STROOP INTERFERENCE AND FACILITATION EFFECTS WITH CHINESE CHARACTERS AND PINYIN

YICHENG QIU, MA

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

MARCH 2023

Abstract

The Stroop task has been used to investigate automatic lexical processing and attentional control. Unlike tasks that involve explicit reading (i.e. the lexical decision task), the Stroop task asks for the names of the ink colours, making the word reading task irrelevant. This thesis investigates the Stroop phenomenon by using Chinese characters and pinyin, and how semantic and phonological information provided by Chinese characters and pinyin can be activated in the Stroop task.

In Chapter 3, three Stroop experiments are presented that investigated the role of sublexical components (i.e. phonetic radicals) in Chinese characters. When the meaning of the target character is irrelevant to the colour words, its phonetic radical, which is a colour word, can still elicit strong Stroop interference and facilitation effects. This suggests the semantic activation of sublexical components in Chinese characters. Additionally, the phonological cues provided by sublexical components can also be activated in the Stroop task.

Chapter 4 Presents a Stroop experiment that explored the decomposed components in Stroop effects. Previous research focussing on multi-stage accounts of Stroop effects confirmed the contribution of response, semantic, and task conflicts. This study provides evidence for the impact of phonological conflicts/facilitation in Stroop effects by using Chinese homophones.

A series of Stroop experiments are presented in Chapter 5 that looked at the impact of using Chinese characters written in pinyin, a Romanization transcription of Chinese characters. The results showed that Chinese characters written in pinyin can activate semantic and phonological information of target characters. The mixed presentation of Chinese characters and pinyin in the final experiment suggested that those two scripts can impact each other's performance in Stroop interference and facilitation effects.

The results of this thesis provide new interpretations of the Stroop phenomenon that it can be decomposed into distinct components, including phonological components. In turn, the Stroop task enables the investigation of the automaticity in reading Chinese words, suggesting the role of phonology in the activation of semantics.

Declaration

I declare that this is my own work conducted during my time as a PhD student at the University of Nottingham. The first three experiments were conducted in the lab at the University of Nottingham. Due to the impact of COVID-19, further experiments were transferred to online data collection.

Three of the chapters in this thesis are being prepared for publication.

Acknowledgement

I would like to express my gratitude to the people who have supported the completion of this work during the past five years. First, thanks to my principal supervisor, Walter van Heuven, who accepted me as his PhD student and was willing to help me out whenever possible. Even during the hardest time when the COVID-19 pandemic hit, we made regular contacts through online meetings. Many thanks to my second supervisors, Rikka Mottonen and Lucy Cragg, and my internal assessor, Ruth Filik, for providing valuable comments on my academic progress.

Then, I would like to offer my gratitude to my family, who supported my pursuit of a PhD degree. Specifically, thanks to my grandpa, who kept an eye on my thesis completion status via video chat every week.

Thanks to my colleagues in B60, my PhD friends, and my roommates. It is with you that I can express my happiness and sadness, and all the trivial things that happened during my PhD life.

My gratitude goes to all the participants who took part in my experiments. It was a difficult job to find native Chinese participants in a foreign country. I was touched by the fact that people were so willing to help even though they may not have been familiar with me at all. Thanks again for your passion for my research and confidence in me.

Last but not least, thanks to Team Delta, who were my mental asylum during the harsh times of the COVID-19 pandemic. It is my honour to make acquaintance with all of you.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction to Chinese Word Recognition

Word recognition is a natural process for a literate person. People may speak various dialects that, in extreme cases, are not mutually intelligible. However, with a standardised written language, they can communicate with each other. Chinese characters are a good example of a standardised script used across China. It developed from oracle bone inscriptions (甲骨文), bronze inscriptions (金文), seal script (篆文), clerical script (隶书), cursive script (草书), to regular script (楷书) (Jiang, 2013). The development of Chinese characters not only relied on social conventions, but also on official standardisation (Xiang, 2010). Modern Chinese characters were simplified and standardised according to the Summary of Simplified Characters (National Language Commission, 1964).

This thesis explores the word recognition of Chinese characters and Chinese characters written in pinyin using behavioural methods in a series of colour naming or colour decision tasks (i.e. Stroop task; Stroop, 1935). The first chapter provides an overview of Chinese visual word recognition, introducing findings from a broad range of tasks, such as semantic decision, phonological decision, and priming task. Section 1.1 of the first chapter provides a general introduction to the Chinese language, discussing the Chinese character and its phonetic system. Section 1.2 reviews the literature that focuses on the recognition of Chinese characters. Section 1.3 provides an overview of findings on sublexical processing in Chinese. Section 1.4 focuses on pinyin and the role of pinyin in reading development and word recognition. Section 1.5 reviews different models/hypotheses about Chinese word recognition, including the direct access hypothesis, the indirect access (phonology-mediated) hypothesis, the interactive framework of both orthography and phonology, and the lexical constituency model. Following the general findings on Chinese word recognition, Chapter 2 presents an introduction to the Stroop task, which is used to examine the automaticity of word processing.

1.1 General Introduction to Chinese

The Chinese language is often categorized as a logographic language, which means that the word itself can be recognized as a symbol that imitates real-life objects (DeFrancis, 1989). However, this is only one of the formation methods in Chinese characters. There are mainly four types of Chinese character formations (Xu, 1963, p. 1). The other two types of formation are "rebus" (phonetic loan characters) and "transformed cognates" (two characters are used as a reference to each other, with the same meaning but different orthography). These two types are more related to usage rather than formation, so they are often omitted. According to Xu's explanation, pictographs and ideograms are "writing/drawing" (文), while compound ideographs and phono-semantic compounds are "character" (字). Combined, they are called "characters" (文字). This may imply that ancient Chinese scholars were aware that compound ideographs and phono-semantic compounds are more artificial in nature.

The first formation method is pictographs, which are pictorial representations of real-life objects and the reason for Chinese being classified as a logographic writing system (e.g. "木" means tree, as it depicts a tree with many branches). The second type is that of ideograms, which are iconic symbols used to express abstract ideas (e.g. "一" means the number one, according to the number of strokes). The third type is compound ideographs that combine two or more characters into a single character (e.g. "林" means trees/forest, which consists of two "木" trees). The final type, which is the most common and most productive one, is phono-semantic compounds that combine semantic and phonetic components (radicals) into a character. Semantic radicals can provide a cue to the meaning or category of a character. For example, the character "银" (silver) contains the semantic radical "钅" (metal), which refers to its category in the real world. Phonetic radicals can provide phonological cues to the pronunciation of the character. The character "清" (clear, /qīng/), for example, has the phonetic radical "青" (cyan, /qīng/), which shares the pronunciation with the colour cyan ("青", /qīng/). Although this is not always correct, for example, the phonetic radical of "银" (silver, /yín/) is "艮" (tough, /gěn/ or /gèn/), whose pronunciation is different from the pronunciation of the whole character. According to Zhou (1978), only about 38% of phonetic radicals have pronunciations consistent with their whole characters. Shu and Anderson (1999) analysed the regularity of Chinese characters that occurred in Chinese language textbooks and found that with the increase of word frequency, regularity decreases. The opacity between the phonemic and orthographic codes is called the orthographic depth and Chinese is categorized as a deep orthography (Frost et al., 1987; Katz & Feldman, 1981). It is argued that in a deep orthography, the complexity of grapheme to phoneme translation would lead to the reliance on the orthographic code that embedded the phonology (Frost et al., 1987).

The Chinese phonology can be decomposed into segments and tones. For example, the character "怜" (/lián/, pity) has the segment /lian/ and a rising tone. When the tone changes, it can result in orthographically and semantically unrelated stimuli (e.g. "练", /liàn/, practice). The use of Romanization transcription of Chinese characters to represent the phonology is called pinyin. This transcription system was proposed by the "Scheme for the Chinese Phonetic Alphabet" and is widely used for teaching and learning pinyin in Mainland China (China, 1958).

Segments can also be divided into initial consonants (声母) and vowels (韵母). There are 21 initial consonants and 39 simple or compound vowels in pinyin, which can be combined into around 400 distinct combinations. For example, the initial consonant /c/ combined with the compound vowel /ao/ becomes cao. Vowels and consonants are the segmental information of Chinese phonology. Suprasegmental information in Chinese phonology refers to the tone. There are four types of tones: high (阴平), rising (阳平), dipping (上声), and falling (去声) (Ding & Rong, 2012). To indicate the tone (suprasegmental information) in pinyin, four diacritics are used. For example, cao can be written as cāo, cáo, cǎo and cào. Sometimes numbers (1-4) are used (e.g. cao1) to indicate tones, but the standard way to indicate tone is to use diacritics. Pinyin with diacritics can reduce the ambiguity in mapping to specific characters, although it is still plenty. For instance, pinyin jīn could represent "金" (jīn, gold), "今" (jīn, the present), or "巾" (jīn, towel). Pinyin without diacritics (tonal information) become more ambiguous in terms of mapping to Chinese characters. For example, pinyin cao could either be "糙" (cāo, rough), "嘈" (cáo, noisy) or "草" (cǎo, grass), each with a different tone and totally unrelated meanings. Segmental (vowels and consonants) and suprasegmental (tones) information provides about 1300 combinations in pinyin (Wang et al., 2008). In contrast, the number of Chinese characters is estimated to be approximately 85,000 (Leng & Wei, 1994), of which 2,500 are the most frequent characters (State Language Commission, 1988). Therefore, there are many more characters than pinyin words, and as a consequence, homophones are very common in Chinese.

In the following section, the discussion about Chinese word recognition will be divided into three parts, namely character, radical, and pinyin. Characters and radicals are part of the logographic writing system, while pinyin is an alphabetic writing system for Chinese phonology.

Research focussing on Chinese word processing has mainly targeted on the whole character. However, research has also suggested that the sublexical components of Chinese characters – radicals are processed first (e.g. Chen & Yeh, 2015; Feldman & Siok, 1997; Yeh et al., 2017). The position and regularity of radicals influence Chinese character reading. Pinyin is the Romanization transcription of Chinese characters. It is an alphabetic writing system for a non-alphabetic language. Pinyin provides no visual (orthographic) information that links to Chinese characters but only phonological information that links to Chinese characters. Thus, the activation of Chinese phonology triggers the activation of the orthography and semantics of Chinese characters. The following literature review will focus on Chinese characters, radicals contained in characters, and characters written in pinyin.

1.2 Chinese Characters

In a shallow orthography like Serbo-Croatian, each grapheme corresponds to each phoneme, achieving a consistent link between phonology and morphology (Frost et al., 1987). In a deep orthography like the Chinese language, the link between orthography and phonology is weak. For example, pinyin jīn could be "金" (jīn, gold), "今" (jīn, the present), or "巾" (jīn, towel). They are pronounced the same but there is no link to the orthography. Wong and Chen (1999) argued that Chinese orthography has direct access to semantics, whereas Perfetti and Zhang (1995) argued that Chinese readers can process not only the semantic information in characters, but also phonological information. They provided participants with two tasks: a semantic decision task where participants must identify if two single characters have similar meanings and a phonological decision task where they are asked to decide if two single characters have the same pronunciation. The tasks included three types of stimuli: synonyms, homophones, and control characters. Participants made vocal responses (yes or no) to each trial to decide the semantic or phonological relationship between the two stimuli. In the phonological decision task, it took longer to reject synonym foils than controls. This effect suggests orthography cannot activate meaning without activating phonology (Guo et al., 2005; Spinks et al., 2000; Tzeng et al., 1977). In the semantic decision task, homophone foils took longer to reject than controls. This is consistent with the view that phonological information is also activated during word reading. Even though the task does not require phonological information retrieval, the activation of phonology was automatic.

As introduced in Section 1.1, there are different formation methods of Chinese characters that can impact how information is utilized in word reading. A study carried out by Leck et al. (1995) showed that different types of characters have a preference for activating visual, phonological, and semantic information. In a semantic categorization task, they used characters from different semantic categories and asked participants to judge whether the stimuli belong to the semantic category. The stimuli could be the target word from the same category or the foil. There are five types of foils: V^+P^+ (visually similar and phonologically identical to the primes), V^+P^- (visually similar but pronounced differently), $V_{R+}P^-$ (visually similar radical but pronounced differently), V^-P^+ (homophones, visually dissimilar, but phonologically identical), and V^-P^- (visually and phonologically dissimilar, controls). The results showed that with phono-semantic compounds, it took longer to reject V^+P^+, $V_{R+}P^-$, and V^+P^- foils than controls, whereas with integrated characters, it took longer to reject V^+P^- foils than controls (V^+P^+ condition is not available for integrated characters). For integrated characters, the word recognition process relied heavily on visual information. For compound characters, both visual and phonological information is used in processing the words. Whereas in Perfetti and Zhang's (1995) work, most stimuli were compound characters, which, according to Leck et al.'s (1995) result cannot represent the phonological processing in all types of Chinese characters. The processing of integrated or compound characters may rely on different sources of information.

As discussed above, the processing of characters also depends on whether they are compound or integrated characters. The homophone density could also affect the processing of characters. Tan and Perfetti (1997) used targets with different homophone densities in a priming task with different SOAs. They divided target homophones into three groups: low homophone density (less than 5 homophones), medium homophone density (8-15 homophones), and high homophone density (more than 20 homophones). Each target homophone was paired with three types of primes: synonyms, homophones of the synonyms, and neutral controls. With short SOAs (129 ms and 243 ms), both synonym primes and homophones of the synonym primes showed significant facilitation relative to controls for targets with low or medium homophone density. When a high homophone density target was used, only synonym primes revealed strong facilitation effects. With a long SOA (500 ms), only synonym primes facilitated word recognition when the target had low, medium, or high homophone density. This study revealed limitations on how phonology mediates access to meaning: 1). it is restricted to primes with fewer than 20 homophones, and 2). when the exposure to primes is sufficient (e.g. 500 ms), only synonyms affected the word recognition process but not homophones of synonyms, whose effect was independent of homophone density.

Perfetti and Tan (1998) investigated the time course of orthographic, semantic, and phonological activation in Chinese characters in the priming task. The orthographic prime was similar to the target but unrelated in semantics and phonology. The control primes consisted of two types: neutral single-character and non-linguistic number symbol (#). The selection of a meaningless symbol has considered its nonword property, so that it would not cause verbal inhibition as neutral characters. The primes were further divided into two groups according to their semantic vagueness: vague and precise primes. This vagueness refers to how precisely the character expresses its meaning. For example, a vague meaning character "何" can be interrogative pronouns, asking about who, what, when or where. A precise meaning character "村" means village. In both vague and precise primes, phonological (homophone) and semantic primes showed facilitation effects compared with both neutral characters and non-linguistic symbol (#). The orthographic primes, however, showed no facilitation effects in 180 ms SOA. Consequent experiments showed that graphic primes became facilitatory at 43 ms SOA but became inhibitory at 85 ms. In the meantime, phonological and semantic primes showed no facilitation effects at 43 ms SOA. By increasing the SOA to 57 ms, the effects of homophones appeared. When SOA came to 85 ms, the semantic primes showed facilitation effects with precise meaning. The results revealed a clear time course for the activation of orthographic, semantic, and phonological information. Phonological activation precedes semantic activation. Graphic information is highly temporal sensitive in that it facilitates the word recognition at short SOAs; however, when the processing time is long enough, this similarity in orthography becomes inhibitory to the recognition of Chinese characters.

Shen and Forster (1999) also observed an orthographic priming effect in the masked lexical decision and naming tasks. Moreover, they argued that phonological information may not be compulsory for accessing to meaning. They tested phonological priming effects in two different tasks: a naming task and a lexical decision task to investigate whether phonological activation is task dependent. Similar to Leck et al.'s (1995) study, they presented simple (integrated) character and compound characters in separate sessions. The primes were orthographically similar to the target (V^+P^-), phonologically identical (V^-P^+), or both orthographically similar and phonologically identical (V^+P^+, unavailable for the simple character condition). In the naming task, they found facilitatory effects for the orthographic priming condition in both simple and compound characters. Phonological priming effects were found in the simple character conditions only. In the lexical decision task, orthographically similar primes took less time to respond than neutral controls using simple characters. Orthographically similar primes and orthographically similar plus phonologically identical primes showed facilitation effects in compound characters. Surprisingly, no phonological priming effects were discovered in either simple or compound character conditions. Thus, they suggested that phonological priming effects were task dependent (naming task) and exclusive to simple characters. These results are consistent with Leck et al.'s (1995) finding that the recognition of a simple character depends on the orthographical information and the recognition of compound characters relies on orthographical, semantic, and phonological information.

The above studies all used word primes that were phonologically identical to the target. However, it is also possible to create phonologically similar primes in Chinese. Wang et al. (2015) investigated the role of segmental and tonal information in Chinese word reading using a priming task with three different SOAs. The priming conditions included primes with segment and tone identical to the target (S^+T^+), segment identical but not the tone (S^+T^-), tone identical but not segment (S^-T^+), and neutral controls (S^-T^-). Across all SOAs (57 ms, 100 ms, and 200 ms), significant facilitation effects were found for the

S^+T^+ condition but not for the other conditions. This suggested that both segmental and tonal information are presented and encoded as an integral unit in reading Chinese characters.

While Wang et al. (2015) concluded that segmental and tonal information are simultaneously activated, Li et al. (2019) argued that segmental information is activated earlier than tonal information during the word recognition process. In a homophone judgement task, participants were asked to judge if the two characters were pronounced the same by pressing the corresponding key buttons. They used the same four conditions as in M. Wang et al.'s (2015) experiment: S^+T^+, S^+T^-, S^-T^+, S^-T^-. Critically, an interference effect was observed in the S^+T^- condition only. This suggested that identical segmental information result in more interference than identical tonal information and, in turn, confirmed that Chinese speakers rely more on segmental information than tonal information.

In summary, studies on Chinese character recognition have confirmed that orthographical, phonological, and semantic information is activated. Importantly, the time course varies: orthographical information is activated earlier than phonological information, and subsequently semantic information is activated. Further investigations focusing on phonology observed that speakers tend to rely more on segmental rather than tonal information when processing the phonology of Chinese characters. The next section focuses on sublexical processing of Chinese characters.

1.3 Chinese Radicals

Among the 3,500 most frequently used Chinese characters, there are 2,305 phono-semantic compounds (65.86%) (Hu et al., 2013). As indicated in Section 1.1, phono-semantic compounds are the most common formation method in modern Chinese. They can be further decomposed into two types of sublexical components: semantic radicals and phonetic radicals. There are four general spatial relationships of radicals: Left-right, e.g. 猜; Top-bottom, e.g. 艺; Closed outside-inside, e.g. 国; Open Outside-inside, e.g. 同.

It is argued that radicals should be treated as the basic level in reading Chinese, as letters are the basic level in alphabetic writing systems (Perfetti et al., 2005)

Semantic radicals usually provide a clue to the meaning or category of a single character. Semantic radicals are often the transformed or abbreviated version of a character, so they may not always provide phonological information (e.g. 水, /shuǐ/, "water" is transformed into "氵", which is less recognizable and has lost its pronunciation). Therefore, the study of semantic radicals mainly focuses on semantic and orthographic activation and usually involves four aspects: the semantic radical transparency of the whole character (whether the meaning or category of semantic radicals can provide a clue to the meaning of the whole character), combinability (how frequently a radical can form a character), surface frequency (the frequency of the character), and the position of the radicals.

Research into semantic radical transparency has attracted much attention. Feldman and Siok (1999) investigated whether semantic radicals can facilitate the processing of characters. A radical's transparency refers to whether the meaning of the radical is semantically related to the whole character. They selected primes that share the same semantic radical with the target and are semantically related (R^+S^+), share the same radical but are not semantically related (R^+S^-), share no radical but are semantically related (R^-S^+), and share no radical nor are semantically related to neutral controls (R^-S^-). The results suggested that when the meaning of a semantic radical is transparent in relation to the meaning of the whole character (R^+S^+), it facilitates processing; when it is opaque (R^+S^-), it inhibits processing.

Surface frequency also has an influence on word recognition when the radicals served as primes are identical to the target words. Ding et al. (2004) used radicals as primes to examine the influence on the processing speed of target characters with different frequencies. Facilitation effects were found with low-frequency target words but not with high frequency. This effect could be explained as mere orthographic similarity between the prime and the target; therefore, a follow-up experiment used a character that is orthographically similar to the radical in the target and found an interference effect, whereas the identical radical still provides facilitation effects. This means the radical is properly processed and facilitate the word processing of compound characters that contain it as a radical; however, when the prime is not identical to the target's radical, it interferes with the word recognition process. Thus, this facilitation effect is not due to orthographic similarity.

The position of semantic radicals can also have an impact on word processing. Feldman and Siok (1997) examined the position and combinability of semantic radicals in a lexical decision task. The stimuli consisted of characters with semantic radicals on the left and semantic radicals on the right. Each semantic radical was further divided into radicals with high-combinability and low-combinability. High-combinability semantic radicals occurred no less than 65 times in a 6,000-character corpus, and low-combinability occurred no more than 36 times. The effect of combinability was observed for semantic radicals on the left but not for semantic radicals on the right. That is, when semantic radicals are of high combinability, semantic radicals on the left are processed faster than semantic radicals on the right. When they are of low combinability, semantic radicals on the left are processed slower than semantic radicals on the right.

In a priming task, Su et al. (2012) found facilitation effects when the prime and the target shared the radical in the same position, whether it was the dominant (a radical that appears more frequently in this position) or subordinate position (a radical that appears less frequently in this position). However, this study mixed different types of phono-semantic compounds, which stimuli were less common to be seen in daily life. Specifically, characters with semantic radicals on the right are less common than characters with semantic radicals on the left, a difference which may benefit the processing of the latter. The finding could be item-specific and not generalizable enough.

It has been argued that semantic radicals play a more dominant role in word recognition than phonetic radicals. Wang et al. (2017) examined the semantic transparency of semantic radicals and the phonological regularity of phonetic radicals in a lexical decision task. The stimuli included characters that have semantic radicals that are transparent with the whole character's meanings and phonetic radicals that are identical to the pronunciation of the whole character (S^+P^+), characters that have transparent semantic radicals but phonetic radicals that are not consistent with the pronunciation of the whole character (S^+P^-), characters that have opaque semantic radicals but regular phonetic radicals (S^-P^+), and neutral control characters (S^-P^-). The results revealed that conditions with transparent semantic radicals were processed faster than those with opaque semantic radicals. Most importantly, there were no regularity effects observed for phonetic radicals. Thus, semantic radicals contribute more than phonetic radicals to word recognition in Chinese characters.

However, increasing attention is being paid to phonetic radicals and it has been argued that phonetic radicals play a more important role than semantic radicals in word reading (Hung et al., 2014). For example, Hung et al. (2014) concluded that phonetic radicals play a more dominant role in early lexical processing than semantic radicals. In a magnetoencephalography (MEG) study, the researchers manipulated the pronunciation of phonetic radicals (regularity) and meaning in relation to the whole character, as well as the orthography of semantic radicals (position and shape) and meaning in relation to the whole character. The homophone judgement task showed that phonetic radical can facilitate the processing of homophones and interfere with non-homophones. The synonym judgement task revealed that semantic radicals facilitate synonyms but interfere with non-synonyms. This could be due to higher combinability in semantic radicals compared to phonetic radicals (a larger pool of selection based on the same semantic radical might be less efficient in providing semantic or phonological information than a phonetic radical that leads to fewer possible combinations of characters). The MEG results showed early and robust effects of the repetition effect of phonetic radicals but not of semantic radicals. Thus, phonetic radicals dominate over semantic radicals in terms of word processing.

The processing advantage of phonetic radicals is often reported in the literature. In an apparent-motion task (where one of the sublexical components is displaced towards different locations. Participants were instructed to indicate which radical has moved). Wang (2006) found that task demands (a naming or lexical decision task) and recognition difficulty (low character frequency) benefit the processing of phonetic radicals, which might involve the cognitive necessity of focusing on the information provided by phonetic radicals. Seidenberg (1985) also found that phonetic radicals can facilitate the processing of compound characters that contain them as phonetic radicals only when the target is of low frequency, indicating that phonological information is more important in the recognition of low-frequency compound characters. A similar conclusion was drawn by Hue (1992) and Lee et al. (2005) who found that regularity effects only occur in low-frequency characters. Furthermore, they argued that phonetic radicals are not recognized as freestanding characters and the processing of phonetic radicals is due to their systematic occurrence within character recognition.

Further evidence from Xu et al. (1999) reported automatic activation of phonology by Chinese characters in a series of semantic relatedness judgement tasks. The prime and target words were not semantically related. They controlled both phonology (homophone and non-homophone) and orthography (orthographically similar or dissimilar), resulting in four distractors in total. However, they failed to find a strong and consistent homophony effect. An examination of their stimuli revealed that orthographically similar distractors share the same phonetic radical with target characters (for 114 out of 121 stimuli). Xu and colleagues also found a significant difference between orthographically dissimilar distractors with different homophony in relation to the target characters but not between orthographically similar distractors. That is, when there is an absence of an identical phonetic radical in two conditions, homophony would lead to slower reaction times. Semantic activation through phonology seems to be automatic in character reading and not restricted to low-frequency words, whereas the activation of the phonological information of phonetic radicals is not supported by the current results.

Y. D. Xu et al. (1999) did not distinguish between the difference in semantic and phonetic radicals in their stimuli. The results are not reliable in predicting semantic and phonological activation in phonetic radicals. Zhou and Marslen-Wilson (1999a) looked at phonological and semantic processing in phonetic radicals in a semantic and phonological relatedness judgement task. Three main types of primes were used: semantic, complex and control. Semantic primes are single characters that are semantically related to the targets. They can be used as phonetic radicals in complex primes, either homophonic or rhyming with complex primes. In Experiment 1, complex primes shared the same phonology as semantic primes; in Experiment 2, they were not homophonic. This additional condition examined whether the priming effects were due to phonological mediation. In terms of phonology, there were two types: regular, where semantic primes and complex primes share the same phonology (both segment and tone) and rhyming, where two primes are rhymed, starting with a different consonant. Three types of SOA (stimulus onset asynchrony) were also used to examine the time course of sublexical semantic activation: 57 ms, 100 ms and 200 ms. The data revealed that semantic activation of phonetic radicals occurred in the first two SOAs, whereas at 200 ms, this effect seemed to disappear or was suppressed. However, due to the strict selection criteria, all the stimuli were of low frequency; thus, sublexical semantic activation was applied to low frequency characters only. The phonological activation of phonetic radicals was also supported by the current study because even though complex primes do not share the same pronunciation as their phonetic radical in Experiment 2, facilitatory effects relative to unrelated control words remained significant for shorter SOAs (57 ms and 100 ms). In summary, phonetic radicals can activate phonological and semantic information when processing characters; however, this facilitatory effect is restricted to low-frequency characters and occurs at short SOAs.

Following Zhou and Marslen-Wilson's work, Lee et al. (2006) carried out a series of event-related potential (ERP) studies using similar ideas to investigate the role of phonetic radicals. The primes were similar: a semantically related, regular phonogram (with the same pronunciation), an irregular phonogram (with a different pronunciation) and a control. Significant N400 effects (a negative-going wave occurring about 400 ms after the onset of the stimuli, which is often related to semantic competition) were found for semantically related primes in all three SOAs: 50 ms, 100 ms and 300 ms. However, both N400 effects totally disappeared in a regular and irregular phonogram at an SOA of 300 ms, indicating that semantic activation of phonetic radicals is temporal and cannot be preserved after 300 ms.

To conclude, the sublexical level in studying Chinese word recognition provides a new perspective on how a character can be decomposed in the mental lexicon. Semantic radicals can facilitate word reading when their meaning is transparent to the target character and interfere with word reading when their meaning is opaque. Phonetic radicals can activate both semantic and phonological information in reading Chinese characters, but it is restricted to low-frequency words. More and more studies suggest that phonetic radicals play a more dominant role than semantic radicals in word recognition of Chinese characters.

The next section discusses the processing of pinyin, a romanization of Chinese characters.

1.4 Pinyin

In early primary school teaching, pinyin is used to teach children the pronunciation of Chinese characters. However, pinyin gradually disappears from textbooks during children's reading development since advanced learners are expected to know the pronunciation of Chinese characters. It is uncommon to see pinyin presented together with Chinese characters in written texts for proficient readers. According to a research report, 85.6% of users have chosen pinyin as the input method for Chinese characters on mobile phones (Huaon.com, 2021). Therefore, Chinese speakers in mainland China are very familiar with pinyin.

1.4.1 Pinyin knowledge and reading development

Research has shown that the acquisition of pinyin is essential in the early learning of the pronunciation of Chinese characters. As early as kindergarten, pinyin provides benefits for children learning Chinese characters. For example, Lin et al. (2010) tested different phonological awareness tasks (e.g. invented pinyin spelling, syllable deletion, phoneme deletion, etc) with 3^{rd} year kindergarten children (around 6 years old). It was found that good pinyin spelling ability is the strongest predictor of Chinese word reading development. Furthermore, Siok and Fletcher (2001) conducted a series of experiments on 1^{st}, 2^{nd}, 3^{rd}, and 5^{th} grade primary school students from mainland China. The task involves phonological awareness (spotting the incorrect spelling of pinyin and tonal differences, sound isolation, and sound-blending), visual processing skills (locating the correct pinyin word), orthographic processing skills (spotting the incorrect character with pinyin provided), testing both orthographic and phonological processing (homophone discrimination, pinyin reading), and reading skills (reading characters without the assistance of pinyin). Children in higher grade performed better in phonological awareness tests, except for the sound-blending test. This is probably because they received less practice of this kind in higher grades. Children in higher grades have less exposure to pinyin; thus, they are outperformed by children at lower grades. Visual processing skills also predict the success at early stages of reading development (as reflected in Grade 1 and Grade 2’s performance), whereas tone awareness and pinyin reading skills are more important at later stages of learning (as reflected in Grade 3 and Grade 5’s performance) because more homophones are introduced.

Because pinyin knowledge is important for the early development of reading Chinese characters, character recognition can also help the development of pinyin knowledge. Zhang et al. (2020) found a reciprocal relationship between pinyin knowledge and character recognition, because they can support each other's growth in the future. This finding is important because their data suggest that early exposure to characters benefits children's learning of pinyin. Because children, especially kindergarten children, may have more chance to be exposed to characters than pinyin in daily life. This exposure can, for example, be through characters printed on cups/bottles, visual words presented on TV/computers, and words presented on shop windows. In mainland China, the formal introduction of pinyin is at the beginning of first grade in primary education (Ministry of Education, 2011). This study suggests that kindergarten children should be encouraged to learn pinyin, as this benefits their reading development.

Many studies have examined how pinyin knowledge can predict early reading development. Pinyin provides information about the phonology of Chinese characters; however, it is unclear to what extent pinyin activates the orthography, phonology, and semantics of Chinese characters when reading pinyin words.

1.4.2 The role of pinyin in word recognition

Evidence that the orthographic form of Chinese characters can be activated while reading pinyin words comes from several studies. For example, Chen et al. (2014) conducted a series of experiments to examine the orthographic activation of Chinese characters in pinyin using a Stroop task to examine the automatic processing of pinyin, and a naming task to examine the semantic processing of pinyin. The stimuli included two-character pinyin primes to limit the activation of homophones. In the Stroop task, participants were first presented with a pinyin prime, followed by a target character that is orthographically similar or dissimilar to the pinyin prime when it is transcribed into a character. They were asked to judge the ink colour of the target character. In the naming task, they were also presented with a pinyin prime, then the target character. The task was to name the target character. In the Stroop task, there were no significant RT differences between target character with similar or dissimilar orthography to the primes. In the naming task, orthographically similar characters were named significantly slower than orthographically dissimilar characters. This means that the orthographic activation of Chinese characters in pinyin is task dependent. During the automatic processing of pinyin, orthographic information may not be activated or fully activated, whereas the controlled processing of pinyin would pre-activate orthographic information of the corresponding characters.

Chen, Perfetti and Leng (2019) further explored the orthographic activation of Chinese characters in pinyin with L2 learners of Chinese. They used the same modified Stroop paradigm as introduced in L. Chen et al. (2014). They did not find any priming effects between orthographically similar characters and control conditions in the intermediate learners of Chinese (who acquired 1200 words). However, an orthographical inhibitory effect was observed in high proficiency learners of Chinese (who acquired 5000 words), similar to what has been found in L. Chen et al. (2014) with native speakers of Chinese. These findings suggest that implicit orthographic activation of pinyin depends on the reading experience. Only for proficient learners and native speakers of Chinese does pinyin activate corresponding characters automatically and unintentionally.

There is evidence that not only does pinyin activate Chinese characters implicitly but also the sublexical components embedded in the characters. For example, Chen, Perfetti, Fang, et al. (2019) investigated whether pinyin can activate sublexical components of characters in native speakers of Chinese. They constructed prime-target conditions involving pinyin primes and target Chinese characters. In the O^+P^+ condition, the corresponding character of the pinyin prime shared the same phonetic radical with the target character and the prime and target had the same pronunciation (e.g. the prime zhū zi (珠子) shared the same phonetic radical and pronunciation with the target character 蛛, /zhū/). In the O^+P^- condition, the pinyin prime shared the same phonetic radical with the target character, but the pronunciation was different (e.g. the prime zhù zi (柱子) shared the same phonetic radical but a different pronunciation with the target character 往, /wǎng/). In the O^-P^- condition, the pinyin prime was different from the target character in terms of orthography and phonology (neutral control). The study involved the semantic judgement task. The results revealed an orthographic priming effect when comparing the O^+P^- and O^-P^- conditions, which indicated that phonetic radicals can be activated with the orthographic cues provided by pinyin. However, there was no significant priming effect in terms of the O^+P^+ and O^+P^- condition, which suggested that the phonological cues provided by phonetic radicals were not activated during pinyin processing. Chang (2018) also investigated how tonal information and the regularity of phonetic radicals (i.e. how consistently a phonetic radical provides the pronunciation of the target character) can influence Chinese character processing in L2 learners. Learners showed significantly better perception and production when the characters were presented with tone diacritics and when the phonetic radical corresponded to the pronunciation of the character.

Although the above studies suggest that a phonetic radical's phonological information cannot be activated via pinyin words, pinyin itself can activate Chinese phonology as shown by the study of Wang et al. (2015). The phonological processing of pinyin and character was investigated in their primed naming task to see how different those two writing systems are. The experiment involved four types of primes: S^+T^+ (prime and target share the same segment and tone with the target), S^+T^- (prime and target share the same segment but the tone of the prime is different from the target), S^-T^+ (prime and target share the same tone but the segment of the prime is different from the target), and S^-T^- (prime and target share no segment and tone with the target). When the prime and target were Chinese characters, a facilitation effect was observed only in the S^+T^+ condition (compared with the S^-T^- condition), but not for S^+T^- or S^-T^+ conditions. This suggests that the phonological information of a character is activated when reading characters aloud. However, when sharing only segmental or tonal information with the target, it is not sufficient to help with the naming process. When the Chinese character primes were replaced with pinyin words, significant facilitation effects were found with only S^+T^+ and S^+T^- conditions when the prime appeared 57 ms and 100 ms before the target character. However, when the SOA was set to 200 ms, all priming conditions showed significant facilitation effects. This means both segmental and suprasegmental information is activated by pinyin primes, and those two components were likely to be separately encoded because the $S^{-}T^{+}$ condition has the facilitation effect only at 200 ms. Tonal information requires more time to process than segmental information; thus, it is accessed later than segmental information.

Li et al. (2019) also investigated the activation of segmental and tonal information by pinyin and Chinese characters in a series of homophone judgement tasks. In their first experiment, stimuli involved characters only, and the results confirmed that both segmental and tonal information help the character recognition process, though tonal information has a weaker representation than segmental information. In their second experiment, pinyin and character primes were used. Interestingly, the same pattern was observed even though half of the stimuli had been replaced with pinyin words. Thus, the segmental and suprasegmental (tonal) information was activated by pinyin words.

To summarize, pinyin knowledge helps with learning to read Chinese characters, and character recognition also helps with the growth of pinyin knowledge. Tasks that involve explicit reading (i.e. the lexical decision task) have found that pinyin can activate the orthography, phonology, and semantics of Chinese characters; however, tasks that involve automatic processing (i.e. the Stroop task) did not find orthographic activation of Chinese characters (Chen et al., 2014). Further investigation on pinyin’s semantic and phonological activation in the automatic processing is needed.

1.5 Theories and Models of Chinese Word Recognition

Various models of Chinese word recognition have been proposed in the literature. These models generally consist of three basic levels of processing: orthography, phonology, and semantics. The main debate in the literature focused on whether the orthography can directly access the semantics (Wong & Chen, 1999) or whether the activation of semantics is mediated by phonology (Perfetti et al., 2005; Perfetti & Tan, 1998; Perfetti

& Zhang, 1995; Tan & Perfetti, 1997) and an interactive framework of both orthography and phonology (Zhou & Marslen-Wilson, 1999b). This debate was also crucial for the universality of models across different writing systems. Many studies argued that alphabetic and logographic languages activate different areas in the brain (Bolger et al., 2005; Siok et al., 2004; Tan et al., 2003). Perfetti and Tan (1998) suggested that phonology is a constituent of Chinese character recognition as in English. However, even within alphabetic languages like English, there was no consistent agreement (Wong & Chen, 1999). This could be due to the fact that orthography and phonology are often confounded in alphabetic languages like English (Shen & Forster, 1999). Logographic languages like Chinese have no confounding between orthography and phonology; thus, the Chinese language would be ideal to examine the contributions of orthography and phonology independently. For example, a character can be orthographically similar and homophonic character (e.g. "哉" is orthographically similar and homophonic to "栽"), an orthographically similar but non-homophonic character (e.g. "截"), an orthographically dissimilar but homophonic to the target character (e.g. "灾"), and an orthographically and phonologically dissimilar character (e.g. "纱"). In an eye-tracking experiment, Wong and Chen (1999) used the stimuli mentioned above to investigate the processing of orthography and phonology in Chinese. The results revealed reliable early activation of orthographic effects and late phonological effects. This favoured the direct access hypothesis that orthography can directly access the semantics without phonological mediation.

An indirect access hypothesis (i.e. semantic access mediated by phonology) was supported by the study of Perfetti and Zhang (1995). This study involved a semantic decision task and a phonological decision task. Participants were asked to decide whether two characters have a similar meaning or the same pronunciation. If meaning is accessed through phonology, then both semantic and phonological interference would occur. The results supported the indirect access hypothesis. Further examination of the time course of phonological and semantic activation using SOAs revealed that phonological activation occurred at an early stage, whereas semantic activation was at a later stage. Tan and Perfetti (1997) also found strong facilitatory effects when using semantic-associated primes and homophones of semantic-associated primes, which supported the indirect access hypothesis. However, this is restricted to words that have relatively few homophones (a detailed discussion of this study can be found in Section 1.2).

Tan and Perfetti's (1997) results were not successfully replicated. Zhou and Marslen-Wilson (1999b) used the same stimuli and procedure as Tan and Perfetti (1997) and set the SOA at 129 ms. Only semantic-associated primes were facilitated but not homophones of semantic-associated primes. Follow-up experiments revealed that homophones that contain regular and consistent phonetic radicals (i.e. the pronunciation of phonetic radical is identical to the whole character) resulted in facilitatory effects. Furthermore, they concluded that orthography can also directly activate semantics, because facilitatory effects were observed with orthographically similar primes and orthographically similar homophone primes. Thus, the authors proposed an interactive framework of phonological and orthographical factors that can both access to semantics.

Perfetti et al. (2005) proposed a model of Chinese word processing called the lexical constituency model. They argued that identifying words involves identifying the constituents in words, namely orthography, phonology, and semantics. Phonology is no longer considered as "an instrument to meaning, but rather as a constituent of a word that constrains the identification process" (p. 56). An illustration of the model is presented in Figure 1.1.

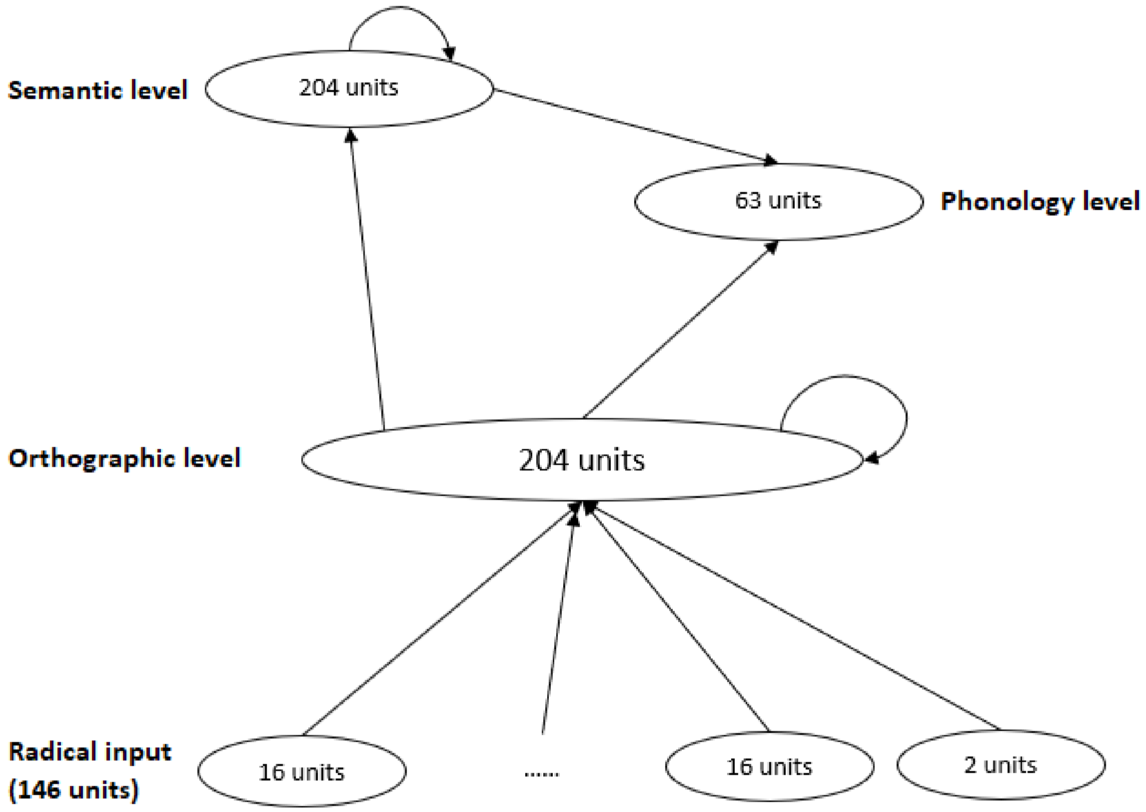


Figure 1.1. The lexical constituency model. Adapted from Perfetti, Liu, and Tan (2005).

A unique feature of this model that is distinct from the models of word processing in alphabetic languages (e.g. Rumelhart et al., 1986) is the basic graphic unit at the radical level. The input level contains radical slots (146 units) to represent the radicals (144 units) and spatial slots (2 units) to represent the composition of characters (there are generally four types of spatial relationships: left-right, top-bottom, closed outside-inside, and open outside-inside). Then, it is fully connected to the orthographic level that contains the representation of the graphic forms of characters.

The orthographic level is also connected to the semantic level, either with precise-meaning characters or vague-meaning characters. Each unit in the orthographic level and semantic level has three connections with the phonological level: onset, vowel, and tone. There are 23 onsets, 34 vowels, and 5 tones in the Chinese language. Unlike a cascade style (word-level phonology can be activated prior to the complete retrieval of letters) in alphabetic languages, the Chinese language is categorized as a threshold style, where each unit in all three levels has a threshold for sending output to another level. Although this model was specialized for the Chinese language, Perfetti et al. (2005) assumed that the difference between writing systems is not whether there are connections to phonology but what units are included. In an alphabetic language, the basic input level is the letter level, whereas in a logographic language like Chinese, the basic input level is the radical level.

Li et al. (2013) also pointed out the importance of segmental and suprasegmental (tonal) information in reading Chinese. They found that both segmental and suprasegmental information will be automatically activated in a Stroop task. More importantly, segmental information plays a more dominant role than suprasegmental information, which is consistent with what Wang et al. (2015) have found. Tonal information facilitated the processing speed only when the segments differed. In Perfetti and Tan's (2005) lexical constituency model, the phonological level includes three sub-levels: onsets, vowels, and tones. However, the activation of each sub-level unit is the same. As suggested by Li et al.'s (2013) study, segmental information should have stronger connections to other levels than suprasegmental information. The detailed distributions and connections within the phonological level can be improved in Perfetti and Tan's (2005) lexical constituency model.

In summary, this section introduced direct and indirect access hypotheses and the lexical constituency model that is based on the meaning-with-phonology hypothesis. Although alphabetic and non-alphabetic languages differed in terms of scripts and the connection between graphic units and phonological units, this should not fundamentally change the role of phonology. In the lexical constituency model, phonology is considered as one of the three constituents along with orthography and semantics. Alphabetic and non-alphabetic languages only differ in the representation of the basic level. The former represents the basic units at the letter level, the latter at the radical level. However, this model does not suggest the direct activation of semantics through radical input, as Yeh et al. (2017) confirmed that phonetic radicals can activate semantics directly. In addition, the above models/hypotheses target on the processing of Chinese characters. There is no model/hypothesis focus on the processing of pinyin words. They are often treated as part of Chinese phonology rather than an actual input unit, which cannot explain the results obtained from pinyin studies (i.e. L. Chen et al., 2014; L. Chen et al., 2019; Li et al., 2019).

Chapter 2: Introduction to the Stroop Task

Chapter 1 discusses Chinese word recognition with a focus on Chinese characters, radicals, and pinyin. Most studies discussed in Chapter 1 involve tasks that require participants to read words. To explore the automaticity of visual word recognition, tasks can be used that do not require word recognition but still involve reading. An example is the Stroop task, which requires people to name the ink colour of words or letter strings.

Section 2.1 provides an introduction to the Stroop task, discussing Stroop interference and facilitation effects that are key to the Stroop phenomenon. Section 2.2 introduces different theories and models of the Stroop task. Section 2.3 discusses multi-stage accounts of the Stroop effects. Section 2.4 discusses the differences between Stroop effects in Stroop tasks that involve vocal and manual responses (response modality effects). The use of non-alphabetic languages in the Stroop task is discussed in Section 2.5 because most Stroop findings in the literature are based on alphabetic languages. This section first discusses Stroop studies involving Chinese characters, radicals, and pinyin. Furthermore, this section also reviews Stroop effects in other non-alphabetic languages, such as Japanese and Arabic. The final section of this chapter (Section 2.6) presents the analysis methods used in this thesis and outlines the overall structure of the thesis.

2.1 Stroop interference and facilitation effects

The original Stroop task designed by J.R. Stroop in 1935 is different from the Stroop tasks used in more recent studies. In his first experiment, J.R. Stroop investigated word naming with different colour words. The participants were asked to read a series of colour words that were printed in incompatible colours (e.g. RED printed in green colour) and to read colour words printed in black ink as the control condition. In the second experiment, colour naming was the focus of interest. The participants were required to name aloud the ink colour of colour words printed in incompatible colours and to name the colour of solid colour squares. The results of the first experiment revealed no significant word naming latency differences between incompatible colour names and colour names in black ink. However, when the task was to name the ink colour of words, there was a significant difference between incompatible colour names and solid colour squares. Specifically, naming the ink colour of incompatible colour names took a longer time than naming the ink colour of solid colour squares. Stroop concluded that word stimuli have stronger associations with reading than colour stimuli to naming. In other words, word reading is more efficient than colour naming, which results in interference when naming the ink colour of colour words that does not match with the ink colour.

Unlike early studies in the Stroop task, recent studies no longer measure the processing time of a list of words. Instead, the processing speed of each individual stimulus is recorded for a more analytical methodology. J.R. Stroop's original study focused on the Stroop interference effect of incompatible colour words. If incompatible colour words can slow down the processing speed, then it is natural to think that compatible colour words would facilitate the processing speed. Dalrymple-Alford and Budayr (1966) were the first to include congruent colour word conditions to investigate the Stroop facilitation effects, although they did not find a faster response time of congruent trials compared to incongruent trials. However, when using individual stimuli instead of a list, Sichel and Chandler (1969) found it is faster to name congruent words than incongruent words.

The Stroop interference effects have been constantly reported by many studies (Dyer, 1973a; Glaser & Glaser, 1982; Klein, 1964; Roelofs, 2012; Spinks et al., 2000). However, the Stroop facilitation effects have been found to be less stable and often referred to as a "fragile" effect (Logan & Zbrodoff, 1998; Macleod, 1991; MacLeod & MacDonald, 2000a).

The strength of the facilitation effect will be affected by the lexicality of the neutral condition, whether a row of Xs or colour-unrelated words is used. Dyer (1973b) and Regan (1978) found significant facilitation effects when using Xs as a neutral control, whereas Dalrymple-Alford (1972) observed facilitation effects only when compared to colour-unrelated word control, but not to Xs. Cohen et al. (1990) and MacLeod (1998) considered facilitation effects to be weak and smaller than interference effects. Roelofs (2012) found no facilitation effects using Xs as a neutral control. Other studies have reported negative facilitation effects when using Xs, which means Xs are responded to faster than congruent stimuli (Logan & Zbrodoff, 1998; Nealis, 1973; Schulz, 1979; Sichel & Chandler, 1969; Vanayan, 1993).

When word and colour information are presented simultaneously, using Xs as a neutral control seems to produce less or even no facilitation effect. In order to examine the time course of Stroop effects, Glaser and Glaser (1982) manipulated the presentation of word and colour information in a Stroop task. To separate the word and colour information for a stimulus onset asynchrony (SOA) setting, they used a separate mode where the target word was surrounded by a vertical rectangle presented in different colours. They found no facilitation effect when the word and colour were presented at the same time (0 ms SOA). However, when the word was presented earlier than the colour rectangle (-400 ms, -300 ms or -100 ms SOAs) or presented very late (400 ms SOA), significant facilitation effects were observed. This means word information can be processed within 100 to 400 ms, with 300 ms as the optimum time for eliciting facilitation effects. Roelofs (2010b) conducted a Stroop task with separated colour and word information in different SOAs using Dutch words. Similar to Glaser and Glaser's results, there were no facilitation effects in 0 ms and 200 ms SOAs, but a significant facilitation effect was found in -400 ms and -200 ms SOAs. A series of manual Stroop tasks using SOAs conducted by Coderre et al. (2013)[1], where Chinese participants were tested in their native language (Chinese), significant facilitation effects also occurred at -400 ms and -200 ms. The findings using SOAs suggested that when word information is presented earlier than colour information, the meaning of words can be processed sufficiently even within 100

[1] The control condition used in this study was a percent sign (%) rather than a row of Xs. This is to match the physical size of a Chinese character.

ms; thus, the colour information presented later would not interfere with the word recognition process.

