by assuming a role for the sub-lexical route in the word production model.
REFERENCE


7, 150-157.


APPENDIX

There were 40 High and 40 low frequency words used in all experiments and 20 pictures used in experiment 1a and 4. (There were also four colors used as both distractors and targets in experiment 3 and 4 and used only as targets in the Stroop tasks of the rest experiments: GREEN, RED, BLUE, YELLOW. There were four pictures used as targets in experiment 2 and 5a: PIG, TREE, CAR, HAT.

<table>
<thead>
<tr>
<th>Picture</th>
<th>HF word</th>
<th>HF word</th>
<th>LF word</th>
<th>LF word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchor</td>
<td>MUSIC</td>
<td>PAPER</td>
<td>CORAL</td>
<td>COMET</td>
</tr>
<tr>
<td>Arrow</td>
<td>PLANT</td>
<td>LIGHT</td>
<td>FROCK</td>
<td>TORCH</td>
</tr>
<tr>
<td>Bell</td>
<td>CITY</td>
<td>EYE</td>
<td>LAVA</td>
<td>EWE</td>
</tr>
<tr>
<td>Candle</td>
<td>STREET</td>
<td>FIELD</td>
<td>STRAIT</td>
<td>POUCH</td>
</tr>
<tr>
<td>Cup</td>
<td>BODY</td>
<td>RIVER</td>
<td>TOFU</td>
<td>TULIP</td>
</tr>
<tr>
<td>Drum</td>
<td>WATER</td>
<td>WOMAN</td>
<td>CAMEL</td>
<td>BACON</td>
</tr>
<tr>
<td>Duck</td>
<td>HOUSE</td>
<td>CHILD</td>
<td>GNOME</td>
<td>CLOWN</td>
</tr>
<tr>
<td>Fish</td>
<td>NIGHT</td>
<td>CHURCH</td>
<td>MARSH</td>
<td>CRUTCH</td>
</tr>
<tr>
<td>Flag</td>
<td>DOG</td>
<td>HAIR</td>
<td>KEY</td>
<td>LOAF</td>
</tr>
<tr>
<td>Foot</td>
<td>TOWN</td>
<td>COURT</td>
<td>VEST</td>
<td>VAULT</td>
</tr>
<tr>
<td>Hat</td>
<td>TABLE</td>
<td>AIR</td>
<td>VALVE</td>
<td>OWL</td>
</tr>
<tr>
<td>Kite</td>
<td>SCHOOL</td>
<td>HAND</td>
<td>SPLEEN</td>
<td>FERN</td>
</tr>
<tr>
<td>Lemon</td>
<td>VOICE</td>
<td>MONEY</td>
<td>CRATE</td>
<td>PATIO</td>
</tr>
<tr>
<td>Necklace</td>
<td>FIRE</td>
<td>FOOD</td>
<td>DOVE</td>
<td>QUIZ</td>
</tr>
<tr>
<td>Pear</td>
<td>GUN</td>
<td>CELL</td>
<td>HOG</td>
<td>CASK</td>
</tr>
<tr>
<td></td>
<td>Pepper</td>
<td>HEAD</td>
<td>WAR</td>
<td>BAIT</td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Pig</td>
<td></td>
<td>NAME</td>
<td>GIRL</td>
<td>BALE</td>
</tr>
<tr>
<td>Pipe</td>
<td></td>
<td>HORSE</td>
<td>WALL</td>
<td>BADGE</td>
</tr>
<tr>
<td>Rope</td>
<td></td>
<td>OFFICE</td>
<td>DOCTOR</td>
<td>ALCOVE</td>
</tr>
<tr>
<td>Tent</td>
<td></td>
<td>CLASS</td>
<td>BOOK</td>
<td>SCALP</td>
</tr>
</tbody>
</table>