Different neutral control conditions would have an impact on the Stroop facilitation effects obtained. While using Xs as a neutral control has produced less stable results, using colour-unrelated word controls has produced much more consistent facilitation effects in many studies (Augustinova & Ferrand, 2012; Brown, 2011; Dalrymple-Alford, 1972; Duncanjohnson & Kopell, 1980; Redding & Gerjets, 1977; Spinks et al., 2000; Yeh et al., 2017). The difference between nonword neutral signs (Xs) and neutral words was further explored by Brown (2011) and explained as the lexicality cost. This is because neutral words are not purely neutral in a Stroop task. The process of reading neutral words, in fact, also competes with naming the ink colour of neutral words; thus, an interference might take place. This lexicality cost is calculated when measuring Stroop interference (neutral signs vs. incongruent words) but deleted when calculating Stroop facilitation based on neutral signs (e.g. Xs). This is why research often reported Stroop facilitation as unreliable and unstable since neutral signs were used as the baseline rather than neutral words.

What Brown defined as the "lexicality cost" was later referred to as task conflict in other literature. Task conflict refers to the conflict between the demands of the task and those of a possible alternative task. In a Stroop paradigm, the task demands the individual to name/match the ink colour of the word, but reading of the word is also activated; thus producing task conflict. Kalanthroff and Henik (2013) argued that task conflict also delays responses to congruent trials, but with proper executive control, it will not become slower than responses to neutral trials. In their study, Kalanthroff and Henik split participants into six equal groups with different inhibitory control levels, and the results revealed that individuals with low inhibitory control had no or even reversed facilitation effects, whereas the group with high inhibitory control elicited large facilitation effects. That is, the ability of executive control is highly correlated with the performance of Stroop facilitation. Good executive control would produce stable and consistent Stroop facilitation effects.

To summarize, strong interference effects are strong and stable, whereas Stroop facilitation effects depend on the selection of a neutral control baseline. It is unclear how Stroop facilitation effects are different from Stroop interference effects. The next section talks about the theories and models of the Stroop task: how Stroop interference and facilitation effects might take place.

2.2 Theories and models of the Stroop task

There are different theories about why Stroop effects occur. The theory of the relative speed of processing explains that the speed of naming words is quicker than the speed of naming colours, thus resulting in competition between these two processes (Cattell, 1886; Fraisse, 1969; Klein, 1964; Morton & Chambers, 1973). However, the relative speed of processing fails to explain the SOA studies. Glaser and Glaser (1982) used the stimulus-onset asynchrony (SOA) design in a Stroop task. The presentation order of word and colour information can be manipulated. They found that even though the colour information is presented 400 ms earlier than the word information, there is no interference effect. If colour information is processed slower than word information, the early presentation of colour information would interfere with the word naming process. This means that the relative processing time of the two dimensions was not the sole cause of Stroop interference effects.

Another prevalent theory on the Stroop task is the automaticity account. The view of automaticity enhances the role of attention in the Stroop effect with the processing of colours requiring more attention than the processing of words. Reading words is automatic and obligatory but naming the colours is not (Logan, 1978; Shiffrin & Schneider, 1977). Macleod (1991) pointed out that the automaticity view cannot explain the differences between colour and word information presented in an integrated or separated manner because the automaticity view would predict similar outcomes of both presentation manners. Kahneman and Chajczyk (1983) also showed that the location of colour information can also influence the Stroop interference effects. Even though word information is far away from the main focus of attention (colour identification), the interference occurs. Besides, Cohen et al. (1990) also argued that the automaticity account is a dichotomy view to classify a dimension as automatic or controlled. The reason why this is problematic is that the target dimension could be replaced, and such a dichotomy view would produce contradictory results. For example, instead of naming colours, MacLeod and Dunbar (1988) trained the participants to associate shapes with colour names. When the task is to name the shapes, strong interference and facilitation effects were observed. But those effects disappeared when the task is to name the ink colour of the shapes. This result contradicted the dichotomy view because colour naming would be instead considered as controlled as it is faster than shape naming. Therefore, Cohen et al. (1990) argued that different tasks have different degrees of automaticity. It is better to view the process as a continuum rather than a dichotomy.

Cohen et al. (1990) integrated and modified some ideas based on the theories discussed above and developed the parallel distributed processing model (PDP). Instead of focusing on the speed of processing, the term "strength" was used; automaticity of different stimuli (words or colours) is not an all-or-none view but one with a gradient. The automaticity of a process is determined by the strength of processing, which can be affected by practice and attention. Figure 2.1 is the network architecture adapted from Cohen et al. (1990).

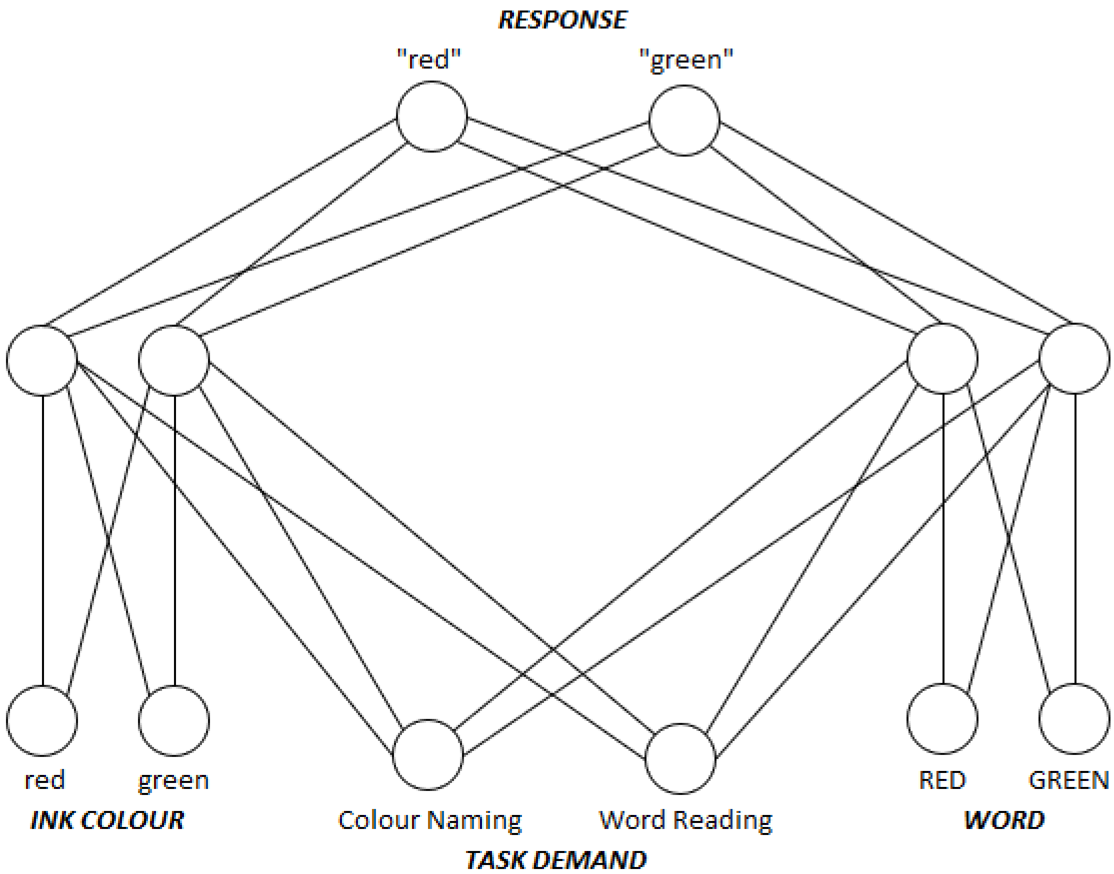


Figure 2.1. Network architecture. Adapted from Cohen et al. (1990).

There are three pathways in this framework: from the bottom to the top, there are input units, intermediate units, and output units. All the input units are connected with the intermediate units, then project to the output units. Sufficient activation to exceed the threshold is required when a stimulus is activated at the input level, gradually passing to the intermediate level before making a response at the output level. At the centre of the framework, two task command units are employed to allocate attention to the intermediate units: one used for colour-naming, and another for word reading. When the task is to name the ink colour of words, the colour-naming units in the task command will be activated, and vice versa. The model also explains the mechanism of Stroop interference and facilitation. If two pathways are active but differ in the intermediate units (e.g. ink colour RED and word GREEN), then interference happens. If the two pathways are activated (e.g. ink colour RED and word RED) in a similar way, facilitation occurs. This view reflects that different processes are considered as competing pathways with different strengths. One process can be seen as automatic in one context and controlled in another. Furthermore, a process does not need to be completed before transitioning into the next units. Although the model is successful in simulating Stroop interference, Stroop facilitation, response-set membership and so on, it cannot account for the results obtained from SOAs, where words are presented earlier than the colour information (Levelt et al., 1999). What is more, when the two pathways have a comparable strength of processing, the model predicts that there is no interference, which is not what has been found by MacLeod and Dunbar (1988).

Cohen et al.'s (1990) PDP model explains that the Stroop interference effects are due to the competition between colour and word information and the Stroop facilitation effects occur because the two dimensions have similar patterns of activation. As discussed in Section 2.1, Stroop facilitation effects raised more questions about why the effects are relatively small and unstable compared to Stroop interference effects. The inadvertent reading hypothesis argues that participants may lack attention at certain stages and forget about the task's goal (Kane & Engle, 2003; MacLeod & MacDonald, 2000b). This lack of attention can be captured when making errors in naming the incongruent conditions because they treat it as a word-reading task instead of a colour-naming task; however, the inadvertent reading can lead to Stroop facilitation effects because it is challenging to discern whether they are making responses to the colour or the word. Roelofs (2010a) argued that inadvertent reading is not the cause of Stroop facilitation effects. He designed a series of Stroop tasks that present colour words in Dutch and English. If Stroop facilitation is due to inadvertent reading, then Stroop facilitation should be observed within a language but not between languages. This is because reading errors can be filtered when presenting colour words in a different language (MacLeod & MacDonald, 2000a). The results of Roelofs (2010a) showed that Stroop facilitation can be obtained both within- and between-languages. This indicates that Stroop facilitation stems from converging information of colour and word, which is called the converging information hypothesis (Cohen et al., 1990; Roelofs, 2010a). Furthermore, Roelofs provided a distributional analysis of quantile plots and showed that Stroop facilitation was present across the quantiles. According to the inadvertent reading hypothesis, Roelofs argued that reading errors should not be present in all congruent trials; otherwise, the Stroop facilitation effects would be much larger than usually observed.

Roelofs (2010a) used the WEAVER++ model to interpret the Stroop interference and facilitation effects (Levelt et al., 1999; Roelofs, 1992, 1997, 2003). The WEAVER++ model includes four stages of processing (see Figure 2.2). In the Stroop task, the perceived colour information enters the conceptual identification stage so that the concept is selected for the perceived colour. Then in the lemma retrieval stage, the concept is used to retrieve a lemma from memory. Lemma is the canonical form in the mental lexicon that derives words into different forms. This process is believed to be statistically driven, as more activated lemma would be favoured than less activated lemma. The next stage is word-form encoding, which consists of morphological, phonological, and phonetic encoding. When the morphological and phonological encoding is completed, a successful articulation of a word requires the articulatory gestures to be prepared, whether it is glottal, nasal, or oral. The perceived word information enters route A and route B in parallel; that is, it activates both lemma retrieval and word-form encoding. When the word information via routes A and B does not correspond to the perceived colour, the interference occurs. When both word and colour information activate similar patterns of processing, the facilitation occurs. Roelofs (2003) also explained the word selection as a goal-oriented process. In the Stroop task, it involves the functioning of word planning and executive control. Word planning has task-dependent production rules (e.g. naming the colour) and task-independent production rules (e.g. lexical selection).

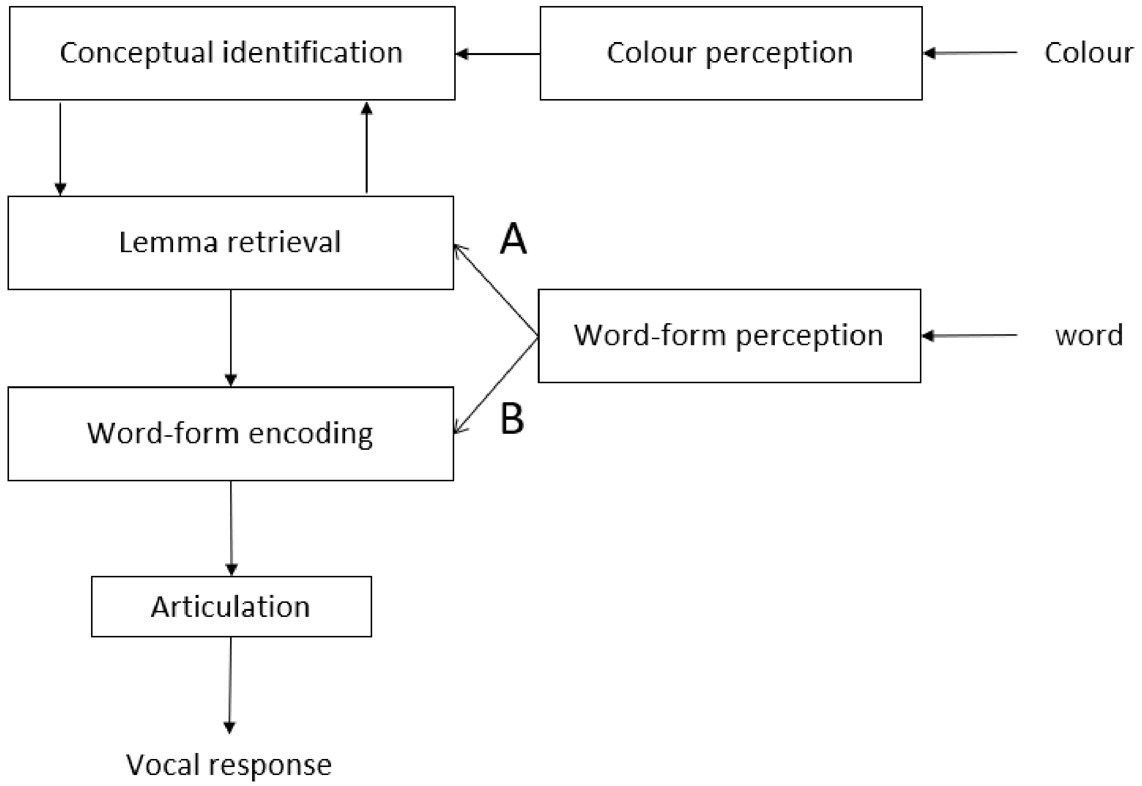


Figure 2.2. The WEAVER++ model for Stroop effects. Adapted from Roelofs (2010a).

Another connectionist model, the Proactive control/task conflict (PC-TC) model, was developed by Kalanthroff et al. (2018). This model is based on the dual-mechanism control (DMC) framework proposed by Braver (2012). The model can deal with the variability in reverse Stroop facilitation (congruent trials responded slower than neutral trials, Kalanthroff & Henik, 2013) through the manipulation of working memory load, percentage of nonword trials, and cues (indicating the type of upcoming stimuli). They argue that this variability in negative facilitation can be attributed to the execution of proactive task control, which varies in ability from individual to individual and from trial to trial for the same individual. Hence, the PC-TC model (see Figure 2.3) was built to account for the variability in negative facilitation. The model has two input layers consisting of colour features (e.g. ink blue) and lexical features (e.g. word blue). These feature nodes activate representations in the response layer. The connections between layers are all excitatory and connections within each layer are inhibitory. For example, the word BLUE would compete with the word GREEN within the lexical features layer.

The input layers are also connected to the task demand layer for colour-naming or word-reading tasks. The task demand layer is modulated by proactive control, a top-down system that regulates processing within the model. Connections between the task demand and input layers are bidirectional. When the proactive control is strong, it resolves the competition within the task layer. When the proactive control is weak, competition occurs within the task layer, resulting in task conflicts. Reactive control will be activated to resolve the task conflict enabling the task-relevant one. Kalanthroff et al. (2018) acknowledged that this model can be more complete by explaining how different variables affect the level of proactive control, including the loop from conflict detection to task control, and how to use task conflict to modulate proactive control.

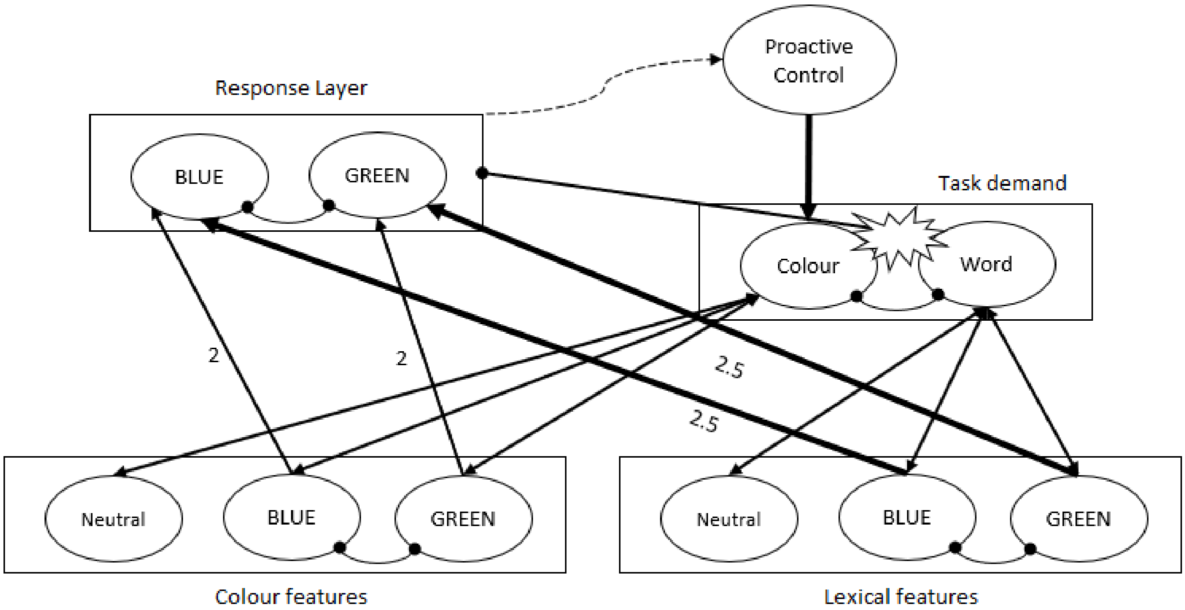


Figure 2.3. Illustration of PC-TC model. Adapted from Kalanthroff et al. (2018).

2.3 Multi-stage accounts of the Stroop effects

Traditionally, Stroop effects (including both Stroop interference and facilitation) are considered to result from a single source. Recent studies using the Stroop task typically select incongruent, congruent, and neutral trials to measure Stroop effects (e.g.

Augustinova et al., 2019; Roelofs, 2012; Spinks et al., 2000). However, more evidence has pointed out that there are multiple sources of Stroop effects, making it possible to decompose Stroop effects into distinct components.

An early version of decomposing Stroop effects was put forward by Klein (1964). They used six types of stimuli to investigate the effects of words in relation to colours: (A) nonsense-syllables (e.g. hjh, evgjc, bhdr, gsxrq), (B) rare English words (e.g. sol, helot, eft, abjure), (C) common English words that have no relation to colour names (e.g. put, heart, take, friend), (D) colour-associated words (e.g. lemon, grass, fire, sky), (E) rare colour words (e.g. tan, purple, grey, black), and (F) common colour words (e.g. yellow, green, red, blue). Klein assumed that nonsense-syllables would provoke more interference than real words; common words more than rare words; colour-associated words more than colour-unrelated words; and colour words more than colour-associated words. Results revealed a gradient increase in RT from Type A to Type F, indicating that the magnitude of interference increases according to its relation to the colours. Statistical analysis showed significant differences between nonsense-syllables and common words, between rare words and common words, between rare words and colour-associated words, and between common words and rare colour words. All other comparisons did not reach significance. Klein's finding was later referred to in the literature as the "semantic gradient effect" (Levin & Tzelgov, 2016; Sharma & McKenna, 1998). Sharma and McKenna (1998) argued that there are four possible components in Stroop effects: being lexical (letters vs. neutral words), semantic relatedness (neutral words vs. colour-associated), semantic relevance (colour-associated vs. colour words [nonresponse set]), and response set membership (colour words [nonresponse set] vs. incongruent colour words). Klein (1964) and Sharma & McKenna's (1998) studies are comparable because they used similar types of stimuli. What is different is that Sharma & McKenna found all four components to be significant in the vocal Stroop task. This confirmed that Stroop effects contain four components rather than being considered as a whole. Furthermore, they suggested that the response set membership effect was due to interference at the late response selection stage rather than in the semantic or lexical systems. De Houwer (2003) explained Stroop effects in terms of stimulus-response compatibility (SRC) and stimulus-stimulus compatibility (SSC). The former assumes the effect takes place at the response selection stage, while the latter suggests the impact on the stimulus-encoding stage. Experiments in De Houwer involved presenting three words on each trial, asking participants to press one of the two keys depending on the identity of the middle word. There were three types of trials: identity trials (all three words are identical, e.g. *blue-blue-blue*), same-response trials (the middle word is different from the flanker words but in the same response set, e.g. *purple-blue-purple*), and different-response trials (the middle word differs from the flanker words but in a different response set, e.g. *green-blue-green*). The results confirmed that Stroop effects can be explained by SSC because RTs for identity trials were shorter than same-response trials. The evidence for SRC in Stroop effects was also found because same-response trials were responded to faster than different-response trials. This is because SRC effects are based on short-term associations. In this study, participants had to learn the arbitrary association of two colour words to make correct responses. Schmidt and Cheesman (2005) replicated De Houwer's study and elaborated that Stroop effects involve both semantic and response competition. They added colour-associated words with identity, same-response, and different-response trials. Colour word distractors revealed the difference between same-response and different-response trials and between identity and same-response trials. This supported De Houwer's finding that the Stroop task involves both semantic interference and response competition. However, the results only showed a difference between identity and same-response trials in colour-associated words, which suggested that the influence of colour-associated words was at the semantic level but not at the response competition level.

De Houwer (2003) and Schmidt & Cheesman's (2005) manipulation on stimuli and response compatibility was specific to the manual Stroop task. There are more general views on the decomposition of Stroop effects that is compatible with other response modalities. Goldfarb and Henik (2007) argued that there are two types of conflicts in Stroop effects: informational and task conflict.

Informational conflict refers to the mismatch between word and colour information in incongruent trials (e.g. blue presented in red colour). Task conflict occurs when one is more familiar with one task than the other. For example, a literate person tends to be more familiar with the word naming task than the colour-naming task. When the task is to name the ink colour of the word, they must overcome the automatic processing of words, which gives rise to interference. They also suggested that congruent conditions might be more conflicting than neutral conditions. To examine this hypothesis, they created an experimental environment where the operation of a control mechanism is reduced by using a high proportion of neutral trials (strings of letters). As indicated before, neutral strings/signs would not induce word reading; thus, task conflict would not occur or be reduced. Another manipulation they utilized to invoke task conflict was the cuing technique. A cuing technique indicating whether a word would appear or not might reduce the use of the control mechanism. The results showed a reversed facilitation effect (neutral conditions faster than congruent conditions) in the non-cued conditions and no facilitation effect in the cued condition. Task conflict is the cause of such a reversed facilitation effect because the non-cued condition was designed to have a high-conflict situation compared to the low-conflict situation in the cued condition. To verify that the findings were not due to the design using neutral signs that produced such effects, they used neutral words in a consequent experiment that should encourage the word reading task. Under the high task conflict control situation, the reversed facilitation effect disappeared in the non-cued condition. In sum, they found information conflict (neutral condition responded slower than congruent condition) and task conflict (non-cued condition found reversed facilitation effect) in Stroop effects.

Levin and Tzelgov (2016) pointed out that informational and task conflicts have different components. Information conflict can be broken down into indirect information conflict (interference caused by colour-associated words, e.g. SKY) and direct informational conflict (interference caused by the colour word itself, e.g. BLUE). Task conflict contains orthographic components (e.g. geometric shapes like a circle or triangle) and lexical components (e.g. a string of letters, ssss). To investigate these four components, different conditions in the Stroop task were contrasted. The direct informational conflict component was measured by comparing colour words with all other stimuli. The indirect informational conflict component was calculated by contrasting colour-associated words with neutral words. The orthographic component was measured by comparing shapes with readable stimuli. The lexical component was calculated by contrasting letter strings with real words. In a series of experiments with native Hebrew speakers, Levin and Tzelgov found stable direct informational conflict and orthographic conflict, whereas indirect informational conflict and lexical conflict were less stable and small.

Entel and Tzelgov (2018) investigated whether the proportion of congruent trials impacts the magnitude of information and task conflict. They used three conditions with a different proportion of congruent and neutral trials: mostly congruent (90% of trials were congruent), mostly neutral (90% of trials were neutral), and equal proportion (an equal proportion of both). A larger facilitation effect was observed in the mostly congruent condition, whereas smaller facilitation effects were found in the mostly neutral and equal proportion conditions. In their study, they defined task conflict as the RT differences between colour words (congruent and incongruent trials) and neutral signs, and informational conflict as the RT differences between incongruent and congruent trials (see Augustinova et al., 2019; Augustinova et al., 2018 for a different definition of task conflict). Thus, they concluded that with the absence of informational conflicts (no incongruent trials), task conflict was not detected when most trials were congruent, and control was not recruited. When incongruent trials were introduced during the practice session, task conflict was observed, as revealed by a negative facilitation effect, and control was recruited. This suggested that with different task demands, it is possible to adjust the control executed.

Following Entel and Tzelgov's (2018) research, Shichel and Tzelgov (2018) argued that informational conflict can be further decomposed into semantic conflict and response conflict. Thus, there are three unique contributions to Stroop effects: task conflict, semantic conflict, and response conflict. To investigate this, they adapted the two-to-one paradigm (two colours assigned to one response key) used by De Houwer (2003). There were four conditions in this design: illegible neutral stimuli (e.g. coloured rectangles), identical (congruent), and two incongruent conditions: same response trials (the word and colour correspond to the same response) and different response trials (the two dimensions indicate different responses). They also manipulated the proportion of neutral conditions within participants, either in low (25%) or high (75%) conditions. They defined these conflicts as follows. Task conflict measures the RT differences between illegible neutrals and colour words. Semantic conflict reflects the RT differences between congruent and incongruent stimuli. Response conflict is defined as the RT differences between same and different response trials. The findings of their experiments revealed significant contributions of all three conflict components to Stroop effects. Importantly, only task conflict was significantly affected by the different proportions of neutral trials: with a higher proportion of neutral trials, task conflict was stronger. This is consistent with what has been found by Entel and Tzelgov (2018).

From Klein's (1964) "semantic gradience effect" to Shichel & Tzelgov's (2018) "three conflict components", more and more literature provides its interpretation on how to decompose the components in the Stroop effects. Augustinova and Ferrand (2014) proposed the semantic Stroop paradigm to investigate the components in the Stroop effects. They distinguished semantic and response conflict using the colour associate words; however, the definition of semantic and response conflict is different from Shichel & Tzelgov's (2018). They argued that semantic conflict should be defined as the RT differences between colour-neutral words and colour-associated words because colour-associated words trigger the meaning associated with colour words and subsequently cause conflict, whereas neutral words have no semantic relatedness to colour words;

thus, they do not lead to semantic conflict. Response conflict was defined as the RT difference between colour-associated words and colour-incongruent words because colour-associated words do not activate motor responses. Crucially, colour-associated words are not in the response set because participants are required to respond to the ink colour of words rather than the words themselves, either by saying the colour or pressing the corresponding button for the colour. Augustinova et al. (2018) and Augustinova et al. (2019) further expanded their semantic Stroop paradigm by adding the third conflict component – task conflict. Different from the definition of Entel & Tzelgov's (2018), task conflict measures the RT differences between neutral strings and neutral words. Compared to neutral strings, neutral words can trigger word reading; thus, they produce task conflict.

The definition of task conflicts is consistent in the literature. However, there are two different interpretations of how to measure semantic and response conflict. In a review by Parris et al. (2022), they justified why semantic and response conflicts should be measured by referencing the colour-associated words instead of response sets. First, semantic conflict should not be measured by the RT difference between congruent trials and same-response incongruent trials. This is because congruent trials and same-response incongruent trials may involve response facilitation, as the two dimensions of the stimuli could lead to the same response (e.g. the word and colour information are compatible, see Hasshim and Parris, 2014, for more details). Second, by the same logic, response conflict measured by the RT differences between same-response and different response incongruent trials would also involve response facilitation. Parris et al. (2022) also defined the response set effect when using non-response set stimuli (e.g. *pink* not used as a response). The RT differences between neutral words and non-response set stimuli are semantic conflict (non-response set effect), and the RT differences between non-response set stimuli and incongruent stimuli as response conflict (response set effect). With those conflict components (and other components not fully listed here), Parris et al. (2022) presented a complete diagram (see Figure 2.4), considering both conflict and facilitation components in the Stroop effects that involve three main parts: information conflict, informational facilitation, and task conflict. They refer to semantic, phonological and response conflicts together as informational conflict, and semantic, phonological and response facilitation as informational facilitation, whereas task conflict becomes a broader term including lexical conflict, orthographic conflict, and negative facilitation effect. Lexical conflict measures the RT difference between repeated letter strings and neutral words, reported as task conflict by Augustinova et al. (2019). Orthographic conflict was reported by Levin and Tzelgov (2016), where repeated letter strings took longer to name than colour shapes. The negative facilitation effect has been discussed in Entel and Tzelgov's (2018) work, where they manipulated the proportion of congruent trials and created a high task conflict condition for congruent trials. Under such a condition, congruent colour information no longer facilitates processing; instead, it takes longer to name congruent trials than repeated letter strings. In sum, Parris et al. (2022) provided a thorough summary of the decomposed components mentioned in the Stroop literature. It is not a complete list of all possible conflict/facilitation components in the Stroop effects, but it is a useful way to combine different conflict/facilitation components into one paradigm. For example, different definition of semantic conflict, response conflict, and task conflict are summarized in Figure 2.4. Each type of stimuli and their effect relative to other types of stimuli can be clearly seen in this figure.

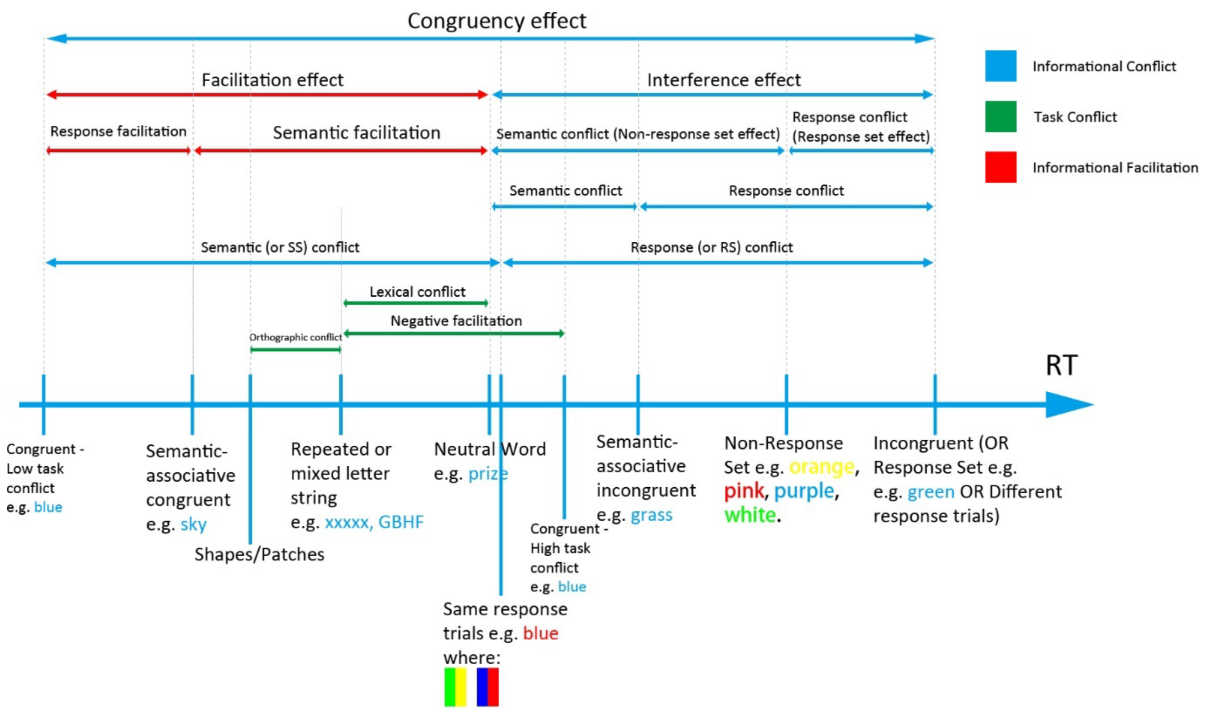


Figure 2.4. The decomposition of Stroop interference and facilitation. Adapted from Parris et al. (2022).

2.4 Response modality effects

The view that Stroop effects can be decomposed into different components helps with understanding the impact of response modality on the Stroop task. The vocal Stroop task is the traditional method first used in Stroop's experiments (Stroop, 1935). The manual version of the Stroop task was compared to the vocal Stroop task for the first time by White (1969). The manual Stroop task requires participants to press a key corresponding to ink colour. Thus, in contrast to the vocal Stroop task, the manual Stroop task does not require an overt verbal response. It was found that the magnitude of Stroop effects was reduced when changing from vocal to manual responses (Neill, 1977; Redding & Gerjets, 1977; White, 1969). This was referred to in the literature as the response modality effect.

Kinoshita et al. (2017) argued that these two response modes of the Stroop task may involve different mechanisms in contributing to the Stroop effects. The vocal Stroop task is essentially a naming task, whereas the manual Stroop task is a classification task. The modality effects can be understood in a broader discussion of Stimulus-Response Compatibility (SRC) (Kornblum, 1992). In a vocal Stroop task, the stimulus can be, for instance, the colour red and its response is "red" in a vocal response. In a manual Stroop task, the response is arbitrarily associated with a "colour red" key press rather than directly responding to the colour name. That is, the manual responses require an extra layer of response by associating the ink colour of words with the corresponding key press.

The response modality effect might be attributed to the differences in the activation of phonological information in the two modes. In the manual Stroop task, colour information triggers the corresponding colour button, with less competition occurring when processing the ink colour and word information in the vocal responses. The view of "less competition in the manual responses" was also strengthened by studies that observed an increase in the facilitation effect in the manual Stroop task (Neill, 1977; Redding & Gerjets, 1977). Naming words in congruent conditions may serve to create more competition when deciding whether to process ink colour or word information first, whereas in the manual responses, this mode selection disappears, and the focus is more on the processing of the ink colour. While the literature suggested a qualitative difference between vocal and manual responses, Parris et al. (2019) found facilitation effects of phonological overlap in both vocal and manual responses, and thus, argued for a quantitative difference between the two response modalities. In a review of the differences between different response modalities, Parris et al. (2022) suggested that the dissociative pattern of significant semantic conflict but reduced response conflict could be due to the indirect measure of response conflict in semantic conflict. That is, response conflict may still be present in the manual responses, but this conflict was measured in semantic conflict.

The response modality effect has also been investigated in relation to the different components in the Stroop effect. For example, Sharma and McKenna (1998) used vocal and manual responses in the Stroop task to investigate the effects of lexical, semantic relatedness, semantic relevance, and response set. The lexical component refers to the RT difference between nonsense syllables and neutral words, which is also called task conflict. The semantic relatedness component measures the RT difference between colour-associated words and neutral words, referred to as semantic conflict in other literature. The semantic relevance component is the RT difference between colour-associated words and colour words that are not included in the response set. The response set membership component refers to the RT difference between colour words that are not included in the response set and incongruent colour words that are included in the response set. It was found that different components behave differently in the two response modalities. In the vocal Stroop task, all four components contribute to the Stroop interference effects. In the manual Stroop task, only response set membership (incongruent words vs. colour words [non-response set]) contributed significantly to the Stroop interference effects. In addition, although the overall Stroop interference effects were larger in the vocal than in the manual responses, the response set membership effect did not differ between the two modalities. They interpreted this result in terms of the response set membership effect taking place at a late response selection stage rather than in the semantic or lexical systems. They concluded that the effects observed in the manual responses take place at the response selection stage.

Risko, Schmidt, and Besner's (2006) findings, however, were inconsistent with those of Sharma & McKenna's (1998). Risko et al. (2006) obtained response set membership effects with colour-associated words (colour-associated words related to a colour in the response set vs. colour-associated words related to a colour not in the response set) in both vocal and manual responses. Therefore, they argued that the manual responses also have access to semantics as the vocal responses. Augustinova et al. (2019) examined the three distinct components in the Stroop task using vocal and manual response modalities. They found task, semantic, and response conflicts contributed to Stroop interference in the vocal responses; however, only semantic and response conflicts contributed to Stroop interference in the manual response. A key finding in Augustinova et al.'s (2019) study was that they also looked at the decomposed facilitation components in Stroop facilitation. The definition of each facilitation component mirrored that of conflict components in Stroop interference. There are two facilitation components: semantic and response facilitation. Semantic facilitation measured the RT differences between neutral words and congruent colour-associated words (e.g. SKY in blue colour). Likewise, the RT differences between congruent colour-associated words and congruent colour words was response facilitation. The contribution of each component in different response modalities also differed in Stroop facilitation. In the vocal responses, both semantic and response facilitation contributed significantly to Stroop facilitation, whereas in the manual responses, only semantic facilitation contributed to Stroop facilitation. This was discussed in terms of different stages of processing in different conflict/facilitation components. In Stroop interference, semantic conflict was unaffected by response modality, whereas response conflict contributed significantly more to the vocal responses than to the manual responses. This means semantic effects did not take place at the response level; otherwise, semantic and response conflicts would have consistent performance in either response modality. Similarly, in the Stroop facilitation, response facilitation was affected by response modality, whereas semantic facilitation was not. This again confirmed that semantic effects did not occur at the response level in the Stroop facilitation.

2.5 Stroop effects with non-alphabetic words

In Chapter 1, the findings on Chinese word recognition were discussed using the lexical decision, naming, homophone judgement, and semantic categorization tasks. In this section, studies using the Stroop task will be discussed that focused on Chinese characters, radicals, and pinyin.

Spinks et al. (2000) investigated the activation of phonology when reading Chinese characters. Unlike many alphabetic languages and Japanese, the Chinese language has four tones. Characters with the same segment but different tones may have totally unrelated meaning and orthography. This brings another possibility in homophone stimuli: different-tone and same-tone homophones. For example, the colour word red "红" is pronounced as /hóng/. The same-tone homophone can be "洪", which means flood or big. The different-tone homophone can be "轰" (/hōng/), which means boom. In this study, same-tone homophones, different-tone homophones, colour-associated words, colour words, neutral control words, and colour patches were included. Compared to neutral words, Spinks et al. (2000) found significant Stroop facilitation effects for colour words, same-tone homophones, different-tone homophones, and colour-associated words. For Stroop facilitation effects relative to colour patches, only colour words reached significance. These results can be explained in terms of task conflicts because neutral words involve word reading whereas colour patches do not (Augustinova et al., 2018). Colour patches were responded to faster than colour neutral words; thus, the RT differences measured against colour patches are much smaller when using different neutral control stimuli. Relative to neutral words, strong Stroop interference effects were observed for colour words, same-tone homophones, and colour-associated words. Relative to colour patches, the Stroop interference effects became larger. However, no Stroop interference was found with different-tone homophones, either comparing to neutral words or to colour patches. Thus, different-tone homophones were processed similarly to neutral words and colour patches. Even though they shared the same segment (e.g. /hong/) with the colour words, the tonal differences in different-tone homophones can lead to the activation of words other than colour words. Same-tone homophones produced robust Stroop interference effects, which suggests that the activation of phonology in the Stroop task is automatic. The provision of tonal information enhanced the magnitude of Stroop interference effects, as fewer potential homophones were activated in same-tone homophones than different-tone homophones.

Guo et al. (2005) further explored the impact of age and reading ability on Stroop effects using Chinese characters. Similar to Spinks et al.'s (2000) design, they used materials from Spinks et al. (2000), including same-tone homophones, colour words, and neutral control words. However, different-tone homophones and colour patches were not included in this study. To investigate the age differences, they recruited three age groups: undergraduate students, sixth-grade primary school students, and third-grade primary school students. They also recruited children with high and low reading abilities, which were judged by non-verbal IQ scores and reading performance. For RT analysis, they only provided undergraduates' performance and found significant Stroop interference and facilitation effects when compared to neutral control words. The RT differences between homophones and neutral words were also significant. Further analyses of the impact of age and reading ability were conducted using error rates as in other studies (Deng & Wei, 2002; Shu et al., 2000). In terms of error rates, they found stronger phonological facilitation in third-grade pupils than in sixth-grade and undergraduates and stronger phonological interference in younger groups than in older groups. Furthermore, children with lower reading ability showed stronger phonological facilitation and interference effects than children with higher reading ability. Reading proficiency affects the activation of phonology in Chinese characters. Instead of more skilled learners yielding more phonological activation, the study showed children with lower reading proficiency yielded more phonological activation. Overall, the study confirmed that phonological activation is automatic during the processing of Chinese characters.

Segmental information can affect the magnitude of Stroop interference and facilitation effects. The role of suprasegmental information (tones) was also investigated. Li et al. (2013) used colour characters, same-tone homophones (S^+T^+), and different-tone homophones (S^+T^-), neutral characters (S^-T^-) to study the activation of segmental and tonal information in the Stroop task. They also included a condition to further examine the tonal information processing: a character with the same tone as the colour character but with a different segment (S^-T^+). Strong facilitation effects were found in all conditions, which is identical to what Spinks et al. (2000) found. Most importantly, the facilitation effect observed in S^-T^+ condition indicated that tonal information has an independent role from the segmental information. When sharing the same segment, the tonal information did not affect the processing speed (S^+T^+ vs. S^-T^+). Tonal information is useful only when the segments are different (S^-T^+ vs. S^-T^-). A limitation of this study was that it did not include incongruent conditions for S^+T^+, S^+T^-, and S^-T^+. Spinks et al. (2000) found Stroop interference effects for S^+T^+ but not for S^+T^-. This, in fact, argued that both segmental and tonal information are essential for processing Chinese characters.

Spinks et al. (2000), Guo et al. (2005), and Li et al.'s (2013) experiments were conducted using the vocal Stroop task. As discussed in Section 2.4, Stroop effects in the vocal Stroop task are usually larger than in the manual Stroop task. This is mainly because of the verbal activation of the vocal responses. If a verbal response is not required, the phonological conflict/facilitation effects are weakened. Therefore, it would be worthwhile to look at the Stroop effects in the manual responses.

Wang et al. (2010) conducted an event-related potential (ERP) study using the manual Stroop task to investigate the activation of semantic and phonological information in Chinese characters. Colour words, colour-associated words, same-tone homophones, neutral words and colour patches were used as materials. The behavioural data showed strong Stroop interference effects with colour words, homophones, and colour-associated words. For the ERP data, they investigated the negativity at around 400 to 500 ms (N450) and the late positivity at about 600 to 700 ms. N450 was suggested as a reliable index of semantic conflict between colour words and responses to colours, whereas late positivity may closely relate to the processing of meaning at later stages (Wang et al., 2010). The findings revealed N450 effects for incongruent colour words and colour-associated words, but not for homophones. A late positivity effect was found for incongruent colour words and homophones, but not for colour-associated words. Thus, they argued that homophones may not activate the meaning of the colour words with the same pronunciation at the early stage but at the later stage when the additional processing of the meaning occurred.

Sun et al. (2016) also investigated the activation of phonology and semantics in children using the manual Stroop task; unlike Spinks et al.'s (2000) and Guo et al.'s (2005), Sun et al. (2016) concluded that phonology may not have an important role in semantic activation in children. Participants were primary school students with a mean age of 9.9 years. Incongruent colour characters, incongruent homophones of the colour characters, neutral words were included in this study. There were four ink colours that children were required to respond to. A classical Stroop interference effect (incongruent colour words vs. neutral) was found to be significant, whereas incongruent homophones did not differ from neutral words. Thus, they concluded that phonology may not play an essential role in Chinese word recognition for children. Sun et al. (2016) and Wang et al.'s (2010) studies using manual responses failed to find homophone interference effect in Chinese character processing. Thus, it does not seem to play an important role in Chinese word recognition. However, this conclusion is based on Stroop experiments with a manual response only, where verbal responses are absent in this response modality. Importantly, the manual Stroop task may not be sensitive enough to detect the phonological activation.

Many studies have explored the processing of Chinese characters using the Stroop task, but only a few studies have investigated the role of sublexical components of Chinese characters in the Stroop task. This could be because the selection of stimuli is rather difficult: a colour character must be a radical that is used in a compound character. For example, characters like blue ("蓝") cannot be used as a radical, so the colour blue would not be used in a Stroop task studying the sublexical components of Chinese characters.

There is a variant of the Stroop task that uses non-colour stimuli. Luo et al. (2014) used a spatial Stroop task to study whether the radicals were activated during Chinese character reading. A spatial Stroop task is achieved first by selecting directional characters (e.g. 上, "up"), then used as radicals of compound characters (e.g. 忐, "nervous" that contain the character "up"). There were no neutral conditions, so Stroop effects were measured by subtracting incongruent conditions from the congruent conditions. It was found that both directional characters and characters that contain directional words as radicals showed significant Stroop effects. Although Stroop effects were observed, the selection of stimuli might be problematic because there are limited possible combinations of directional characters for "up" and "down", and the compound characters that contained directional characters as radicals are orthographically similar and often used as compounds to express nervousness: "忐" and "忑".

Considering Luo et al.'s (2014) results, they might be less generalizable to Chinese character recognition, Luo et al. (2015) designed a new Stroop task employing colour-related radicals. The simple characters they used were black ("黑") and white ("白"). They can be combined into compound characters in different positions. For example, they can be located on the left side of the character (e.g. "默" or "皈"), on the right (e.g. "伯" or "拍"), or on the top (e.g. "墨"). Some of the compound characters are of low frequency and others are of high frequency. The data revealed significant Stroop effects for low-frequency compound characters but not for high-frequency compound characters. In addition, the results were obtained with four complex characters (two for high-frequency, and two for low-frequency), which may not be generalizable to Chinese character recognition. As discussed in Section 1.3, the position of the radicals could impact the processing speed, with certain positions prevailing over other positions during word reading. In Luo et al.'s (2015) study, colour-related radicals were located either on the left, right, or other less common positions. If the position of radicals was not controlled, it also affects which type of radical is investigated. For example, the compound character "伯" used colour-related radicals as phonetic radicals, whereas "皈" used it as semantic radicals. It would be difficult to know whether it is semantic or phonetic radical that posed an impact on the sublexical processing of Chinese characters.

Thus, it is better to manipulate the type of radicals (semantic or phonetic) and their position when studying sublexical processing in Chinese characters.

Yeh et al. (2017) investigated the role of phonetic radicals in processing Chinese characters in the Stroop task. Usually, phonetic radicals are considered as carrying phonological information, whereas semantic information is handled by semantic radicals. In Yeh et al.'s study, colour words can be used as phonetic radicals in compound characters, so the activation of semantic information can be investigated. Because phonetic radicals may also carry phonological cues to the whole characters, there includes a Valid-Radical condition in which the phonetic radical contained in this character shares the same pronunciation as the whole character (e.g. colour cyan "青" is pronounced as /qīng/, when used in Valid-Radical "清", the same pronunciation remains), and the Invalid-Radical condition where the phonetic radical does not have the same pronunciation as the whole character (e.g. colour cyan "青" is pronounced as /qīng/, when used in the Invalid-Radical "猜", the pronunciation changed to /caī/). Figure 2.5 provided an illustration of the radical stimuli.

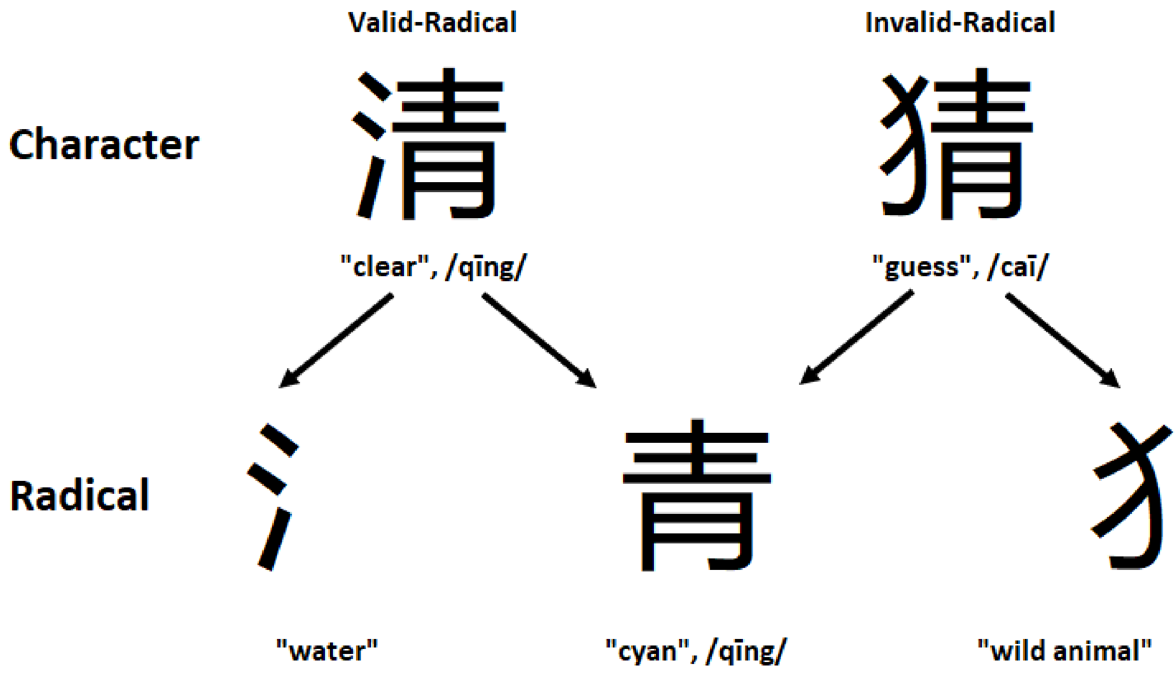


Figure 2.5. An illustration of radical stimuli used in Yeh et al.'s (2017) study.

Neutral controls were included in this study, so that Stroop interference and facilitation effects could be calculated. Strong interference effects were found in the Colour-Character, Valid- and Invalid-Radical conditions. However, the Valid-Radical condition did not differ significantly from the Invalid-Radical condition in terms of Stroop interference effects. Strong facilitation effects were observed in the Colour-Character and Invalid-Radical conditions only. These results indicate that phonetic radicals embedded in the character can still activate semantics associated with the radical and produce significant Stroop interference effects in Valid- and Invalid-Radical conditions. Surprisingly, the Stroop interference effects observed for the Valid- and Invalid-Radical conditions did not differ from each other, which suggests that the phonological information of phonetic radicals is not activated in the Stroop task. In other tasks, however, the role of phonology provided by phonetic radicals is confirmed, though restricted to low-frequency characters (Hue, 1992; Lee et al., 2005; Seidenberg, 1985). A more detailed exploration of the role of phonetic radicals will be addressed in Chapter 3.

Pinyin provides the phonological information of Chinese characters but does not provide the orthography of Chinese characters directly. The recognition of Chinese characters through pinyin could be mediated by phonology. If so, Stroop effects would still be present using pinyin colour words as its phonology would activate the corresponding colour words. As far as I am aware there are only two studies that included pinyin words in a Stroop task. Chen et al.'s (2014) work used a modified Stroop task that mixed the priming task and the Stroop task, and the orthographic information of pinyin is not activated in the Stroop task. Liu and Weng (2007) looked at the processing of Chinese characters, pinyin, and English words in the Stroop task. Furthermore, they investigated how Stroop effects would change with more practice. The study included only incongruent trials, and colour patches were used as a control condition to calculate the Stroop interference effects. The results showed that the rate of interference gradually decreased from the first to the fourth block, except English stimuli received more interference in the last block. In block 1 and

2, each stimulus was significantly different from each other, with characters invoking more Stroop interference effects than Pinyin and Pinyin invoking more than English. In block 3, the interference effects of English stimuli were significantly less than the other two, but there was no difference between characters and pinyin. In block 4, three types of stimuli were not significantly different from each other. Thus, with more practice, Stroop interference effects became similar for different types of stimuli including both alphabetic and logographic languages. Although this study involves a mixed presentation of different stimuli, it suggests that pinyin words can result in similar Stroop interference effects as characters. Liu and Weng's study featured the Stroop interference effects only. The Stroop facilitation effects may also be explored using pinyin to compare the performance of Chinese characters.

Overall, the majority of studies using the Stroop task to investigate Chinese language processing used Chinese characters, which elicited strong Stroop interference and facilitation effects similar to alphabetic languages (Saalbach & Stern, 2004). Recently, studies have focused on sublexical components of the Chinese language and have produced promising results that suggest sublexical components also aid in the processing of the whole character (Yeh et al., 2017). A less explored field is the study of pinyin processing. This unique alphabetic transcription of Chinese has been shown to activate a character's orthography, semantics, and phonology in tasks that require the explicit reading (i.e. the lexical decision task), as well as in tasks that do not require the explicit reading of words (i.e. the Stroop task).

After discussing different patterns of Stroop effects in Chinese characters, radicals, and pinyin, the next section will discuss other non-alphabetic languages (using Japanese and Arabic as examples) and their automatic processing in the Stroop task.

Sumiya and Healy (2004) investigated the Stroop effects comparing English and Japanese languages in Japanese-English bilinguals. They used Katakana and Hiragana in Japanese to create similar and dissimilar colour words from the English. Katakana is used for loan words from other languages and is a syllabic transcription into Japanese. For example, the colour word “green” can be phonologically transcribed into Katakana as グリーン (gu-ri-n), whereas Hiragana is normally used to write colour terms (e.g. green is みどり[mi-do-ri]). Participants received one of the two types of Japanese script (Katakana or Hiragana) first before the English stimuli, so that the performance of Japanese kana would not be affected by English. The results showed that Katakana stimuli were responded to more slowly than Hiragana stimuli. English stimuli were responded to more slowly compared to Hiragana, but not compared to Katakana. Most importantly, the Stroop interference effect was larger in Katakana stimuli than in Hiragana stimuli. Katakana and Hiragana share no orthographic similarity with English (completely different scripts), but Katakana has more phonological similarity with English since it is a transcription of English. These results showed that there is unintentional phonological processing in Japanese.

Another writing system used for Japanese is called kanji. Kanji is adopted from Chinese characters, so it is a logographic language system in Japanese. Thus, both a phonetic script (kana) and a logographic script (kanji) exist in Japanese. Kanji normally has two different readings: *on* reading (音読) and *kun* reading (訓読). *On* reading is the “sound” reading that is based on the Chinese sound, and *kun* reading is the “meaning” reading that is based on native Japanese terms (Iwasaki, 2002). Coderre et al. (2008) examined the Stroop effects with words written in kanji and kana to investigate the effects of scripts in the same language using functional magnetic resonance imaging (fMRI). The Stroop experiment involved kana (hiragana) and kanji words. The behavioural data revealed that there was no significant difference between kana and kanji in the Stroop interference effects (incongruent vs. neutral words). The neuroimaging data showed significant activation in the left inferior parietal lobe for Stroop effects with kana words, and significant activation in the left inferior frontal gyrus using kanji words. This pattern of result is consistent with the findings that the difference between those two areas of activation is the difference between alphabetic and logographic languages (Bolger et al., 2005; Siok et al., 2004; Tan et al., 2003). Although the behavioural data did not show the difference between the phonetic (e.g. kana) and logographic (e.g. kanji) writing systems, fMRI results suggest that those two scripts elicit different activation in the brain.

In addition to Stroop interference and facilitation effects, studies have also investigated different conflict components in Japanese. Non-alphabetic languages like Chinese and Japanese have abundant homophones, so it is easier to investigate phonological components by using homophone conditions in Japanese, while the studies of alphabetic languages have to use pseudo-homophones to achieve a similar effect. For example, Levitt et al. (2015) investigated the Stroop effects with kanji homophones. These homophones share the same pronunciation with kanji colour words, but they have a different meaning and orthography (e.g. the colour word "white" in kanji is 白 /shi-ro/, the kanji homophone 城 is also pronounced as /shi-ro/. Four types of stimuli were used: kanji colour words, kanji homophones, hiragana and neutral sign (X). The experiment revealed significant Stroop interference effects with kanji colour words and hiragana, but no Stroop interference effects with kanji homophones. Significant Stroop facilitation effects were found only with kanji colour words. The absence of strong Stroop effects using kanji homophones suggests that the activation of the phonology of kanji might not be automatic. Interestingly, the syllabic transcription of kanji – hiragana resulted in significant Stroop interference effects. Thus, a script that enables direct access to phonology leads to Stroop interference effects.

Another way to activate phonology directly in Japanese could be the use of Romaji. Romaji is the alphabetic transcription of kana. For example, the colour word "red" is 赤 in kanji, あか in hiragana, and /aka/ in Romaji. Yoshihara et al. (2021) used the Romaji transcribed from hiragana that contains an initial phoneme which corresponds to the initial phoneme of colour words in Japanese (e.g. the neutral word "head" is pronounced as /a-ta-ma/ in Japanese, which shares the same initial phoneme with the colour word "red" that is pronounced as /a-ka/). Significant Stroop effects (congruent vs. incongruent trials) were found, where congruent trials were responded to faster than incongruent trials using Romaji transcribed from hiragana. The match of initial phoneme to the colour words helped the processing of word recognition but hindered when there is a mismatch of initial phoneme to the colour words. When they used the same Romaji and transcribed it into katakana (e.g. /a-ta-ma/ matches the initial phoneme of the colour word "red", and its katakana is written as アタマ), the phoneme-based effect disappeared. This suggests that orthographic information can have an impact on the Stroop effects. This study looked at the Stroop effects only. It would have been useful if the study included a neutral control condition, so that Stroop interference and facilitation effects could be examined.

To summarize, Japanese contains different types of scripts, including Katakana, Hiragana, Kanji, and Romaji. Kana represents the alphabetic writing system, and Kanji represents the logographic system, which elicits different activation in the brain (Sumiya & Healy, 2004). Another non-alphabetic language that I am going to introduce is Arabic. There are two forms of Arabic: the literary form called Fusha, often used in literature and official documents, and the spoken form called Ammia, which consists of local dialects (Asaad & Eviatar, 2013). This is similar to the distinction of Kanji and Kana in Japanese, where the latter (Kana) was borrowed and developed from the former (Kanji) (Iwasaki, 2002). The complexity of Arabic orthography comes from the letters' visual and phonological neighbours. Many letters are similar or even identical in structure and can only be distinguished by the presence, location, and number of dots. For example, the letter /θ/ (ت) and /t/ (ث) share the same structure but differ in the number of dots on the top. Compared to Chinese characters and Japanese Kanji, the recognition of Arabic letters seems to rely even more heavily on the trivial differences in orthography. Due to this characteristic of the Arabic language, Asaad and Eviatar (2013) examined the effects of orthographic complexity of Arabic in a vocal Stroop task. Three types of cards were presented: (1) "Control" card, neutral words written in different ink colours; (2) "Classic

Stroop" card, colour words written in different colours other than themselves, i.e. incongruent colour words; (3) "Distorted Stroop" card, the same colour words written in different colours other than themselves, but the shape was distorted. For example, the colour red (احمر) can be written in incorrect forms (احم ر, اح مر, اح م ر). Participants were from the first grade, third grade, fifth grade, and undergraduate students. The results showed significant Stroop interference effects in all age groups and in both classic and distorted stimuli conditions. Stronger Stroop interference effects were found in classic than distorted stimuli except for the first graders. This means that from third grade on, Arabic speakers can process the Arabic words in an automatic way. The distortion in letters would affect their recognition of words and thus cause less Stroop interference effects than the normally presented colour words.

Different languages have different characteristics in word recognition. When studying word recognition in a particular language/script, those characteristics should be taken into consideration. For example, Pinyin and Romaji emphasize more on the phonological activation of words, while Chinese characters, Japanese Kanji, and Arabic focus on the orthographic representation of words/letters.

The next section will focus on the analysis methods used in this thesis and it will present an outline of the thesis.

2.6 Current Thesis

2.6.1 Analysis methods used in the thesis

In this section, I will discuss the approaches used in this thesis to analyse the data obtained in the Stroop experiments. The section consists of two parts: the mean reaction time (RT) analysis and the distributional RT analysis. A mean RT analysis is the most common way to analyse response time data. I will present a review of the traditional analysis of variance (ANOVA) and linear mixed-effects modelling. Furthermore, this section will discuss the distributional analysis, which helps to capture the changes in the magnitude of effects across the whole range of reaction times.

2.6.1.1 Mean RT analysis

In psycholinguistics, analysis of variance (ANOVA) has traditionally been used to analyse reaction times (e.g. De Houwer, 2003; Goldfarb & Henik, 2007; Hutchison, 2011; Schmidt & Besner, 2008; Wu et al., 2012). Data are typically analysed in terms of subject (F_1) and items (F_2).

Psycholinguistic experiments are conducted using a small group of participants; however, the researchers' interest is in terms of how generalizable a study is. This is also true for items selected for a particular experiment because they are generally randomly sampled from a larger pool of items. The language materials used in an experiment are a sample, not an exhaustive list, because conducting an experiment with the complete list is practically impossible to implement (Baayen et al., 2008). When significant results are found in both F_1 and F_2 analysis, it meets the $F_1 \times F_2$ criterion that the subject and item are random samples that can be applied to a larger population (Raaijmakers et al., 1999). However, this criterion does not come without flaws. Locker et al. (2007) pointed out that either F_1 or F_2 analysis ignores systematic variability that is attributed to individual subjects or items. Baayen et al. (2008) argued that the traditional interpretation of F_1 and F_2 analysis that the effects are significant in both F_1 and F_2 and thus applied to all subjects and all items is incorrect. Individual subjects and items may diverge dramatically from the average means. This is where mixed-effect models are introduced to the field of psychology. Mixed-effects models treat both subjects and items as crossed, independent, random effects. They require no prior averaging of data, so it is possible to take other potentially relevant covariates into consideration (Baayen et al., 2008). In addition, covariates to the subjects and items can also be considered in mixed-effect models.

Section 2.5 discussed Yeh et al.'s (2017) study which investigated the role of phonetic radicals. Yeh et al. (2017) analysed their Stroop data using linear mixed-effect models (LMMs). First, they started their linear effects modelling with a full model:

> lmer(RT ~ congruency * character type + (1 + congruency + character type + con gruency:character type | subject) + (1 + congruency + character type + congruen cy:character type | pair) + (1 + congruency + character type + congruency:charac ter type | colour))

A full model involves the full fixed structure (main effects and interactions) as well as a full random structure (random intercepts and slopes for the fixed factors and their interactions). A full model is dependent on the design of the experiment. Certain fixed factors added to the random structure would not make sense and the dataset would not converge on such a complex model; therefore, they need to reduce the model until it converges successfully:

> lmer(RT ~ congruency * character type + (1 + congruency | subject) + (1 | pair) + (1 | colour))

Then, a Likelihood-Ratio Test was conducted to investigate whether removing the random slope "congruency" from the model affected the model's goodness-of-fit. The simplified model turned out not to be significantly different from the MMP-Model, so the simplified model was used. In Yeh et al.'s (2017) study, the neutral-control condition was the reference condition for Stroop interference (neutral vs. incongruent) and facilitation effects (neutral vs. congruent); therefore *p*-values were corrected using Dunnett's methods (Dunnett, 1955). Exploratory comparisons were adjusted by Holm-Bonferroni Correction (Holm, 1979).

However, there are some issues with Yeh et al.'s (2017) analysis method. Firstly, they averaged the reaction times across the two blocks before running LMMs, which led to fewer observations per participant per condition. Secondly, the LMM included "subject", “pair” and “colour” as random effects, which differ considerably in terms of the levels of each factor: 28, 27, and 3. It has been suggested that random effects in LMMs with less than 5 levels should be avoided (Bolker, 2015; Bolker et al., 2009). Thirdly, Yeh et al. (2017) analysed the data using raw RTs in LMMs. Shapiro-Wilk’s test of normality showed that the data were not normally distributed ($p < 0.001$). It is not recommended to use skewed data in LMMs because it could distort the outcome as it affected the estimate of the mean (Lo & Andrews, 2015). Therefore, it is better to transform the raw RTs to meet the Gaussian assumptions of normality before running LMMs (e.g. log transformation or inverse RT). Instead of distorting the data by transforming the raw RTs, Lo and Andrews (2015) suggested using generalized linear mixed-effect models (GLMMs) that allow raw RTs to be assessed, because GLMMs allow the researcher to select a distribution (e.g. Gamma or Inverse Gaussian) that fits the raw RTs better than a normal distribution (Gaussian) used in LMMs.

Based on the above issues, a re-analysis of Yeh et al.’s (2017) data was conducted. The dataset was reorganized, and data were not averaged across blocks. After trimming the data based on the original criteria (RTs smaller than 200 ms and larger than 1200 ms were discarded), the best model for the data was determined by comparing different linear mixed effects models. First, to test which type of distribution fits best with Yeh et al.’s (2017) data, three models were used:

(1) Gaussian: lmer(RT ~ (1|subject))

(2) Gamma: glmer(RT ~ (1|subject), family = Gamma(link = "identity"))

(3) Inverse Gaussian: glmer(RT ~ (1|subject), family = inverse.gaussian(link = "identity"))

The models will be compared in R (4.2.2) using “compare_performance” function provided by the package “Performance” (0.10.2) (Lüdecke et al., 2021). The “compare_performance” function provides performance scores for each model based on

AIC (Akaike's Information Criterion, [Akaike, 1973]), AICc (second-order variant of AIC, [Hurvich and Tsai, 1989]), BIC (Bayesian Information Criterion, [Akaike, 1978a,b; Schwarz, 1978]) indices. All indices are normalized and the mean value of all indices in each model was calculated for the performance score. Results showed that the Inverse Gaussian distribution has the highest performance scores among the three models. Figure 2.6 shows the comparison of model performance indices. Larger values indicate better model performance, and vice versa. It is obvious that the Inverse Gaussian model outperforms other models in most cases.

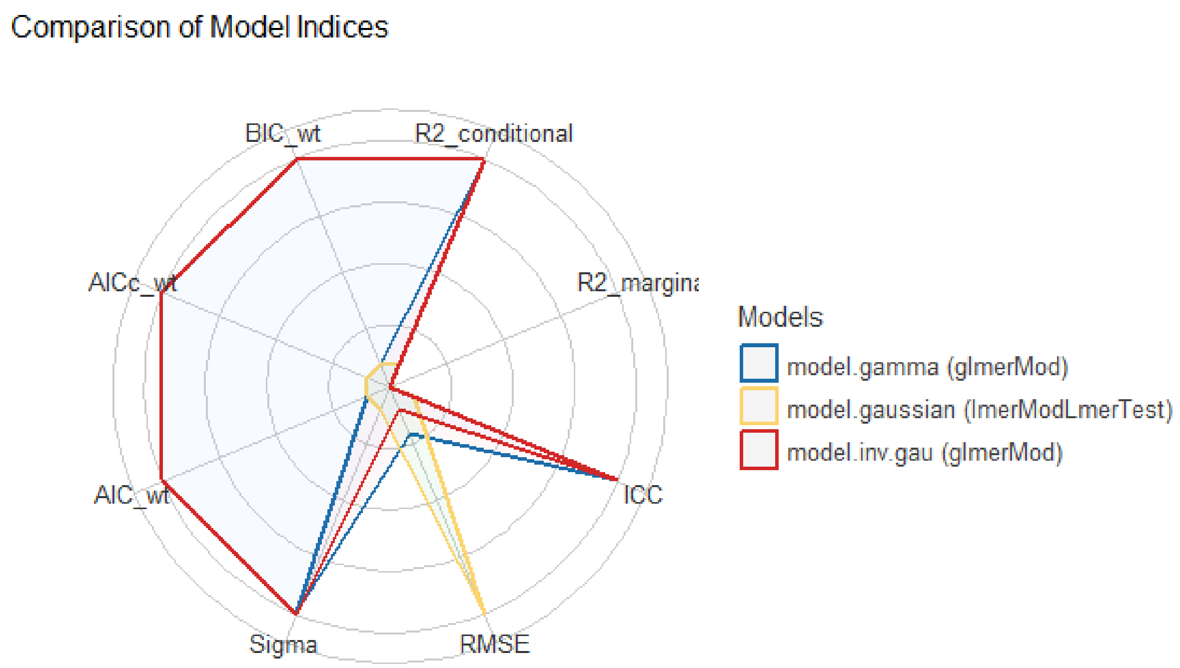


Figure 2.6. Comparison of model indices for different distribution types.

After choosing the Inverse Gaussian distribution for Yeh et al.'s (2017) data, fixed and random effects should be decided. Table 2.1 listed five different combinations of fixed and random effects. As suggested by Burnham (2002), to find the best model to use is not to find the "true model". It is to find the approximating model that can best make inferences from the data. The candidate models provided in Table 2.1 are not an exhaustive list of models. These are the possible models that have been used in the literature and a final model will be selected if it fits the data best.

Model No.	Model Structure
1	RT ~ congruency * character type + (1|subject) + (1|pair) + (1|colour)
2	RT ~ congruency * character type + (1|subject) + (1|item)
3	RT ~ congruency * character type + (1|subject)
4	RT ~ congruency * character type + colour + (1|subject) + (1|item)
5	RT ~ congruency * character type + colour + (1|subject)

Table 2.1. Models with different combinations of fixed and random effects. Factor "congruency" and "character type" used simple coding; factor "colour" used deviation coding.

Model 1 is identical to that of Yeh et al. (2017). Model 2 is nested within Model 1. They share the same fixed effects but have different random effects. Model 3 is nested within Models 1 and 2, considering only "subject" as the random effect. Model 4 has "colour" as a covariant because it only has three levels, and should be avoided as a random factor (Bolker, 2015; Bolker et al., 2009). Model 5 is nested within Model 4, considering only "subject" as the random effect. Each model was fitted with the Inverse Gaussian distribution as it fits Yeh et al.'s (2017) data best. The models will be compared again using "compare_performance" function. Table 2.2 indicates that Model 4 achieves better performance scores than the other four models. The performance score is a result of several information criteria (i.e. AIC, AICc, BIC). They indicate the quantitative difference between the models and should not be interpreted as qualitative differences (i.e. significant) (Burnham, 2002). Model selection is not about selecting a model that is "significantly" better than the other model. The "performance" package provides quantitative evidence for a better model based on various information criteria and should not be used to interpret the qualitative difference.

Name	R2 (marg.)[2]	AIC weights[3]	AICc weights	BIC weights	Performance-Score
Model 4	0.657	0.866	0.865	0.248	84.71%
Model 1	0.487	0.083	0.083	0.465	56.29%
Model 5	0.678	0.051	0.051	0.287	56.22%
Model 2	0.509	4.29E-31	4.37E-31	4.74E-29	11.09%
Model 3	0.534	9.98E-33	1.02E-32	2.17E-29	3.34%

Table 2.2. Performance scores for the inverse Gaussian distribution models.

Following Barr et al. (2013), a maximal model appropriate for the design should be adapted if possible. A full model based on Model 4 was constructed:

```
glmer(RT ~ congruency * character type + colour + (1+congruency+character typ
e+congruency:character type|subject) + (1+congruency+character type+congruen
cy:character type|item), family = inverse.gaussian(link = "identity"), control=glme
rControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
```

However, the full model did not converge; therefore, the random structures are simplified until it successfully converges:

```
glmer(RT ~ congruency * character type + colour + (1+character type|subject) +
(1+ character type|item), family = inverse.gaussian(link = "identity"), control=glm
erControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
```

By plotting residuals against predicted RT, it is possible to visualize the difference between Yeh et al.'s (2017) original model and the final model selected using the approach above (see Figure 2.7). It is clear that the residual plot for Yeh et al.'s (2017)

---

[2] The marginal R2 considers only the variance of the fixed effects Lüdecke, D., Ben-Shachar, S. M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139 .

[3] Weights are computed as w = exp(-0.5 * delta_ic) / sum(exp(-0.5 * delta_ic)), where delta_ic is the difference between the model's IC (Information Criterion) value and the smallest IC value in the model set ibid..

model shows a heteroscedastic (fan-shaped) pattern, whereas the residual plot for the final model reveals a more homoscedastic pattern.

Figure 2.7. Plots of the residuals again predicted RT from Yeh et al.'s (2017) original model and the final model used in the thesis.

Tables 2.3 and 2.4 provide detailed summaries of Yeh et al.'s (2017) model (LMMs) and the final model proposed in this analysis (GLMMs). The interaction between Stroop interference and Colour-Character vs. Invalid-Radical was significant in Yeh et al.'s (2017) model but showed a trend in the final model.

Random Effects			
Group	Name	Variance	SD
subject	(Intercept)	3728.4	61.06
pair	(Intercept)	104.1	10.2
colour	(Intercept)	1347	36.7
Number of observations: 1505, groups: subject, 28; pair, 27; colour, 3			

Fixed Effects					
	Estimates	SE	df	t	p
(Intercept)	631.3558	24.4022	3.32199	25.873	6.01e-05 ***
congruency2	-36.4838	8.17805	364.3646	-4.461	1.09e-05 ***
congruency3	57.95523	6.2903	1008.444	9.213	< 2e-16 ***
character type2	-0.09931	8.94285	25.49367	-0.011	0.9912
character type3	-24.0664	8.94142	25.48596	-2.692	0.0124 *
congruency2:character type2	15.58368	20.04892	366.266	0.777	0.4375
congruency3:character type2	-22.9947	15.41576	1005.756	-1.492	0.1361
congruency2:character type3	11.95619	20.04601	365.1547	0.596	0.5513
congruency3:character type3	-40.1484	15.40772	1008.234	-2.606	0.0093 **

Model Fit		
AIC	BIC	Log Likelihood
18452	18522	-9213.2

Family: Gaussian (identity)
Model equation: RT ~ congruency * character type + (1 | subject) + (1 | pair) + (1 | colour)
Note:
congruency2: neutral vs. congruent (i.e. Stroop facilitation effects)
congruency3: neutral vs. incongruent (i.e. Stroop interference effects)
character type2: Colour-Character vs. Valid-Radical
character type3: Colour-Character vs. Invalid-Radical
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Table 2.3. Yeh et al.'s (2017) model summary.

Random Effects					
Group	Name	Variance	SD	Correlation	
subject	(Intercept)	9.36E+02	30.590738		
	character type2	2.45E+02	15.664345	0.22	
	character type3	3.30E+02	18.173687	-0.48	0.42
item	(Intercept)	1.73E-01	0.416378		
	character type2	6.73E-01	0.820034	-0.73	
	character type3	3.38E+02	18.390743	-0.64	0.94
Number of observations: 2852, groups: subject, 28; item, 18					

Fixed Effects				
	Estimates	SE	t	p
(Intercept)	643.7457	15.1329	42.539	< 2e-16***
congruency2	-34.2996	8.1082	-4.23	2.33E-05***
congruency3	53.4461	8.0507	6.639	3.16E-11***
character type2	-0.6873	9.2338	-0.074	0.9407
character type3	-27.347	12.3698	-2.211	0.0271*
colour1	13.7267	3.4317	4	6.33E-05***
colour2	26.4922	3.5073	7.553	4.24E-14***
congruency2:character type2	19.065	13.4434	1.418	0.1561
congruency3:character type2	-22.58	14.6513	-1.541	0.1233
congruency2:character type3	16.7533	15.9	1.054	0.292
congruency3:character type3	-31.0816	16.1876	-1.92	0.0548.

Model Fit		
AIC	BIC	Log Likelihood
35814.3	35927.4	-17888.1

Family: inverse.gaussian (identity)
Model equation: RT ~ congruency * character type + colour + (1 + character type | subject) + (1 + character type | item)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Note:
congruency2: neutral vs. congruent (i.e. Stroop facilitation effects)
congruency3: neutral vs. incongruent (i.e. Stroop interference effects)
character type2: Colour-Character vs. Valid-Radical
character type3: Colour-Character vs. Invalid-Radical
colour1: colour cyan vs. mean
colour2: colour red vs. mean
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Table 2.4. Final model summary. "bobyga" was chosen as the optimizer.

Next, multiple comparisons were conducted on those two models using the "emmeans" package (1.8.4-1). The *p*-values were corrected using Dunnett's method (Dunnett, 1955). The Stroop interference and facilitation effects are presented in Table 2.5.

		Yeh et al.'s	Final Model
Stroop Interference Effects			
Colour-Incongruent vs. Neutral	RT differences	79***	72***
Invalid-Radical-Incongruent vs. Neutral	RT differences	39***	42*
Valid-Radical-Incongruent vs. Neutral	RT differences	56***	52***
Stroop Facilitation Effects			
Neutral vs. Colour-Congruent	RT differences	46**	48***
Neutral vs. Invalid-Radical-Congruent	RT differences	34*	30*
Neutral vs. Valid-Radical-Congruent	RT differences	30^{+}	27^{+}

Table 2.5. Stroop interference and facilitation effects of each character type. RT differences as a function of character type and congruency from Yeh et al.'s (2017) model and the final model.

Although the overall pattern remained the same when analysing the data with the final model, the *p*-values for Stroop interference effects were smaller in the Invalid-Radical condition in Yeh et al.'s (2017) model than in the final model.

In conclusion, this section used Yeh et al.'s (2017) data as an example to show that GLMMs provided a better fit for the current Stroop data without the transformation of raw RTs. To avoid the need to transform raw RTs, analyses reported in this thesis will be conducted using GLMMs.

2.6.1.2 Distributional analysis

Mean RT analysis averages the RTs by subjects and conditions, making it impossible to observe changes in certain effects across the entire RT distribution. It has been argued that the effects of attentional inhibition are greatest at the tail of the RT distribution (Bub et al., 2006; Ridderinkhof et al., 2005; Roelofs et al., 2011; Sharma et al., 2010). The dynamics of attentional control are likely lost when only focusing on mean RT analysis.

To investigate the impact of experimental manipulations across the entire RT distribution, delta plots can be used (Ridderinkhof et al., 2005; Roelofs et al., 2011). Delta plots are effective tools for revealing the magnitude of a particular effect as a function of response speed. They can be directly derived from cumulative density functions (i.e. quantile plots). Figure 2.8 shows the possible interpretation of reaction distribution effects in the Stroop task by Roelofs et al. (2011).

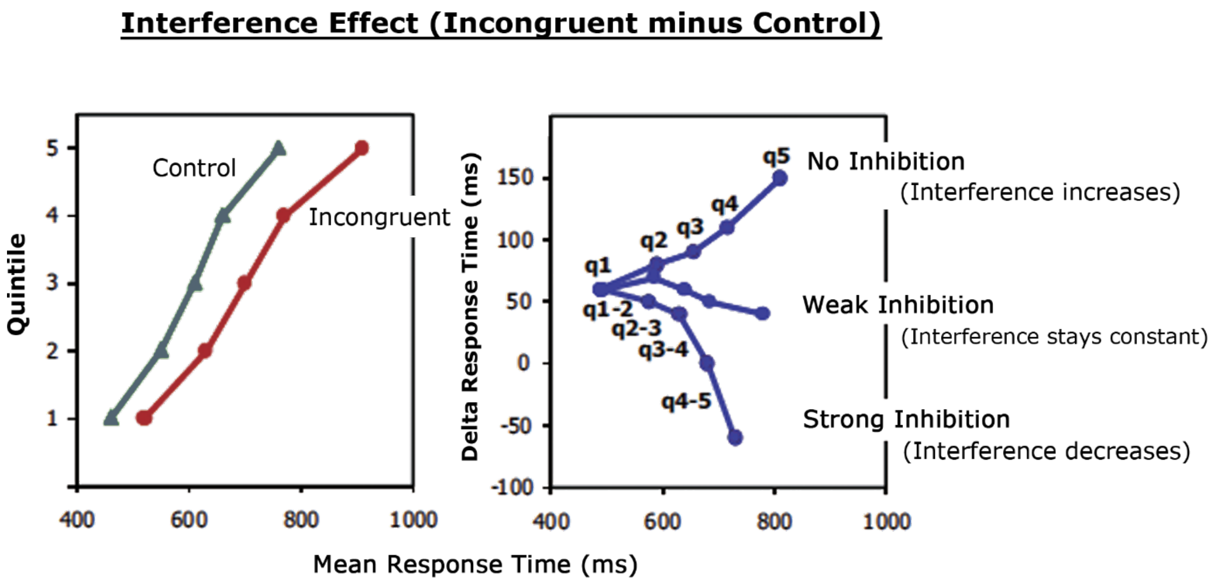


Figure 2.8. The left-hand panel shows cumulative distribution curves for response times in incongruent and control conditions. The right-hand panel presents delta plots showing the magnitude of the interference effect (deltas) as a function of five quintiles, and the amount of inhibition (no, weak or strong). The term "q1" relates to quintile 1 and so forth; q1-2 is the segment connecting quintiles 1 and 2 etc.

In Figure 2.8, inhibition refers to active/willed inhibition as defined by Aron (2007), which suppresses the irrelevant response, stimulus, or memory. In a vocal Stroop task, inhibition takes place when the word information should be suppressed to name the ink colour of the words. In a manual Stroop task, inhibition resolves the competition between the irrelevant word recognition task and the relevant colour-matching task. Delta plots can clearly illustrate the inhibition applied throughout the recognition process, which information is less obvious in quantile plots. Delta plots reflect "the effect of an experimental factor tends to increase as a function of RT" (Roelofs, 2011, p. 2). The effects are stronger for long RTs than short RTs. When this rationale is applied to the Stroop interference effect, an upward trend in the delta plot shows that interference increases with slower reaction times, implying that no inhibition is applied to resolve the conflict between the incompatible word and colour information; a levelling-off curve indicates that weak inhibition is preventing the interference from growing larger; a downward trend shows strong inhibition in cognitive control and the interference decreases.

Likewise, an illustration for facilitation effects in the Stroop paradigm was created (Figure 2.9). Instead of inhibition, Stroop facilitation was interpreted in terms of enhancement.

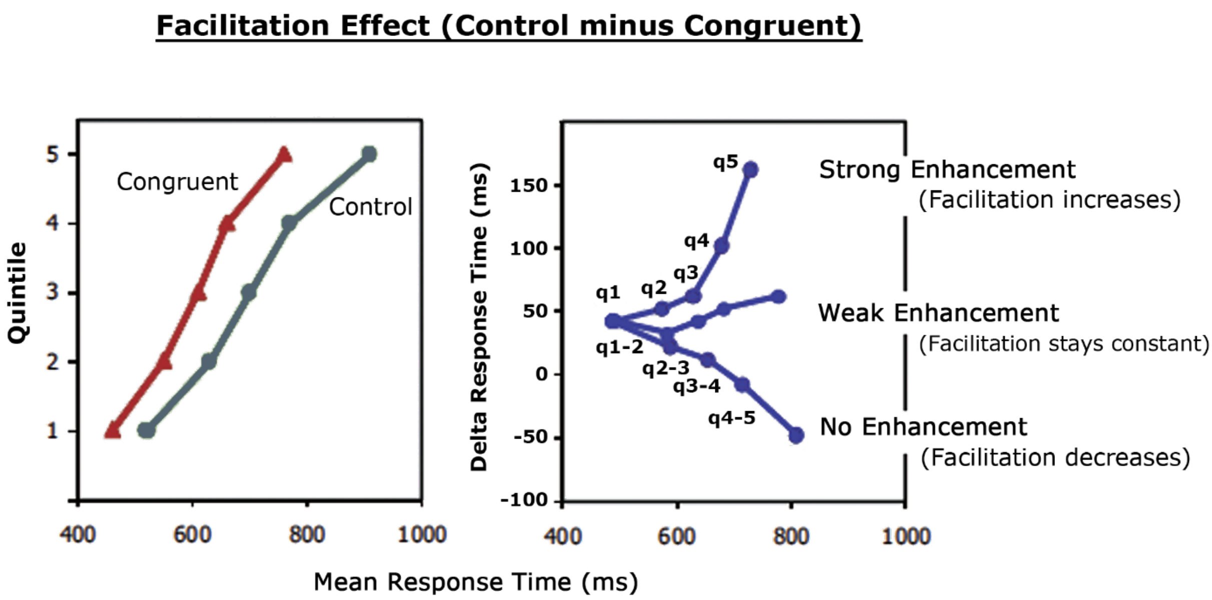


Figure 2.9. The left-hand panel shows cumulative distribution curves for response times in congruent and control conditions. The right-hand panel presents delta plots showing the magnitude of the facilitation effect (deltas) as a function of five quintiles, and the amount of enhancement (no, weak or strong).

This time, with the increasing RT, the upward curve reveals an increase in facilitation. When facilitation increases, strong enhancement is applied to converge the colour and word information because the perception of colour and word information is at different levels as suggested by Roelofs (2010a). A levelling off curve means that there is weak enhancement; a downward trend shows that no enhancement is applied to converge the colour and word information, so facilitation decreases.

Delta plots can help investigate the differences at the tail of the distribution (late effects). For example, Bub et al. (2006) conducted a Stroop task with elementary-school children. Previous studies found that children have greater Stroop interference effects than adults (Carter et al., 1995; Comalli et al., 1962; Guttentag & Haith, 1978; Vurpillot & Ball, 1979). The results of the delta-plot analysis revealed that younger children do have the ability to suppress word reading. However, this suppression is stronger for younger children than for older children at the tail of the distribution.

In addition to age differences, delta plots can be used to investigate response modality effects in the Stroop task. Kinoshita et al. (2017) used several types of distractors (e.g. pseudoword, consonant string, nonalphabetic symbols, a row of Xs, incongruent colour words) to investigate the rationale of the Stroop task. In a vocal Stroop task, a word-likeness gradience was observed, with the strongest interference effects revealed in the real words and pseudowords, followed by consonant strings and nonalphabetic symbol strings. This pattern, however, was not observed in the manual Stroop task. Thus, they argued that the differences between vocal and manual responses lay in the vocal response requirement. That is, it reflects the specific goal of the task that requires participants to pay different levels of attention to it.

In light of the Stroop studies using distributional analysis, it would be interesting to conduct this analysis with the data from Yeh et al. (2017). Furthermore, it will illustrate the approach that will be used in the experiments reported in this thesis.

Because attentional control is largest at the tail of the RT distribution, no data trimming was performed on Yeh et al.’s (2017) data, an approach commonly seen in the distributional analysis (Pratte et al., 2010; Ridderinkhof et al., 2005). Next, the RTs of each subject for each condition were rank ordered and divided into five quintiles. The mean RT was calculated for each quintile in each condition. The delta plots for the Stroop interference effect in the Colour-Character, Invalid-Radical, and Valid-Radical conditions were obtained by computing the RT difference between each experimental condition and their neutral control conditions (e.g. Colour-Character vs. Neutral-Control). An illustration of the quantile and the delta plots is presented in Figure 2.10.

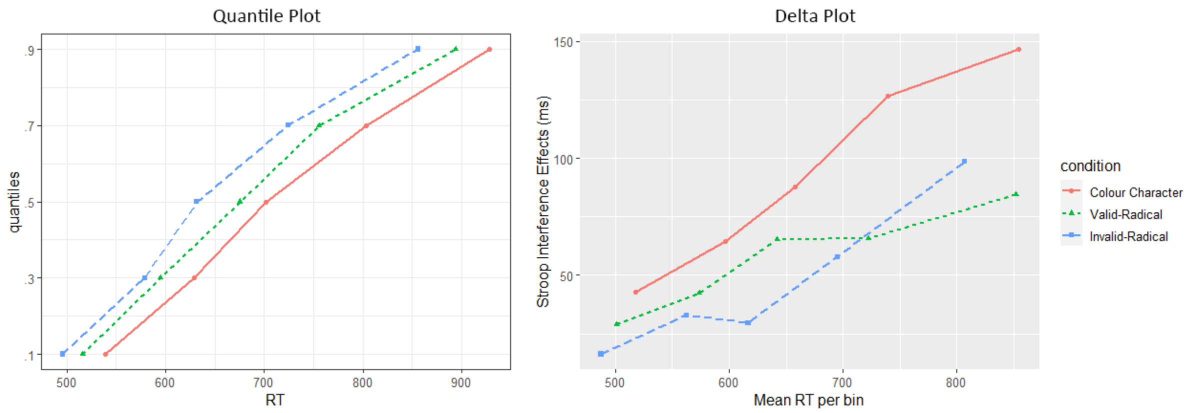


Figure 2.10. Quantile plots and delta plots based on Yeh et al.'s (2017) data.

From the delta plots, it is clear that all conditions show Stroop interference effects, although it is not clear whether there is a linear increase of Stroop interference effects across the quantiles. As suggested by Pratte (2021), Burle et al. (2014), and Grant (1956), an orthogonal-polynomial-contrast trend analysis can be used to quantify the shape of the delta plots. A significant linear component indicates a linear trend for the observed effects. A significant quadratic component indicates the delta plot presents an initial increase followed by a decrease. After coding "quantiles" as orthogonal-polynomial contrasts, a one-way ANOVA was performed for each condition. The results suggest that the delta plot for the Colour-Character, Valid-Radical, and Invalid-Radical conditions showed positive linear trends (Colour-Character: $F(1,135) = 14.601$, $p < 0.001$, $d = 0.72$; Valid-Radical: $F(1,135) = 7.427$, $p < 0.01$, $d = 0.52$; Invalid-Radical condition: $F(1,135) = 12.296$, $p < 0.001$, $d = 0.66$) and no quadratic trends (Colour-Character: $F(1,135) = 0.023$, $p = 0.881$, $d = 0.03$; Valid-Radical: $F(1,135) = 0.04$, $p = 0.843$, $d = -0.04$; Invalid-Radical condition: $F(1,135) = 1.552$, $p = 0.215$, $d = 0.24$).

According to Roelofs et al.'s (2011) interpretation, a constant increase in Stroop interference effects across the quantiles indicates no inhibition that prevents the interference from growing larger. By analysing the delta plots, it is clear that no inhibition was applied to resolve the conflict between colour and word information in Stroop effects. These findings cannot be concluded from the mean RT analysis.

Therefore, distributional analysis will be used in this thesis in order to capture the amounts of inhibition/enhancement applied when conducting the Stroop task.

2.6.2 Thesis outline

This thesis investigates Chinese word recognition using the Stroop task and how Stroop effects can be decomposed into conflict/facilitation components. The experiments involve Chinese characters and pinyin and investigate the automatic activation of characters, radicals, and pinyin when conducting a colour naming or colour decision task (Stroop task). Furthermore, the thesis focuses on how response modality can affect different conflict/facilitation components.

The next chapter (Chapter 3) focuses on automatic sublexical processing in Chinese when participants complete a Stroop task. This research focuses on the semantic and phonological activation of radicals when embedded in Chinese characters. In addition, previous research used vocal Stroop task that requires explicit verbal responses. This chapter used both vocal and manual Stroop tasks to investigate whether response modality can affect the semantic and phonological activation of radicals.

Chapter 4 focuses on character processing in Chinese and explores how Stroop interference and facilitation effects can be further decomposed into different components (see the discussion in Section 2.3). Most importantly, the influences of decomposed components may differ for response modalities used (see the discussion in Section 2.4), which is the key to studying the mechanism behind the Stroop phenomenon. Using Chinese homophones in the experiment reported in this chapter, it is possible to investigate phonological components of the Stroop effects. Homophones are difficult to create in alphabetic languages like English or French.

Chapter 5 focuses on the effect of pinyin in the Stroop task. Most models/hypotheses on Chinese word recognition focused on Chinese characters. Characters can also be written in pinyin; however, there is limited research that investigated Chinese characters written in pinyin in the Stroop task, which means the automaticity of reading pinyin has not been thoroughly explored. In a series of Stroop tasks, this chapter investigates the tonal information, phonological activation, and semantic activation with Chinese characters written in pinyin. Both Stroop interference and facilitation effects will be investigated. These effects will then be decomposed and compared with Chinese characters to see how similar or different they are when processing characters and pinyin.

The General Discussion (Chapter 6) summarizes the findings of the thesis and discusses distinct conflict/facilitation components in the Stroop effects using Chinese characters, radicals, and characters written in pinyin and how response modality effect can affect those components. This chapter also explores the mechanisms behind the Stroop task from a non-alphabetic language like Chinese. Implications for theories of Chinese word processing and models of Stroop will be discussed as well as the limitations of the current work.

Chapter 3: The Role of Phonetic Radicals in Visual Word Recognition

3.1 Introduction

As introduced in Section 1.3, phono-semantic compounds are the most common formation method in modern Chinese. They consist of semantic and phonetic radicals, which are the sublexical components of Chinese characters. Radicals are of great interest to the research of word recognition as they can affect the processing at the word level.

The semantic radical in phono-semantic compounds usually provides a cue to the meaning of the word and on their own many semantic radicals are unpronounceable. According to Gao et al. (1993), 46.15% of semantic radicals have a direct link to the meaning of the word, 14.1% have an indirect association, and 39.74% have no association with the meaning of the word.

The phonetic radical in phono-semantic compounds often provides phonological cues to the pronunciation of the whole character. According to Zhou (1978), only about 38% of phonetic radicals have pronunciations consistent with their phono-semantic compounds. Because phonetic radicals are pronounceable, they can also occur as a freestanding character and carry meanings. That is, they also carry meaning as a whole character ("艮" is a freestanding character of "银"). Phonetic radicals that can be used as freestanding characters make up around 81% of all other characters (Zhou, 2003).

As can be seen from the linguistic properties of radicals, their roles are not always consistent with their linguistic definition. This leads to the question of the role of radicals in psychological perspectives. One argument is that the character is primarily processed as a whole, and radicals are not activated or less activated (Liu et al., 1996; Yu et al., 1990). Another view argues for the decomposition of characters, where characters are first decomposed into radicals (Chen & Yeh, 2015; Feldman & Siok, 1997; Taft & Zhu, 1997; Taft et al., 1999). Radicals will be the primary processing unit, and semantic radicals may provide meaning and phonology, depending on the linguistic property of the individual semantic radical. Phonetic radicals may also provide phonology as well as meaning to the recognition process.

The next section will provide an overview of studies that have investigated the role of phonetic radicals in visual word recognition. An overview of the role of phonetic radicals will be provided first, followed by the discussion of its role in providing phonology and semantics.

The role of phonetic radicals is worth investigating, as more and more literature reports that they have a more important role in word recognition than semantic radicals. In an apparent-motion task in which one of the sublexical components is displaced towards different locations, participants were instructed to indicate whether the semantic or phonetic radical had moved. Wang (2006) found participants could detect the apparent motion more efficiently in phonetic radicals rather than in semantic radicals when the character is of low frequency. This could occur because phonetic radicals provide a consistent pronunciation to the whole character, especially in low-frequency words. Hung et al. (2014) tested the processing of phonetic and semantic radicals in priming tasks. Data revealed that phonetic radicals produced more radical repetition effects than semantic radicals, suggesting that phonetic radicals might have a processing advantage over semantic radicals. This could be due to the higher combinability of semantic radicals, compared to phonetic radicals. When composed into phono-semantic compounds, there are fewer semantic radicals available, which means they will be used more frequently than phonetic radicals. When searching for a phono-semantic compound, it might be more efficient to use phonetic radicals which lead to fewer possible combination of characters. Based on the findings of Hung et al. (2014), phonetic radicals play a more dominant role in early lexical processing than semantic radicals.

Phonetic radicals can activate phonological information when processing characters. Zhou and Marslen-Wilson (1999a) confirmed that phonetic radicals in low-frequency compound characters are decomposed and activate phonological representations. In their priming study, target characters (e.g. 族, /zu2/) were preceded by complex characters (粹, /cui4/) that contain phonetic radicals (卒, /zu2/) that are pronounced the same as the targets. The results showed significant priming effects for low-frequency complex primes, but not for high-frequency complex primes. This means phonetic radicals in phono-semantic compounds facilitate the processing of target that are homophonic to the phonetic radical, providing evidence for the phonological activation of phonetic radicals.

Seidenberg (1985) also found that naming is facilitated only when the target is of low frequency, indicating that phonological information is more important in the recognition of low-frequency characters. A similar conclusion was also drawn by Hue (1992) and Lee et al. (2005), who found that regularity effects only occur in low-frequency characters. Various studies found that phonetic radicals can provide phonological cues in word recognition, but this effect seems to be limited to low-frequency characters.

In addition to providing phonological cues for the pronunciation of Chinese characters, more and more research has investigated whether phonetic radicals activate semantics, because around 81% of Chinese characters can be used as freestanding phonetic radicals (Zhou, 2003). Zhou and Marslen-Wilson (1999b) explored the phonological and semantic processing of phonetic radicals in a semantic and phonological relatedness judgement task. Three types of primes were used: semantic, complex and control. Semantic primes are single characters and are semantically related to the targets. Complex primes are complex characters that contain the semantic prime as a phonetic radical, thus semantically related to the target at the radical level. In Experiment 1, complex primes shared the same phonology as semantic primes; in Experiment 2, they were pronounced differently. Three types of stimulus onset asynchrony (SOA) were also used to examine the time course of sublexical semantic activation: 57 ms, 100 ms and 200 ms. The findings revealed phonological and semantic activation of phonetic radicals with short SOAs (57 and 100 ms), whereas this effect seemed to disappear or was suppressed with an SOA of 200 ms. The phonological and semantic information of phonetic radicals was activated under short SOA conditions. This is an important finding as phonetic radicals are linked to semantic as well as phonological information. However, this applies to low-frequency stimuli only, as it is difficult to find sufficient high-frequency stimuli in sublexical studies, where strict selection criteria were implemented.

Following Zhou and Marslen-Wilson’s (1999b) work, Lee et al. (2006) carried out a series of event-related potential (ERP) studies to investigate the role of phonetic radicals in a priming task. The primes were similar to Zhou and Marslen-Wilson's (1999b): a semantically related regular phonogram (with the same pronunciation), an irregular phonogram (with a different pronunciation), and a control. Significant N400 effects (a negative-going wave occurring about 400 ms after the onset of the stimuli, which is often related to semantic competition) were found for semantically related primes in all three SOAs: 50 ms, 100 ms, and 300 ms. However, both N400 effects disappeared with the regular and irregular phonograms at an SOA of 300 ms, indicating that the semantic activation of phonetic radicals is temporal and cannot be preserved after 300 ms.

Yeh et al. (2017) investigated whether phonetic radicals activate their meaning using a Stroop paradigm. Their vocal Stroop task included the Valid- and Invalid-Radical conditions. The Valid-Radical condition has a phonetic radical that shares the same pronunciation as the whole character (e.g. 清, /qīng/, “clear” contains a phonetic radical 青, /qīng/, meaning “cyan”). The Invalid-Radical contains a phonetic radical that does not have the same pronunciation as the whole character (e.g. 猜, /cāi/, meaning “guess”).

Yeh et al. (2017) found facilitation and interference effects in Colour-Character, Valid-Radical, and Invalid-Radical conditions, except for a marginal facilitation effect in the Valid-Radical condition (see Figure 3.1 for a summary of the findings). These findings suggest that phonetic radicals activate semantics because conditions that include colour characters as phonetic radicals elicited significant interference effects. It was expected that the Valid-Radical condition would show more Stroop interference because the pronunciation of its phonetic radical and the colour character is identical. However, no significant difference between the amount of Stroop interference was found between the Invalid- and Valid-Radical conditions. This result suggests that the phonological activation of phonetic radicals is not strong or fully completed when reading Chinese characters, which contrasts with what Zhou & Marslen-Wilson (1999a) found.

Yeh et al.'s (2017) study, however, involved only limited numbers of trials, especially in the congruent conditions: 18 congruent trials in two blocks; that is, only 6 trials in each character type and 2 trials for each character type and each colour (3 colours). The purpose of setting fewer congruent trials was to lower the possibility of participants detecting the goal of the experiment (studying the radicals), but this leads to smaller facilitation effects (congruent vs. neutral trials) based on a smaller number of trials.

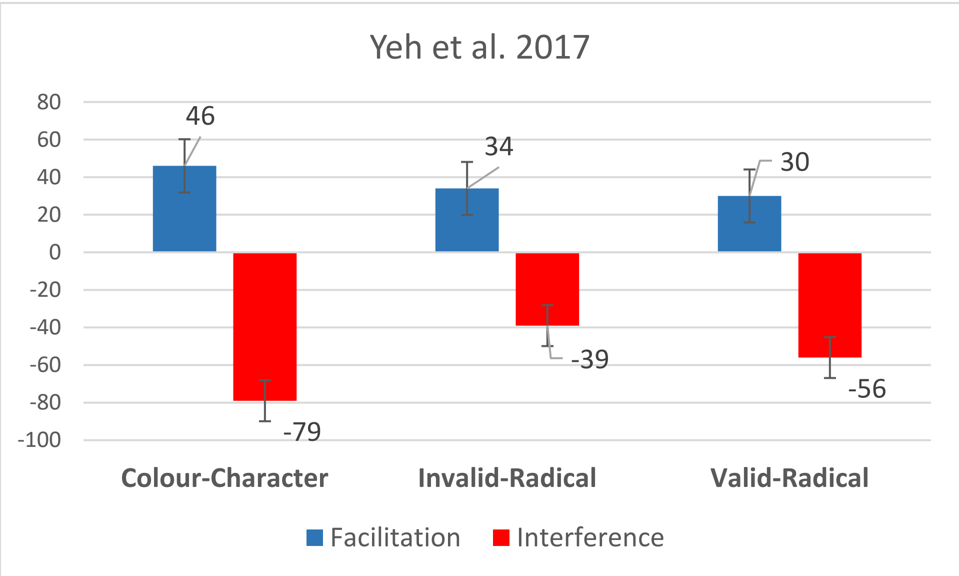


Figure 3.1. Summary of Yeh et al.'s (2017) results in Experiment 1. RT differences = Neutral trials minus Non-Neutral trials. A positive value refers to facilitation effect (faster than neutral conditions), a negative value denotes interference effect (slower than neutral conditions). Error bars indicate one standard error from the mean.

Another aspect that has not been mentioned in Yeh et al.'s (2017) study is the response modality. The Stroop effect is commonly measured in two response modalities: vocal (naming the ink colour) and manual (pressing the key button for the corresponding ink colour). The response modality effect refers to the reduced interference observed in manual responses, compared to vocal responses (Neill, 1977; Redding & Gerjets, 1977; White, 1969). This might be attributed to how phonological information is activated in the two modes. The manual Stroop task is essentially a colour categorization task. There are no verbal responses, and thus it leads to less competition than the vocal Stroop task. The view of "less competition in the manual responses" was also strengthened by studies that observed an increase in the facilitation effect in the manual Stroop task (Neill, 1977; Redding & Gerjets, 1977). Naming words in congruent conditions may serve to create less competition when deciding whether to process ink colour or word information first, whereas in manual responses, this mode selection disappears, and the focus is more on the processing of the ink colour.

Although the results of Yeh et al. (2017) suggested that phonetic radicals can activate semantics in a vocal Stroop task, it remains unclear whether the activation of semantics was the result of the vocal response that involves the activation of language representations. To explore the impact of the use of vocal responses in the activation of semantics by phonetic radicals, it would be useful to use a manual Stroop task that does not require any language because only a manual response is needed. In particular, it would enable the investigation of the activation of phonetic radicals when no explicit vocal responses are required. Stroop studies using Chinese characters have observed Stroop interference effects in manual responses (Wang et al., 2010). However, as far as I know, no manual Stroop task has looked at the sublexical processing in Chinese characters. Phonetic radicals may still activate semantics and phonology in manual responses, but as there is less competition compared to vocal responses, the interference effects would be smaller.

To summarize, research into phonetic radicals confirmed their role in phonological activation (mainly in low-frequency characters) and discovered that they can be treated as freestanding characters, thus leading to semantic activation (Lee et al., 2006; Zhou &

Marslen-Wilson, 1999b). Further research on how different response modalities will affect the phonological and semantic activation of phonetic radicals will be addressed in the current study.

3.1.1 Present study

The first experiment focuses on whether the findings of Yeh et al. (2017) can be replicated. Experiment 1 will use the vocal Stroop task, whereas Experiment 2 will use the manual Stroop task. This enables the investigation of the impact of response modality on phonological and semantic activation by phonetic radicals.

3.2 Experiment 1 – The Vocal Stroop Task

3.2.1 Methods

3.2.1.1 Participants

Forty-three students recruited from the University of Nottingham, UK participated in this experiment (mean age = 23, range = 19–30, females = 26). All were native speakers of Chinese from mainland China. Two participants were excluded from the study: one for failing to complete the experiment and another because he was not born in mainland China. Participants had normal or corrected-to-normal vision. Participation was voluntary, and each participant received an inconvenience allowance after completing the experiment. The studies were approved by the ethics committee in the Department of Psychology at the University of Nottingham (Ethics approval number: S1117R). The experiments were conducted in accordance with applicable research subject guidelines. All participants signed informed consent forms prior to data collection.

3.2.1.2 Stimuli and Design

The stimuli were presented on a 24-inch LCD monitor (refresh rate: 120 Hz) using DMDX software (Forster & Forster, 2003). As in Yeh et al.'s (2017) design, the background colour was set as grey (RGB: 150,150,150). Characters presented on the screen used the Kai font (楷书), size 20, and in three different colours: cyan (RGB: 0,255,255), yellow (RGB: 255,255,0) or red (RGB: 255,0,0). Colour patches (rectangles of the same size as the character) were also presented in the experiment. A fixation point was indicated by a grey dot "●" (RGB: 128,128,128) in the centre of the screen. Using a grey dot rather than a fixation cross was because the strokes of some Chinese character may overlap with the "+" sign. For example, the characters "黄", "理", and "轩" contain cross or cross-like strokes.

The same stimuli were used as in Yeh et al.'s (2017) first experiment (see Table 3.1 for a quick summary, and Appendix A and B for details). The list contained a mixture of both high- and low-frequency characters. Stimuli consisted of three character types: Colour-Character, Valid-Radical and Invalid-Radical conditions. The Colour-Character condition contained single characters that referred directly to the colours (e.g. 青, /qīng/, meaning "cyan"). The Invalid-Radical condition consisted of complex characters that contained the colour characters as phonetic radicals but with a different pronunciation and a different meaning (e.g. 猜, /cāi/, "guess" has the phonetic radical 青, /qīng/, "cyan" but is distinct both in meaning and pronunciation). The Valid-Radical condition was made up of complex characters that embedded the colour characters as phonetic radicals, with the same pronunciation but a different meaning (e.g. 清, /qīng/, "clear" is pronounced the same as 青, /qīng/, "cyan" but has a dissimilar meaning). Each character type was accompanied by a Neutral-Control character that matched in frequency and stroke count. Neutral-Control characters and their radicals had no relation to colour names, either semantically or phonetically. The Colour-Character condition had single characters as a control condition, while Invalid- and Valid-Radical conditions had complex characters as neutral controls. Yeh et al. (2017) presented stimuli in Traditional Chinese characters. The present study recruited participants from mainland China, all the stimuli were presented in Simplified Chinese. This change affected three neutral characters (諄 into 谆; 軒 into 轩; and 帳 into 帐). The experimental character – colour "yellow" (黃 in Traditional Chinese) is written slightly different from 黄 in Simplified Chinese, which also affected the presentation of its Invalid-Radical condition (横 and 横) and Valid-Radical condition (潢 and 潢).

Condition	Character	Meaning and pronunciation	Neutral-Control	Meaning and pronunciation
Colour-Character				
Cyan	青	cyan, /qing1/	具	Tool, /ju4/
Yellow	黄	yellow, /huang2/	曾	Already, /ceng2/
Red	朱	red, /zhu1/	丢	Discard, /diu1/
Valid-Radical				
Cyan	清	clear, /qing1/	理	Reason, /li3/
Yellow	潢	pond, /huang2/	谆	Iterate, /zhun1/
Red	珠	pearl, /zhu1/	轩	Pavilion, /xuan1/
Invalid-Radical				
Cyan	猜	guess, /cai1/	帐	Tent, /zhang4/
Yellow	横	horizontal, /heng2/	榜	Placard, /bang3/
Red	殊	different, /shu1/	勒	Strangle, /le4/

Table 3.1. Stimuli used in Experiment 1.

For each of the three character types, there were two congruency conditions: congruent and incongruent. In congruent conditions, characters were presented in colours that matched the whole character in the Colour-Character condition and matched the phonetic radicals in the Invalid- and Valid-Radical conditions. In incongruent conditions, characters were presented in colours that were inconsistent with the whole characters in the Colour-Character conditions and inconsistent with the phonetic radicals in the Invalid- and Valid-Radical conditions.

For each character type, there were three characters and their controls, so 18 different characters were used in total: 3 conditions × (3 characters + 3 controls). The original experimental design included two blocks of 60 trials. Each block contained 54 character trials and six colour-patch trials. Within the character trials, there were nine congruent trials (3 conditions × 1 colour × 3 characters); 18 incongruent trials (3 conditions × 2 colours × 3 characters); 27 Neutral-Control trials (3 Neutral-Control characters × 3 colours × 3 conditions); and six colour patches (3 colours × 2 trials). Two extra blocks were added to the original design to increase the number of trials. Thus, in total the experiment contained 240 trials (60 trials × 4 blocks) for each participant. Trials were presented in a pseudo-randomized order.

As suggested by Brysbaert and Stevens (2018), a minimum of 1,600 observations per condition would be recommended in a repeated measure study. There are 1,548 observations for congruent trials (43 participants × 4 blocks × 9 trials), 3,096 for incongruent trials (43 participants × 4 blocks × 18 trials), and 4,644 for neutral trials (43 participants × 4 blocks × 27 trials). The number of observations in the current study should ensure a properly powered experiment.

3.2.1.3 Procedure

Participants were tested individually in a dimly lit room. They were asked to name aloud the ink colour of the presented characters. A microphone with a voice key was used to measure the naming response onset latencies. The accuracy and latencies of responses were checked afterwards using the CheckVocal program (Protopapas, 2007). The presentation sequence and timing differed slightly from Yeh et al.'s (2017) study. In Yeh et al.'s study, a fixation point was first presented for 347 ms, then replaced by the target character. After the participant responded vocally, a feedback tone was presented for 50 ms. In the current study, participants were first shown a fixation point for 500 ms, followed by a blank screen for 300 ms. The target character then appeared on the screen for a maximum of 3000 ms. Characters disappeared as soon as the participant made a vocal response. There was an intertrial interval of 1000 ms. Twenty-four practice trials were presented prior to the experimental trials. The stimuli used in the practice blocks were not presented in the experimental blocks, except for the three colour characters (青, 黄 and 朱) and three colour patches.

3.2.2 Results

3.2.2.1 Mean RT analysis

Incorrect responses were discarded (2.88%). RTs below 200 ms and above 1500 ms were also discarded before the analyses (1.24%).

A series of generalized linear mixed-effect models (GLMMs) analyses were carried out using R (4.2.2) and the packages "lme4" (1.1-31), "emmeans" (1.8.4-1) and "lmerTest" (3.1-3). Fixed factor "congruency" used simple coding so that neutral trials will be compared to congruent and incongruent trials. Fixed factor "character type" also used simple coding that the Colour-Character condition will be compared to the Valid- and Invalid-Radical conditions. Fixed factor "colour" used deviation coding so that each colour will be compared to the mean of all colours.

First, different types of distributions were compared, and the results of the performance analyses revealed that the Inverse Gaussian fitted the data best. Next, four models with different fixed and random structures were compared. A summary of Information Criterion for these models is shown in Table 3.2.

Name	R2 (marg.)	AIC weights	AICc weights	BIC weights	Performance-Score
Model 4	0.624	0.705	0.706	0.988	87.56%
Model 3	0.623	0.295	0.294	0.012	65.53%
Model 2	0.54	3.10E-56	3.11E-56	1.47E-54	13.24%
Model 1	0.543	4.36E-56	4.39E-56	7.02E-53	0.40%

Model Structure:
Model 1: glmer(RT ~ congruency * character type + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 2: glmer(RT ~ congruency * character type + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 3: glmer(RT ~ congruency * character type + colour + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 4: glmer(RT ~ congruency * character type + colour + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

Table 3.2. Summary of Information Criterion for models used in Experiment 1.

Model 4 is the best fit for the data. Based on Model 4, a full model was constructed as follows:

```
glmer(RT ~ congruency * character type + colour + (1+congruency+character typ
e+congruency:character type|subject), family = inverse.gaussian(link = "identity
"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
```

However, the full model did not converge. Therefore, the random structures were simplified until the model successfully converged, and the final model was:

```
glmer(RT ~ congruency * character type + colour + (1|subject), family = inverse.g
aussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list
(maxfun=2e5)))
```

A summary of the final model can be found in Appendix I. The means of reaction time (RT) are presented in Table 3.3. RT differences between the three character types are presented in Figure 3.2.

		Experiment 1		
		Colour-Character	**Invalid-Radical**	**Valid-Radical**
Congruent	RT (SE)	680 (14.1)	670 (14.4)	665 (14.7)
	ER (SE)	1.02 (0.45)	1.63 (0.57)	0.61 (0.35)
Incongruent	RT (SE)	824 (14.0)	735 (14.1)	756 (14.5)
	ER (SE)	6.71 (0.80)	2.64 (0.51)	5.79 (0.75)
Control	RT (SE)	700 (13.9)	692 (14.0)	696 (13.6)
	ER (SE)	2.24 (0.38)	2.03 (0.37)	1.90 (0.36)

Table 3.3. Mean reaction times (RT, in milliseconds), error rates (ER, %), and standard errors (SE, in parentheses) as a function of congruency and character type from the GLMMs in Experiment 1.

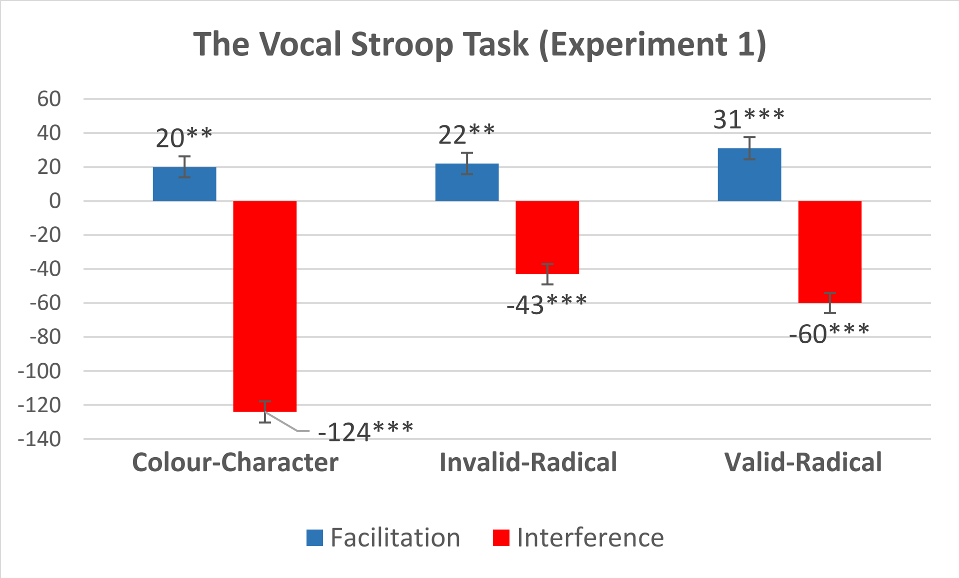


*Figure 3.2. Bar chart results of Experiment 1. RT differences = Neutral trials minus non-Neutral trials. Positive values indicate Stroop facilitation effects, and negative values refer to Stroop interference effects. Error bars represent the standard error from the mean. ***p < 0.001; **p < 0.01.*

Following the approach from Yeh et al. (2017), planned comparisons were conducted using the “emmeans” package (1.8.4-1). Neutral conditions were compared to either congruent or incongruent conditions within each character type. The *p*-values were adjusted using Dunnett’s method (Dunnett, 1955). Further exploratory comparisons were corrected for multiple comparisons using Holm-Bonferroni correction (Holm, 1979).

Interference effects (incongruent minus neutral) were found in all three conditions (Colour-Character condition: M = 124 ms, SE = 6.21 ms, z = 19.988, $p < 0.001$; Invalid-Radical condition: M = 43 ms, SE = 6.07 ms, z = 7.108, $p < 0.001$; Valid-Radical condition: M = 60 ms, SE = 5.96 ms, z = 10.051, $p < 0.001$). Pairwise comparisons of Stroop interference effects revealed that interference effects in the Colour-Character conditions (124 ms) were stronger than in the Invalid-Radical condition (43 ms; difference = 81 ms, SE = 8.33 ms, z = 9.727, $p < 0.001$) and in the Valid-Radical condition (60 ms; difference = 64 ms, SE = 7.40 ms, z = 8.685, $p < 0.001$). The Valid-Radical condition (60 ms) showed a trend for stronger interference effect than the

Invalid-Radical condition (43 ms; difference = 17 ms, SE = 8.79 ms, z = 1.913, p = 0.06).

Compared to Neutral-Control conditions, significant facilitation effects (neutral minus congruent) were found in all three conditions (Colour-Character condition: M = 20 ms, SE = 6.17 ms, z = 3.254, p = 0.002; Invalid-Radical condition: M = 22 ms, SE = 6.32 ms, z = 3.456, p = 0.001; Valid-Radical condition: M = 31 ms, SE = 6.58 ms, z = 4.677, $p < 0.001$). Pairwise comparisons were not significant for the differences between each character type condition in Stroop facilitation effects (Colour-Character vs. Valid-Radical: M = 11 ms, SE = 8.44 ms, z = 1.264, p = 0.618; Colour-Character vs. Invalid-Radical: M = -2 ms, SE = 7.80 ms, z = -0.224, p = 0.823; Valid-Radical vs. Invalid-Radical: M = 9 ms, SE = 8.97 ms, z = 0.996, p = 0.639).

3.2.2.2 Error analysis

The model for error analysis is identical to the one used in mean RT analysis, except for using the Gaussian distribution:

lmer(ER ~ congruency * character type + colour + (1|subject))

Error rates were similar between congruent and neutral control conditions in all three character types. Error rates were higher for the incongruent condition in the Colour-Character and Valid-Radical conditions relative to the neutral control condition, but not in the Invalid-Radical condition (Colour-Character: M = 4.5%, SE = 0.68%, $t(2163)$ = 6.60, $p < 0.001$; Valid-Radical: M = 3.9%, SE = 0.68%, $t(2163)$ = 5.750, $p < 0.001$; Invalid-Radical: M = 0.6%, SE = 0.68%, $t(2163)$ = 0.90, p = 0.568).

3.2.2.3 Distributional analysis

Following the procedure described in Section 2.6.1.2, quantile plots and delta plots based on untrimmed data from Experiment 1 for Stroop interference and facilitation effects were presented in Figures 3.3 and 3.4.

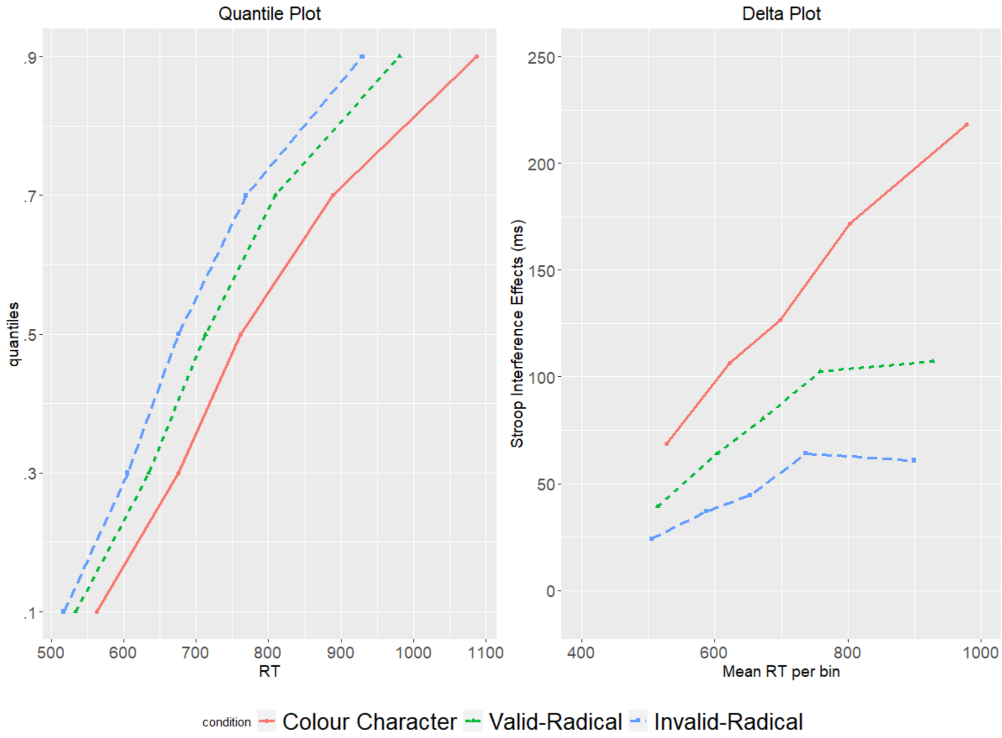


Figure 3.3. Quantile plots and delta plots for Stroop interference effects in Experiment 1.

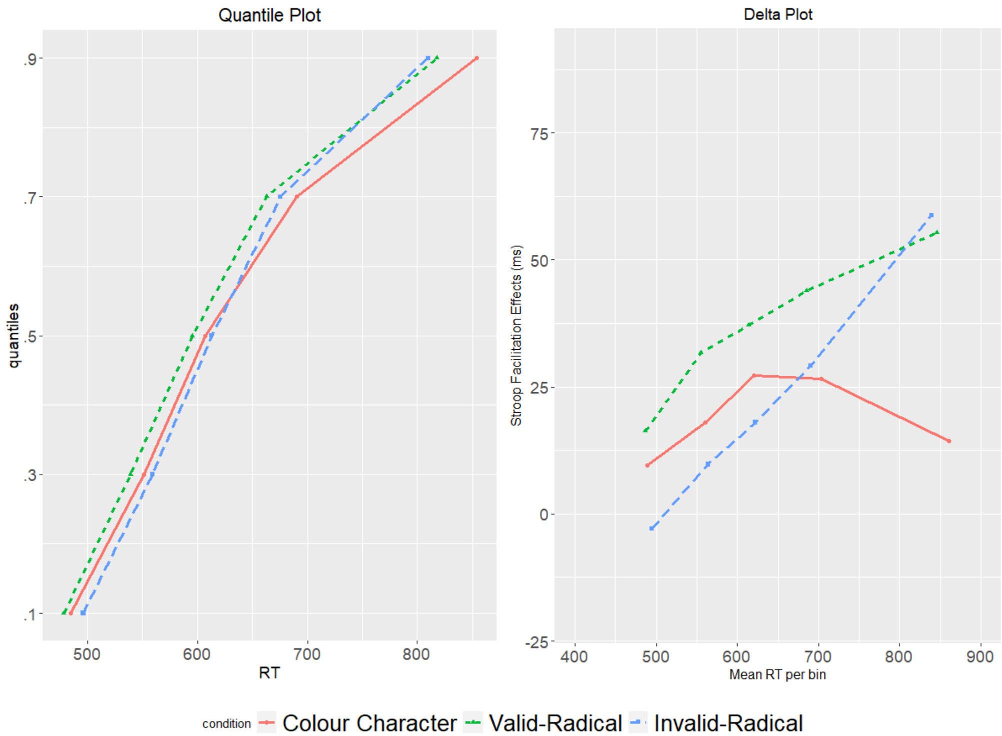


Figure 3.4. Quantile plots and delta plots for Stroop facilitation effects in Experiment 1.

Analyses were separated for Stroop interference (incongruent trials vs. neutral words) and Stroop facilitation effects (congruent trials vs. neutral words). Then, a series of one-way ANOVAs were performed for each character type in Stroop interference and Stroop facilitation.

For Stroop interference effects, the delta plots for all three character types showed positive linear trends (Colour-Character: $F(1,205) = 59.649$, $p < 0.001$, $d = 1.19$; Valid-Radical: $F(1,205) = 13.693$, $p < 0.001$, $d = 0.57$; Invalid-Radical: $F(1,205) = 4.263$, $p < 0.05$, $d = 0.32$), and no quadratic trends (Colour-Character: $F(1,205) = 0.579$, $p = 0.448$, $d = 0.12$; Valid-Radical: $F(1,205) = 0.378$, $p = 0.539$, $d = -0.09$; Invalid-Radical: $F(1,205) = 0.126$, $p = 0.724$, $d = -0.05$).

For Stroop facilitation effects, the delta plots for the Valid- and Invalid-Radical conditions showed linear trends as the effects increased across the quantiles (Valid-Radical: $F(1,205) = 5.392$, $p = 0.021$, $d = 0.36$; Invalid-Radical: $F(1,205) = 12.699$, $p < 0.001$, $d = 0.55$), and no quadratic trends (Valid-Radical: $F(1,205) = 0.021$, $p = 0.885$, $d = -0.02$; Invalid-Radical: $F(1,205) = 0.609$, $p = 0.436$, $d = 0.12$). The delta plot for Colour-Character revealed no linear trend ($F(1,205) = 0.263$, $p = 0.609$, $d = 0.08$) nor quadratic trend ($F(1,205) = 1.459$, $p = 0.228$, $d = -0.19$).

3.2.3 Discussion

Consistent with Yeh et al.'s (2017) results, Experiment 1 showed significant interference effects for the Colour-Character, Invalid-Radical, and Valid-Radical conditions. This confirms that phono-semantic compounds are decomposed into radicals, and that phonetic radicals activate semantics. Although it is possible that phonological information provided by phonetic radicals is sufficient to cause Stroop interference effects, the source of this interference should be based on the semantic level rather than the phonological level. This is because the phonology of phonetic radicals do not conflict with the colour names. It is the semantics of phonetic radicals that related to colour names that caused interference. Interestingly, pairwise comparison revealed significant Stroop interference differences between all three character conditions except for the comparison between the Invalid-Radical and Valid-Radical conditions. Thus, similar to Yeh et al. (2017), there was no difference in Stroop interference between the Invalid-Radical and Valid-Radical conditions, which means phonetic radicals provide little or no pronunciation cues in the Stroop paradigm.

Strong Stroop facilitation effects were also observed in all three character types. Yeh et al. (2017) found facilitation effects in the Colour-Character and Invalid-Radical conditions but a marginally significant effect for the Valid-Radical condition. This difference from Yeh et al. (2017) could be explained by the increase in the number of trials used in the present experiment. Thus, with more power it becomes clear that Stroop facilitation effects also occur in the Valid-Radical condition.

The distributional analysis of Experiment 1 provided a new perspective for interpreting the results. Stroop interference effects for all three character types revealed linearly increasing trends in the delta plots. As suggested by Roelofs et al. (2011), an upward trend in delta plots reflects that no inhibition was applied to resolve the competition between colour and word information in the Stroop task, while a levelling-off showed weak inhibition was applied to keep the effects constant. In this study, no inhibition was executed in Stroop interference effects, so the effects became larger with the increase of RTs.

Stroop facilitation effects for Valid- and Invalid-Radical conditions showed linearly increasing trends, but the Colour-Character condition showed a levelling-off curve. Stroop facilitation effects are interpreted as enhancement applied to converge the colour and word information (see Section 2.6.1.2). Strong enhancement is applied when Stroop facilitation increases, and weak enhancement is applied when Stroop facilitation stays constant. The levelling-off trend in Stroop facilitation for the Colour-Character condition suggests weak enhancement was applied to converge colour and word information, while the upward curve in Stroop facilitation for Invalid- and Valid-Radical conditions shows strong enhancement. This is the benefit of using delta plots as the performance of each condition can be captured across the quantiles, which information is lost or hard to capture in the mean RT analysis.

Positive linear trends with large, medium, and small effects were observed for the Colour-Character, Valid-Radical, and Invalid-Radical conditions in Stroop interference, respectively. This indicated that it is more difficult to maintain attentional control in slower RTs, especially for the Colour-Character condition. The Valid-Radical had an identical phonology provided by phonetic radicals, and thus, its effect size of linear trend was larger than the Invalid-Radical condition. On the contrary, the Stroop facilitation for the Colour-Character reported a very small effect size, which is inconsistent with the literature and should be further investigated in the following experiments. The Invalid-Radical condition showed a better linear trend than the Valid-Radical condition.

As suggested by Neill (1977), Redding and Gerjets (1977) and White (1969), a reduced Stroop interference effect would occur when switching from vocal responses to manual responses because an overt vocal response is not required in the manual Stroop task. In contrast, there might be an increase of Stroop facilitation effect, because of the absence of overt verbal responses. Thus, in the next experiment, the manual Stroop task will be used with the same set of stimuli to examine whether there is a reduced Stroop interference effect and an increased Stroop facilitation effect. Crucially, Yeh et al. (2017) did not explore the role of phonetic radicals in the manual Stroop task. Thus, the experiment will explore whether phonetic radicals are able to activate semantics and/or phonology when no overt vocal responses are required for the task.

3.3 Experiment 2 – The Manual Stroop Task

3.3.1 Methods

3.3.1.1 Participants

Forty-two students from the same subject pool as Experiment 1 participated in Experiment 2 (mean age = 23.83, range = 21–31, females = 32), none of whom had participated in Experiment 1. The studies were approved by the ethics committee in the Department of Psychology at the University of Nottingham (Ethics approval number: S1117R).

3.3.1.2 Stimuli, Design, and Procedure

The stimuli, design and procedure were identical to Experiment 1, with the only difference being that participants performed a manual Stroop task rather than a vocal Stroop task. Twenty-four practice trials were prepared to ensure that participants were familiarized with the matching of colours and key buttons.

The number of observations per condition is identical to Experiment 1.

3.3.2 Results

3.3.2.1 Mean RT analysis

The data were analysed in the same way as in Experiment 1. Incorrect responses were discarded (2.44%), and RTs below 200 ms and above 1500 ms were also discarded before the analyses (0.87%). Fixed factor "congruency" used simple coding so that neutral trials will be compared to congruent and incongruent trials. Fixed factor "character type" also used simple coding that the Colour-Character condition will be compared to the Valid- and Invalid-Radical conditions. Fixed factor "colour" used deviation coding so that each colour will be compared to the mean of all colours.

First, different types of distributions were compared, and the results of the performance analyses revealed that the Inverse Gaussian fitted the data best. Next, four models with different fixed and random structures were compared. A summary of Information Criterion for these models is shown in Table 3.4.

Name	R2 (marg.)	AIC weights	AICc weights	BIC weights	Performance-Score
Model 3	0.136	0.83	0.83	0.122	87.64%
Model 4	0.139	0.17	0.17	0.864	53.59%
Model 1	0.107	8.80E-06	8.85E-06	0.002	26.44%
Model 2	0.11	2.01E-06	2.03E-06	0.012	2.20%

Model Structure:
Model 1: glmer(RT ~ congruency * character type + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 2: glmer(RT ~ congruency * character type + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 3: glmer(RT ~ congruency * character type + colour + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 4: glmer(RT ~ congruency * character type + colour + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

Table 3.4. Summary of Information Criterion for models used in Experiment 2.

Model 3 is the best fit for the data. Based on Model 3, a full model was constructed as follows:

> glmer(RT ~ congruency * character type + colour + (1+congruency+character type+congruency:character type|subject) + (1+congruency+character type+congruency:character type|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

However, the full model did not converge. Therefore, the random structures were simplified until the model successfully converged, and the final model was:

> glmer(RT ~ congruency * character type + colour + (1+character type|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

A summary of the final model can be found in Appendix J. The means of reaction time (RT) are presented in Table 3.5. RT differences between the three character types are presented in Figure 3.5.

		Experiment 2		
		Colour-Character	Invalid-Radical	Valid-Radical
Congruent	RT (SE)	612 (12.9)	630 (13.2)	626 (13.4)
	ER (SE)	1.19 (0.48)	2.18 (0.65)	1.98 (0.62)
Incongruent	RT (SE)	674 (13.3)	649 (13.1)	652 (13.0)
	ER (SE)	4.07 (0.62)	1.98 (0.44)	2.68 (0.51)
Control	RT (SE)	645 (12.8)	631 (13.0)	639 (13.0)
	ER (SE)	2.12 (0.37)	2.12 (0.37)	2.18 (0.38)

Table 3.5. Mean reaction times (RT, in milliseconds), error rates (ER, %), and standard errors (SE, in parentheses) as a function of congruency and character type from the GLMMs in Experiment 2.

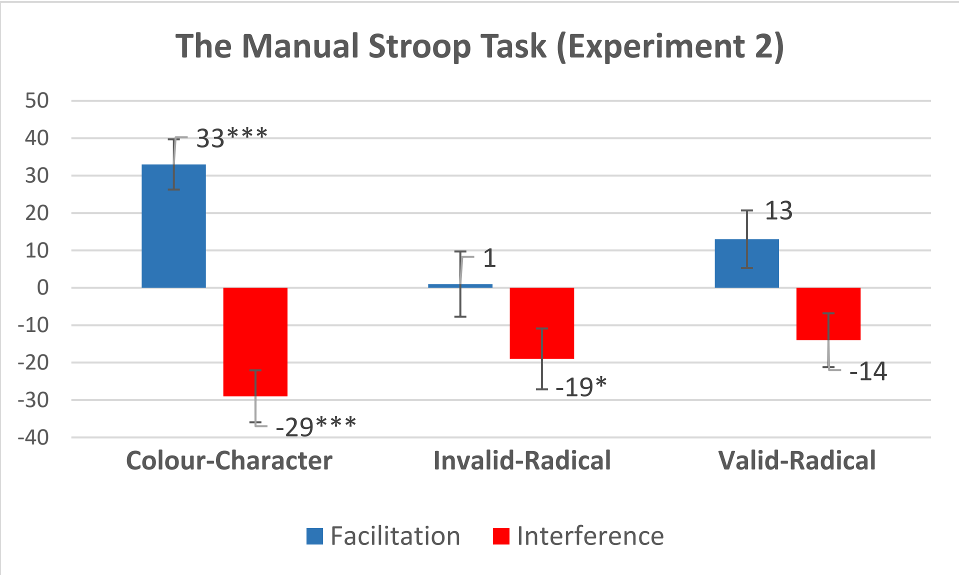


*Figure 3.5. Bar chart results of Experiment 2. RT differences = Neutral trials minus non-Neutral trials. Positive values indicate Stroop facilitation effects, and negative values refer to Stroop interference effects. Error bars represent the standard error from the mean. ***p < 0.001; **p < 0.01; *p < 0.05.*

Interference effects were found in the Colour-Character condition (M = 29 ms, SE = 6.95 ms, z = 4.160, *p* < 0.001), in the Invalid-Radical condition (M = 19 ms, SE = 8.13 ms, z = 2.278, *p* = 0.043), but not in the Valid-Radical condition (M = 14 ms, SE = 7.20 ms, z = 1.871, *p* = 0.1133). Importantly, no differences were found between these conditions (ps > 0.2). A significant facilitation effect was only found in the Colour-Character condition (M = 33 ms, SE = 6.73 ms, z = 4.868, *p* < 0.001). There were no significant facilitation effects found in the Invalid-Radical condition (M = 1 ms, SE = 8.73 ms, z = 0.149, *p* = 0.976) or in the Valid-Radical condition (M = 13 ms, SE = 7.70 ms, z = 1.654, *p* = 0.177).

3.3.2.2 Error analysis

The model for error analysis is identical to the one used in mean RT analysis, except for using the Gaussian distribution:

> lmer(ER ~ congruency * character type + colour + (1+character type|subject) + (1|item))

Error rates were similar between congruent and neutral control conditions in all three character types. Error rates were higher for the incongruent condition in the Colour-Character condition relative to the neutral control condition, but not in the Valid- and Invalid-Radical conditions (Colour-Character: M = 2.0%, SE = 0.71%, $t(17.3) = 2.753$, $p = 0.026$; Valid-Radical: M = 0.5%, SE = 0.71%, $t(17.3) = 0.7$, $p = 0.707$; Invalid-Radical: M = -0.13 %, SE = 0.71%, $t(17.3) = -0.187$, $p = 0.966$). No other effects were observed.

3.3.2.3 Distributional analysis

Quantile plots and delta plots based on untrimmed data from Experiment 2 for Stroop interference and facilitation effects were presented in Figure 3.6 and 3.7.

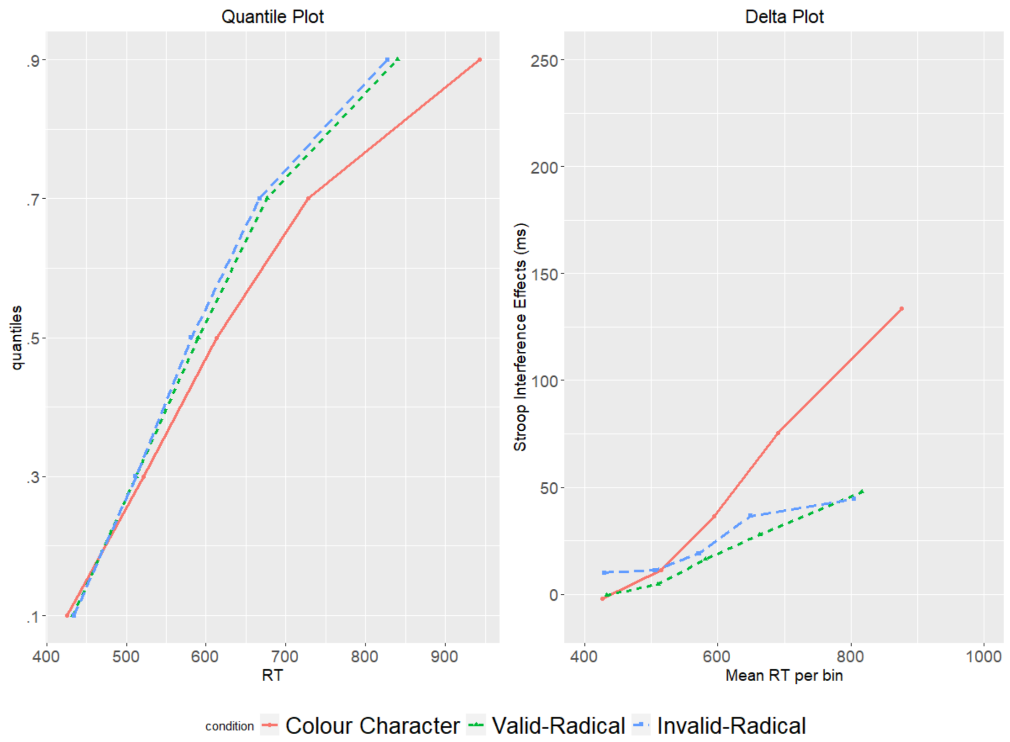


Figure 3.6. Quantile plots and delta plots for Stroop interference effects in Experiment 2.

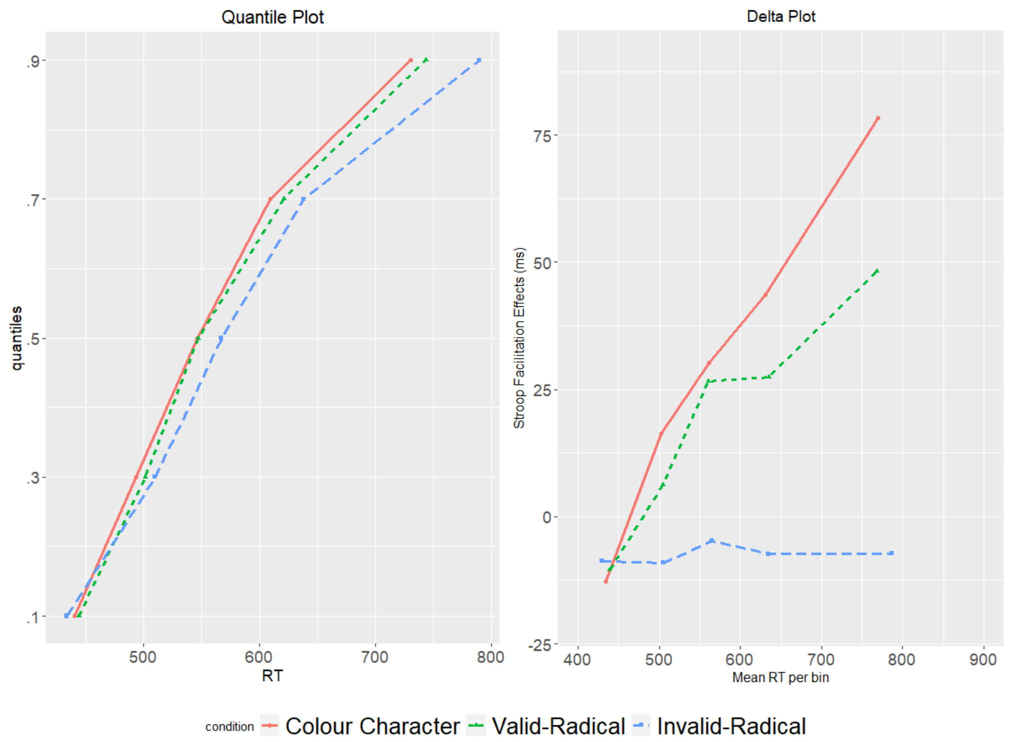


Figure 3.7. Quantile plots and delta plots for Stroop facilitation effects in Experiment 2.

Analyses were separated for Stroop interference (incongruent trials vs. neutral words) and Stroop facilitation effects (congruent trials vs. neutral words). Then, a series of one-way ANOVAs were performed for each character type in Stroop interference and Stroop facilitation.

For Stroop interference effects, the delta plots for all three character types showed positive linear trends (Colour-Character: $F(1,205) = 44.898$, $p < 0.001$, $d = 1.03$; Valid-Radical: $F(1,205) = 8.754$, $p = 0.003$, $d = 0.46$; Invalid-Radical: $F(1,205) = 5.586$, $p = 0.019$, $d = 0.36$), and a trend for a quadratic trend in the Colour-Character condition but not in the other two conditions (Colour-Character: $F(1,205) = 3.034$, $p = 0.083$, $d = 0.27$; Valid-Radical: $F(1,205) = 0.384$, $p = 0.536$, $d = 0.10$; Invalid-Radical: $F(1,205) = 0.236$, $p = 0.628$, $d = 0.07$).

For Stroop facilitation effects, the delta plots for the Colour-Character and Valid-Radical conditions showed linear trends as the effects increased across the quantiles (Colour-Character: $F(1,205) = 20.657$, $d = 0.70$, $p < 0.001$; Valid-Radical: $F(1,205) = 9.77$, $p = 0.002$, $d = 0.48$), and no quadratic trends (Colour-Character: $F(1,205) = 0.04$, $p = 0.842$, $d = 0.03$; Valid-Radical: $F(1,205) = 0.047$, $p = 0.828$, $d = -0.03$). The delta plot for the Invalid-Radical condition revealed no linear trend ($F(1,205) = 0.011$, $p = 0.918$, $d = 0.02$) nor quadratic trend ($F(1,205) = 0.012$, $p = 0.914$, $d = -0.02$).

3.3.3 Discussion

Stroop interference effects were found in the Colour-Character and Invalid-Radical conditions in the manual Stroop task but not in the Valid-Radical condition. When phonetic radicals do not correspond to the pronunciation of the whole character (the Invalid-Radical condition), they led to significant Stroop interference effects but there was no significant Stroop interference effect when they correspond to the pronunciation (the Valid-Radical condition). This indicates that phonetic radicals without corresponding pronunciation to the whole character can cause Stroop interference even though there are no explicit vocal responses required. That is, phonetic radicals can activate semantics in the manual Stroop task.

Overall, the reduced Stroop interference effects observed in the manual Stroop task compared to the vocal Stroop task are supported by the results of Neill (1977), Redding and Gerjets (1977), and White (1969). An increase in Stroop facilitation occurred only for the Colour-Character condition, as predicted by Neill (1977) and Redding and Gerjets (1977). This phenomenon was not observed for colour-related radical conditions; instead, Stroop facilitation effects disappeared when using colour-related radical conditions.

In both Experiment 1 and Experiment 2, no significant difference in Stroop interference was found between the Invalid- and Valid-Radical conditions. Although there was a trend in the vocal responses, the Valid-Radical condition (60 ms) triggered more Stroop interference than the Invalid-Radical condition (43 ms). This finding is consistent with the findings of Yeh et al. (2017). Thus, the phonological cues provided by phonetic radicals were not activated in the vocal Stroop task. The current experiment using the manual Stroop task confirmed that phonetic radicals activate phonological cues. Studies using tasks other than the Stroop task (Hue, 1992; Lee et al., 2005; Seidenberg, 1985) have reported that phonetic radicals can provide phonological information, although limited to low-frequency characters. The colour-related radical conditions in the current study were of low frequency, but no phonological information was activated when processing those characters.

The distributional analysis of Stroop interference effects revealed that the Colour-Character, Invalid-, and Valid-Radical conditions were unaffected by changing the response modality to manual responses (compared to vocal responses in Experiment 1) because they continued to show upward trends in the delta plots. Even under the condition where overt verbal responses were not required, there was no inhibition applied to resolve the competition between the colour categorization task and word reading task, resulting in greater Stroop interference effects with an increase in RTs.

For Stroop facilitation effects, the delta plot analysis revealed a levelling-off trend for the Invalid-Radical condition and linear increasing trends for the Colour-Character and Valid-Radical conditions. In mean RT analysis, the Invalid- and Valid-Radical condition produced no significant Stroop facilitation effects; however, the trend analysis revealed that strong enhancement was applied to the Valid-Radical condition to converge the colour and word information, whereas weak enhancement was applied to the Invalid-Radical condition. This difference was not observed in the vocal Stroop task (Experiment 1), where the Valid- and Invalid-Radical conditions both had strong enhancement applied.

A positive linear trend with a large effect size was observed for the Colour-Character condition, as in the vocal Stroop task. The effect size for the linearity of the Valid-Radical condition in the manual responses became smaller compared to the vocal responses and did not differ from that of the Invalid-Radical condition, which was also reflected in the mean RT analysis. For Stroop facilitation effects, the effect size for the linearity of the Colour-Character was medium, as strong facilitation effects were reflected in the mean RT analysis. The Valid-Radical condition showed a small effect in linearity.

Because phonological activation by phonetic radicals is not consistent in terms of mean RT analysis and distributional analysis, increasing the number of trials might help to clarify this effect. Furthermore, the way colour information was presented might impact the findings because the amount of colour information presented is small (integrated mode) and differs depending on the number of strokes in Chinese characters. Therefore, the next experiment will increase the number of trials and the presentation of the colour information.

3.4 Experiment 3 – The Vocal Stroop Task

3.4.1 Methods

3.4.1.1 Participants

Forty students from the same pool as Experiments 1 and 2 participated (mean age = 23.52, range = 20–31, females = 27). None of the students had participated in Experiments 1 or 2. The studies were approved by the ethics committee in the Department of Psychology at the University of Nottingham (Ethics approval number: S1206).

3.4.1.2 Stimuli, Design, and Procedure

The stimuli were identical to Experiment 1 and 2. The presentation of stimuli had the following changes. Firstly, a separate mode rather than an integrated mode was used. In this mode, the word is in black (RGB: 0,0,0), first surrounded by a grey square (RGB: 204,204,204), and then with a colour-filled rectangle that will present one of the three colours (see Figure 3.8 for an example). By separating the colour information from the character itself, the character has a better presentation style. The colour yellow was slightly adjusted for better viewing (from RGB: 255,255,0 to RGB: 240,240,0). The fixation dot "●" was now in black (RGB: 0,0,0) rather than in grey (RGB: 128,128,128).

Figure 3.8. An example of separate presentation of Stroop stimuli.

Another difference was that the total number of trials was increased from 240 to 432. The colour patches were deleted, so the number of trials in each block has decreased from 60 to 54. The number of blocks has been increased from 4 to 8. The stimuli and procedure remained identical to Experiment 1.

The experiment is properly powered with 2,808 observations for congruent trials (39 participants × 8 blocks × 9 trials), 5,616 for incongruent trials (39 participants × 8 blocks × 18 trials), and 8,424 for neutral trials (39 participants × 8 blocks × 27 trials).

3.4.2 Results

3.4.2.1 Mean RT analysis

One participant was excluded due to high error rates (22%). Incorrect responses were discarded (3.43%). RTs below 200 ms and above 1500 ms were also discarded before the analyses (0.76%). Fixed factor "congruency" used simple coding so that neutral trials will be compared to congruent and incongruent trials. Fixed factor "character type" also used simple coding that the Colour-Character condition will be compared to the Valid- and Invalid-Radical conditions. Fixed factor "colour" used deviation coding so that each colour will be compared to the mean of all colours.

First, different types of distributions were compared, and the results of the performance analyses revealed that the Inverse Gaussian fitted the data best. Next, four models with different fixed and random structures were compared. A summary of Information Criterion for these models is shown in Table 3.6.

Name	R2 (marg.)	AIC weights	AICc weights	BIC weights	Performance-Score
Model 4	0.634	1	1	1	87.50%
Model 1	0.436	2.10E-165	2.10E-165	9.82E-164	12.88%
Model 2	0.437	2.47E-165	2.48E-165	5.40E-162	7.49%
Model 3	Fail to converge				

Model Structure:
Model 1: glmer(RT ~ congruency * character type + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 2: glmer(RT ~ congruency * character type + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 3: glmer(RT ~ congruency * character type + colour + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 4: glmer(RT ~ congruency * character type + colour + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

Table 3.6. Summary of Information Criterion for models used in Experiment 3.

Model 4 is the best fit for the data. Based on Model 4, a full model was constructed as follows:

glmer(RT ~ congruency * character type + colour + (1+congruency+character type+congruency:character type|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

However, the full model did not converge. Therefore, the random structures were simplified until the model successfully converged, and the final model was:

glmer(RT ~ congruency * character type + colour + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

A summary of the final model can be found in Appendix K. The means of reaction time (RT) are presented in Table 3.7. RT differences between the three character types are presented in Figure 3.9.

		Experiment 3		
		Colour-Character	**Invalid-Radical**	**Valid-Radical**
Congruent	RT (SE)	612 (11.3)	623 (11.2)	604 (10.8)
	ER (SE)	0.43 (0.21)	1.18 (0.35)	1.18 (0.35)
Incongruent	RT (SE)	716 (10.6)	676 (11.5)	686 (11.3)
	ER (SE)	7.64 (0.61)	3.95 (0.45)	6.30 (0.56)
Control	RT (SE)	657 (11.0)	652 (10.8)	650 (10.9)
	ER (SE)	2.53 (0.30)	2.46 (0.29)	2.74 (0.31)

Table 3.7. Mean reaction times (RT, in milliseconds), error rates (ER, %), and standard errors (SE, in parentheses) as a function of congruency and character type from the GLMMs in Experiment 3.

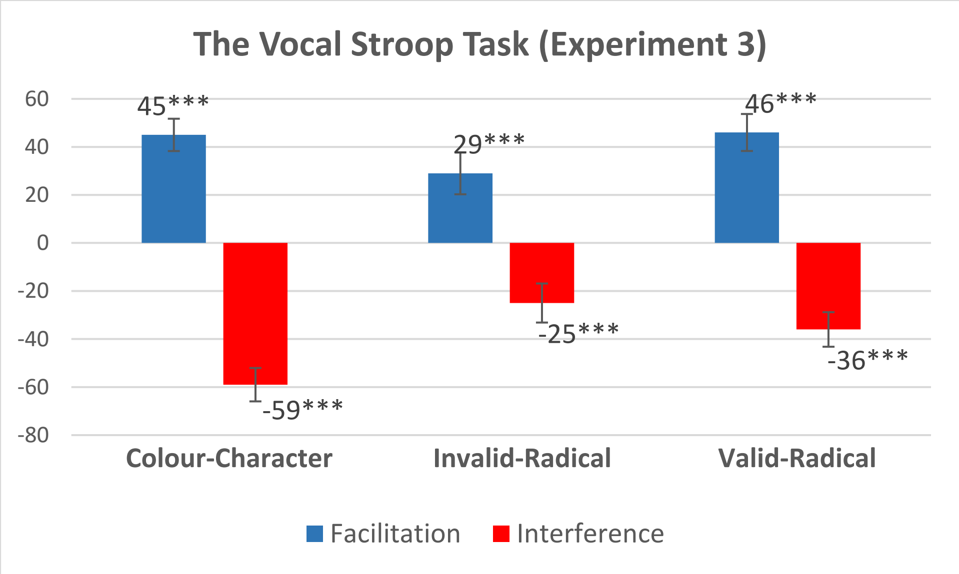


*Figure 3.9. Bar chart results of Experiment 3. RT differences = Neutral trials minus non-Neutral trials. Positive values indicate Stroop facilitation effects, and negative values refer to Stroop interference effects. Error bars represent the standard error from the mean. ***p < 0.001.*

Compared to neutral control conditions, interference effects were found in all three conditions (Colour-Character condition: M = 59 ms, SE = 4.01 ms, z = 14.621, $p < 0.001$; Invalid-Radical condition: M = 25 ms, SE = 4.05 ms, z = 6.058, $p < 0.001$; Valid-Radical condition: M = 36 ms, SE = 4.07 ms, z = 8.944, $p < 0.001$). Pairwise comparisons showed that interference effects were stronger in the Colour-Character condition (59 ms) than in the Invalid-Radical condition (25 ms; difference = 34 ms, SE = 5.51 ms, z = 6.197, $p < 0.001$) and in the Valid-Radical condition (36 ms; difference = 23 ms, SE = 5.19 ms, z = 4.298, $p < 0.001$). The Valid-Radical condition (36 ms) showed a stronger interference effect than the Invalid-Radical condition (25 ms; difference = 11 ms, SE = 5.54 ms, z = 2.138, $p = 0.033$).

Significant facilitation effects were found in all three conditions (Colour-Character condition: M = 45 ms, SE = 3.93 ms, z = 11.475, $p < 0.001$; Invalid-Radical condition: M = 29 ms, SE = 4.56 ms, z = 6.247, $p < 0.001$; Valid-Radical condition: M = 46 ms, SE = 4.37 ms, z = 10.536, $p < 0.001$). Pairwise comparisons were significant for the differences between the Colour-Character and Invalid-Radical conditions and between the Valid-Radical and Invalid-Radical conditions in Stroop facilitation effects (Colour-

Character vs. Invalid-Radical: M = 16 ms, SE = 5.69 ms, z = 2.928, p = 0.01; Valid-Radical vs. Invalid-Radical: M = 17 ms, SE = 6.23 ms, z = 2.814, p = 0.01).

3.4.2.2 Error analysis

The model for error analysis is identical to the one used in mean RT analysis, except for using the Gaussian distribution:

lmer(ER ~ congruency * character type + colour + (1|subject))

Error rates were similar between congruent and neutral control conditions in all three character types, except for the Colour-Character condition (M = 2.1%, SE = 0.72%, $t(2057) = 2.913$, $p = 0.007$). Error rates were higher for the incongruent condition in all three character types relative to the neutral control condition (Colour-Character: M = 5.1%, SE = 0.57%, $t(2057) = 8.961$, $p < 0.001$; Valid-Radical: M = 3.6%, SE = 0.57%, $t(2057) = 6.245$, $p < 0.001$; Invalid-Radical: M = 1.5%, SE = 0.57%, $t(2057) = 2.623$, $p = 0.017$).

3.4.2.3 Distributional analysis

Quantile plots and delta plots based on untrimmed data from Experiment 3 for Stroop interference and facilitation effects are presented in Figure 3.10 and 3.11.

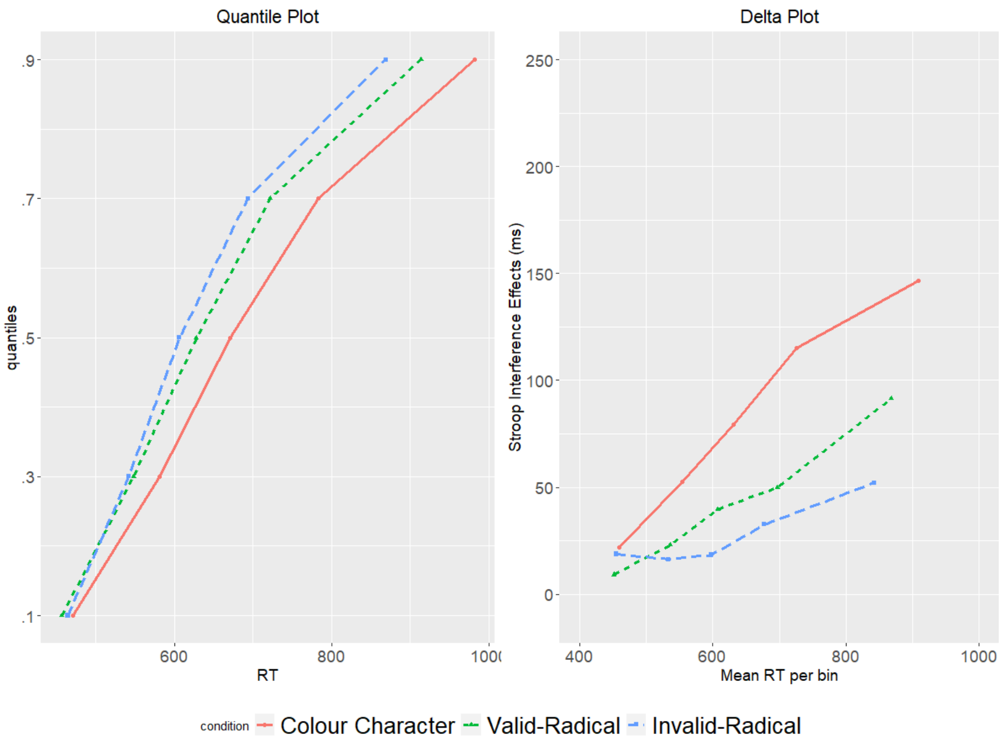


Figure 3.10. Quantile plots and delta plots for Stroop interference effects in Experiment 3.

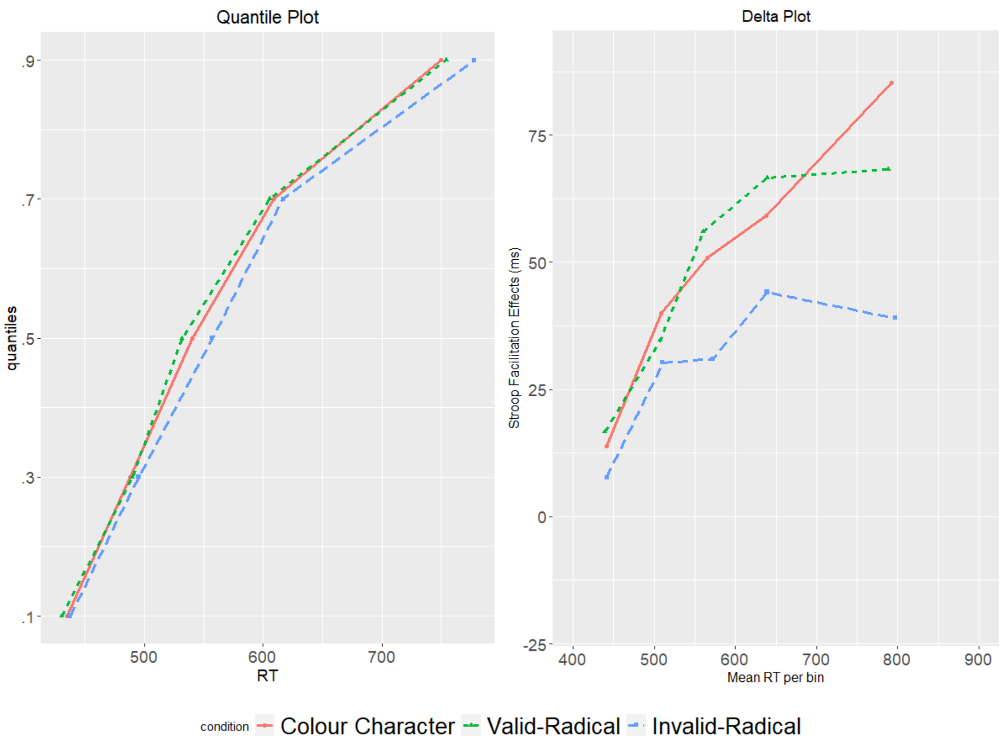


Figure 3.11. Quantile plots and delta plots for Stroop facilitation effects in Experiment 3.

The analyses were separated for Stroop interference (incongruent trials vs. neutral words) and Stroop facilitation effects (congruent trials vs. neutral words). Then, a series of one-way ANOVAs were performed for each character type in Stroop interference and Stroop facilitation.

For Stroop interference effects, the delta plots for all three character types showed positive linear trends (Colour-Character: $F(1,190) = 43.544$, $p < 0.001$, $d = 1.06$; Valid-Radical: $F(1,190) = 27.482$, $p < 0.001$, $d = 0.84$; Invalid-Radical: $F(1,190) = 9.949$, $p = 0.002$, $d = 0.51$), and a trend for quadratic trend in the Invalid-Radical condition but not in the other two (Colour-Character: $F(1,190) = 0.038$, $p = 0.846$, $d = 0.03$; Valid-Radical: $F(1,190) = 1.307$, $p = 0.254$, $d = 0.18$; Invalid-Radical: $F(1,190) = 3.257$, $p = 0.073$, $d = 0.29$).

For Stroop facilitation effects, the delta plots for all three character types showed linear trends as the effects increased across the quantiles (Colour-Character: $F(1,190) = 24.749$, $p < 0.001$, $d = 0.80$; Valid-Radical: $F(1,190) = 19.12$, $p < 0.001$, $d = 0.70$; Invalid-Radical: $F(1,190) = 6,185$, $p = 0.014$, $d = 0.40$), and no quadratic trends (Colour-Character: $F(1,190) = 0.006$, $p = 0.938$, $d = -0.01$; Valid-Radical: $F(1,190) = 1.417$, $p = 0.235$, $d = -0.19$; Invalid-Radical: $F(1,190) = 1.393$, $p = 0.239$, $d = -0.19$).

3.4.3 Discussion

Experiments 1 and 3 are both vocal Stroop tasks. Compared to Experiment 1, Experiment 3 again found significant interference effects for all three character conditions. The numeric values of Stroop interference effects observed in Experiment 3 almost halved compared to Experiment 1. This could be the result of the new presentation mode (separated mode) which makes it easier to discern colour and word information when they are not overlapping in a separate presentation mode.

The critical difference between the Valid-Radical and Invalid-Radical conditions was significant (Stroop interference: $p = 0.033$; Stroop facilitation: $p = 0.01$). The phonological activation of phonetic radicals was supported by the results of Experiment 3.

Stroop facilitation effects were again observed in all three character types, as in Experiment 1, although there was no difference between each character type.

The delta plot analysis of Stroop interference effects showed the same pattern as in Experiment 1. In both vocal Stroop tasks, all three character types showed upward trends with the increase of RTs. The difference lies in the Stroop facilitation effects, where the Colour-Character showed weak enhancement in Experiment 1 and strong enhancement in Experiment 3. Because more trials were introduced in Experiment 3, it is assumed that the Stroop facilitation effects were stabilized for the Colour-Character condition in Experiment 3, and this was reflected in the trend analyses.

The effect size of linearity in Colour-Character condition of Stroop interference effects remained large as in Experiment 1 and 2. The Valid-Radical condition also had a larger effect size than the Invalid-Radical condition as in Experiment 1. The effect size for each character type of Stroop facilitation effects followed the same pattern as in Stroop interference effects.

3.5 General Discussion

3.5.1 Semantic activation of phonetic radicals

In the present study, strong interference effects were observed for the Valid- and Invalid-Radical conditions in the vocal Stroop tasks (Experiments 1 & 3). This is consistent with the findings of Yeh et al. (2017) who found that phonetic radicals can activate semantic information. The current study used a manual Stroop task (Experiment 2) to examine whether phonetic radicals can activate semantics even though no overt vocal response was required for the task. The results showed that only phonetic radicals without corresponding pronunciation of the whole character (i.e. Invalid-Radical condition) activate semantics in the manual Stroop task.

The above conclusion challenges conventional thought that phonetic radicals carry only phonological information and no semantic information. A possible explanation for this could be that the contribution of meaning/pronunciation was unequal from sematic and phonetic radicals. Semantic radicals were relatively few in number but highly combinable. A disadvantage of high combinability is that semantic radicals cannot always convey the exact meaning. For example, the semantic radical "犭" denotes animals. This meaning was preserved in characters like "狼" (wolf, pronounced as /láng/), "狮" (lion, /shī/), "猫" (cat, /māo/) and so on. However, characters like "猜" (guess, /cāi/) and "独" (lonely, /dú/) no longer have clear meaning related to animals. In addition, it can be seen from these examples that semantic radicals do not usually determine how those phono-semantic compounds are pronounced, whereas phonetic radicals often do. Phonetic radicals, on the other hand, are less combinable but more recognizable and larger in number: "良" (kind, /liáng/) in "狼", "师" (teacher, /shī/) in "狮", and "苗" (seed, /miáo/) in "猫". Zhou (2003) calculated that about 81% of phonetic radicals can be used as freestanding characters, therefore most phonetic radicals have their own meaning and pronunciation. Furthermore, about 38% of phonetic radicals are pronounced the same as their phono-semantic compounds (Zhou, 1978). In sum, the way phonetic radicals provide the meaning and pronunciation cues is more reliable than semantic radicals. Phonetic radicals are less combinable but more recognizable as freestanding characters, and thus, facilitate the activation of their meaning and pronunciation.

3.5.2 Phonological activation of phonetic radicals

The activation of phonology by phonetic radicals was supported by the current study. This role was investigated by pairwise comparison of Stroop interference effects between the Valid- and Invalid-Radical conditions because only one condition provides an identical pronunciation to the Colour-Character and the other does not. In Experiment 1, there were tendencies that more Stroop interference was caused by the Valid-Radical condition. In Experiment 3, with more trials introduced, this difference became significant. In addition, Stroop facilitation effects between the Valid- and Invalid-Radical conditions were also significant in Experiment 3. This does not agree with the findings of Yeh et al. (2017), where they found no difference between the Valid- and Invalid-Radical conditions. With the manual responses (Experiment 2), strong Stroop interference effects were observed in the Invalid- but not in the Valid-Radical condition, which confirms the phonological activation in phonetic radicals.

3.5.3 Response modality effects

Although Experiment 1 and Experiment 2 are between-subject design, it is clear from the numerical values that Stroop interference effects reduced dramatically when shifting from vocal to manual responses. Such reduction in Stroop effects was called the response modality effect (Neill, 1977; Redding & Gerjets, 1977; White, 1969).

In addition, the radical-related conditions also showed a different pattern for vocal and manual responses: significant Stroop interference effects were found for the Valid-Radical conditions in the vocal but not in the manual responses. This suggests the semantic information of phonetic radicals can be activated in the vocal responses but not in the manual responses.

The results of Stroop facilitation effects provides a clearer pattern: the radical-related conditions (i.e. the Valid- and Invalid-Radical conditions) do not elicit strong facilitation effects in the manual responses; however, strong facilitation effects were observed in the vocal responses. The observed Stroop interference and facilitation effects suggest smaller effects found in the manual responses compared to vocal responses.

3.6 Conclusion

The data reported in this chapter confirmed that phonetic radicals activate semantics in a vocal Stroop task (as in Yeh et al.'s study, 2017), and more importantly, semantics are also activated in the manual Stroop task in which no explicit activation of phonology is required (no overt vocal response needed). The activation of phonology by phonetic radicals was present in both the vocal Stroop and the manual Stroop task.

The distributional analysis showed that radical-related conditions have no inhibition applied to resolve the colour and word information as the Colour-Character condition.

One limitation to the experiments in this chapter is that they are between-subject design, which means it is difficult to compare the results of vocal and manual responses directly. The next chapter introduces a within-subject design Stroop task when investigating the distinct components in Stroop effects and the response modality effect.

Chapter 4: Distinct Components of Stroop effects

4.1 Introduction

The vocal Stroop task was used in Stroop's (1935) original experiments as explained in the Introduction of this thesis (Chapter 2). Participants in Stroop's original experiment had to name aloud the ink colour of incongruent colour words (e.g. red in green colour) and solid squares in different colours. The solid squares were used as a control condition. The differences in reaction times between those two conditions are referred to as Stroop interference. To investigate whether the reaction time differences result from either simultaneous memory retrieval or conflict from two sources (colour and word information), Keele (1972) conducted a Stroop task with colour words, non-colour words, scrambled letters, mixed letter-like forms, and pure Gibson forms. Results showed that non-colour words were processed quicker than colour words, but not different from control conditions (e.g. non-words or symbols). Non-colour words must have influenced memory retrieval processes because responses to incongruent colour words revealed interference effects. Colour words responses were affected by competition between colour and word information. Keele (1972) did not find response time differences between nonword controls and non-colour words in the manual Stroop task; however, Klein (1964) observed a difference between nonsense syllables and common words in the vocal Stroop task. This means the response modality can affect Stroop effects. The response modality effect refers to the difference in Stroop effects (e.g. interference and facilitation) between manual and vocal responses.

Many studies have reported that vocal responses resulted in larger Stroop interference effects than manual responses (Augustinova et al., 2019; Fennell & Ratcliff, 2019; Neill, 1977; Redding & Gerjets, 1977; Sharma & McKenna, 1998; Zahedi et al., 2019). Kinoshita et al. (2017) argued that vocal and manual response modes of the Stroop task require different mechanisms that contribute differently to Stroop effects because the vocal Stroop task is a naming task, whereas the manual Stroop task is a categorization task. The modality effects can be understood in terms of the broader discussion of Stimulus-Response Compatibility (Kornblum, 1992). In a vocal Stroop task, the stimulus and the response are directly connected (e.g. the word "red" written with the ink colour red). In contrast, in a manual Stroop task, pressing a response key is arbitrarily associated with a particular colour. Thus, manual responses require an extra step of stimulus-response mappings that associate the colour of words to the corresponding response keys.

The difference between the vocal and manual response modality is crucial for interpreting the Stroop effects. In the traditional single-stage accounts (see also Section 2.2), the response competition is a goal-referenced selection of verbal actions, rather than a "blind" selection of responses that reach first (Roelofs, 2003). Stroop interference, as Roelofs (2003) stated, "lies within the language production system. Interference should remain if lexical entries are needed to mediate a button-press response" (p. 115). This assumes that the Stroop interference occurs solely from response competition and that the lexical processing of words and colours is not involved in this process. Furthermore, according to Roelof's model (see Section 2.4), the outcomes for vocal and manual responses should be the same because response modality is mediated by lexical entries, and lexical entries do not affect the magnitude of the Stroop interference.

In contrast, multi-stage accounts of Stroop effects (Augustinova & Ferrand, 2014; Neely & Kahan, 2001; Schmidt & Cheesman, 2005; Sharma & McKenna, 1998) argue that the locus of Stroop effects is different for vocal and manual responses. These accounts assume that there are multiple simultaneous sources of conflicts underlying Stroop interference effects. Support for these accounts comes, for example, from Klein (1964) who reported a "semantic gradient" effect in the Stroop task. A semantic gradient refers to the semantic relationship between words and colours. For example, colour associated words (e.g. SKY) are associated with a particular colour word (e.g. BLUE), whereas neutral words are not associated with a specific colour (e.g. SEAT). Based on Klein (1964) and Neely & Kahan's (2001) work, Augustinova et al. (2018) extended the *semantic Stroop paradigm* to investigate task conflicts.

The idea that task conflicts play a role in the vocal Stroop task originates from the automaticity view of the Stroop task. According to this view, word reading is automatic, whereas ink colour naming is not (see Macleod, 1991 for a detailed review of the automaticity view). In the vocal Stroop, the relevant task is colour naming, and the irrelevant task is word reading. The tendency to read words instead of naming the ink colour of words produced task conflicts (Goldfarb & Henik, 2006, 2007; Kalanthroff et al., 2013). Therefore, using word stimuli (e.g. colour words, colour associated words, neutral words) in the Stroop task leads to task conflict because of the automaticity of word reading.

To measure the magnitude of task conflicts, Augustinova et al. (2018) contrasted colour neutral stimuli with a row of Xs (e.g. XXXX). Neutral words were expected to trigger word reading, whereas a row of Xs is unpronounceable and meaningless and would therefore not trigger word reading (see Figure 4.1).

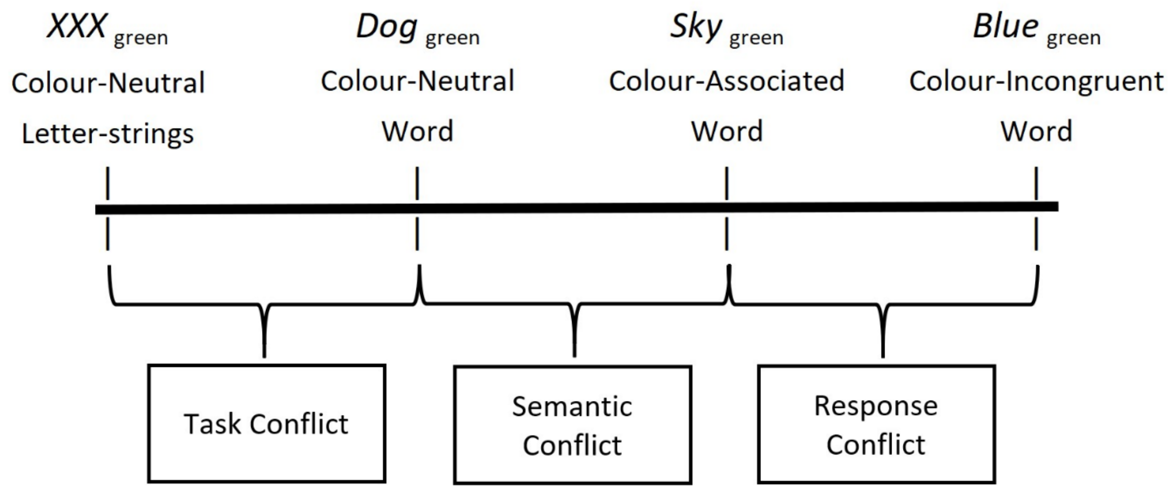


Figure 4.1. The subtractive logic of semantic Stroop interference effects. Adapted from Augustinova et al. (2018).

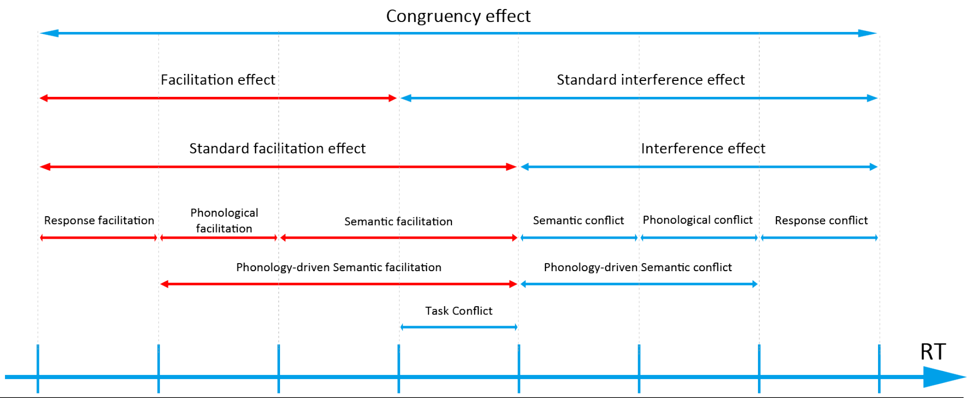
Semantic conflict is measured by subtracting the reaction times of incongruent colour associated words (e.g. *LAKE*$_{\text{green}}$, word Lake written on a blue colour) from neutral words (e.g. *SEAT*$_{\text{green}}$). Augustinova et al. (2018) argued, just like Roelofs (2003) that colour-associated words do not activate (pre)-motor responses linked to the associated colour, because the word *LAKE* does not exist in the response set, whereas the word *BLUE* activates (pre-)motor responses to *BLUE*. Therefore, responses to incongruent colour words will be slower than to incongruent colour associated words. The difference between colour-associated words and incongruent colour words is thus called response conflict.

Augustinova et al. (2018) found strong semantic and response conflicts in both the manual and vocal Stroop tasks, whereas task conflicts occurred only in the vocal Stroop task. However, the response modality was a between-subject factor in their study. In order to compare those conflict components in the manual and vocal Stroop tasks, Augustinova et al. (2019) used a within-subject design. Results showed that response and semantic conflicts were present in both the vocal and manual Stroop tasks. Task conflicts, however, appeared only in the vocal task. Furthermore, response and task conflicts were stronger in the vocal than in the manual task (response modality effect) due to a reduction of the contribution of response and task conflicts. In contrast, semantic conflicts were similar in both response modalities, so they did not contribute to the response modality effect. These findings suggested that Stroop effects are not solely due to response competition as suggested by Roelofs (2003).

In addition to semantic, response, and task conflicts, the role of phonological conflicts in the Stroop task has also been studied. Besner and Stolz (1998) included pseudohomophones of colour words: "wred", "bloo", "yeloe", "grene" in a manual Stroop task. The results revealed significant Stroop interference effects for pseudohomophones relative to the baseline (xxxx) and neutral words, indicating that phonology is automatically activated in the Stroop task. Further evidence was provided by Parris et al.

(2019) who used words with phonemic overlap at the initial and end position of colour words (e.g. “rack” to “red” is initial phonemic overlap, “cud” to “red” is end phonemic overlap) in both vocal and manual Stroop tasks. Results confirmed the role of phonology in Stroop facilitation. Furthermore, vocal responses resulted in greater Stroop facilitation than manual responses. However, Parris et al.’s (2019) study did not investigate the role of phonology in Stroop interference and how it is potentially modulated by response modality.

Parris et al. (2022) argued that measuring phonological conflicts using pseudohomophones may be confounded by orthographic overlap with the base words (e.g. bloo vs. blue). Thus, to avoid this confound, it would be best to use heterographic homophones with no orthographic overlap to investigate phonological conflicts. There are a few heterographic homophones or pseudohomophones in English; however, there are many homophones in Chinese. For example, in Chinese there are many homophones of the colour word green “绿” (pronounced as /lǜ/), which have totally different meanings and orthography: “虑” (consider), “律” (restrain), “率” (rate), “滤” (filter) and so on. Thus, in a language with plenty of homophones, phonological conflicts should be more prevalent.



Conditions	Congruent	Homophone congruent	Semantic-associative congruent	Neutral Sign	Neutral Word	Semantic-associative incongruent	Homophone incongruent	Incongruent
Examples in Chinese	蓝	岚	天	xxxxx	奖	草	虑	绿
English Translation	blue	mist	sky	xxxxx	prize	grass	consider	green
Properties	+ Word Reading + Phonological + Semantic + Response	+ Word Reading + Phonological (+ Semantic)	+ Word Reading + Semantic		+ Word Reading	+ Word Reading + Semantic	+ Word Reading + Phonological (+ Semantic)	+ Word Reading + Phonological + Semantic + Response

Figure 4.2. Decomposed Stroop conflict/facilitation components in Chinese characters.

The Stroop paradigm used by Augustinova et al. (2018) can be extended with additional conditions to investigate phonological effects as well. Figure 4.2 illustrates the Stroop condition based on Chinese stimuli that can be utilized to investigate the different components of Stroop interference and facilitation. Comparisons between the different conditions make it possible to investigate conflict and facilitation components. Task conflicts can be measured by the RT difference between neutral signs and neutral words as neutral signs do not involve word reading whereas neutral words trigger word reading. Colour-associated words can activate the semantic representation of colour words, which neutral words do not; thus, leading to semantic conflict and facilitation. The RT difference between colour-associated words and homophones in Chinese is assumed to be phonological conflict. In Chinese, homophones activate the phonology of colour words. However, this phonological activation is also assumed to activate the semantics of colour words because otherwise there would be no impact on Stroop performance. Therefore, as indicated in the "Properties" of Figure 4.2, the difference between colour-associated words and homophones is phonological conflict. Similarly, it is assumed that response conflict and facilitation should be measured by contrasting homophones with colour words.

Some of the conditions illustrated in Figure 4.2 have been used in previous Stroop research with Chinese words. Spinks et al. (2000) used Chinese homophones and colour-associated Chinese characters to investigate the role of phonology and semantics in a vocal Stroop task. The experimental design was very similar to that of Augustinova et al.’s (2018; 2019). However, unlike Augustinova et al. (2019), Spinks et al. (2000) did not distinguish between task, response, semantic and phonological conflicts. Although it is not possible to analyse whether the conditions significantly differ because the raw data was not provided, a numerical estimation of each conflict component is possible. The impact of phonology in the manual Stroop task with Chinese characters was also investigated in an ERP study by Wang et al. (2010). The stimuli in this study consisted of colour characters, colour-associated characters, homophones, and neutral characters. Similar to Spinks et al.’s (2000) study, this study also did not report the different types of conflicts, and the raw data is not available. The different components illustrated in Figure 4.2 can be calculated from the mean RTs provided in both studies and the numerical differences can be compared to Augustinova et al.’s (2019). The results of these calculations are presented in Figures 4.3 (conflict components) and 4.4 (facilitation components).

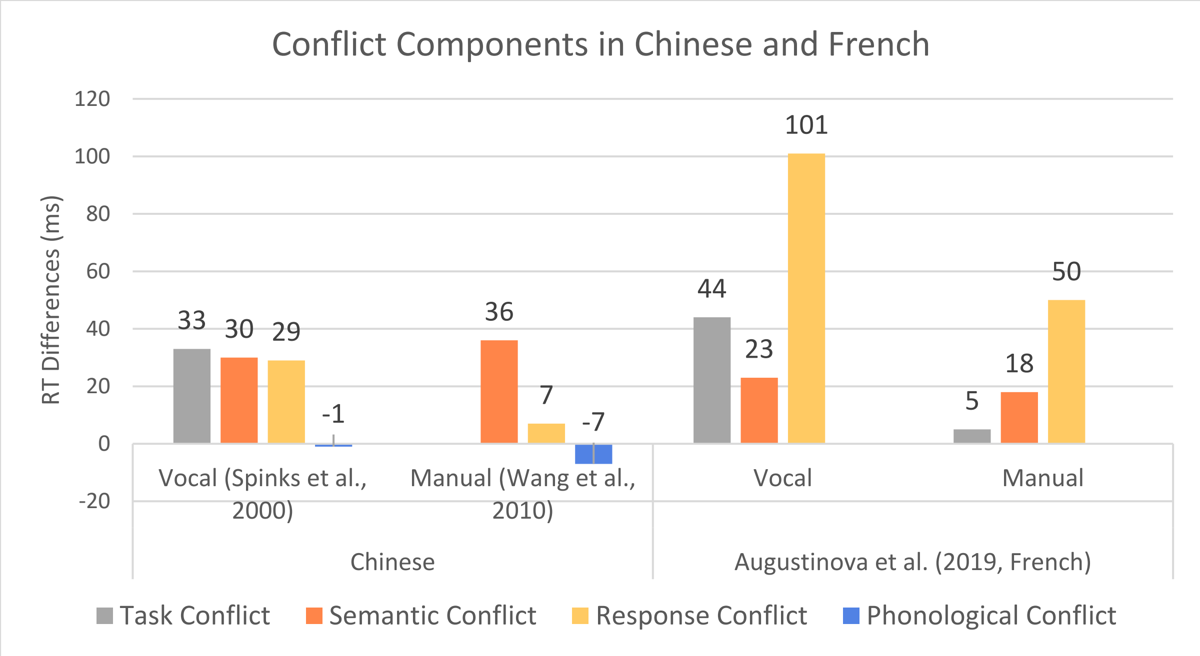


Figure 4.3. Conflict components in Chinese and French in vocal and manual Stroop tasks.

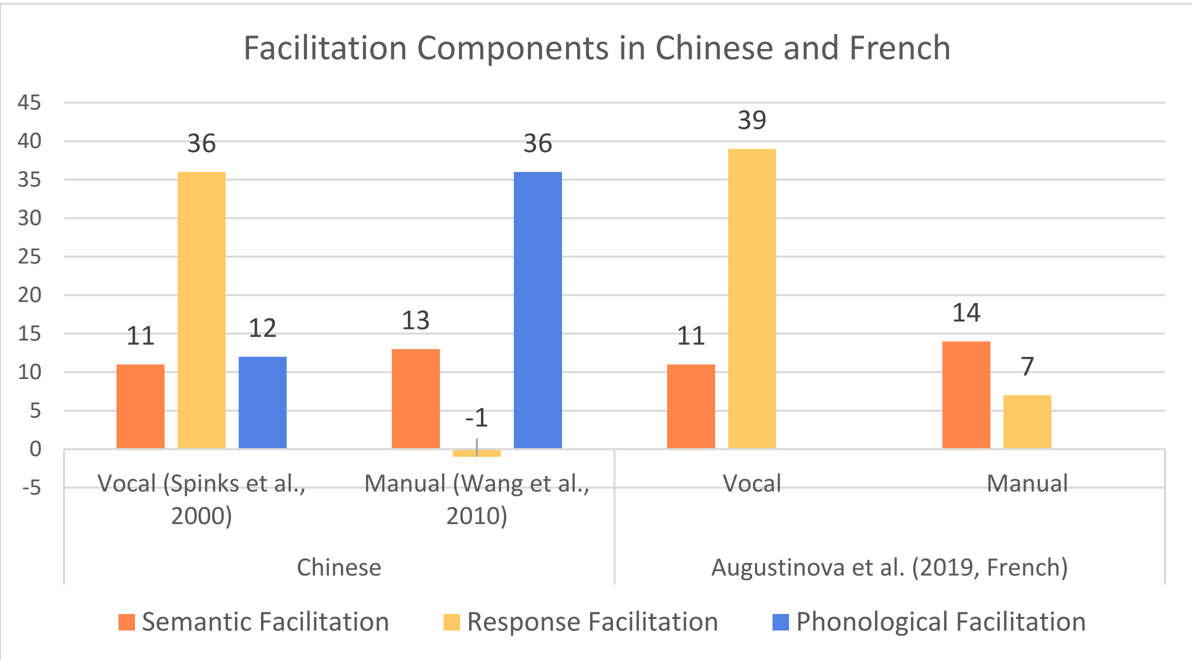


Figure 4.4. Facilitation components in Chinese and French in vocal and manual Stroop tasks.

As can be seen in the bar charts of Figure 4.3, the numerical differences seem to suggest that there might be semantic conflicts in both vocal and manual Stroop tasks with Chinese characters. Response conflicts are reduced from vocal to manual responses, as is the case in French. Furthermore, it is unclear whether there are phonological conflicts in the data from Spinks et al. (2000) and Wang et al. (2010).

For facilitation components, response facilitation is greatly reduced for manual responses in both languages. Furthermore, a numerically stronger phonological facilitation is observed in manual responses compared to vocal responses.

Stroop studies with Chinese stimuli have explored phonological components using homophones. However, they did not decompose Stroop effects into distinct components, so the magnitude of those phonological components remains unknown. Most importantly, as far as I am aware there are no Stroop studies with Chinese that used a within-subject design to study vocal and manual Stroop tasks.

The present study investigates for the first time the role of phonological conflicts and facilitation in Stroop interference and facilitation in a vocal and manual Stroop task with Chinese stimuli using a within-subject design. The findings will also be compared with Augustinova et al. (2019) who investigated task, semantic and response conflicts in an alphabetic language (French) to investigate the impact of script. Semantic conflicts were not affected by response modality in Augustinova et al. (2019). However, task conflicts and response conflicts were affected by response modality. Based on the numerical differences from Spinks et al. (2000) and Wang et al. (2010), it is expected that similar results would be found in the current study because those components should not be language-specific. Phonological conflicts are predicted to be larger in vocal than in manual responses because the vocal responses will trigger more phonological activation. Similarly, for Stroop facilitation, semantic facilitation may not be affected by response modalities, but response facilitation was affected. Phonological facilitation would be varied by response modality as well.

4.2 Methods

4.2.1 Participants

Forty participants were recruited from the University of Nottingham, UK (mean age = 24.68, range = 21–33, females = 29). All were native speakers of Chinese from mainland China. Participants had normal or corrected-to-normal vision. All participants signed an informed consent form prior to data collection and received an inconvenience allowance for participating in the experiment. The experiment was approved by the ethics committee at the School of Psychology, University of Nottingham (Ethics approval number: S1206R).

4.2.2 Stimuli and Design

The stimuli were adapted from Spinks et al.'s (2000) study. Only the colour words green, yellow, and blue were included in the experiment because the colour red (/hóng/) and yellow (/huáng/) used by Spinks et al. (2000) are pronounced very similarly. The control character 华 (/huá/, magnificent) was changed to 炭 (/tan4/, charcoal) because its pronunciation alliterates with the colour yellow (/huáng/).

The following character types were included in the task: Colour-Characters, Colour-Associated words, and Homophones. The Colour-Character condition contained characters that referred directly to colours (e.g. "绿", /lǜ/, meaning "green"). The Colour-Associated condition consisted of characters that are associated with the colour word (e.g. "草", /cao3/, "grass" is associated with the colour green). The Homophones condition consisted of characters with the same pronunciation as colour words (e.g. "虑", /lü4/, "consider" is pronounced the same as "绿", /lǜ/, "green" but has a completely different meaning). Table 4.1 presents the stimuli used in Experiment 4. Detailed linguistic properties of these stimuli can be found in Appendix D.

Condition	Character	Meaning	Pronunciation
Colour words			
	绿	green	/lü4/
	黄	yellow	/huang2/
	蓝	blue	/lan2/
Associated Colour words			
	草	grass	/cao3/
	金	golden	/jin1/
	天	sky	/tian1/
Homophone words			
	虑	ponder	/lü4/
	皇	emperor	/huang2/
	岚	mist	/lan2/
Colour Neutral words			
	贯	pass through	/guan4/
	奖	prize	/jiang3/
	炭	charcoal	/tan4/

Table 4.1. Stimuli used in Experiment 4.

Each character type was either congruent (e.g. blue in blue colour) or incongruent (e.g. blue in green colour) with the surrounding coloured rectangle. Neutral words were matched with character stimuli in terms of frequency and stroke count. The neutral sign was a percent sign (%) that had the same length as a one-word character.

The experiment used an 8 (stimulus type: colour-character incongruent, colour-character congruent, colour-associated incongruent, colour-associated congruent, homophone-incongruent, homophone-congruent, neutral word, neutral sign) × 2 (response modality: vocal, manual) within-subject design. The order of the two response modalities was counterbalanced across participants. There were 48 trials in each condition. In total, there were 384 trials in each response modality (768 trials for each participant), and these were divided into eight blocks of 48 trials. In each block, each condition was repeated twice (8 conditions × 3 colours × 2 repetitions = 48 trials).

The experiment is properly powered with 1,920 observations for each condition and each response modality (40 participants × 8 blocks × 6 trials).

4.2.3 Procedure

The stimuli were presented on a 24-inch LCD monitor (refresh rate: 120 Hz) using DMDX software (Forster & Forster, 2003). The stimuli were presented in black (RGB: 0,0,0), surrounded by a grey square (RGB: 204,204,204) and with a colour-filled rectangle representing one of three colours: green (RGB: 0,255,0), yellow (RGB: 240,240,0) and blue (RGB: 0,11,255). The background colour was also set to grey (RGB: 204,204,204). Characters presented on the screen used the Kai font (楷书). A fixation dot "●" (RGB: 0,0,0) was used because strokes of some Chinese characters may overlap with the "+" sign (e.g. the character "草" and "皇" contain cross or cross-like strokes).

Participants were tested individually in a dimly lit room. They were asked to either name the ink colour of the characters (vocal Stroop task) or press the key on the keyboard that corresponded with the correct ink colour of the characters presented on the screen (manual Stroop task). In the vocal response modality, a microphone with a voice key was used to measure the naming response latencies and to record the naming response. The accuracy and latencies of responses were subsequently checked using the CheckVocal program (Protopapas, 2007). In each trial, a fixation dot was shown for 500 ms, followed by a blank screen for 300 ms. The target character then appeared on the screen for a maximum of 3000 ms or until the participant responded. Characters disappeared as soon as the participant responded. There was an intertrial interval of 1000 ms. For the vocal Stroop task, participants first conducted 32 practice trials. For the manual Stroop task, 128 key-matching practice trials were used to train participants in the key-colour correspondences (MacLeod, 2005; Augustinova et al., 2019).

4.3 Results

Incorrect responses were discarded (3.27%) for the response time analyses. RTs below 200 ms and above 1500 ms were also discarded before the analyses (0.73%).

Correct responses were then analysed using generalized linear mixed-effect models (GLMMs) analyses in R (4.2.2) using the packages "lme4" (1.1-31), "emmeans" (1.8.4-1) and "lmerTest" (3.1-3). Fixed factor "condition" used simple coding so that neutral trials will be compared to all other trials. Fixed factor "mode" also used simple coding that manual responses will be compared to vocal responses. Fixed factor "colour" used deviation coding so that each colour will be compared to the mean of all colours.

First, different types of distributions were compared, and the results of the performance analyses revealed that the Inverse Gaussian fitted the data best. Next, four models with different fixed and random structures were compared. A summary of Information Criterion for these models is shown in Table 4.2.

Name	R2 (marg.)	AIC weights	AICc weights	BIC weights	Performance-Score
Model 4	0.645	1	1	1	87.50%
Model 1	0.548	6.15E-53	6.16E-53	3.88E-51	28.60%
Model 2	0.576	2.07E-80	2.07E-80	8.25E-77	3.65%
Model 3	Fail to converge				

Model Structure:
Model 1: glmer(RT ~ condition * mode + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 2: glmer(RT ~ condition * mode + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 3: glmer(RT ~ condition * mode + colour + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 4: glmer(RT ~ condition * mode + colour + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

Table 4.2. Summary of Information Criterion for models used in Experiment 4.

Model 4 is the best fit for the data. Based on Model 4, a full model was constructed as follows:

```
glmer(RT ~ condition * mode + colour + (1+condition+mode+condition:mode|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
```

However, the full model did not converge. Therefore, the random structures were simplified until the model successfully converged, and the final model was:

```
glmer(RT ~ condition * mode + colour + (1|subject), family = inverse.gaussian(lin
k = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e
5)))
```

A summary of the final model can be found in Appendix L. Further planned comparisons were conducted using the "emmeans" package (1.8.4-1). P-values are corrected using Holm-Bonferroni correction.

Table 4.3 provides a summary of the descriptive statistics and the decomposition of each interference and facilitation components presented in Table 4.4.

Stimulus Type (examples)		**Vocal**	**Manual**
Colour-Incongruent words	RT (SE)	680 (6.57)	636 (6.54)
蓝绿 (*BLUE*green)	ER (SE)	8.28 (0.63)	5.36 (0.51)
Associated Colour-Incongruent words	RT (SE)	625 (6.41)	608 (6.62)
天绿 (*SKY*green)	ER (SE)	1.77 (0.30)	3.85 (0.44)
Homophone-Incongruent words	RT (SE)	641 (6.33)	607 (6.33)
岚绿 (*MIST*green)	ER (SE)	6.25 (0.55)	4.22 (0.46)
Colour-Neutral words	RT (SE)	593 (5.95)	585 5.92)
奖绿 (*PRIZE*green)	ER (SE)	1.25 (0.25)	3.18 (0.40)
Colour-Neutral signs	RT (SE)	585 (6.63)	585 (6.56)
%绿 (*%*green)	ER (SE)	1.20 (0.25)	4.06 (0.45)
Colour-Congruent words	RT (SE)	571 (6.29)	583 (6.48)
蓝蓝 (*BLUE*blue)	ER (SE)	0.78 (0.20)	2.81 (0.38)
Associated Colour-Congruent words	RT (SE)	602 (6.47)	586 (6.34)
天蓝 (*SKY*blue)	ER (SE)	1.77 (0.30)	2.92 (0.38)
Homophone-Congruent words	RT (SE)	570 (6.47)	583 (6.22)
岚蓝 (*MIST*blue)	ER (SE)	0.52 (0.16)	4.06 (0.45)

Table 4.3. Mean reaction times (RT, in milliseconds), error rates (ER, %), and standard errors (SE, in parentheses) as a function of stimulus type from the GLMMs.

Stroop-like effects		Vocal		Manual	R.M.
Stroop Interference (Colour-Incongruent – Colour Neutral Signs)	RT diff.	95***	>	51***	44***
Response Conflict (Colour-Incongruent – Homophone-Incongruent)	RT diff.	39***	>	29***	10^{+}
Phonological Conflict (Homophone-Incongruent – Associated-Incongruent)	RT diff.	16***	>	1	15**
Semantic Conflict (Associated-Incongruent – Colour Neutral Words)	RT diff.	32***	>	23***	9*
Task Conflict (Colour Neutral Words – Colour Neutral Signs)	RT diff.	8^{+}	≈	0	8^{+}
Stroop Facilitation (Colour Neutral Words – Colour-Congruent)	RT diff.	22***	>	2	20***
Semantic Facilitation (Colour Neutral Words – Associated-Congruent)	RT diff.	-9*	<	0	-9*
Phonological Facilitation (Associated-Congruent – Homophone-Congruent)	RT diff.	32***	>	2	30***
Response Facilitation (Homophone-Congruent – Colour-Congruent)	RT diff.	-2	≈	0	-2

*Table 4.4. Stroop-like effects (in milliseconds). ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < $^{+}$p < 0.10; R.M., response modality effect; RT diff., reaction time differences.*

4.3.1 Mean RT analysis

4.3.1.1 Stroop interference effects

Strong Stroop interference effects (relative to neutral signs) occurred in both vocal and manual responses (vocal: M = 95 ms, SE = 4.85 ms, z = 19.568, $p < 0.001$; manual: M = 51 ms, SE = 4.39 ms, z = 11.623, $p < 0.001$). The distinct conflict components were all significant in both response modalities except for phonological conflict. Response conflict was significant in both vocal (M = 39 ms, SE = 4.42 ms, z = 8.749, $p < 0.001$) and manual (M = 29 ms, SE = 4.25 ms, z = 6.799, $p < 0.001$). Phonological conflict was significant in vocal (M = 16 ms, SE = 4.30 ms, z = 3.80, $p < 0.001$), but not in manual (M = 1 ms, SE = 4.04 ms, z = 0.227, $p = 1$). Semantic conflict was significant in both vocal (M = 32 ms, SE = 3.55 ms, z = 8.934, $p < 0.001$) and manual (M = 23 ms, SE = 3.36 ms, z = 6.817, $p < 0.001$). Task conflicts showed a trend in the vocal but not in the manual (vocal: M = 8 ms, SE = 3.82 ms, z = 2.126, p = 0.067; manual: M = 0 ms, SE = 3.48 ms, z = 0.028, p = 1).

The overall Stroop interference was larger in the vocal than in the manual Stroop tasks (M = 44 ms, SE = 6.07 ms, z = 7.219, $p < 0.001$). Further analyses revealed that phonological conflict and semantic conflict were significantly larger in the vocal than in the manual responses (response conflict: M = 10 ms, SE = 5.39 ms, z = 1.806, p = 0.071; phonological conflict: M = 17 ms, SE = 5.32 ms, z = 3.244, p = 0.001; semantic conflict: M = 9 ms, SE = 3.99 ms, z = 2.217, p = 0.027). There was a trend for more task conflicts in the vocal than in the manual responses (M = 8 ms, SE = 4.52 ms, z = 1.773, p = 0.076).

4.3.1.2 Stroop facilitation effects

Stroop facilitation effects (relative to neutral words) were significant in the vocal Stroop task (M = 22 ms, SE = 3.16 ms, z = 6.925, $p < 0.001$), but not in the manual Stroop task (M = 2 ms, SE = 3.17 ms, z = 0.552, p = 1). In fact, only phonological facilitation in the vocal Stroop task was significant (M = 32 ms, SE = 3.76 ms, z = 8.614, $p < 0.001$). Semantic facilitation showed a negative effect in the vocal Stroop task, M = 9 ms, SE = 3.23 ms, z = 2.736, p = 0.019.

Overall, Stroop facilitation was larger in the vocal than in the manual Stroop tasks (M = 20 ms, SE = 3.71 ms, z = 5.46, $p < 0.001$). Further analysis of each facilitation component revealed that phonological facilitation was significantly larger in the vocal than in the manual responses: M = 30 ms, SE = 4.81 ms, z = 6.246, $p < 0.001$. The variation in response modality resulted in a slower response for semantic facilitation in the vocal responses (M = -9 ms, SE = 3.50 ms, z = 2.347, p = 0.019).

4.3.2 Error analysis

The model for error analysis is identical to the one used in mean RT analysis, except for using the Gaussian distribution:

lmer(ER ~ stimulus type * response modality + colour + (1|subject))

In the vocal responses, error rates were higher for Stroop interference effects, response conflict, and phonological conflicts but not for other conflict components (Stroop interference: M = 7%, SE = 0.63%, $t(3783) = 11.186$, $p < 0.001$; response conflict: M = 2%, SE = 0.63%, $t(3783) = 3.232$, $p = 0.02$; phonological conflict: M = 4.5%, SE = 0.63%, $t(3783) = 7.126$, $p < 0.001$; semantic conflict: M = 0.5%, SE = 0.63%, $t(3783) = 0.829$, $p = 1$; task conflict: M = 0.05%, SE = 0.63%, $t(3783) = 0.083$, $p = 1$).

In the manual responses, error rates were higher for Stroop interference effects but not for other conflict components (Stroop interference: M = 2.2%, SE = 0.63%, $t(3783) = 3.480$, $p = 0.013$).

In both vocal and manual responses, error rates were similar for all Stroop facilitation and facilitation components.

4.3.3 Distributional analysis

4.3.3.1 Stroop interference effects

Distributional analyses were conducted to observe the distribution of each conflict/facilitation component and how they contribute to the overall Stroop interference/facilitation effects across different quantiles. Quantile plots and delta plots based on untrimmed data from Experiment 4 for Stroop interference effects in the vocal and manual responses are presented in Figures 4.5 and 4.6.

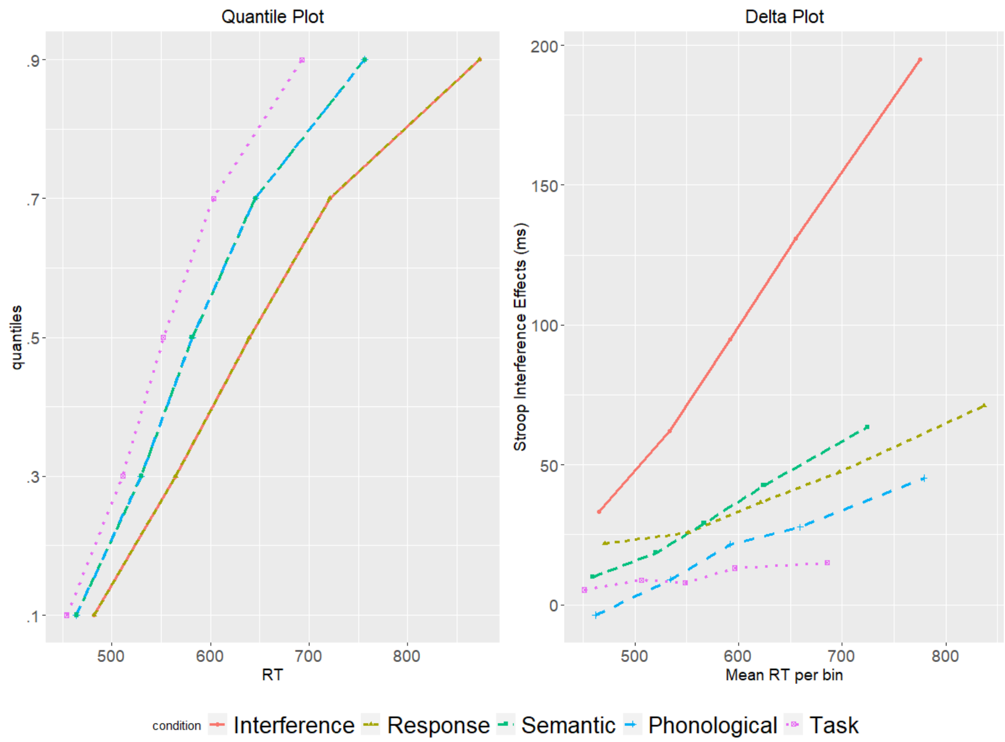


Figure 4.5. Quantile plots and delta plots for Stroop interference effects in the vocal responses in Experiment 4.

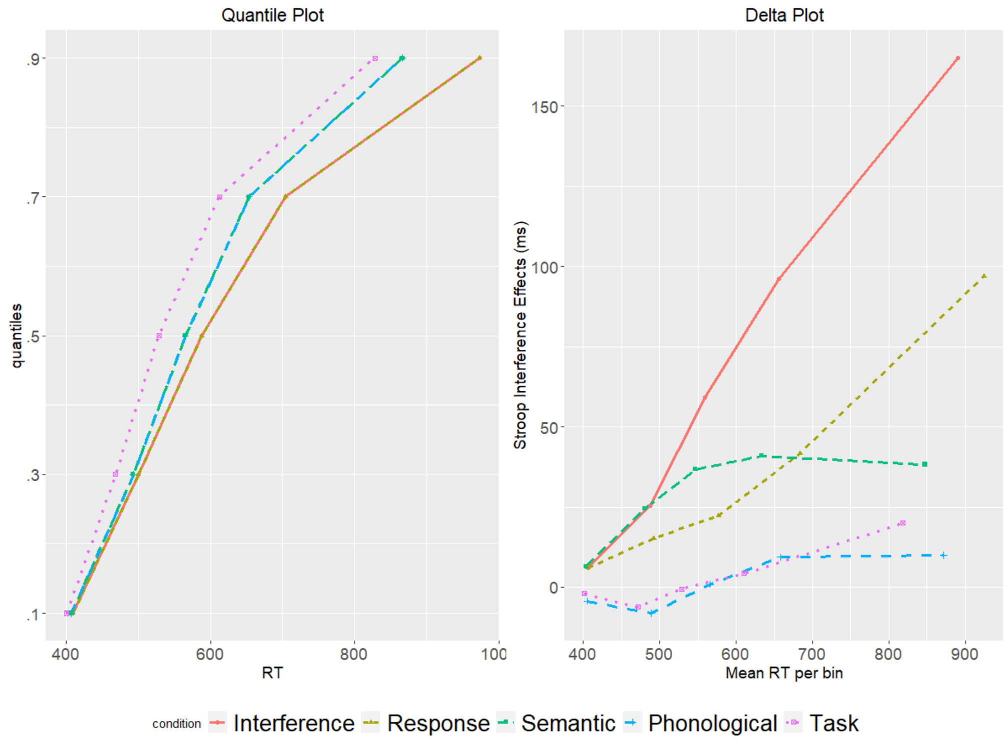


Figure 4.6. Quantile plots and delta plots for Stroop interference effects in the manual responses in Experiment 4.

Analyses were separated for vocal and manual responses. Then, a series of one-way ANOVAs were performed for each conflict component in vocal and manual responses.

For Stroop interference effects in the vocal responses, the delta plots for Stroop interference, response, phonological, and semantic conflicts showed positive linear trends (Stroop interference: $F(1,195) = 115.591$, $p < 0.001$, $d = 1.70$; response conflict: $F(1, 195) = 18.365$, $p < 0.001$, $d = 0.68$; phonological conflict: $F(1,195) = 22.228$, $p < 0.001$, $d = 0.75$; semantic conflict: $F(1,195) = 50.332$, $p < 0.001$, $d = 1.12$), and a trend for a quadratic trend in Stroop interference but not in the others (Stroop interference: $F(1,195)= 2.861$, $p = 0.092$, $d = 0.27$; response conflict: $F(1,195) = 1.44$, $p = 0.232$, $d = 0.19$; phonological conflict: $F(1,195) = 0.01$, $p = 0.922$, $d = 0.02$; semantic conflict: $F(1,195) = 1.539$, $p = 0.216$, $d = 0.20$). The delta plot for task conflicts revealed no linear component ($F(1,195) = 1.682$, $p = 0.196$, $d = 0.21$) nor quadratic component ($F(1,195) = 0.02$, $p = 0.887$, $d = 0.02$).

For Stroop interference effects in the manual responses, the delta plots for Stroop interference and semantic conflicts showed positive linear trends (Stroop interference: $F(1,195) = 69.95$, $p < 0.001$, $d = 1.32$; semantic conflict: $F(1,195) = 5.537$, $p = 0.02$, $d = 0.37$), and a trend for quadratic trend in Stroop interference but not in semantic conflict (Stroop interference: $F(1,195)= 3.473$, $p = 0.064$, $d = 0.29$; semantic conflict: $F(1,195) = 1.517$, $p = 0.22$, $d = -0.19$). The delta plot for response conflict showed a positive linear trend ($F(1,195) = 26.232$, $p < 0.001$, $d = 0.81$), as well as a quadratic trend ($F(1,195) = 4.757$, $p = 0.03$, $d = 0.34$). The delta plot for phonological conflict revealed no linear component ($F(1,195) = 1.627$, $p = 0.204$, $d = 0.20$) or quadratic component ($F(1,195) = 0.038$, $p = 0.845$, $d = 0.03$). The delta plot for task conflict revealed a trend for linear components ($F(1,195) = 3.237$, $p = 0.074$, $d = 0.28$) but no quadratic component ($F(1, 195) = 1.185$, $p = 0.278$, $d = 0.17$).

4.3.3.2 *Stroop facilitation effects*

Quantile plots and delta plots based on untrimmed data from Experiment 4 for Stroop facilitation effects in vocal and manual responses were presented in Figures 4.7 and 4.8.

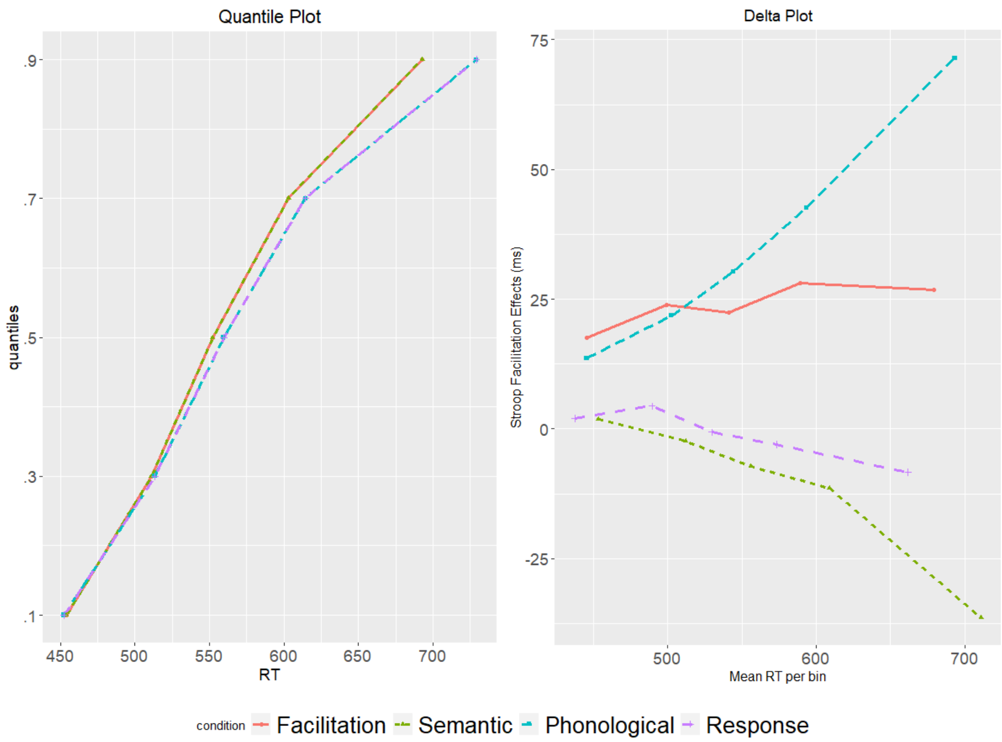


Figure 4.7. Quantile plots and delta plots for Stroop facilitation effects in vocal responses in Experiment 4.

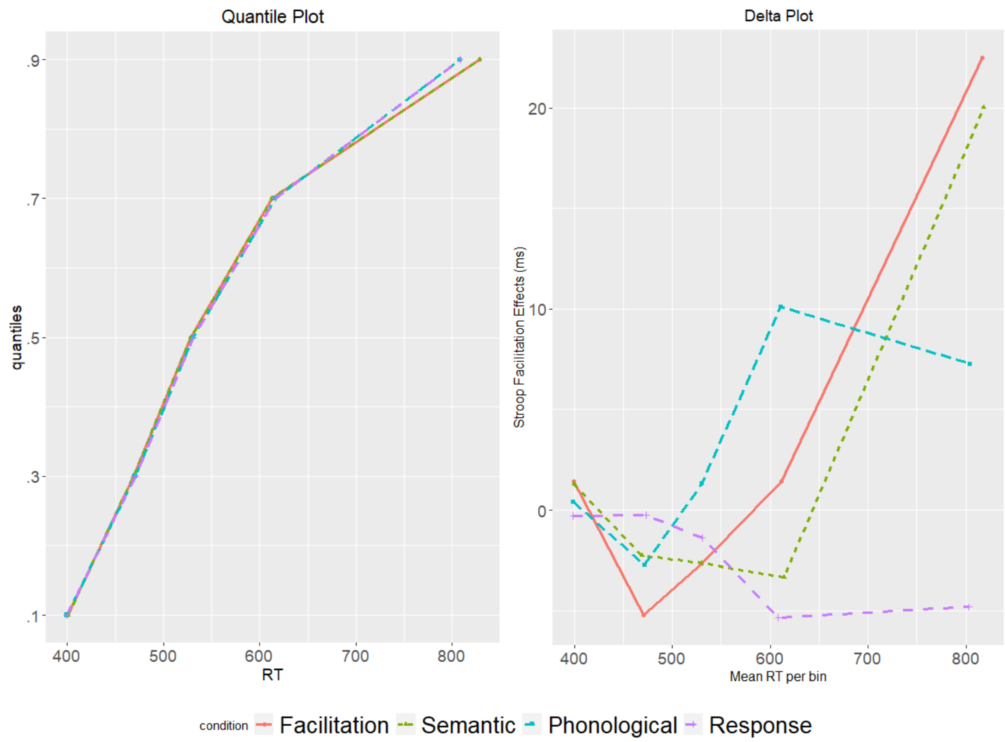


Figure 4.8. Quantile plots and delta plots for Stroop facilitation effects in manual responses in Experiment 4.

Analyses were separated for vocal and manual responses. Then, a series of one-way ANOVAs were performed for each facilitation component in vocal and manual responses.

For Stroop facilitation effects in vocal responses, the delta plot for phonological facilitation showed a positive linear trend ($F(1,195) = 36.731$, $p < 0.001$, $d = 0.96$), and a trend for quadratic trend ($F(1,195) = 2.879$, $p = 0.091$, $d = 0.27$), whereas the delta plot for semantic facilitation showed a negative linear trend ($F(1,195) = 17.457$, $p < 0.001$, $d = -0.67$), and a trend for quadratic trend ($F(1,195) = 2.791$, $p = 0.096$, $d = -0.26$). The delta plot for Stroop facilitation showed no linear trend ($F(1,195) = 1.621$, $p = 0.204$, $d = 0.20$) nor quadratic trend ($F(1,195) = 0.148$, $p = 0.7$, $d = -0.06$). The delta plot for response facilitation showed no linear trends ($F(1,195) = 2.921$, $p = 0.089$, $d = -0.27$) but a trend for quadratic trend ($F(1,195) = 0.433$, $p = 0.511$, $d = -0.10$).

For Stroop facilitation effects in manual responses, the delta plots for all facilitation components showed no linear trends (Stroop facilitation: $F(1,195) = 2.559$, $p = 0.111$, d

= 0.25; semantic facilitation: $F(1,195) = 1.003$, $p = 0.318$, $d = 0.16$; phonological facilitation: $F(1,195) = 0.464$, $p = 0.496$, $d = 0.11$; response facilitation: $F(1,195) = 0.169$, $p = 0.682$, $d = -0.06$), nor quadratic trends (Stroop facilitation: $F(1,195) = 2.492$, $p = 0.116$, $d = 0.25$; semantic facilitation: $F(1,195) = 1.556$, $p = 0.214$, $d = 0.20$; phonological facilitation: $F(1,195) = 0.01$, $p = 0.909$, $d = 0.02$; response facilitation: $F(1,195) = 0.002$, $p = 0.963$, $d = -0.01$).

4.4 Discussion

The response modality effects observed in the current study are consistent with the literature indicating vocal responses receive stronger Stroop interference than manual responses (Augustinova et al., 2019; Fennell & Ratcliff, 2019; Neill, 1977; Redding & Gerjets, 1977; Sharma & McKenna, 1998; Zahedi et al., 2019). Importantly, stronger Stroop facilitation was found in vocal responses, which is consistent with Augustinova et al. (2019).

The Chinese homophones included in this experiment enabled investigation of the phonological components in Stroop effects. Results revealed a different pattern for vocal and manual responses. For Stroop interference effects, the reduced contribution of phonological and semantic conflicts led to smaller Stroop interference effects in the manual responses compared to vocal responses. The response modality effect of Stroop interference in Augustinova et al.’s (2019) study was due to a lesser contribution of task and response conflicts.

For Stroop facilitation, only phonological facilitation contributed to the response modality effect in Chinese characters. This is different from Augustinova et al. (2019) who found that response facilitation solely contributed to the response modality effect in French.

The results of this study suggest that phonological components should be taken into consideration when using Chinese characters in the Stroop task. Parris et al. (2022) expressed concerns that measuring phonological conflicts using pseudohomophones would also involve orthographic conflict. In Chinese, however, homophones do not overlap in orthography with colour words. Therefore, Chinese homophones are ideal to measure the role of phonology in Stroop tasks.

The magnitude of phonological conflicts calculated from the vocal Stroop data collected by Spinks et al. (2000) was -2 ms and -7ms from Wang et al.'s manual Stroop data (2010). In the present study, the magnitude of phonological conflicts was 16 ms and 1 ms for the vocal and manual Stroop tasks, respectively. Phonological conflicts were affected by response modality, as stronger effects were found in vocal than manual responses.

Phonological facilitation based on Spinks et al. (2000) and Wang et al.'s (2010) data was 12 ms and 36 ms for vocal and manual responses, respectively. The current study showed 32 ms and 2 ms in phonological facilitation for vocal and manual responses, and this difference was significant. Thus, phonological facilitation was stronger in vocal than in manual responses.

Based on the results of Stroop interference and facilitation, phonological components are affected by the response modality. Phonological facilitation was evident with the vocal responses but not with the manual responses. This is consistent with the assumption that vocal responses involve an overt verbal reaction, thus leading to stronger interference/facilitation when assisted by phonological cues provided by homophones.

Figure 4.9 presents an overview of the magnitude of the distinct conflict components of Stroop interference in Augustinova et al. (2019) and the present study. The pattern looks very similar in both languages. For manual responses, response conflicts are the largest, followed by semantic conflicts and task conflicts. However, the pattern for the vocal responses between the two languages was rather different: in Augustinova et al.'s study, task conflicts were larger than semantic conflicts.

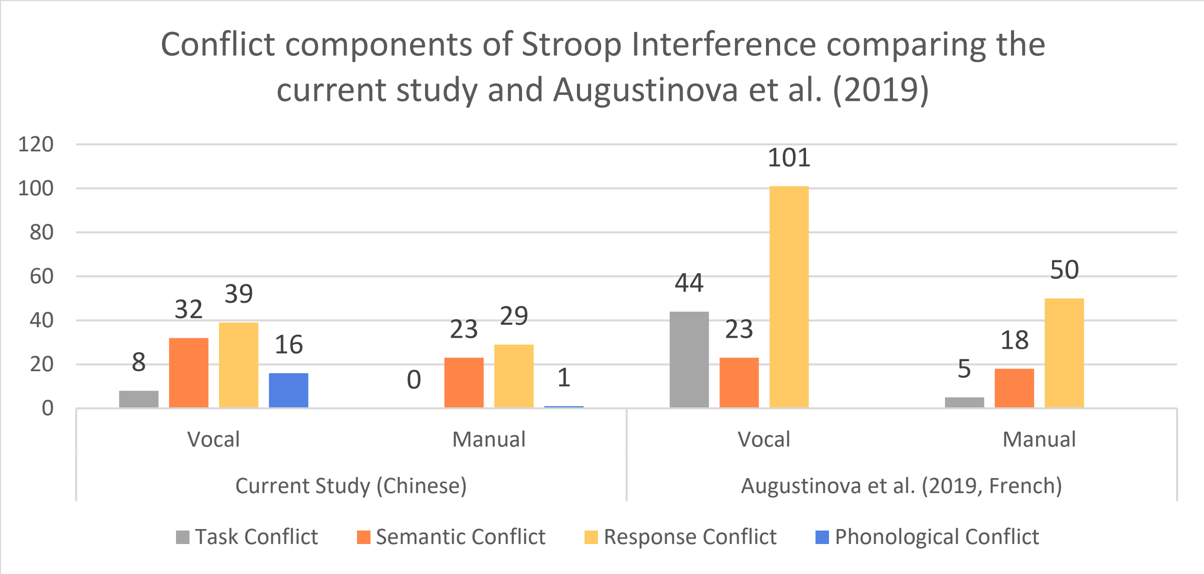


Figure 4.9. Conflict components of Stroop Interference comparing Experiment 4 and Augustinova et al. (2019) in both vocal and manual tasks.

Significant semantic conflicts and response conflicts were found in both modalities. Semantic conflicts were affected by response modality (i.e. larger effects in the vocal responses than in the manual responses), but response conflicts were not. In contrast, Augustinova et al. (2019) found that response conflicts were affected by response modality but not semantic conflicts.

In terms of Stroop facilitation, response facilitation was not observed in either response modality, whereas Augustinova et al. (2019) found this for the vocal responses. Augustinova et al. (2019) found semantic facilitation in both response modalities, but semantic facilitation was unaffected by response modality. In contrast, the present study observed negative semantic facilitation in the vocal responses and no semantic facilitation in the manual responses.

Only a trend was found for task conflicts in the vocal responses but not in the manual responses, whereas Augustinova et al. (2019) found significant task conflicts in the vocal Stroop tasks only as well as stronger task conflicts in the vocal than in the manual responses.

This discrepancy between the current study and Augustinova et al. (2019) in terms of task conflicts could be related to the differences in the neutral sign between the present study and Augustinova et al.'s (2019) (percent sign "%" vs. a row of Xs "XXXX" in Augustinova et al., 2019). A percentage sign was used in the present study to match the length of a Chinese character and to ensure that participants were not exposed to any alphabetic characters. However, the percentage sign does contain some meaning, unlike the row of Xs. This could potentially result in a reduction of the difference between neutral words and signs, i.e. a smaller task conflict as seen in the current study. Future research could use colour patches as neutral signs, like other studies (Spinks et al., 2000; Wang et al., 2010; Yeh et al., 2017).

Another factor that may impact the results of task conflicts could be the use of colour neutral words 炭 (charcoal, /tan4/). This item could be associated to colour black, though this colour is not in the response set. However, if this item does cause semantic conflict in the Stroop task, then stronger task conflicts should be observed, which is inconsistent with the current results. Therefore, the use of colour neutral word charcoal would not pose a negative impact on the findings.

The influence of response modality on the distinct components of Stroop interference is crucial for the debate about single- or Multi-stage accounts of Stroop interference. In the present study, the magnitude of Stroop interference was significantly larger in vocal than in manual responses. This contradicts Roelof's (2003) model, which predicts that vocal and manual responses have the same outcome, because Stroop interference is the result of response competition. The current results support multi-stage account in which phonological and semantic conflicts contribute to Stroop interference and phonological facilitation contributes to Stroop facilitation. Sharma and McKenna (1998) assumed that there is no privileged access to words in manual responses; thus, semantic conflicts were absent with manual responses. The present data revealed that both vocal and manual responses involve semantic conflicts, with vocal responses carrying stronger influence on semantic conflicts than manual responses.

Augustinova et al. (2019) argued that if semantic conflicts and response conflicts occur at the same level, then their magnitude should be the same for each response modality. However, their findings showed that semantic conflicts were not affected by response modalities, whereas response conflicts were affected, indicating that semantic conflicts are independent of the response level.

In the present study, phonological conflicts and semantic conflicts were both affected by response modality. In Augustinova et al.'s argument (2019), they are not independent of each other. However, this logic could be problematic when there are multiple levels of comparison. For example, task conflicts and response conflicts were found to be affected by response modality in Augustinova et al.'s results (2019). This did not necessarily mean that task conflicts and response conflicts were based on the same level of processing. The same logic can be applied here. Thus, although the three conflict components were all affected by response modality, it is difficult to discern whether they were from the same level or not.

The distributional analysis supports the observed mean RT analysis. In vocal responses, it is difficult to maintain a high level of suppression for response conflict, phonological conflict, and semantic conflict. This means that response, phonological, and semantic conflict contributed to Stroop interference effects. For manual responses, phonological conflicts disappeared, whereas response and semantic conflicts still showed no inhibition. This corresponds to the findings of the mean RT analysis that phonological conflicts do not contribute to Stroop interference effects in manual responses.

For Stroop facilitation effects, the delta plot analysis showed that there is strong enhancement applied to phonological facilitation, and no enhancement applied to semantic facilitation in vocal responses, which is identical to the observed effects in mean RT analysis.

It is expected that large effect size can be found in Stroop interference effects in both vocal and manual Stroop tasks. Response conflict showed medium effects in both response modalities, whereas semantic and phonological conflict showed large and medium effects in the vocal responses but small effects in the manual responses. On the other hand, small effects were found in the Stroop facilitation with both response modalities. Phonological facilitation had a large effect size in the vocal but a very small effect in the manual responses.

4.5 Conclusion

The study presented in this chapter revealed that phonological components are one of the distinct components in the Stroop task with Chinese characters. Robust semantic conflicts and response conflicts were found in both vocal and manual responses, as observed in Augustinova et al.'s (2019) study. These results support the idea that multi-stage accounts of Stroop effects are independent of the language used.

The response modality effect in Stroop interference was due to the reduced contribution of phonological conflicts and semantic conflicts in manual responses compared to vocal responses.

The response modality effect in Stroop facilitation was also due to the reduced contribution of phonological facilitation in manual responses compared to vocal responses. This means that phonological activation is stable and can be found in both Stroop interference and facilitation.

In Augustinova et al.'s (2019) study, task conflicts were affected by response modality. In the present study, task conflicts were not found in either response modality. This was discussed in terms of the selection of different neutral signs, which could be language-specific. To investigate this further, additional research is needed using different scripts.

The distributional analyses provided a similar pattern to what was observed in the mean RT analysis, confirming that response modality effects are attributed to the reduced phonological conflict/facilitation and semantic conflict.

The current study supports multi-stage accounts of Stroop effects by decomposing them into different conflict/facilitation components. In addition to response, semantic and task conflicts, phonological conflicts should be taken into consideration when Stroop effects are studied in Chinese.

Chapter 5: Distinct Components of the Stroop effects using Pinyin

5.1 Introduction

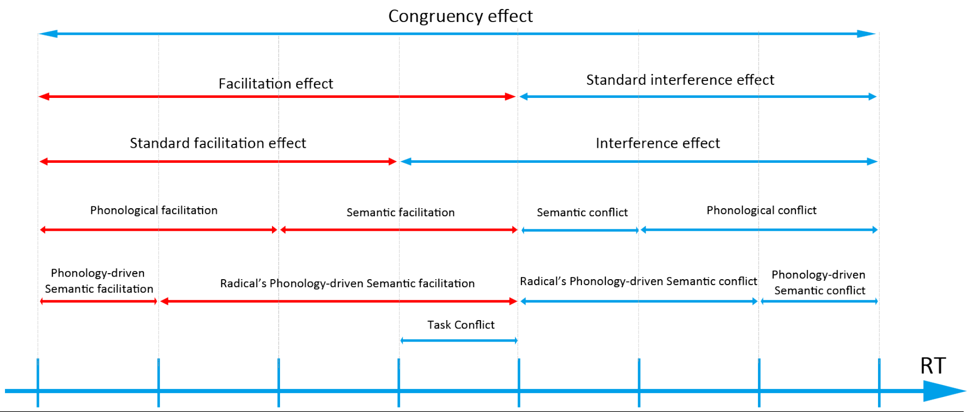
Chapter 1 discussed studies that explored different aspects of pinyin, such as the orthographic, phonological, and semantic activation of Chinese characters via pinyin. Pinyin is the Romanization of Chinese characters, representing the phonological units in Chinese characters but without direct access to the orthography of Chinese characters.

Pinyin consists of segmental and suprasegmental information. However, research has shown that suprasegmental information is activated later than segmental information (Li et al., 2019; Wang et al., 2015). Pinyin is also considered an orthographic mediator to access the meaning of characters. Crucially, the sublexical components of Chinese characters have been found to be activated through pinyin (Chen, Perfetti, & Leng, 2019). However, studies using pinyin mainly involved explicit reading tasks (e.g. priming tasks combed with lexical decision task). The automatic processing of pinyin can be investigated in a Stroop task. Interestingly, Chen et al. (2014) reported that the orthographic information of pinyin is not fully activated.

Because of the mixed findings about the phonological and semantic activation of Chinese characters by pinyin, it is important to investigate this further in a Stroop task. Furthermore, no studies have explored the distinct conflict/facilitation components in Chinese characters written in pinyin.

In Section 2.3, multiple-stage accounts of the Stroop effects were introduced (Augustinova & Ferrand, 2014; Augustinova et al., 2019; Augustinova et al., 2018; Neely & Kahan, 2001; Schmidt & Cheesman, 2005; Sharma & McKenna, 1998). This account explains the Stroop effects as resulting from different sources/levels, which include response conflict/facilitation, semantic conflict/facilitation, and task conflict. The Stroop experiment reported in Chapter 4 (Experiment 4) investigated not only response, semantic, and task components but also phonological components using Chinese characters. Experiment 4 included the following eight conditions: incongruent colour-character, congruent colour-character, incongruent colour-associated, congruent colour-associated, incongruent homophones, congruent homophones, neutral words, and neutral signs. By comparing these conditions, response, semantic, phonological, and task components were investigated. The results revealed that Stroop interference effects were attributed to response, phonological, and semantic conflicts. However, in contrast to Augustinova et al. (2019), no significant task conflicts were observed in Experiment 4.

Colour words in Chinese characters activate (pre-)motor responses to colour, as they are directly linked to semantics and orthography. Unlike Chinese characters, pinyin words directly link to phonological representations and then indirectly activate semantic representation. Pinyin words are essentially homophonic to colour words, thus not only colour words but also many words with no association with colours. (Pre-)motor responses may not be activated as is the case with colour-associated words. Therefore, a revision of Figure 4.2 was presented in Figure 5.1. This explains the conflict/facilitation components of Stroop effects in pinyin.



Conditions	Homophone congruent	Invalid-Radical congruent	Semantic-associative congruent	Neutral Sign	Neutral Word	Semantic-associative incongruent	Invalid-Radical incongruent	Homophone incongruent
Examples in Pinyin	lán	cāi	tiān	xxxxx	jiǎng	cǎo	cāi	lǜ
Chinese Character	蓝	猜	天	xxxxx	奖	草	猜	绿
English Translation	blue	guess	sky	xxxxx	prize	grass	guess	green
Properties	+ Word Reading + Phonological (+ Semantic)	+ Word Reading + Radical's Phonological (+ Radical's Semantic)	+ Word Reading + Semantic		+ Word Reading	+ Word Reading + Semantic	+ Word Reading + Radical's Phonological (+ Radical's Semantic)	+ Word Reading + Phonological (+ Semantic)

Figure 5.1. Decomposed Stroop conflict/facilitation components in pinyin.

Phonological conflicts/facilitation are calculated using the RT difference between colour words in pinyin and colour-associated words in pinyin. As indicated in Figure 5.1 (see "Properties"), colour words in pinyin (homophones) activate phonological representations and then trigger the semantics of colour words that caused Stroop interference or facilitation effects. Colour-associated words in pinyin do not share the same phonology as colour words, and therefore, only semantic relatedness to colour words triggers semantic conflict/facilitation compared to neutral words. The sublexical activation of Chinese characters in pinyin is also indicated in this figure. For example, the Invalid-Radical condition "猜" (/cāi/, meaning "guess") contains the colour word cyan "青" (/qīng/) as its phonetic radical, but the pronunciation of the whole character is different from its phonetic radical.

Evidence for the activation of phonetic radicals through pinyin comes from a lexical decision task but not a Stroop task (Chen et al., 2014). If the Invalid-Radical condition in pinyin can still activate phonetic radicals, then the difference between the Invalid-Radical condition in pinyin and neutral words is radical's phonology-driven semantic conflict/facilitation because radical's phonological information and the subsequent semantic activation of colour words are absent in neutral words. The difference between homophones in pinyin and the Invalid-Radical condition in pinyin is the character's phonology-driven semantic conflict/facilitation. This is because homophones activate the colour words from the character level rather than the sublexical level.

The processing of Chinese characters presented in pinyin is less investigated in the literature. As far as I am aware, only one study included Chinese words written in pinyin in a Stroop task (Liu & Weng, 2007). The experiment included Chinese characters, pinyin, and English stimuli to investigate whether different scripts have an impact on Stroop interference effects. The results revealed that with repeated training through blocks, Stroop interference effects for all three types of stimuli became smaller and the differences between them disappeared. The findings are interesting; however, the study has a number of limitations. For example, the experiment only included a total of 192 trials; thus only 48 trials in four conditions (characters, pinyin, English, and colour patches). Furthermore, all stimuli were presented in the incongruent condition (except for colour patches), so no Stroop facilitation has been investigated. Thirdly, all three types of stimuli were mixed in one block. Therefore, it is hard to tell the exact magnitude of Stroop effects in pinyin.

In this chapter, a series of manual Stroop tasks is presented that involves pinyin stimuli. Experiment 5 investigates whether Stroop interference and facilitation effects occur with Chinese words written in pinyin. Furthermore, this experiment investigates the impact of tonal information by presenting the pinyin words with and without diacritics, as diacritics are used to indicate the tone of Chinese characters. As discussed in Section 1.1, without tonal information, pinyin is even more ambiguous in terms of the mapping to specific Chinese characters (e.g. colour words). Thus, it is expected that removing tonal information reduces Stroop effects compared to pinyin with tonal information.

Experiment 6 investigates whether semantic and phonological activation occurs with Chinese colour-associated words that are written in pinyin (e.g. sky or grass written in pinyin). Because pinyin activates all characters with the same pronunciation, the magnitude of semantic conflicts is expected to be relatively small compared to using Chinese characters.

The activation of phonetic radicals by Chinese words written in pinyin is investigated in Experiment 7. An Invalid-Radical character contains a radical that can be a colour word. When it is transcribed into pinyin, its orthographical information is lost. The activation of phonetic radicals is very indirect: pinyin first activates the phonology of target characters, then the phonetic radicals embedded in characters are activated.

Experiment 8 will directly compare Stroop effects with Chinese characters and pinyin in a single experiment. According to Liu and Weng's (2007) study, a mixed script presentation reduced Stroop interference effects. Unlike Liu and Weng, the experiment also includes congruent conditions to study the impact of script on Stroop facilitation effects.

5.2 Experiment 5

5.2.1 Methods

5.2.1.1 Participants

Twenty-four participants were recruited online (mean age = 23.38, range = 18–33, females = 15). All were native speakers of Chinese from mainland China and had normal or corrected-to-normal vision. Participation was voluntary and they had the opportunity to enter a prize draw after completing the experiment. The experiment was approved by the ethics committee at the School of Psychology, University of Nottingham, UK (Ethics approval number: S1289R 3.1) and all participants read and agreed to the statements on the electronic consent form prior to data collection.

5.2.1.2 Stimuli and Design

The stimuli were all presented in pinyin, for example, the colour character (黄, meaning "yellow") was presented as huáng in pinyin. Three colours were used: green, yellow, and blue. Four types of stimuli were included: incongruent colour characters written in pinyin (also treated as the incongruent homophone condition, e.g. huáng, "yellow", presented in blue colour), congruent colour characters written in pinyin (also treated as the congruent homophone condition, e.g. huáng presented in yellow colour), neutral characters written in pinyin (e.g. xún has no phonological or semantic association with colours), and neutral signs (e.g. xxxxx). Chinese words written in pinyin do not directly refer to colour words, so colour characters presented in pinyin will be considered as homophones. Neutral characters are matched with each colour character stimulus in terms of length, stroke count, frequency, and family size (when written in pinyin).

Neutral signs are a string of five x's that have the same length as the longest stimulus, which is huáng. Crucially, pinyin stimuli are presented with or without tonal information.

Table 5.1 presents the stimuli used in Experiment 5. Detailed linguistics properties of these stimuli can be found in Appendices E and F.

Pinyin	Possible Meaning 1	Possible Meaning 2	Possible Meaning 3	Possible Meaning 4	Possible Meaning 5
Colour characters in pinyin (homophones)					
/lv4/	green	law	filter	consider	
/huang2/	yellow	emperor	phoenix	jade	bright
/lan2/	blue	holdback	eupatorium	basket	railing
Neutral characters in pinyin					
/po1/	very	splash	slope		
/qiong2/	poor	red jade	high		
/xun4/	instruct	interrogate	modest	tame	sacrifice

Table 5.1. Stimuli used in Experiment 5.

The experiment is a 4 (stimulus type: homophone-incongruent vs. homophone-congruent vs. neutral pinyin vs. neutral sign) × 2 (tonal information: with vs. without tonal information) within-subject design. There are 72 trials in each block, with 18 trials in each stimulus type. There are three blocks for with and without tones conditions, that is, 432 trials in total. The order of presenting with or without tonal information stimuli was counterbalanced across participants. In each block, each stimuli type has the same number of trials by manipulating repetition: incongruent stimuli repeated three times (2 colours × 3 pinyin × 3 repetitions = 18 trials), congruent six times (1 colour × 3 pinyin × 6 repetitions = 18 trials), neutral pinyin twice (3 colours × 3 pinyin × 2 repetitions = 18 trials), and neutral signs six times (3 colours × 1 pinyin × 6 repetitions = 18 trials).

This experiment serves as a pilot study for further investigation on pinyin words, so the number of observations is slightly below the recommended 1,600 observations per condition: 22 participants × 3 blocks × 18 trials = 1,188 observations.

5.2.1.3 Procedure

The experiment was carried out online using pavlovia.org. The stimuli were presented in black ink (RGB: 0,0,0) and surrounded by a grey square (RGB: 204,204,204), with a colour-filled rectangle representing one of three colours: green (RGB: 0,255,0), yellow (RGB: 240,240,0) and blue (RGB: 0,11,255). The background colour was also set to grey (RGB: 204,204,204). Words presented on the screen used the Arial font. A fixation point was indicated by a "+" sign (RGB: 0,0,0) in the centre of the screen.

Participants were asked to press the corresponding colour buttons when they saw the correct ink colour of the words presented on the screen. In each trial, a fixation cross was shown for 500 ms, followed by a blank screen for 300 ms. The target stimuli then appeared on the screen for a maximum of 3000 ms or until the participant responded. Stimuli disappeared as soon as the participant responded. There was an inter-trial interval of 1000 ms.

For the practice session, 48 key-matching practice trials were used to teach participants the key-colour correspondences (MacLeod, 2005). Feedback about the correctness of their response was provided after each practice trial. The practice session only includes neutral signs with the colours used in the experiment.

5.2.2 Results

5.2.2.1 Mean RT analysis

Two participants were excluded due to high error rates (19%, 32%). Incorrect responses were discarded (4.71%) for the response time analyses. RTs below 200 ms and above 1700 ms were discarded before the analyses (1.58%). Fixed factor "condition" used simple coding so that neutral trials will be compared to all other trials. Fixed factor "tone" also used simple coding that pinyin with tonal information condition will be compared to pinyin without tonal information condition. Fixed factor "colour" used deviation coding so that each colour will be compared to the mean of all colours.

First, different types of distributions were compared, and the results of the performance analyses revealed that the Inverse Gaussian fitted the data best. Next, four models with different fixed and random structures were compared. A summary of Information Criterion for these models is shown in Table 5.2.

Name	R2 (marg.)	AIC weights	AICc weights	BIC weights	Performance-Score
Model 4	0.393	1	1	1	90.80%
Model 1	0.123	5.29E-55	5.30E-55	1.84E-53	15.91%
Model 2	0.133	2.43E-58	2.44E-58	2.93E-55	0.46%
Model 3	Fail to converge				

Model Structure:
Model 1: glmer(RT ~ condition * tone + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 2: glmer(RT ~ condition * tone + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 3: glmer(RT ~ condition * tone + colour + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 4: glmer(RT ~ condition * tone + colour + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

Table 5.2. Summary of Information Criterion for models used in Experiment 5.

Model 4 is the best fit for the data. Based on Model 4, a full model was constructed as follows:

glmer(RT ~ condition * tone + colour + (1+condition+tone+condition:tone|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

However, the full model did not converge. Therefore, the random structures were simplified until the model successfully converged, and the final model was:

glmer(RT ~ condition * tone + colour + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

A summary of the final model can be found in Appendix M. As in previous chapters, the planned comparisons were conducted using the “emmeans” package (1.8.4-1). P-values were corrected using Holm-Bonferroni method. Table 5.3 provided a summary of the descriptive statistics and results of final model and further contrasts are presented in Table 5.4.

Stimulus Type (examples)		**With Tone**	**Without Tone**
Homophone-Incongruent words	RT (SE)	694 (13.8)	685 (14.0)
/lán/$_{\text{/lǜ/}}$ (*BLUE*green)	ER (SE)	5.90 (0.68)	4.40 (0.60)
Colour-Neutral words	RT (SE)	670 (14.1)	676 (13.8)
/pō/$_{\text{/lǜ/}}$ (*VERY*green)	ER (SE)	4.98 (0.63)	3.80 (0.56)
Colour-Neutral signs	RT (SE)	658 (13.6)	655 (14.0)
xxxxx$_{\text{/lǜ/}}$ (*XXXXX*green)	ER (SE)	4.98 (0.63)	4.48 (0.60)
Homophone-Congruent words	RT (SE)	644 (13.7)	648 (13.9)
/lán/$_{\text{/lán/}}$ (*BLUE*blue)	ER (SE)	3.04 (0.50)	3.97 (0.57)

Table 5.3. Mean reaction times (RT, in milliseconds), error rates (ER, %), and standard errors (SE, in parentheses) as a function of stimulus type from the GLMMs in Experiment 5.

Stroop effects		With Tone		Without Tone	M. D.
Stroop Interference relative to neutral signs (Homophone-Incongruent – Colour Neutral Signs)	RT diff.	37***	≈	30***	7
Stroop Interference relative to neutral words (Homophone-Incongruent – Colour Neutral Words)	RT diff.	24***	≈	9	15^{+}
Stroop Facilitation relative to neutral signs (Colour Neutral Signs – Homophone-Congruent)	RT diff.	14^{+}	≈	6	8
Stroop Facilitation relative to neutral words (Colour Neutral Words – Homophone-Congruent)	RT diff.	25***	≈	27***	-2
Task Conflict (Colour Neutral Words – Colour Neutral Signs)	RT diff.	13^{+}	≈	21**	9

*Table 5.4. Stroop effects in Experiment 5 (in milliseconds). ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < ^{+}p < 0.1; M.D., mean differences; RT diff., reaction time differences.*

Further contrast analyses revealed detailed comparisons between each stimulus type when they were presented with or without tonal information. For Stroop interference effects relative to neutral signs, the RT differences between homophone-incongruent conditions and neutral signs were significant for the with and without tonal information conditions (with tone: M = 37 ms, SE = 6.68 ms, z = 5.479, p < 0.001; without tone: M = 30 ms, SE = 6.77 ms, z = 4.489, p < 0.001). For Stroop interference effects relative to neutral words, the RT differences were significant for Chinese words written in pinyin with tonal information (M = 24 ms, SE = 6.14 ms, z = 3.922, p < 0.001) but not when there was no tonal information presented (M = 9 ms, SE = 6.31 ms, z = 1.525, p = 0.255).

Task conflicts (neutral signs vs. neutral words) showed a trend when there was tonal information and were significant when there was no tonal information provided (with tone: M = 13 ms, SE = 6.02 ms, z = 2.082, *p* = 0.07; without tone: M = 21 ms, SE = 6.09 ms, z = 3.410, *p* = 0.002).

There was no difference between tonal and without tonal conditions in Stroop interference relative to neutral signs and task conflicts (all *p*s > 0.3). There was a tendency that more Stroop interference relative to neutral words was received with tonal conditions compared to without tones (differences: 15 ms, SE = 7.70 ms, z = 1.875, *p* = 0.061).

For Stroop facilitation effects relative to neutral signs, the RT differences between homophone-congruent conditions and neutral signs showed a trend when presented with tones and were not significant when presented without tones (with tone: M = 14 ms, SE = 6.35 ms, z = 2.088, *p* = 0.074; without tone: M = 6 ms, SE = 6.44 ms, z = 0.990, *p* = 0.322). For Stroop facilitation effects relative to neutral words, the RT differences were significant in both tonal conditions (with tone: M = 25 ms, SE = 5.75 ms, z = 4.488, *p* < 0.001; without tone: M = 27 ms, SE = 5.92 ms, z = 4.588, *p* < 0.001). There was no difference between tonal and without tonal conditions in Stroop facilitation relative to neutral signs or neutral words (all *p*s > 0.4).

5.2.2.2 Error analysis

The model for error analysis is identical to the one used in mean RT analysis, except for using the Gaussian distribution:

lmer(ER ~ condition * tone + colour + (1|subject))

In both with and without tonal conditions, error rates were similar for all Stroop interference and facilitation components.

5.2.2.2.1 Distributional analysis

5.2.2.2.1.1 Stroop interference effects

Quantile plots and delta plots based on untrimmed data from Experiment 5 for Stroop interference effects in pinyin with and without tones are presented in Figure 5.2 and 5.3.

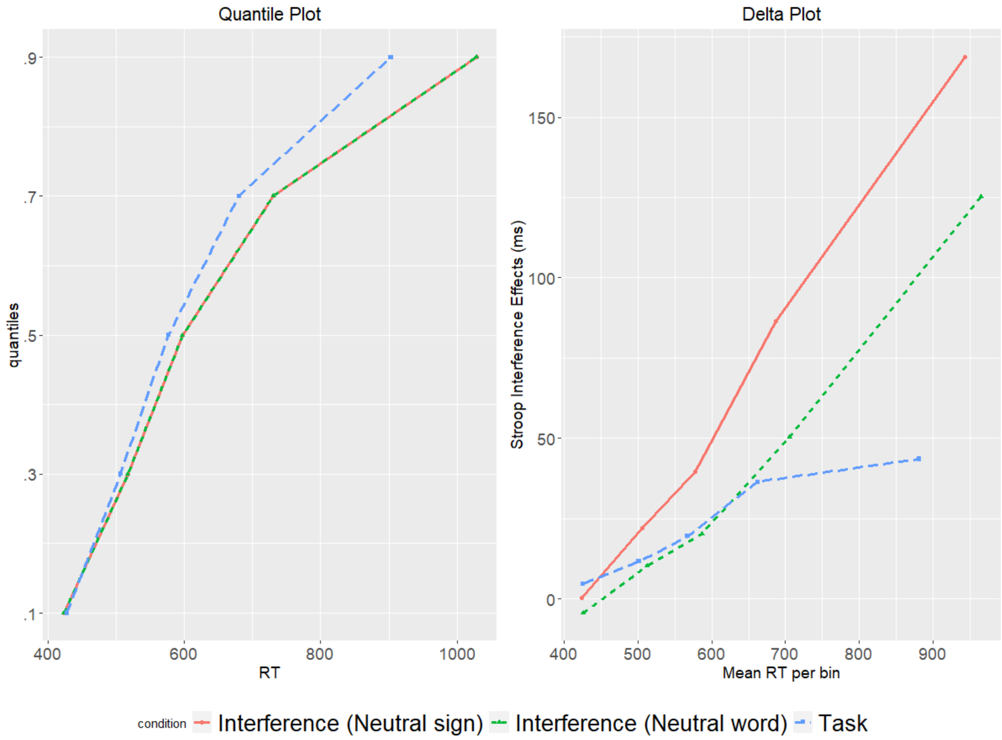


Figure 5.2. Quantile plots and delta plots for Stroop interference effects with tonal information in Experiment 5.

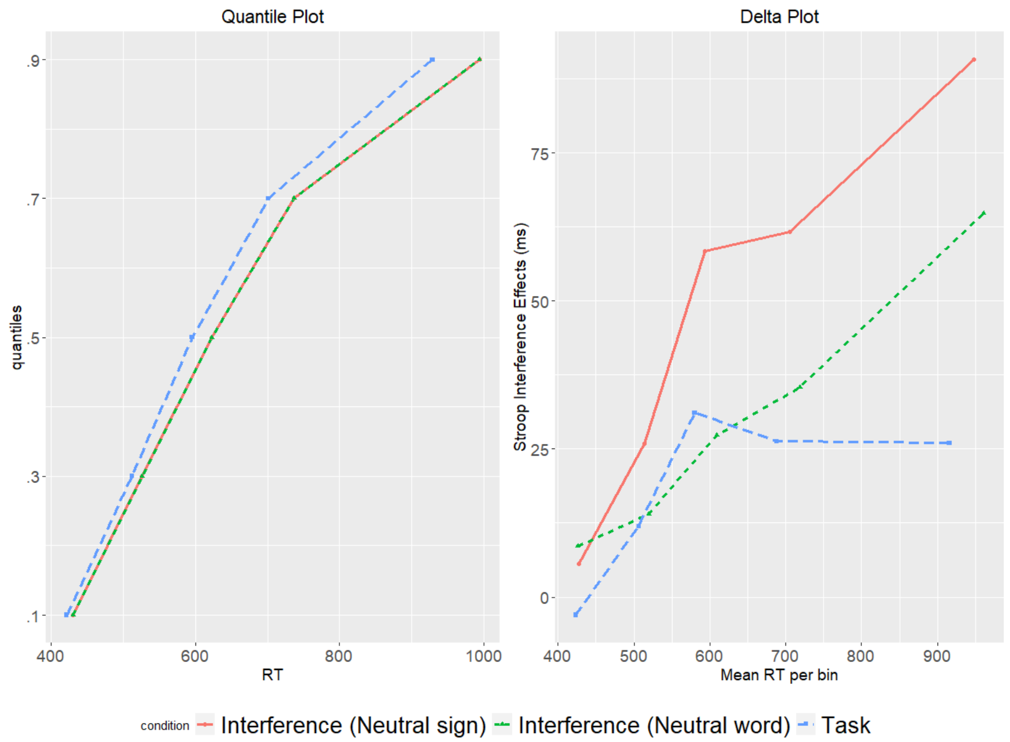


Figure 5.3. Quantile plots and delta plots for Stroop interference effects without tonal information in Experiment 5.

Analyses were separated for pinyin with and without tone conditions. Then, a series of one-way ANOVAs were performed for each conflict component in pinyin with and without tone conditions.

For Stroop interference effects in pinyin with tones, the delta plots for Stroop interference relative to neutral signs and neutral words showed positive linear trends (Stroop interference relative to neutral signs: $F(1,105) = 37.356$, $p < 0.001$, $d = 1.30$; Stroop interference relative to neutral words: $F(1,105) = 18.632$, $p < 0.001$, $d = 0.92$). Trends for quadratic trends were found for Stroop interference relative to neutral signs and neutral words (Stroop interference relative to neutral signs: $F(1,105) = 4.13$, $p = 0.056$, $d = 0.41$; Stroop interference relative to neutral words: $F(1,105) = 3.279$, $p = 0.089$, $d = 0.37$). The delta plot for task conflict showed a trend for the linear component ($F(1,105) = 4.734$, $p = 0.054$, $d = 0.42$) but no quadratic component ($F(1,105) = 0.023$, $p = 0.879$, $d = 0.03$).

For Stroop interference effects in pinyin without tones, the delta plots for Stroop interference relative to neutral signs and Stroop interference relative to neutral words showed positive linear trends (Stroop interference relative to neutral signs: *F*(1,105) = 10.398, *p* = 0.002, *d* = 0.69; Stroop interference relative to neutral words: *F*(1,105) = 3.98, *p* = 0.049, *d* = 0.43) and no quadratic trends (Stroop interference relative to neutral signs: *F*(1,105) = 0.024, *p* = 0.877, *d* = -0.03; Stroop interference relative to neutral words: *F*(1,105) = 0.292, *p* = 0.60, *d* = 0.12). The delta plot for task conflicts showed no linear trend (*F*(1,105) = 2.433, *p* = 0.122, *d* = 0.33) nor quadratic trend (*F*(1,105) = 0.987, *p* = 0.323, *d* = -0.21).

5.2.2.2.1.2 Stroop facilitation effects

Quantile plots and delta plots based on untrimmed data from Experiment 5 for Stroop facilitation effects in pinyin with and without tones are presented in Figure 5.4 and 5.5.

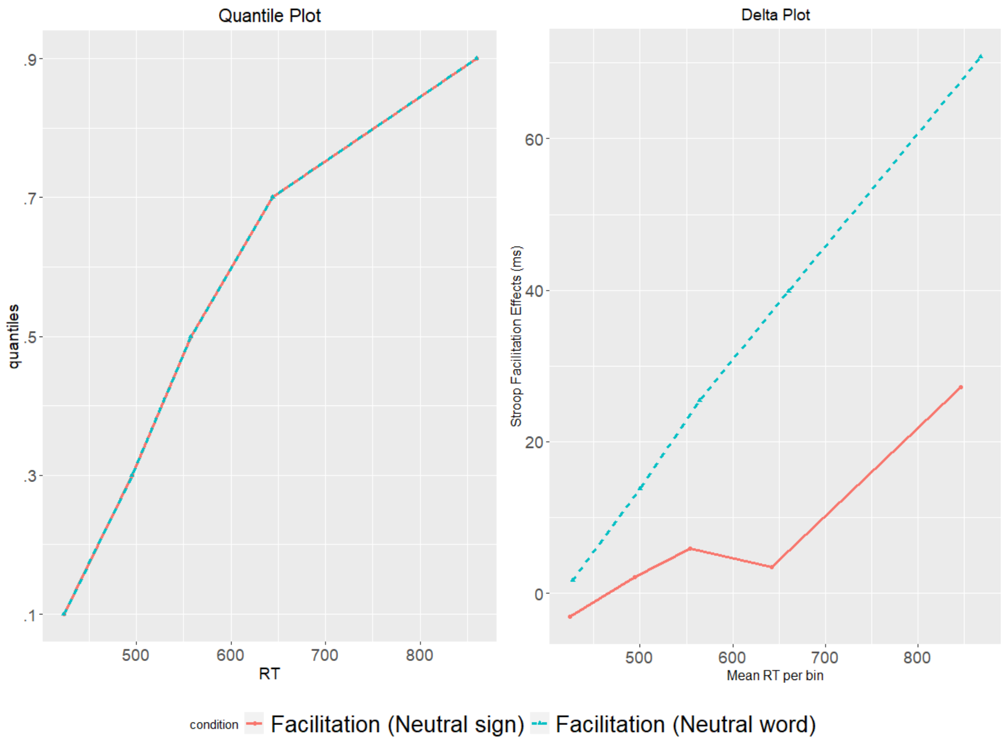


Figure 5.4. Quantile plots and delta plots for Stroop facilitation effects with tonal information in Experiment 5.

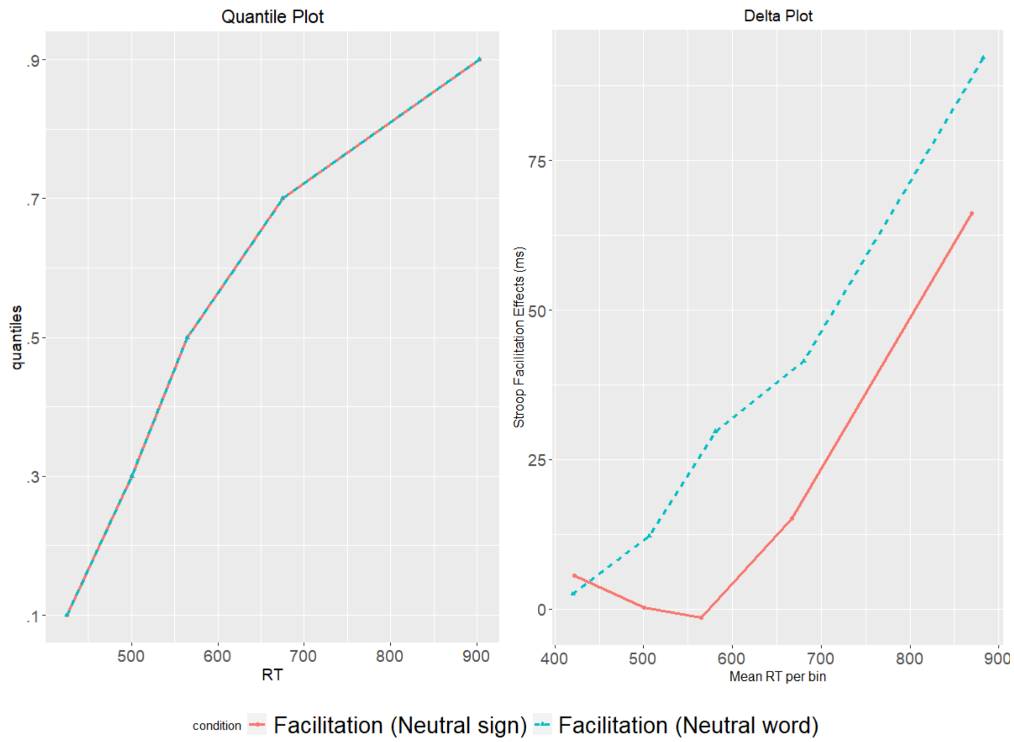


Figure 5.5. Quantile plots and delta plots for Stroop facilitation effects without tonal information in Experiment 5.

Analyses were separated for pinyin with and without tone conditions. Then, a series of one-way ANOVAs were performed for each facilitation component in pinyin with and without tone conditions.

For Stroop facilitation effects in pinyin with tones, the delta plot for Stroop facilitation relative to neutral words showed a positive linear trend ($F(1,105) = 17.269$, $p < 0.001$, $d = 0.89$), and no quadratic trend ($F(1,105) = 0.745$, $p = 0.40$, $d = 0.18$). The delta plots for Stroop facilitation relative to neutral signs showed no linear trend ($F(1,105) = 1.351$, $p = 0.25$, $d = 0.25$) nor quadratic trend ($F(1,105) = 0.241$, $p = 0.625$, $d = 0.10$).

For Stroop facilitation effects in pinyin without tones, the delta plots for Stroop facilitation relative to neutral signs showed a linear trend ($F(1, 105) = 9.767$, $p = 0.002$, $d = 0.67$) and a quadratic trend ($F(1,105) = 6.469$, $p = 0.012$, $d = 0.54$). The delta plot for Stroop facilitation relative to neutral words showed a positive linear trend ($F(1, 105)$

= 18.022, $p < 0.001$, $d = 0.91$), and no quadratic trend ($F(1,105) = 1.728$, $p = 0.19$, $d = 0.28$).

5.2.3 Discussion

Stroop interference effects relative to neutral signs were found with Chinese words written in pinyin both with and without tonal information conditions. The findings replicated the results of Liu & Weng (2007) based on colour patches as neutral controls and most importantly, in a completely pinyin context (without Chinese characters or English words).

Stroop interference effects relative to neutral words were significant when tonal information was present with the pinyin but not when pinyin was presented without tonal information. There was trend that tonal information would produce stronger Stroop interference effects; however, it is difficult to make a strong claim about the difference between presenting with and without tones.

Stroop facilitation effects relative to neutral signs were absent when presented with or without tonal information. Stroop facilitation effects relative to neutral words were found with pinyin stimuli that contained tonal and without tonal information. These results are consistent with Dalrymple-Alford's (1972) observation that facilitation effects occur only when compared to neutral words rather than neutral signs. In fact, when neutral signs are used as the neutral control often no or even negative facilitation effects are reported (Logan & Zbrodoff, 1998; Nealis, 1973; Roelofs, 2012; Schulz, 1979; Sichel & Chandler, 1969; Vanayan, 1993).

Because pinyin without tonal information links to more possible Chinese characters, it is predicted that reduced Stroop effects would be observed relative to pinyin words with tonal information. Although numerically the tonal information condition has larger Stroop interference effects than the condition without tonal information (Stroop interference relative to neutral signs: 37 ms vs. 30 ms, Stroop interference relative to neutral words:

24 ms vs. 9 ms), the comparisons between these conditions were not significant or showed only a trend towards significance.

Both Stroop interference and facilitation effects have been observed in Experiment 5 with Chinese words written in pinyin. The observed Stroop interference and facilitation effects are similar to those obtained with Chinese characters (e.g. Experiment 4) and English words (e.g., Dalrymple-Alford, 1972).

The distributional analysis showed that the Stroop interference effects (relative to neutral signs) with Chinese words written in pinyin had no inhibition applied to resolve the conflict between colour and word information. Task conflicts in pinyin with and without diacritics conditions both showed weak inhibition applied. Stroop facilitation (relative to neutral words) with Chinese words written in pinyin had strong enhancement applied to converge the colour and word information.

The next experiment focuses on decomposed conflict/facilitation components using pinyin stimuli. Because the presentation of tonal information does not greatly impact the Stroop interference relative to neutral words, all stimuli in the subsequent experiments are presented with tonal information, which is the normal form in which pinyin is presented.

5.3 Experiment 6

5.3.1 Methods

5.3.1.1 Participants

Thirty-two participants were recruited through Prolific (mean age = 26.16, range = 20–34, females = 21), all of whom were native speakers of Chinese from mainland China, with normal or corrected-to-normal vision. Each participant received an inconvenience allowance after completing the experiment, which was approved by the ethics committee at the School of Psychology, University of Nottingham, UK (Ethics approval number:

S1289R Chair Approval). All participants had read and agreed to the statements on the informed consent form prior to data collection.

5.3.1.2 Stimuli and Design

In Experiment 6, the stimuli were the same as in Experiment 5, with the addition of an extra condition and three more neutral words in pinyin. Three colours were used: green, yellow, and blue.

The experiment contained six types of stimuli, four of which were the same as in Experiment 5: incongruent colour characters written in pinyin (also treated as incongruent homophone condition, e.g. huáng, "yellow", presented in blue), congruent colour characters written in pinyin (also treated as the congruent homophone condition, e.g. huáng presented in yellow), neutral characters written in pinyin (e.g. xún has no phonological or semantic association with colours), and neutral signs (e.g. xxxxx). Two new types of stimuli were added: incongruent colour-associated character written in pinyin (e.g. cǎo, "grass", presented in yellow), and congruent colour-associated character written in pinyin (e.g. cǎo, "grass", presented in green). Neutral characters in pinyin were matched with each colour character stimulus in terms of length, stroke count, frequency, and family size. Neutral signs were strings of five x's that had the same length as the longest stimulus, which was huáng.

Table 5.5 presented the stimuli used in Experiment 6, and detailed linguistic properties of these stimuli could be found in Appendices E and F.

Pinyin	Possible Meaning 1	Possible Meaning 2	Possible Meaning 3	Possible Meaning 4	Possible Meaning 5
Colour characters in pinyin (homophones)					
/lv4/	green	law	filter	consider	
/huang2/	yellow	emperor	phoenix	jade	bright
/lan2/	blue	holdback	eupatorium	basket	railing
Colour associated characters in pinyin					
/cao3/	grass				
/jin1/	gold	today	a unit of weight (=1/2 kilogram)		
/tian1/	sky	add			
Neutral characters in pinyin					
/po1/	very	splash	slope		
/qiong2/	poor	red jade	high		
/xun4/	instruct	interrogate	modest	tame	sacrifice
/xiu1/	repair	rest	shy	make a din	
/rou4/	meat				
/zhua1/	scratch				

Table 5.5. Stimuli used in Experiment 6.

The experiment contained six stimulus types: homophone-incongruent, homophone-congruent, colour-associated-incongruent, colour-associated-congruent, neutral pinyin, and neutral signs. There were 144 trials in each block, with 18 trials in each experimental condition (including homophone-incongruent, homophone-congruent, colour-associated-incongruent, and colour-associated-congruent conditions) and 36 trials in both neutral pinyin and neutral signs, so that the trials in each experimental condition had the same number of trials as in neutral control conditions. There were three blocks, and each participant received a total of 432 trials. In each block, each experimental condition had the same number of trials by manipulating repetition: homophone-incongruent stimuli were repeated three times (2 colours × 3 pinyin × 3 repetitions = 18 trials), homophone-congruent six times (1 colour × 3 pinyin × 6 repetitions = 18 trials), colour-associated-incongruent stimuli were repeated three times (2 colours × 3 pinyin × 3 repetitions = 18 trials), colour-associated-congruent six times (1 colour × 3 pinyin × 6 repetitions = 18 trials), and both neutral control conditions had the same number of trials by manipulating repetition: neutral pinyin twice (3 colours × 6 pinyin × 2 repetitions = 36 trials), and neutral signs twelve times (3 colours × 1 pinyin × 12 repetitions = 36 trials).

A sufficient number of observations were included in this experiment: 1,728 observations for congruent and incongruent trials (32 participants × 3 blocks × 18 trials), and 3,456 for neutral trials (32 participants × 3 blocks × 36 trials).

5.3.1.3 Procedure

Same as in Experiment 5.

5.3.2 Results

5.3.2.1 Mean RT analysis

Incorrect responses were discarded (2.73%) for the response time analyses. RTs below 200 ms and above 1700 ms were discarded before the analyses (0.96%). Fixed factor "condition" used simple coding so that neutral trials will be compared to all other trials. Fixed factor "colour" used deviation coding so that each colour will be compared to the mean of all colours.

First, different types of distributions were compared, and the results of the performance analyses revealed that the Inverse Gaussian fitted the data best. Next, four models with different fixed and random structures were compared. A summary of Information Criterion for these models is shown in Table 5.6.

Name	R2 (marg.)	AIC weights	AICc weights	BIC weights	Performance-Score
Model 4	0.425	0.731	0.731	0.991	95.42%
Model 3	0.425	0.269	0.269	0.009	67.25%
Model 2	0.111	1.61E-70	1.61E-70	3.93E-67	13.92%
Model 1	0.103	2.77E-70	2.77E-70	1.59E-68	0.20%

Model Structure:
Model 1: glmer(RT ~ condition + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"))
Model 2: glmer(RT ~ condition + (1|subject), family = inverse.gaussian(link = "identity"))
Model 3: glmer(RT ~ condition + colour + (1|subject) + (1|item), family = inverse.gaussian(link = "identity"))
Model 4: glmer(RT ~ condition + colour + (1|subject), family = inverse.gaussian(link = "identity"))

Table 5.6. Summary of Information Criterion for models used in Experiment 6.

Model 4 is the best fit for the data. Based on Model 4, a full model was constructed as follows:

```
glmer(RT ~ condition + colour + (1+condition|subject), family = inverse.gaussian
(link = "identity"))
```

However, the full model did not converge. Therefore, the random structures were simplified until the model successfully converged, and the final model was:

```
glmer(RT ~ condition + colour + (1|subject), family = inverse.gaussian(link = "ide
ntity"))
```

A summary of the final model can be found in Appendix N. Table 5.7 provided a summary of the descriptive statistics and results of final model and further contrasts are presented in Table 5.8.

Stimulus Type (examples)		
Homophone-Incongruent words	RT (SE)	649 (9.55)
/lán/$_{/lǜ/}$ (*BLUE*green)	ER (SE)	3.25 (0.43)
Colour-Neutral words	RT (SE)	651 (9.22)
/pō/$_{/lǜ/}$ (*VERY*green)	ER (SE)	2.58 (0.27)
Colour-Neutral signs	RT (SE)	627 (9.30)
xxxxx$_{/lǜ/}$ (*XXXXX*green)	ER (SE)	2.40 (0.26)
Homophone-Congruent words	RT (SE)	623 (9.49)
/lán/$_{/lán/}$ (*BLUE*blue)	ER (SE)	2.66 (0.39)
Colour-Associated-Incongruent words	RT (SE)	650 (9.82)
/tiān/$_{/lǜ/}$ (*SKY*green)	ER (SE)	3.01 (0.41)
Colour-Associated-Congruent words	RT (SE)	630 (9.68)
/tiān/$_{/lán/}$ (*SKY*blue)	ER (SE)	1.97 (0.33)

Table 5.7. Mean reaction times (RT, in milliseconds), error rates (ER, %), and standard errors (SE, in parentheses) as a function of stimulus type from the GLMMs in Experiment 6.

Stroop-like effects		
Stroop Interference relative to neutral signs (Homophone-Incongruent – Colour Neutral Signs)	RT diff.	22***
Stroop Interference relative to neutral words (Homophone-Incongruent – Colour Neutral Words)	RT diff.	-2
Phonological Conflict (Homophone-Incongruent – Colour-Associated-Incongruent)	RT diff.	-1
Semantic Conflict (Colour-Associated-Incongruent – Colour Neutral Words)	RT diff.	-1
Task Conflict (Colour Neutral Words – Colour Neutral Signs)	RT diff.	24***
Stroop Facilitation relative to neutral signs (Colour Neutral Signs – Homophone-Congruent)	RT diff.	4
Stroop Facilitation relative to neutral words (Colour Neutral Words – Homophone-Congruent)	RT diff.	28***
Semantic Facilitation (Colour Neutral Words – Colour-Associated-Congruent)	RT diff.	21***
Phonological Facilitation (Colour-Associated-Congruent – Homophone-Congruent)	RT diff.	7

*Table 5.8. Stroop-like effects in Experiment 6 (in milliseconds). ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $0.05 < {}^{+}p < 0.1$; RT diff., reaction time differences.*

For Stroop interference effects relative to neutral signs, the RT difference between homophone-incongruent conditions and neutral signs was significant (M = 22 ms, SE = 4.70 ms, z = 4.699, p < 0.001). For Stroop interference effects relative to neutral words, the RT difference was not significant (M = 2 ms, SE = 4.66 ms, z = 0.459, p = 0.647). In terms of phonological conflict (homophone-incongruent vs. colour-associated-incongruent) and semantic conflict (colour-associated-incongruent vs. neutral words), there were no significant results (phonological conflict: M = -1 ms, SE = 5.50 ms, z = -0.141, p = 0.888; semantic conflict: M = -1 ms, SE = 4.59 ms, z = -0.296, p = 0.767). Task conflicts were significant (M = 24 ms, SE = 3.76 ms, z = 6.436, p < 0.001).

For Stroop facilitation effects relative to neutral signs, the RT difference between homophone-congruent conditions and neutral signs was not significant (M = 4 ms, SE = 4.48 ms, z = 0.841, p = 0.40). For Stroop facilitation effects relative to neutral words, the RT difference was significant (M = 28 ms, SE = 4.56 ms, z = 6.129, p < 0.001). For semantic facilitation effects, the RT difference between neutral words and colour-associated-congruent conditions was significant (M = 21 ms, SE = 4.27 ms, z = 4.933, p < 0.001). For phonological facilitation effects, the RT difference between colour-associated-congruent and homophone-congruent conditions was not significant (M = 7 ms, SE = 5.07 ms, z = 1.364, p = 0.173).

5.3.2.2 Error analysis

The model for error analysis is identical to the one used in mean RT analysis, except for using the Gaussian distribution:

lmer(ER ~ condition + colour + (1|subject))

Error rates were similar for all Stroop interference and facilitation components.

5.3.2.3 Distributional analysis

Quantile plots and delta plots based on untrimmed data from Experiment 6 for Stroop interference and facilitation effects are presented in Figures 5.6 and 5.7.

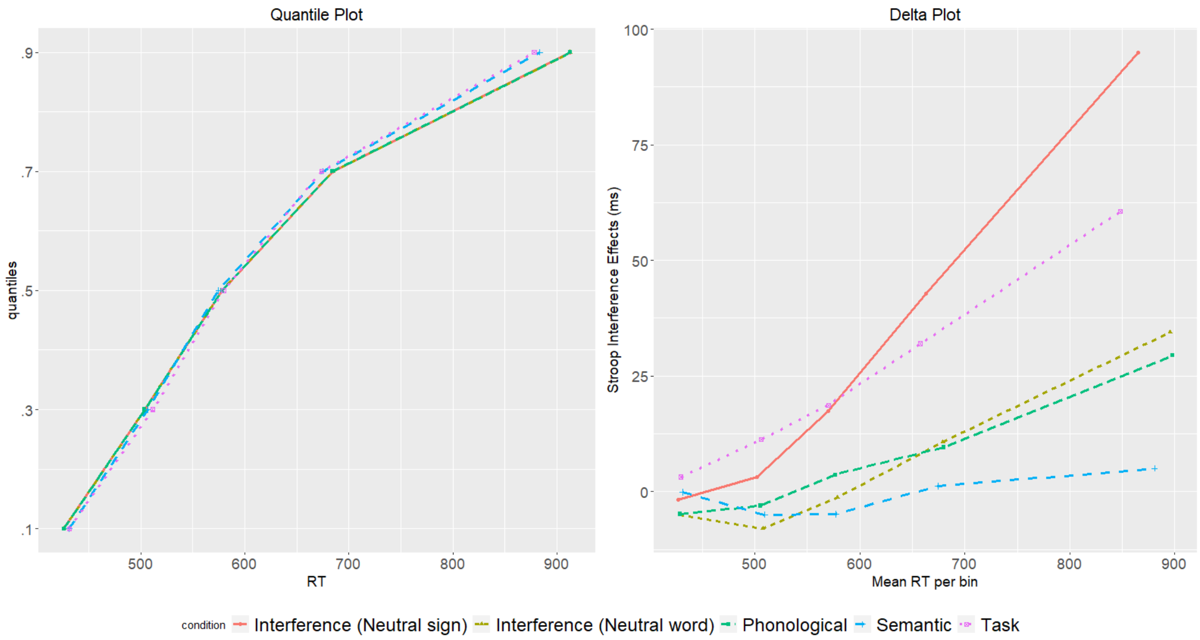


Figure 5.6. Quantile plots and delta plots for Stroop interference effects in Experiment 6.

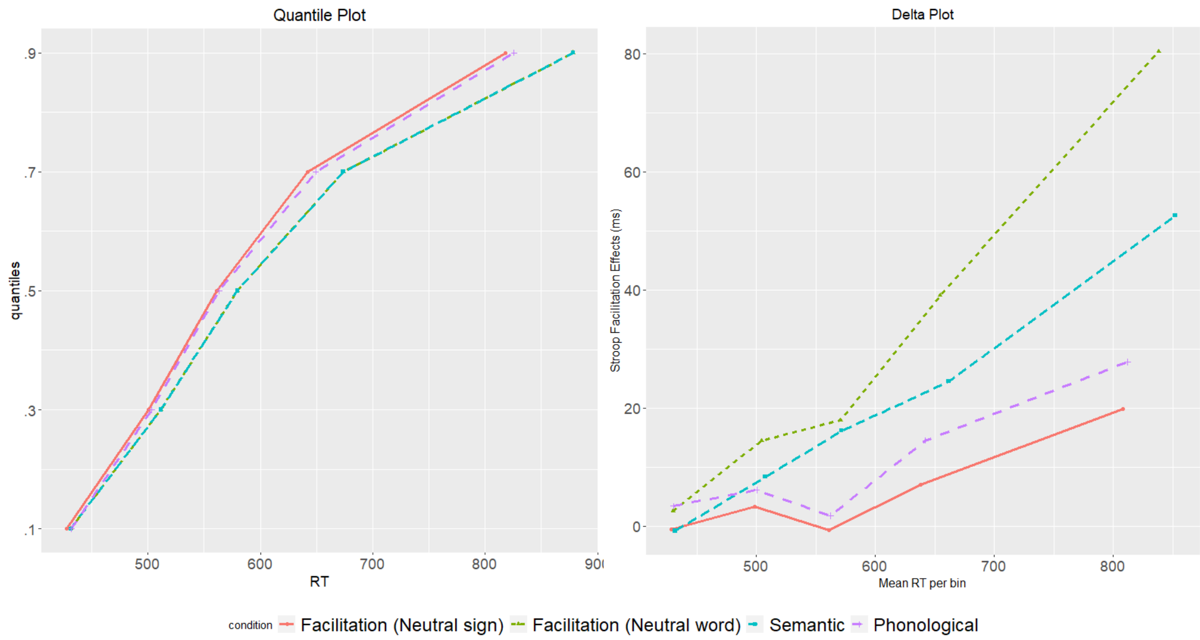


Figure 5.7. Quantile plots and delta plots for Stroop facilitation effects in Experiment 6.

Analyses were separated for Stroop interference. Then, a series of one-way ANOVAs were performed for each character type in Stroop interference and Stroop facilitation.

For Stroop interference effects, the delta plots for Stroop interference relative to neutral signs, Stroop interference relative to neutral words, phonological, and task conflicts showed positive linear trends (Stroop interference relative to neutral signs: $F(1,155) = 41.704$, $p < 0.001$, $d = 1.14$; Stroop interference relative to neutral words: $F(1,155) = 7.884$, $p = 0.035$, $d = 0.50$; phonological conflict: $F(1,155) = 6.778$, $p = 0.01$; task conflict: $F(1,155) = 20.749$, $p < 0.001$, $d = 0.46$, $d = 0.81$), and a quadratic trend for Stroop interference relative to neutral signs but not for other conflict components (Stroop interference relative to neutral signs: $F(1,155)= 6.773$, $p = 0.01$, $d = 0.46$; Stroop interference relative to neutral words: $F(1,155) = 2.547$, $p = 0.113$, $d = 0.28$; phonological conflict: $F(1,155) = 1.707$, $p = 0.193$, $d = 0.23$; task conflict: $F(1,155) = 1.079$, $p = 0.3$, $d = 0.18$). The delta plot for semantic conflicts revealed no linear component ($F(1,155) = 0.642$, $p = 0.424$) or quadratic component ($F(1,155) = 0.445$, $p = 0.506$).

For Stroop facilitation effects, the delta plots for Stroop facilitation relative to neutral words and semantic facilitation showed positive linear trends (Stroop facilitation relative to neutral words: $F(1,155) = 17.86$, $p < 0.001$, $d = 0.75$; semantic facilitation: $F(1,155) = 7.288$, $p = 0.008$, $d = 0.48$), and no quadratic trends (Stroop facilitation relative to neutral words: $F(1,155) = 1.059$, $p = 0.305$, $d = 0.18$; semantic facilitation: $F(1,155) = 0.081$, $p = 0.777$, $d = 0.05$). The delta plots for Stroop facilitation relative to neutral signs and phonological facilitation showed no linear trends (Stroop facilitation relative to neutral signs: $F(1,155) = 1.385$, $p = 0.241$, $d = 0.21$; phonological facilitation: $F(1,155) = 2.089$, $p = 0.15$, $d = 0.26$) or quadratic trends (Stroop facilitation relative to neutral signs: $F(1,155) = 0.132$, $p = 0.716$, $d = 0.06$; phonological facilitation: $F(1,155) = 0.46$, $p = 0.499$, $d = 0.12$).

5.3.3 Discussion

Strong Stroop interference was observed relative to neutral signs and Stroop facilitation was observed relative to neutral words in Experiment 6. Stroop interference relative to neutral words disappeared in Experiment 6 compared to Experiment 5. This finding is unexpected because most studies have reported Stroop interference effects using neutral words in a manual Stroop task (Augustinova et al., 2019; Fennell & Ratcliff, 2019; Redding & Gerjets, 1977; Wang et al., 2010; Zahedi et al., 2019). In Experiment 5, Stroop interference effects relative to neutral words were observed with tonal information. Because Experiments 5 and 6 only differed in the inclusion of semantic-associated words, it could be that this condition influenced the Stroop interference effects.

Although the current study argued that the RT difference between homophone trials and colour-associated trials should be phonological conflict/facilitation, it is acknowledged that potential response conflict/facilitation is included in phonological components. This is because one of the meanings of homophones corresponds to the colour words. It is impossible to eliminate the response components from using homophones in pinyin.

however, this response component is weaker compared to the response component elicited by colour characters, because colour characters have a one-to-one correspondence to the colour names, whereas homophones in pinyin may lead to other non-colour-related meanings.

Task conflicts were present in this experiment, whereas no task conflicts were found in Experiment 4 with Chinese characters and in Experiment 5 with Chinese words written in pinyin with tones. Redding and Gerjets (1977) and Augustinova et al. (2019) did not observe task conflicts in the manual responses, but they observed them in vocal responses.

In Experiment 6, a colour-associated condition was included to investigate phonological and semantic conflicts/facilitation. However, the results of the manual Stroop task revealed no phonological conflicts/facilitation with Chinese words written in pinyin. Phonological conflicts/facilitation were also not found with Chinese words written in pinyin (phonological facilitation was found in Experiment 4 with Chinese characters). Semantic conflicts were not found with Chinese words written in pinyin either. Several manual Stroop studies have reported strong semantic conflicts using colour-associated conditions, including English and Chinese characters (Augustinova et al., 2019; Augustinova et al., 2018; Sharma & McKenna, 1998; results of Experiment 4). Presenting Chinese words in pinyin, however, did not elicit phonological and semantic conflicts. This means that either the phonological or semantic information was not activated by pinyin, or the magnitude of activation was not strong enough to impact the colour categorization task.

Although colour-associated words written in pinyin eliminated conflict components, facilitation components were enhanced. The results of Experiment 6 revealed a significant effect of semantic facilitation. Thus, colour-associated-congruent words written in pinyin resulted in faster responses than neutral words. The semantic information carried by colour-associated words written in pinyin was successfully activated in congruent conditions. In contrast, significant semantic facilitation was not reported in Experiment 4 with Chinese characters, however, Augustinova et al. (2019) reported it with French words.

The contrast between the results of Stroop interference and facilitation suggests potential differences in how pinyin and Chinese characters are processed. Chinese characters link directly to meaning and phonology. In contrast, the orthography of pinyin does not directly link to semantics but to phonology. Pinyin words are essentially homophones to characters. The current results are consistent with the idea that the Stroop interference effects with pinyin stimuli are not strongly associated with word meaning. In contrast, pinyin is directly associated with phonology; therefore, it does not require the conversion from character to phonology, which leads to faster responses when pinyin and colour information are matched in congruent conditions.

The distributional analyses support the findings of the mean RT analyses. Crucially, the delta plot for semantic facilitation showed an upward trend, indicating strong enhancement was applied to converge the colour and word information. For task conflicts, there is no inhibition applied, so the conflicts became larger across the quantiles.

Having explored phonological and semantic conflict/facilitation components in pinyin words, the next experiment investigates the activation of sublexical components of Chinese words written in pinyin.

5.4 Experiment 7

5.4.1 Methods

5.4.1.1 Participants

Thirty-seven participants were recruited through Prolific (mean age = 23.83, range = 20–34, females = 26). All were native Chinese speakers from mainland China, and participants had normal or corrected-to-normal vision. Each participant received an inconvenience allowance after completing the experiment, which was approved by the ethics committee at the School of Psychology, University of Nottingham, UK (Ethics approval number: S1417 Chair Approval of Minor Amendments). All participants read and agreed to the statements on the informed consent form prior to data collection.

5.4.1.2 Stimuli and Design

In Experiment 7, which explored the role of phonetic radicals in using pinyin, three new colour words that can be used as phonetic radicals were selected: cyan, yellow, and red. The experiment contained six types of stimuli: incongruent colour characters written in pinyin (also treated as the incongruent homophone condition, e.g. huáng, "yellow", presented in red colour), congruent colour characters written in pinyin (also treated as the incongruent homophone condition, e.g. huáng presented in yellow colour), neutral character written in pinyin (e.g. jù has no phonological or semantic association with colours), and neutral signs (e.g. xxxxx). Two new types of stimuli are incongruent invalid-radical characters written in pinyin (e.g. cāi, "guess", presented in yellow colour), and congruent invalid-radical characters written in pinyin (e.g. cāi, "grass", presented in cyan colour). Neutral characters in pinyin are matched with each colour character stimulus in terms of length, stroke count, frequency, and family size. Neutral signs are strings of five x's that had the same length as the longest stimulus, which is huáng. As in Experiment 6, pinyin words were presented with tonal information.

Table 5.9 presents the stimuli used in Experiment 7, and detailed linguistic properties of these stimuli can be found in Appendices G and H.

Pinyin	Possible Meaning 1	Possible Meaning 2	Possible Meaning 3	Possible Meaning 4	Possible Meaning 5
Colour characters in pinyin (homophones)					
/qing1/	cyan	clear	light/little	bend/lean	
/huang2/	yellow	emperor	phoenix	jade	bright
/zhu1/	red	pig	pearl	stub	
Invalid-Radical condition in pinyin					
/cai1/	guess				
/heng2/	horizontal	permanent	weight/measure	purlin	
/shu2/	different	book	transport/defeat	comb	uncle
Neutral characters in pinyin					
/ju4/	tool	sentence	assemble	rely on	distance
/ceng2/	already	layer/tier			
/diu1/	discard				
/zhang4/	tent	swell	rely on	hold	
/bang3/	placard	bind/tie			
/le4/	strangle				

Table 5.9. Stimuli used in Experiment 7.

The experiment contained six types of stimuli: homophone-incongruent, homophone-congruent, invalid-radical -incongruent, invalid-radical-congruent, neutral pinyin, and neutral signs. There were 144 trials in each block, with 18 trials in each experimental condition (including homophone-incongruent, homophone-congruent, invalid-radical-incongruent, and invalid-radical-congruent conditions) and 36 trials in both neutral pinyin and neutral signs, so that the trials in experimental conditions had the same amount as in neutral control conditions. There were 3 blocks, meaning each participant received 432 trials in total. In each block, each experimental condition has the same number of trials by manipulating repetition: homophone-incongruent stimuli repeated three times (2 colours × 3 pinyin × 3 repetitions = 18 trials), homophone-congruent six times (1 colour × 3 pinyin × 6 repetitions = 18 trials), invalid-radical-incongruent stimuli repeated three times (2 colours × 3 pinyin × 3 repetitions = 18 trials), invalid-radical-congruent six times (1 colour × 3 pinyin × 6 repetitions = 18 trials); both neutral control conditions had the same number of trials by manipulating repetition: neutral pinyin twice (3 colours × 6 pinyin × 2 repetitions = 36 trials), and neutral signs twelve times (3 colours × 1 pinyin × 12 repetitions = 36 trials).

There were 1,998 observations for congruent and incongruent trials (37 participants × 3 blocks × 18 trials) and 3,996 observations for neutral trials (37 participants × 3 blocks × 36 trials), which met the recommendation of 1,600 observations per condition.

5.4.1.3 Procedure

Same as in Experiments 5 and 6.

5.4.2 Results

5.4.2.1 Mean RT analysis

Incorrect responses were discarded (3.53%) for the response time analyses. RTs below 200 ms and above 1700 ms were discarded before the analyses (1.39%). Fixed factor "condition" used simple coding so that neutral trials will be compared to all other trials. Fixed factor "colour" used deviation coding so that each colour will be compared to the mean of all colours.

First, different types of distributions were compared, and the results of the performance analyses revealed that the Inverse Gaussian fitted the data best. Next, four models with different fixed and random structures were compared. A summary of Information Criterion for these models is shown in Table 5.10.

Name	R2 (marg.)	AIC weights	AICc weights	BIC weights	Performance-Score
Model 4	0.205	0.731	0.731	0.992	87.50%
Model 3	0.205	0.269	0.269	0.008	59.32%
Model 2	0.041	6.95E-41	6.96E-41	4.28E-39	13.25%
Model 1	0.045	6.71E-41	6.73E-41	1.87E-37	3.07%

Model Structure:
Model 1: glmer(RT ~ condition + (1|subject) + (1|item), family = inverse.gaussian(link = "identity") , control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 2: glmer(RT ~ condition + (1|subject), family = inverse.gaussian(link = "identity") , control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 3: glmer(RT ~ condition + colour + (1|subject) + (1|item), family = inverse.gaussian(link = "identity") , control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 4: glmer(RT ~ condition + colour + (1|subject), family = inverse.gaussian(link = "identity") , control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

Table 5.10. Summary of Information Criterion for models used in Experiment 7.

Model 4 is the best fit for the data. Based on Model 4, a full model was constructed as follows:

glmer(RT ~ condition + colour + (1+condition|subject), family = inverse.gaussian (link = "identity"))

However, the full model did not converge. Therefore, the random structures were simplified until the model successfully converged, and the final model was:

glmer(RT ~ condition + colour + (1|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

A summary of the final model can be found in Appendix O. Table 5.11 provided a summary of the descriptive statistics and results of final model and further contrasts are presented in Table 5.12.

Stimulus Type (examples)		
Homophone-Incongruent words	RT (SE)	678 (10.8)
/lán/$_{\text{/lǜ/}}$ (*BLUE*green)	ER (SE)	3.77 (0.43)
Colour-Neutral words	RT (SE)	669 (10.6)
/pō/$_{\text{/lǜ/}}$ (*VERY*green)	ER (SE)	3.36 (0.29)
Colour-Neutral signs	RT (SE)	659 (10.4)
xxxxx$_{\text{/lǜ/}}$ (*XXXXX*green)	ER (SE)	3.38 (0.29)
Homophone-Congruent words	RT (SE)	651 (10.5)
/lán/$_{\text{/lán/}}$ (*BLUE*blue)	ER (SE)	2.87 (0.37)
Invalid-Radical-Incongruent words	RT (SE)	675 (10.8)
/cāi/$_{\text{/lán/}}$ (GUESSblue)	ER (SE)	3.31 (0.40)
Invalid-Radical-Congruent words	RT (SE)	661 (10.8)
/cāi/$_{\text{/qīng/}}$ (GUESScyan)	ER (SE)	2.91 (0.38)

Table 5.11. Mean reaction times (RT, in milliseconds), error rates (ER, %), and standard errors (SE, in parentheses) as a function of stimulus type from the GLMMs in Experiment 7.

Stroop-like effects		
Stroop Interference relative to neutral signs (Homophone-Incongruent – Colour Neutral Signs)	RT diff.	19***
Stroop Interference relative to neutral words (Homophone-Incongruent – Colour Neutral Words)	RT diff.	9
Phonology-driven Semantic Conflict (Homophone-Incongruent – Invalid-Radical-Incongruent)	RT diff.	4
Semantic Conflict (Invalid-Radical) (Invalid-Radical -Incongruent – Colour Neutral Words)	RT diff.	6
Task Conflict (Colour Neutral Words – Colour Neutral Signs)	RT diff.	10*
Stroop Facilitation relative to neutral signs (Colour Neutral Signs – Homophone-Congruent)	RT diff.	7^{+}
Stroop Facilitation relative to neutral words (Colour Neutral Words – Homophone-Congruent)	RT diff.	17***
Semantic Facilitation (Invalid-Radical) (Colour Neutral Words – Invalid-Radical-Congruent)	RT diff.	8
Phonology-driven Semantic Facilitation (Invalid-Radical-Congruent – Homophone-Congruent)	RT diff.	9

*Table 5.12. Stroop-like effects in Experiment 7 (in milliseconds). ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < $^{+}$p < 0.1; RT diff., reaction time differences.*

For Stroop interference effects relative to neutral signs, the RT difference between the homophone-incongruent conditions and neutral signs was significant (M = 19 ms, SE = 4.52 ms, z = 4.303, *p*< 0.001). For Stroop interference effects relative to neutral words, it was not significant (M = 9 ms, SE = 4.42 ms, z = 2.062, *p* = 0.259). In terms of phonology-driven semantic conflict (homophone-incongruent vs. invalid-radical-incongruent) and radical's semantic conflict (invalid-radical-incongruent vs. neutral word), there were no significant results (phonology-driven semantic conflict: M = 4 ms, SE = 5.38 ms, z = 0.653, *p* = 1; radical's semantic conflict: M = 6 ms, SE = 4.45 ms, z = 1.260, *p* < 0.624). Task conflicts were significant (M = 10 ms, SE = 3.47 ms, z = 2.981, *p* = 0.026).

For Stroop facilitation relative to neutral signs, the RT difference between homophone-congruent conditions and neutral signs was not significant (M = 7 ms, SE = 4.26 ms, z = 1.663, *p* = 0.386). Stroop facilitation effects relative to neutral words were significant (M = 17 ms, SE = 4.16 ms, z = 4.195, *p* < 0.001). The radical's semantic facilitation effect (neutral word minus invalid-radical-congruent) was not significant: M = 8 ms, SE = 4.03 ms, z = 2.085, *p* = 0.259. Phonology-driven semantic facilitation (invalid-radical-congruent minus homophone-congruent) was not significant: M = 9 ms, SE = 4.87 ms, z = 1.851, *p* = 0.321.

5.4.2.2 Error analysis

The model for error analysis is identical to the one used in mean RT analysis, except for using the Gaussian distribution:

lmer(ER ~ condition + colour + (1|subject))

Error rates were similar for all Stroop interference and facilitation components.

5.4.2.3 Distributional analysis

Quantile plots and delta plots based on untrimmed data from Experiment 7 for Stroop interference and facilitation effects are presented in Figures 5.8 and 5.9.

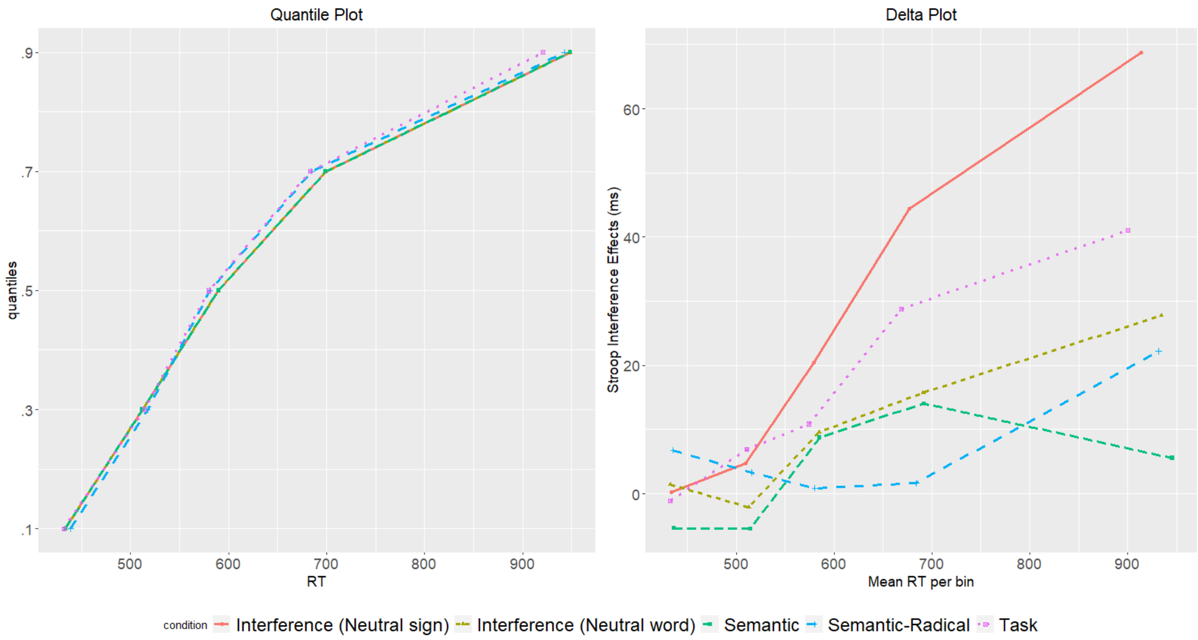


Figure 5.8. Quantile plots and delta plots for Stroop interference effects in Experiment 7.

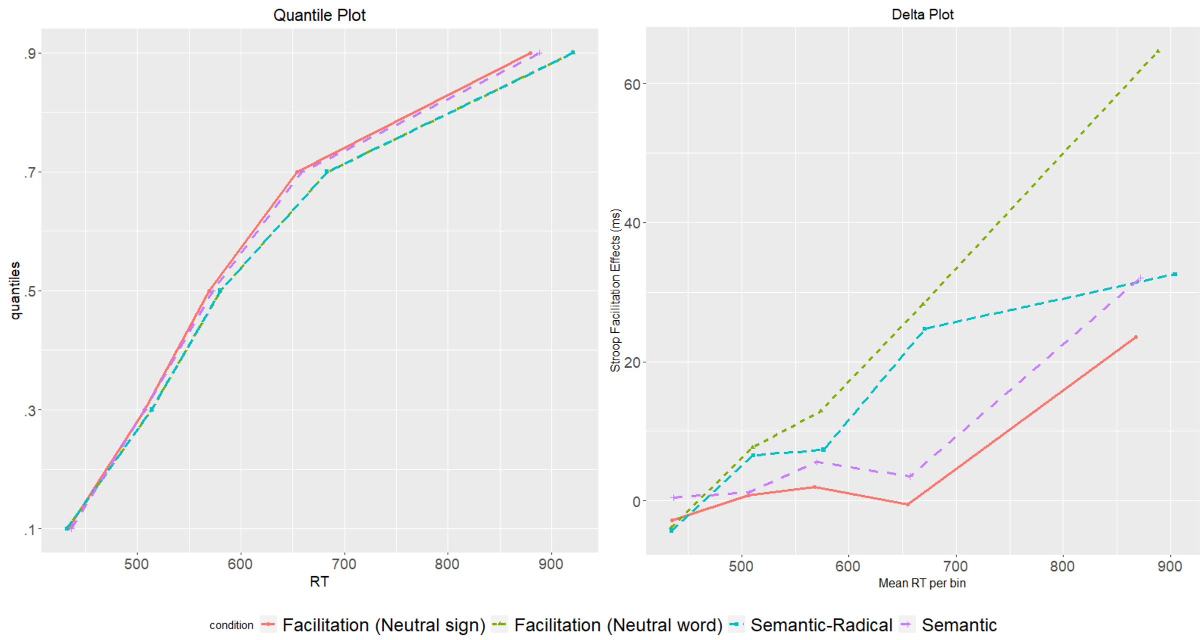


Figure 5.9. Quantile plots and delta plots for Stroop facilitation effects in Experiment 7.

Analyses were separated for Stroop interference and Stroop facilitation effects. Then, a series of one-way ANOVAs were performed for each character type in Stroop interference and Stroop facilitation.

For Stroop interference effects, the delta plots for Stroop interference relative to neutral signs and task conflicts showed positive linear trends (Stroop interference relative to neutral signs: $F(1,180) = 16.104$, $p < 0.001$, $d = 0.66$; task conflict: $F(1,180) = 11.54$, $p < 0.001$, $d = 0.56$), and no quadratic trends (Stroop interference relative to neutral signs: $F(1,180)= 0.847$, $p = 0.359$, $d = 0.15$; task conflict: $F(1,180) = 0.369$, $p = 0.544$, $d = 0.10$). The delta plots for Stroop interference relative to neutral words, semantic conflict, and radical's semantic conflict revealed no linear component (Stroop interference relative to neutral words: $F(1,180) = 2.543$, $p = 0.113$, $d = 0.26$; semantic conflict: $F(1,180) = 1.008$, $p = 0.317$, $d = 0.17$; radical's semantic conflict: $F(1,180) = 0.648$, $p = 0.422$, $d = 0.13$) nor quadratic component (Stroop interference relative to neutral words: $F(1,180) = 0.237$, $p = 0.627$, $d = 0.08$; semantic conflict: $F(1,180) = 0.279$, $p = 0.598$, $d = -0.09$; radical's semantic conflict: $F(1,180) = 1.427$, $p = 0.234$, $d = 0.20$).

For Stroop facilitation effects, the delta plots for Stroop facilitation relative to neutral words and radical's semantic facilitation showed positive linear trends (Stroop facilitation relative to neutral words: $F(1,180) = 28.212$, $p < 0.001$, $d = 0.87$; radical's semantic facilitation: $F(1,180) = 5.34$, $p = 0.022$, $d = 0.38$), and a trend for quadratic trend in Stroop facilitation relative to neutral words but no quadratic trend in radical's semantic facilitation (Stroop facilitation relative to neutral words: $F(1,180) = 2.91$, $p = 0.09$, $d = 0.28$; radical's semantic facilitation: $F(1,180) = 0.051$, $p = 0.822$, $d = 0.04$). The delta plots for Stroop facilitation relative to neutral signs and semantic facilitation showed no linear trends (Stroop facilitation relative to neutral signs: $F(1,180) = 2.64$, $p = 0.106$, $d = 0.27$; semantic facilitation: $F(1,180) = 2.494$, $p = 0.116$, $d = 0.26$) nor quadratic trends (Stroop facilitation relative to neutral signs: $F(1,180) = 1$, $p = 0.319$, $d = 0.16$; semantic facilitation: $F(1,180) = 1.007$, $p = 0.317$, $d = 0.17$).

5.4.3 Discussion

As in Experiments 5 and 6, strong Stroop interference relative to neutral signs was observed. Stroop interference relative to neutral words was not reported in Experiment 7, which was consistent with Experiment 6. It is difficult to interpret why this effect disappeared in Experiments 6 and 7 and occurred in Experiment 5, because the designs of the experiments are very similar.

For Stroop facilitation relative to neutral signs, there were no significant effects as in Experiments 5 and 6. Stroop facilitation relative to neutral words was significant as in Experiments 5 and 6. Thus, with invalid-radical conditions included, the basic Stroop effects remained strong and stable, which suggests that Chinese words written in pinyin carry semantic and phonological information as Chinese character do.

Task conflicts were found in Experiment 7, which is consistent with what was found in Experiment 6. This pattern was observed in previous studies with vocal responses (Augustinova et al., 2019; Redding & Gerjets, 1977) but not in manual responses (Experiment 4).

In the present experiment, the invalid-radical word written in pinyin was used to investigate whether sublexical information of Chinese words written in pinyin can still be activated. This activation process is rather indirect: a colour-word is embedded in the invalid-radical word while the word is written in pinyin rather than using Chinese characters, which means there is no orthographic information in terms of Chinese characters. The findings revealed that there was no phonology-driven semantic conflict/facilitation and radical's semantic conflict/facilitation. The sublexical components of Chinese words written in pinyin cannot be activated.

The distributional analysis showed that there was no inhibition applied to Stroop interference relative to neutral signs and task conflicts. Stroop facilitation relative to neutral words showed strong enhancement applied to converge the colour and word information.

This chapter so far has explored various aspects of Chinese words written in pinyin in the Stroop task, which includes tonal information, meaning association, and the activation of sublexical components in pinyin. As an input unit, pinyin can activate semantics and phonology of Chinese characters, as characters do. Liu and Weng (2007) mixed the presentation of Chinese characters, pinyin, and English in a Stroop task and found that the Stroop interference effects became smaller. The next experiment includes the investigation of both Stroop interference and facilitation effects when presenting both Chinese characters and Chinese words written in pinyin in one study.

5.5 Experiment 8

5.5.1 Methods

5.5.1.1 Participants

Thirty participants were recruited through Prolific (mean age = 28.3, range = 19–37, females = 23). All were native Chinese speakers from mainland China. Participants had normal or corrected-to-normal vision. Each participant received an inconvenience allowance after completing the experiment. The experiment was approved by the ethics committee at the School of Psychology, University of Nottingham, UK (Ethics approval number: S1420 Chair Approval of Minor Amendments). All participants read and agreed to the statements on the electronic consent form prior to data collection.

5.5.1.2 Stimuli and Design

The stimuli were presented either using Chinese characters or pinyin. For example, the word 黄 (meaning "yellow") was presented as huáng in pinyin. Three colours were used: green, yellow, and blue.

Stimuli were presented congruent, incongruent or neutral with the ink colour: incongruent colour character (e.g. 黄, "yellow", presented in blue colour), congruent colour character (e.g. 黄 presented in yellow colour), neutral character (e.g. 颇 has no phonological or semantic association to colours), neutral sign matched for character (e.g. a pseudo-character □), incongruent colour character in pinyin[4] (e.g. huáng, "yellow", presented in blue colour), congruent colour character in pinyin (e.g. huáng presented in yellow colour), neutral character in pinyin (e.g. xiū has no phonological or semantic association to colours), and neutral signs (e.g. xxxxx). Neutral characters and pinyin were matched with each colour word stimulus in terms of length, stroke count, frequency, and family size. The neutral sign matched for the character is a pseudo-character. It contains two identical radicals that have no pronunciation or meaning when combined. Neutral signs matched for pinyin are strings of five x's that have the same length as the longest stimulus, that is huáng. Pinyin stimuli are presented with tonal information by using diacritics.

---

[4] In Exp 5, 6 and 7, this condition was treated as homophone condition. In Exp 8, due to further comparison to its character counterpart and there involves no semantic or phonological conflict in the current experiment, this condition is treated as a colour word condition.

Table 5.13 presents the stimuli used in Experiment 8.

Character	Pinyin	Possible Meaning 1	Possible Meaning 2	Possible Meaning 3	Possible Meaning 4	Possible Meaning 5
Colour words						
绿	/lv4/	green	law	filter	consider	
黄	/huang2/	yellow	emperor	phoenix	jade	bright
蓝	/lan2/	blue	holdback	eupatorium	basket	railing
Neutral words						
颇	/po1/	very	splash	slope		
穷	/qiong2/	poor	red jade	high		
训	/xun4/	instruct	interrogate	modest	tame	sacrifice

Table 5.13. Stimuli used in Experiment 8.

The experiment involved a 4 (stimuli type: incongruent vs. congruent vs. neutral word vs. neutral sign) × 2 (script: character vs. pinyin) within-subject design. There were 144 trials in each block. Each participant received 3 blocks, that is, 432 trials in total. There are two scripts, and four conditions in each. In each block, incongruent stimuli repeated three times (2 colours × 3 word × 3 repetitions = 18 trials), congruent six times (1 colour × 3 word × 6 repetitions = 18 trials), neutral word twice (3 colours × 3 word × 2 repetitions = 18 trials), and neutral sign six times (3 colours × 1 word × 6 repetitions = 18 trials).

There were 1,620 observations for each script (character and pinyin) for each condition (30 participants × 3 blocks × 18 trials).

5.5.1.3 Procedure

The procedure was the same as in Experiments 5, 6, and 7. The only difference was that the fixation cross was changed to a fixation dot "●" (RGB: 0,0,0). This was done to avoid the possible priming effect brought by the fixation cross when using Chinese characters.

5.5.2 Results

5.5.2.1 Mean RT analysis

Incorrect responses were discarded (3.31%) for the response time analyses. RTs below 200 ms and above 1700 ms were discarded before the analyses (1.12%). Fixed factor “condition” used simple coding so that neutral trials will be compared to all other trials. Fixed factor “script” also used simple coding that character condition will be compared to pinyin condition. Fixed factor “colour” used deviation coding so that each colour will be compared to the mean of all colours.

First, different types of distributions were compared, and the results of the performance analyses revealed that the Inverse Gaussian fitted the data best. Next, four models with different fixed and random structures were compared. A summary of Information Criterion for these models is shown in Table 5.14.

Name	R2 (marg.)	AIC weights	AICc weights	BIC weights	Performance-Score
Model 4	0.402	0.731	0.731	0.991	88.43%
Model 3	0.402	0.269	0.269	0.009	60.26%
Model 1	0.16	3.08E-77	3.09E-77	1.71E-75	14.17%
Model 2	0.157	4.46E-78	4.48E-78	1.01E-74	1.06E-75

Model Structure:
Model 1: glmer(RT ~ condition * script + (1|subject) + (1|item), family = inverse.gaussian(link = "identity") , control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 2: glmer(RT ~ condition * script + (1|subject), family = inverse.gaussian(link = "identity”) , control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 3: glmer(RT ~ condition * script + colour + (1|subject) + (1|item), family = inverse.gaussian(link = "identity") , control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
Model 4: glmer(RT ~ condition * script + colour + (1|subject), family = inverse.gaussian(link = "identity") , control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))

Table 5.14. Summary of Information Criterion for models used in Experiment 8.

Model 4 is the best fit for the data. Based on Model 4, a full model was constructed as follows:

glmer(RT ~ condition * script + colour + (1+condition+script+condition:script|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer= "bobyqa", optCtrl=list(maxfun=2e5)))

However, the full model did not converge. Therefore, the random structures were simplified until the model successfully converged, and the final model was:

```
glmer(RT ~ condition * script + colour + (1+condition|subject), family = inverse.gaussian(link = "identity"), control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
```

A summary of the final model can be found in Appendix P. Table 5.15 provided a summary of the descriptive statistics and results of final model and further contrasts are presented in Table 5.16.

Stimulus Type (examples)		**Character**	**Pinyin**
Homophone-Incongruent words	RT (SE)	719 (12.4)	693 (12.6)
/lán/$_{\text{/lǜ/}}$ (*BLUE*green)	ER (SE)	3.72 (0.47)	2.85 (0.41)
Colour-Neutral words	RT (SE)	672 (11.0)	674 (11.2)
/pō/$_{\text{/lǜ/}}$ (*VERY*green)	ER (SE)	3.77 (0.47)	3.40 (0.45)
Colour-Neutral signs	RT (SE)	674 (11.4)	668 (11.2)
xxxxx$_{\text{/lǜ/}}$ (*XXXXX*green)	ER (SE)	2.91 (0.42)	2.48 (0.39)
Homophone-Congruent words	RT (SE)	658 (11.2)	658 (10.9)
/lán/$_{\text{/lán/}}$ (*BLUE*blue)	ER (SE)	2.78 (0.41)	2.48 (0.39)

Table 5.15. Mean reaction times (RT, in milliseconds), error rates (ER, %), and standard errors (SE, in parentheses) as a function of stimulus type from the GLMMs in Experiment 8.

Stroop effects		Character		Pinyin	Mean Diff.
Stroop Interference relative to neutral signs (Homophone-Incongruent – Colour Neutral Signs)	RT diff.	45***	>	25*	20**
Stroop Interference relative to neutral words (Homophone-Incongruent – Colour Neutral Words)	RT diff.	47***	>	19*	28***
Task Conflict (Colour Neutral Words – Colour Neutral Signs)	RT diff.	-2	≈	6	-8
Stroop Facilitation relative to neutral signs (Colour Neutral Signs – Homophone-Congruent)	RT diff.	16*	≈	10	6
Stroop Facilitation relative to neutral words (Colour Neutral Words – Homophone-Congruent)	RT diff.	13*	≈	16*	-3

*Table 5.16. Stroop effects in Experiment 8 (in milliseconds). ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < $^{+}$p < 0.1; Mean Diff., Mean differences; RT diff., reaction time differences.*

Follow-up comparisons were conducted using the "emmeans" package (1.8.4-1). Following Augustinova et al.'s (2019) procedure, uncorrected p-values were reported, and conclusions were based on that. Corrected p-values were also included using the Holm-Bonferroni method. The differences of Stroop effects in two scripts were exploratory analyses, and uncorrected p-values were provided.

Further contrast analyses of GLMMs revealed detailed comparisons between each stimulus type and script. For Stroop interference effects relative to neutral signs, the RT differences between colour-incongruent conditions and neutral signs were significant for both character and pinyin stimuli (character: M = 45 ms, SE = 8.83 ms, z = 5.118, $p < 0.001$; pinyin: M = 25 ms, SE = 8.63 ms, z = 2.904, $p = 0.015$). For Stroop interference effects relative to neutral words, the RT differences were also significant for both character and pinyin stimuli (character: M = 47 ms, SE = 7.26 ms, z = 6.540, $p < 0.001$; pinyin: M = 19 ms, SE = 6.89 ms, z = 2.814, $p = 0.015$). Task conflicts (neutral signs vs. neutral words) were not significant in either script (character: M = -2 ms, SE = 5.42 ms, z = 0.427, $p = 0.67$; pinyin: M = 6 ms, SE = 5.39 ms, z = 1.050, $p = 0.294$).

There was no difference between scripts in task conflicts M = -8 ms, SE = 5.45 ms, z = 1.463, $p = 0.143$). For Stroop interference effects relative to neutral signs and neutral words, the script differences were significant (vs. neutral signs: M = 20 ms, SE = 7.31 ms, z = 2.753, $p = 0.006$; vs. neutral words: M = 28 ms, SE = 6.33 ms, z = 4.438, $p < 0.001$).

For Stroop facilitation, the RT difference between colour-congruent conditions and neutral signs was significant for character stimuli (M = 16 ms, SE = 5.72 ms, z = 2.752, $p = 0.018$) but not for pinyin stimuli (M = 10 ms, SE = 5.84 ms, z = 1.715, $p = 0.173$). For Stroop facilitation relative to neutral words, the RT differences were significant in both scripts (character: M = 13 ms, SE = 5.13 ms, z = 2.621, $p = 0.018$; pinyin: M = 16 ms, SE = 5.14 ms, z = 3.048, $p = 0.012$). There was no difference between character and pinyin stimuli in Stroop facilitation relative to neutral signs or neutral words (all *p*s > 0.4).

5.5.2.2 Error analysis

The model for error analysis is identical to the one used in mean RT analysis, except for using the Gaussian distribution:

lmer(ER ~ condition * script + colour + (1+condition|subject))

In both the Chinese character and pinyin conditions, error rates were similar for all Stroop interference and facilitation components.

5.5.2.3 Distributional analysis

5.5.2.3.1 <u>Stroop interference effects</u>

Quantile plots and delta plots based on untrimmed data from Experiment 8 for Stroop interference effects in Chinese characters and characters written in pinyin are presented in Figure 5.10 and 5.11.

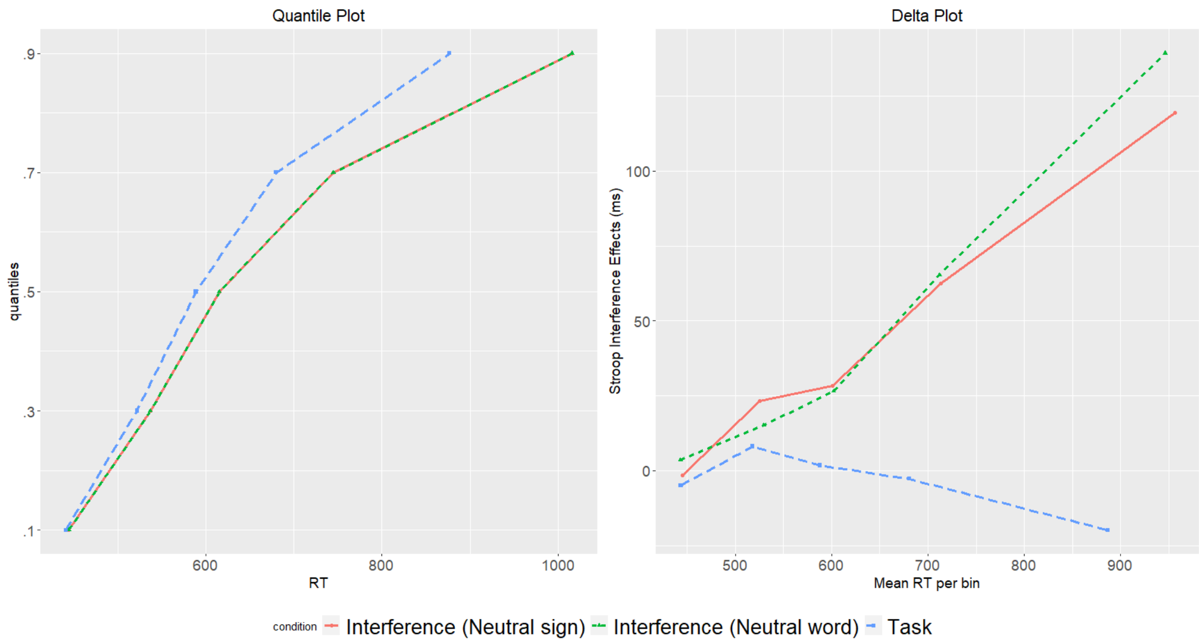


Figure 5.10. Quantile plots and delta plots for Stroop interference effects with Chinese character in Experiment 8.

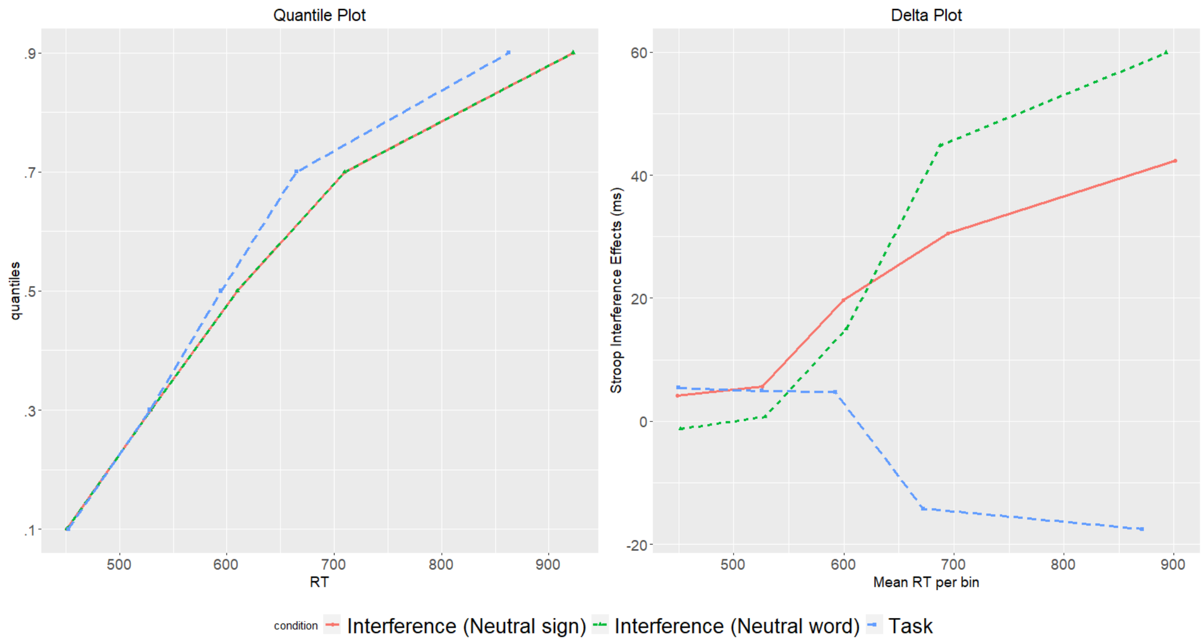


Figure 5.11. Quantile plots and delta plots for Stroop interference effects with pinyin in Experiment 8.

Analyses were separated for character and pinyin words. Then, a series of one-way ANOVAs were performed for each conflict component in character and pinyin words.

For Stroop interference effects in Chinese characters, the delta plots for Stroop interference relative to neutral signs and Stroop interference relative to neutral words showed positive linear trends (Stroop interference relative to neutral signs: $F(1,145) = 40.848$, $p < 0.001$, $d = 1.17$; Stroop interference relative to neutral words: $F(1,145) = 47.975$, $p < 0.001$, $d = 1.26$). A quadratic trend was found for Stroop interference relative to neutral words, and a trend for quadratic trend for Stroop interference relative to neutral signs (Stroop interference relative to neutral words: $F(1,145)= 7.638$, $p = 0.006$, $d = 0.50$; Stroop interference relative to neutral signs: $F(1,145) = 3.236$, $p = 0.074$, $d = 0.33$). The delta plot for task conflict showed no linear trend ($F(1,145) = 0.951$, $p = 0.331$, $d = -0.18$), nor quadratic trend ($F(1,145) = 1.397$, $p = 0.239$, $d = -0.22$).

For Stroop interference effects in Chinese characters written in pinyin, the delta plots for Stroop interference relative to neutral signs and Stroop interference relative to neutral words showed positive linear trends (Stroop interference relative to neutral signs: $F(1,145) = 7.477$, $p = 0.007$, $d = 0.50$; Stroop interference relative to neutral words: $F(1,145) = 17.001$, $p < 0.001$, $d = 0.75$) and no quadratic trends (Stroop interference relative to neutral signs: $F(1,145)= 0.157$, $p = 0.692$, $d = 0.07$; Stroop interference relative to neutral words: $F(1,145) = 0.764$, $p = 0.383$, $d = 0.16$). The delta plot for task conflict showed no linear trend ($F(1,145) = 2.15$, $p = 0.145$, $d = -0.27$), nor quadratic trend ($F(1,145) = 0.215$, $p = 0.644$, $d = -0.08$).

5.5.2.3.2 Stroop facilitation effects

Quantile plots and delta plots based on untrimmed data from Experiment 8 for Stroop facilitation effects in Chinese characters and characters written in pinyin are presented in Figure 5.12 and 5.13.

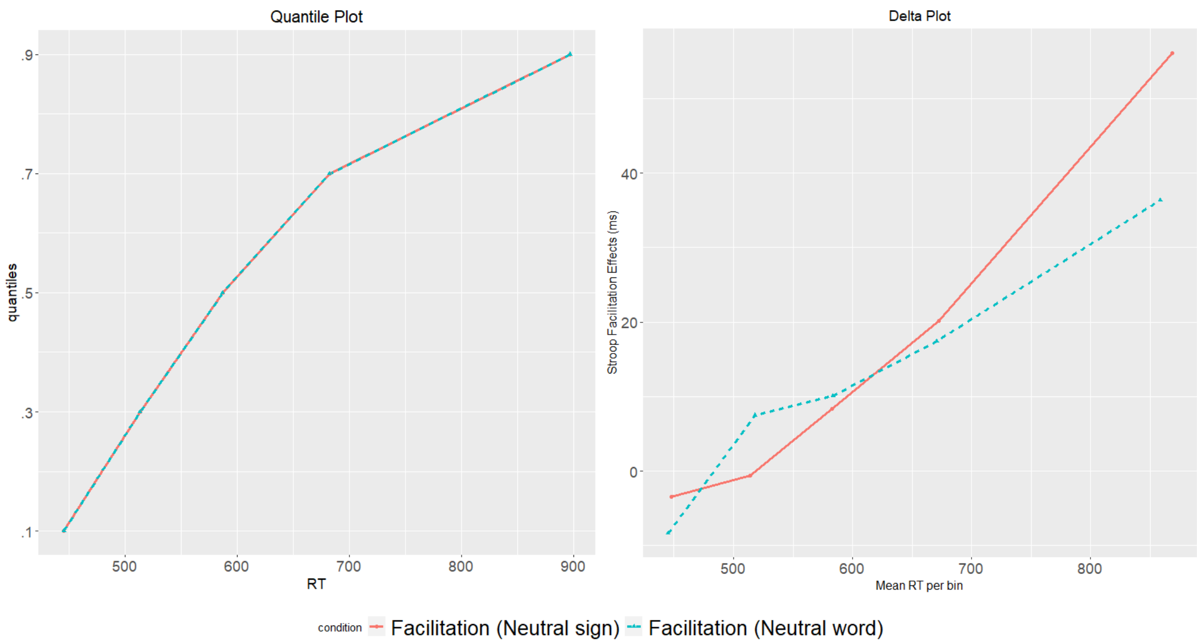


Figure 5.12. Quantile plots and delta plots for Stroop facilitation effects with Chinese character in Experiment 8.

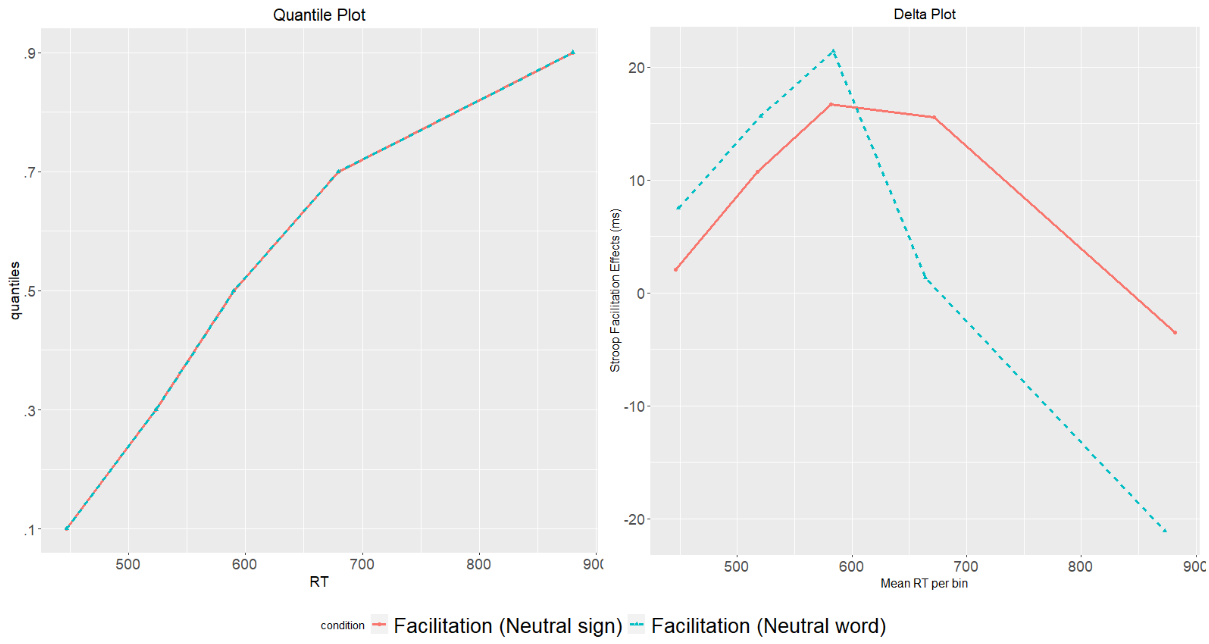


Figure 5.13. Quantile plots and delta plots for Stroop facilitation effects with pinyin in Experiment 8.

The analyses were separated for character and pinyin words. Then, a series of one-way ANOVAs were performed for each facilitation component in character and pinyin words.

For Stroop facilitation effects in Chinese characters, the delta plot for Stroop facilitation relative to neutral words and Stroop facilitation relative to neutral signs showed positive linear trends (Stroop facilitation relative to neutral words: $F(1,145) = 8.666$, $p = 0.004$, $d = 0.54$; Stroop facilitation relative to neutral signs: $F(1,145) = 19.278$, $p < 0.001$, $d = 0.80$) and a quadratic trend for Stroop facilitation relative to neutral signs but not for Stroop facilitation relative to neutral words (Stroop facilitation relative to neutral signs: $F(1,145) = 3.357$, $p = 0.069$, $d = 0.34$; Stroop facilitation relative to neutral words: $F(1,145) = 0.073$, $p = 0.787$, $d = 0.05$).

For Stroop facilitation effects in Chinese characters written in pinyin, the delta plot for Stroop facilitation relative to neutral words and Stroop facilitation relative to neutral signs showed no positive linear trends (Stroop facilitation relative to neutral words: $F(1,145) = 3.198$, $p = 0.076$, $d = -0.33$; Stroop facilitation relative to neutral signs: $F(1,145) = 0.03$, $p = 0.864$, $d = -0.03$) and a quadratic trend for Stroop facilitation relative to neutral words but not for Stroop facilitation relative to neutral signs (Stroop facilitation relative to neutral words: $F(1,145) = 3.386$, $p = 0.068$, $d = -0.34$; Stroop facilitation relative to neutral signs: $F(1,145) = 2.064$, $p = 0.153$, $d = -0.26$).

5.5.3 Discussion

In the previous experiments, the Stroop effects and their decomposed components in characters and pinyin were investigated separately. Experiment 8 involved a mixed presentation of two scripts: Chinese characters and pinyin. Theoretically, colour words in Chinese characters involve response, phonological, and semantic components (see Experiment 4 for details), whereas pinyin contains phonological and semantic components only. This is because pinyin does not point towards a specific character. (Pre-)motor responses should not be triggered by pinyin because it is similar to the colour-associated condition that has no corresponding response to ink colours in the Stroop task. Thus, no response conflict/facilitation is produced from colour words in pinyin. Indeed, based on the results of Experiment 4 (character, manual), Experiment 5, 6, and 7 (pinyin, manual), it is obvious that Chinese words written in pinyin resulted in less Stroop interference (relative to neutral signs) compared to characters. However, stronger Stroop facilitation can be observed in Chinese words written in pinyin. In Liu and Weng's (2007) study, they combined the presentation of Chinese characters, pinyin, and English in one design, and observed the Stroop interference differences between each script from significant to non-significant. This implies that the mixture of different scripts has an impact on the processing of the other scripts. Thus, the current study focused on how character and pinyin would affect each other's Stroop interference and facilitation effects.

The results revealed that in a mixed presentation of both scripts, strong Stroop interference relative to neutral signs and neutral words was found. The script difference between character and pinyin was also significant because characters resulted in more Stroop interference than pinyin. This pattern is consistent with studies that investigated

Stroop interference effects with Chinese characters (e.g. Experiment 4, Spinks et al., 2000; Wang et al., 2010) and Chinese words written in pinyin (e.g. Experiments 5, 6, and 7).

Stroop facilitation effects relative to neutral signs were significant for characters but not for pinyin words. However, Stroop facilitation effects relative to neutral words were found with both character and pinyin stimuli. Stroop facilitation effect relative to neutral words was found to be non-significant in Experiment 4 with Chinese characters; however, Experiments 5, 6, and 7 with Chinese words written in pinyin revealed strong Stroop facilitation effects. The mixture of characters and pinyin in one Stroop task facilitated the processing of congruent colour characters.

No task conflicts were found either with character or pinyin stimuli. Experiment 4 also showed that Chinese characters do not produce task conflicts. This is consistent with the findings in French using manual responses (Augustinova et al., 2019). However, pinyin stimuli produce significant task conflicts in Experiments 5, 6, and 7, which involved only pinyin stimuli, unlike Experiment 8. Thus, the mixed presentation of the two scripts eliminated the task conflicts of pinyin words.

The distributional analyses revealed that strong inhibition was applied to resolve the competition between colour and word information in Stroop interference effects with either character or pinyin stimuli. The pattern in Stroop facilitation was rather different: strong enhancement was found in Stroop facilitation effects (either relative to neutral words or neutral signs) with Chinese characters, whereas weak enhancement was found with Chinese characters written in pinyin. The results indicated that the mixed presentation of both Chinese characters and pinyin can lead to weaker enhancement applied to pinyin stimuli in Stroop facilitation, because in Experiments 5 to 7, there were always strong enhancement applied when using Chinese characters written in pinyin.

5.6 General Discussion

Stroop interference effects relative to neutral signs were found in all experiments reported in this chapter. Thus, effects were found when there was no tonal information provided in pinyin stimuli (Experiment 5), when colour-associated pinyin stimuli were added (Experiment 6), when invalid-radical pinyin stimuli were added (Experiment 7), and when Chinese characters and pinyin were mixed (Experiment 8). This effect is strong and often reported in manual Stroop tasks using alphabetic or non-alphabetic languages (e.g., Augustinova et al., 2019; Augustinova et al., 2018; Fennell & Ratcliff, 2019; Wang et al., 2010; Zahedi et al., 2019). Stroop interference effects relative to neutral words were relatively smaller than when compared to neutral signs. This is because neutral signs do not trigger word reading, whereas neutral words and colour words trigger word reading.

Stroop interference effects relative to neutral signs include Stroop interference relative to neutral words (neutral words vs. colour words) and task conflicts (neutral signs vs. neutral words). Stroop interference disappeared when there was no tonal information provided in pinyin stimuli (Experiment 5), when colour-associated pinyin stimuli were added (Experiment 6), and when invalid-radical pinyin stimuli were added (Experiment 7). In Experiment 5, pinyin without tonal information was responded to slower than neutral signs but not to neutral words. This means that pinyin without tonal information still triggers word reading; however, it is weak to activate related semantic and phonological information of a colour word and thus makes no difference to a neutral word.

In Experiment 6, incongruent colour words did not produce Stroop interference compared to neutral words, but Stroop interference was found when compared to neutral signs. Further examination of the data revealed that task conflicts were present and contributed to the Stroop interference effect. Incongruent colour words, however, were similar to the performance of neutral words (649 ms vs. 651 ms) in Experiment 6.

This could be attributed to the use of colour-associated words in this experiment, as incongruent colour-associated words were not different from incongruent colour words or neutral words (650 ms vs. 649 ms vs. 651 ms). Consistent with the conclusion of Experiment 6, in Experiment 7, incongruent colour words written in pinyin were not significantly slower than neutral words but were slower than neutral signs.

It should be acknowledged that the neutral words /qiong2/ used in Experiment 5, 6 and 8 contain a colour-associated meaning 'red jade'. Such potential semantic association should not greatly affect the findings, because the neutral word is presented in pinyin, which also points to other colour-unrelated meanings like 'poor' and 'high'. Furthermore, different sets of neutral words were used in Experiment 7 and the Stroop interference effects, Stroop facilitation effects, and task conflicts remained similar to the results of Experiment 5 and 6.

In Experiments 5 to 8, Stroop facilitation effects relative to neutral signs disappeared when using Chinese words written in pinyin. This effect was also not found with Chinese characters in a manual Stroop task (Experiment 4) and in other alphabetic languages (Logan & Zbrodoff, 1998; Nealis, 1973; Roelofs, 2012; Schulz, 1979; Sichel & Chandler, 1969; Vanayan, 1993). Significant Stroop facilitation effects relative to neutral words were found in all experiments reported in this chapter. This indicates that Chinese words in pinyin can facilitate the colour categorization process. In contrast, Stroop facilitation has not been found with Chinese characters in the manual Stroop task (see Experiment 4; Coderre et al., 2013). However, Stroop facilitation has been reported in alphabetic languages (Augustinova et al., 2019; Brown et al., 2002; Fennell & Ratcliff, 2019; Redding & Gerjets, 1977; Zahedi et al., 2019).

Because Chinese words written in pinyin elicited strong facilitation effects, whereas characters did not, this suggests that the processing of pinyin is different. Colour words written in pinyin can only lead to Stroop interference when semantic information is activated by phonology. However, the data in this chapter shows that phonology alone does not produce Stroop interference. This could be because pinyin refers only to the phonology of Chinese characters. Because there are many homophones in Chinese, it is not specific enough to activate the meaning of crucial Chinese words. In contrast, the semantic information does not need to be fully activated by pinyin stimuli to produce Stroop facilitation. Thus, the identical phonology of pinyin and colour characters would be sufficient to facilitate the process.

Overall, Stroop interference effects relative to neutral signs were found with Chinese words written in pinyin. This indicates that pinyin can activate semantics in the Stroop task just as with Chinese characters and in other alphabetic languages. Pinyin has a direct link to phonology; thus, the Stroop facilitation effects found in pinyin were stronger than with characters and were comparable to alphabetic languages.

The distributional analysis of Experiments 5 to 8 provided a useful way to compare the patterns with Chinese characters (Experiment 4). The Stroop interference effects with pinyin stimuli were numerically smaller than those with characters, but both had strong inhibition applied to resolve the competition between colour and word information. Strong task conflicts were observed with Chinese characters written in pinyin because no inhibition was applied, which is absent with Chinese characters.

The linearity of Stroop interference relative to neutral signs and neutral words showed consistent medium to large effects across the four experiments, whereas the linearity of Stroop facilitation relative to neutral signs showed small effects in pinyin stimuli and large effect in character stimuli. The linearity of Stroop facilitation relative to neutral words revealed medium to large effects, except for pinyin stimuli when mixed presented in Experiment 8. The trends in task conflicts were positively linear with medium and large effects in Experiments 5 to 7; however, the trend was no longer linear when mixing both character and pinyin stimuli in Experiment 8.

The following discussion would focus on the key features of each experiment. In Experiment 5, the impact of tonal information in pinyin was investigated. Diacritics provides tonal information to pinyin. This reduces the number of characters corresponding to the phonology activated by pinyin. Thus, pinyin with tonal information should result in more interference than pinyin without tonal information because the more specific the pinyin is, the stronger the association to colour words. The current results do not strongly claim that there is a difference between presenting with or without tonal information conditions, though there is a tendency for more Stroop interference relative to neutral words when presenting with tones compared to without tones.

In Experiment 6, colour-associated words written in pinyin were added to investigate the semantic and phonological components in Stroop effects. Experiment 4 showed that phonological components should be considered as one of the Stroop components when processing Chinese characters. Thus, response, phonological, and semantic components together contributed to the Stroop effects with Chinese characters. When investigating the impact of Chinese words written in pinyin, this paradigm used in Experiment 4 was adjusted for Experiments 5 to 7. Response components are absent in pinyin, because pinyin is reflecting the phonology that refers to many different Chinese words and therefore it does not directly link to a motor response because all possible words (homophones) are likely to be activated. Based on the results of Experiment 6, there is no evidence that phonological or semantic conflict was produced by pinyin words in the manual Stroop task.

The Stroop interference effects observed with pinyin can largely be attributed to task conflicts. For Stroop facilitation effects, it was found that semantic facilitation contributes to the Stroop facilitation effects. In Experiment 4, strong semantic conflict was observed in the manual responses, but no semantic facilitation was found when using Chinese characters. Experiment 6 used colour-associated words written in pinyin and the findings indicated the reversed pattern for semantic conflict and facilitation. This again showed that pinyin has different effects on Stroop interference and facilitation. When presented in the incongruent condition, pinyin is recognized as a word not a string of meaningless symbols. This characteristic only guarantees that incongruent words in pinyin are responded to slower than neutral signs but not neutral words. When presented in the congruent condition, colour-associated meaning makes a difference to neutral words. It helped congruent words in pinyin be responded to faster than neutral words as semantic information is relevant to colours.

In Experiment 7, the invalid-radical condition was used to examine the radical's semantic component and phonology-driven semantic components. Only task conflicts contributed to the Stroop interference effect. Characters with phonetic radicals cannot be activated when presented in pinyin.

In Experiment 8, Chinese characters and pinyin were used in one Stroop task: character and pinyin to investigate the impact of script. Task conflicts were strong in the manual Stroop task using pinyin words (Experiments 5, 6, and 7). Previous studies (Augustinova et al., 2019; Augustinova et al., 2018; Sharma & McKenna, 1998) have found task conflicts in the vocal Stroop task only. Experiment 4 revealed no task conflicts in either vocal or manual responses using Chinese characters. The mixed presentation of Chinese characters and pinyin no longer produced task conflicts for pinyin stimuli, which means that the character presented in the same list has an impact on pinyin processing.

Stroop facilitation effects were absent in the manual responses using Chinese characters (Experiment 4). However, with the presence of pinyin, Chinese characters resulted in strong Stroop facilitation effects. Thus, pinyin affected the impact of characters on Stroop facilitation.

5.7 Conclusion

The current chapter examined the impact of pinyin stimuli on the manual Stroop task. Strong Stroop interference and facilitation effects were found with Chinese words written in pinyin, which indicates that pinyin activate phonology and semantics, just like Chinese characters.

Further investigations of the components of Stroop effects revealed that implicit activation from colour-associated words in pinyin and invalid-radical words in pinyin was evident in Stroop facilitation but not in Stroop interference.

Chapter 6: General Discussion

This chapter summarizes and discusses the findings reported in this thesis. First, a summary of the findings of the Stroop experiments in terms of the activation of radicals is provided (Section 6.1). Next, the Stroop data obtained with Chinese characters (Section 6.2), Chinese characters written in pinyin (Section 6.3), and implications from the distributional analyses (Section 6.4) that interpret the mean RT results from a different perspective. In addition, the use of Chinese characters has also provided insights into the Stroop task (Section 6.5) that explains the Stroop effects from a non-alphabetic language perspective. Section 6.6 talks about the implications of this thesis on the theories and models of Chinese word recognition and the Stroop task. Section 6.7 discusses the limitations of this thesis and provides suggestions for future research.

6.1 Chinese word recognition and activation of radicals

Chapter 3 investigated the sublexical processing of Chinese characters in the Stroop task. The experiments in Chapter 3 focused on the semantic activation of phonetic radicals, phonological activation of phonetic radicals, and response modality effects in compound characters that contain phonetic radicals. Compared to Yeh et al.'s (2017) study, Chapter 3 included manual Stroop task which asks for no explicit verbal responses to investigate the response modality effect. The results showed that phonetic radicals can still be activated in a manual Stroop task.

6.1.1 Semantic activation of phonetic radicals

In Experiments 1 and 3 of Chapter 3, strong Stroop interference effects were observed with the Valid- and Invalid-Radical conditions providing the evidence for semantic activation of phonetic radicals in the vocal Stroop task. Because the Stroop task does not require explicit reading of words, the findings confirm the automaticity of the effects. In the vocal Stroop task, the relevant task is to name the ink colour of the words and not to overtly name the words. If semantic information provided by phonetic radicals could be ignored, then there should be no Stroop interference effects with the Valid- or Invalid-

Radical conditions. However, the strong Stroop interference effects suggest that semantic activation of phonetic radicals is automatic.

In the manual Stroop task (Experiment 2), only the Invalid-Radical condition showed strong Stroop interference effects and not the Valid-Radical condition. A possible explanation for this is that the irregular phonetic radicals (the Invalid-Radical condition) result in more Stroop interference than the regular phonetic radicals (the Valid-Radical condition). Thus, for the Valid-Radical condition, the phonological information provided by phonetic radicals may not interfere with automatic reading. In contrast, when the pronunciation of phonetic radical does not match with the pronunciation of the whole character, it causes more Stroop interference. The semantic activation of phonetic radicals in the manual responses was also automatic but restricted to conditions in which phonetic radicals do not have the same pronunciation as the whole characters.

The above conclusion in terms of semantic activation of phonetic radicals is consistent with literature that used a semantic relatedness judgement task and the Stroop task (Lee et al., 2006; Yeh et al., 2017; Zhou & Marslen-Wilson, 1999b), indicating that semantic activation of phonetic radicals can be found in tasks that require explicit reading of the words as well as in tasks in which the focus is not reading the words (the Stroop task).

The current results support the decomposition views of Chinese word reading that assume radicals are processed first (Chen & Yeh, 2015; Feldman & Siok, 1999; Saito et al., 1998; Taft & Zhu, 1997; Taft et al., 1999; Yeh & Li, 2004). Although phonetic radicals can have no phonological or semantic cues to the whole character, they can still trigger strong Stroop interference effects. The results also support the lexical constituency model proposed by Perfetti et al. (2005) that assumes the basic level of word recognition in Chinese should be the radicals.

6.1.2 Activation of phonology by phonetic radicals

The current study found evidence for the activation of phonology by phonetic radicals. There was a tendency that the Valid-Radical condition had a stronger Stroop interference effect than the Invalid-Radicals in Experiment 1 (vocal responses). When the number of trials doubled in Experiment 3 (vocal responses), the difference became significant in both Stroop interference and facilitation effects. In the manual responses (Experiment 2), strong Stroop interference effects were observed in the Invalid- but not in the Valid-Radical conditions, which is opposite to the tendency shown with vocal responses. Such a contrast in two response modalities suggests that with vocal responses, a regular phonetic radical (the Valid-Radical condition) can elicit more Stroop interference effects than an irregular one (the Invalid-Radical condition). When verbal responses were not required as in the manual Stroop task, a reversed pattern was found, indicating the irregularity of phonetic radical causes interference; thus, phonetic radicals activate phonological cues in the manual Stroop task. A possible explanation of these findings is that the Valid-Radical condition contains phonological cues that are the same in terms of both the character and the radical level. In contrast, in the Invalid-Radical condition, the phonological cue of the character differs from that of the radical level. When verbal responses are required, two identical phonological cues in the Valid-Radical condition enhance phonological activation together, whereas only phonological cues from the radical level in the Invalid-Radical condition interferes with word recognition. When verbal responses are not required, the identical phonological cues provided by the Valid-Radical condition do not enhance phonological activation. Instead, the two different phonological cues provided by the Invalid-Radical condition cause difficulty in word reading and therefore result in Stroop interference effects.

The current results are not consistent with what Yeh et al. (2017) observed with vocal responses. Phonological cues provided by phonetic radicals were not activated in the data they collected. However, in tasks that require the target character to be named, phonetic radicals have been found to activate phonology, although evidence for this has been obtained only with low-frequency characters (Hue, 1992; Lee et al., 2005; Seidenberg, 1985; Zhou & Marslen-Wilson, 1999a). This can be explained by the fact that the regularity of phonetic radicals decreases with higher word frequencies (Shu & Anderson, 1999).

6.1.3 Response modality effects

The experiments reported in Chapter 3 are the first experiments that investigated response modality effects in the Stroop task using radical-related conditions in Chinese. Compared to Yeh et al.'s (2017) results with the vocal Stroop task, both Yeh et al.'s (2017) results and Experiment 1 found significant Stroop interference and facilitation effects in the Colour-Character and Invalid-Radical conditions. The only difference was that Yeh et al. (2017) observed a tendency for the Valid-Radical condition in Stroop facilitation (p = 0.06), whereas in Experiment 1 and Experiment 3, a strong Stroop facilitation effect was found in the Valid-Radical condition (ps < 0.001). This suggests that the Valid-Radical condition can facilitate the naming of colour words as well as the Invalid-Radical condition, though no difference was found between those two radical-related conditions. However, when the manual Stroop task was used, a different pattern was found. The Valid-Radical condition did not show Stroop interference effects. This finding has been discussed in Section 6.1.1 and 6.1.2. Another difference in terms of response modality was that both the Invalid- and Valid-Radical conditions did not show Stroop facilitation effects. This means that phonetic radical's semantic and phonological cues were not fully activated or not strongly enough to facilitate the processing of colour words as the Colour-Character condition does.

In sum, response modality did not affect the performance of the Colour-Character condition; however, the performance of the Invalid- and Valid-Radical conditions was affected by different response modalities.

6.2 Chinese character recognition

Experiment 4 investigated conflict/facilitation components in the Stroop interference and facilitation effects using Chinese characters. While it is difficult to find homophones in alphabetic languages, there are abundant homophones in Chinese characters, which is ideal to investigate the phonological components and crucially, there is no orthographic overlap (Parris et al., 2022). The Stroop paradigm used by Augustinova et al. (2018; 2019) was extended to investigate phonological conflict/facilitation components. This section discusses response, phonological, semantic conflict/facilitation, and task conflicts as well as how these conflict/facilitation components are modulated by response modality.

6.2.1 Response conflict/facilitation

Response conflict refers to the response delay brought about by (pre-)motor responses (Augustinova & Ferrand, 2014). In the context of the Stroop task, an incongruent colour word could trigger response conflict because the colour words are in the response set required for naming in the case of a vocal Stroop and in the case of a manual Stroop task, which are part of the categories associated with specific keys.

Schmidt and Cheesman (2005) found no response conflict in incongruent colour-associated words (e.g. SKY presented in red colour) because colour-associated words would not trigger (pre-)motor responses. Augustinova and Ferrand (2014) defined the RT differences between the incongruent colour words and incongruent colour-associated words as response conflict.

The present thesis proposes a different way to measure response conflict when Chinese characters are used in a Stroop task. Response conflict in this case is defined as the difference between incongruent colour words and incongruent homophones of colour words. Thus, homophones of colour words rather than colour-associated words should be the reference for measuring response conflict because the orthography and phonology can be separated in Chinese, whereas those two aspects are often confounded in alphabetic languages. Therefore, the influence of phonology should be taken into consideration when defining the decomposed components of Stroop effects. An illustration of this subtractive logic is presented in Figure 6.1. The change in the definition of response conflict does not mean that incongruent colour-associated Chinese words activate (pre-)motor responses. Rather, the RT difference between the incongruent colour words and incongruent colour-associated words contains two distinct components: response conflict and phonological conflict.

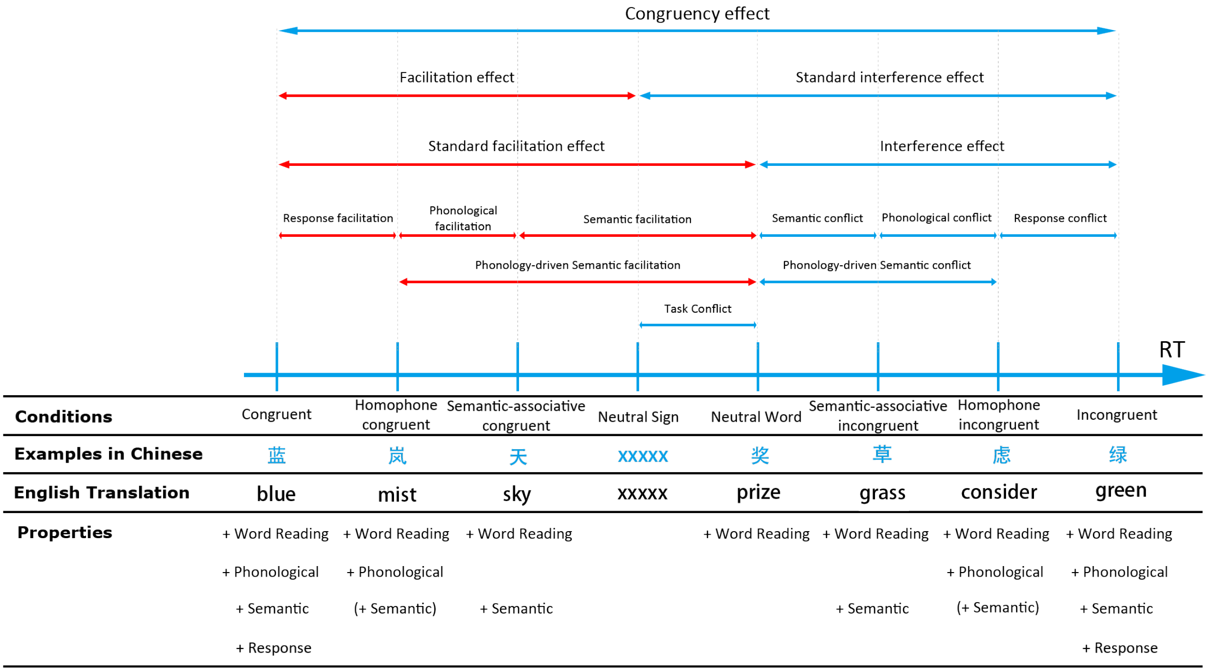


Figure 6.1. Decomposed Stroop conflict/facilitation components in Chinese characters.

Stroop data from Spinks et al. (2000) and Wang et al. (2010) were used to look at response and phonological conflicts using the new definitions (see Section 4.1 for details). Although the statistical analyses could not be conducted, there was numerically a response conflict of 29 ms in the vocal Stroop task (based on Spinks et al.'s data) and 7 ms in the manual Stroop task (based on Wang et al.'s data). Experiment 4 of this thesis revealed strong response conflicts in both response modalities (vocal: 39 ms, manual: 27 ms). Experiment 4 of this thesis is the first experiment as far as I am aware that treated response modality as a within-subject factor when investigating the decomposed Stroop components with Chinese characters. Although the calculation of response conflict is different from what has been suggested by Augustinova and colleagues, the results would be identical if using the original definition of response conflict (incongruent colour-associated words vs. incongruent colour words).

Response facilitation was calculated using a similar logic. Thus, the RT difference between congruent homophones and colour-associated words in Chinese was used to measure response facilitation. In Augustinova et al.'s (2019) study with French stimuli, they found strong response facilitation in the vocal but not in the manual responses. With Chinese stimuli, a 36 ms response facilitation was observed in the vocal responses (based on Spinks et al.'s study), but not in the manual responses (based on Wang et al.'s study). However, no response facilitation in either response modality was found in Experiment 4 of this thesis. It is difficult to explain why Experiment 4 did not reveal response facilitation in the vocal responses, in contrast to Augustinova et al. (2019). Further research is needed to investigate whether response facilitation occurs in the vocal responses.

6.2.2 Phonological conflict/facilitation

In this thesis, I proposed that the RT difference between incongruent colour-associated words and incongruent colour words consists of response conflict and phonological conflict. The reason why there is a phonological conflict, which is missing from Augustinova et al.'s (2014) paradigm, is that there are abundant homophones in Chinese characters which would greatly influence how words are processed and therefore phonological components should be taken into consideration. As illustrated in Figure 6.1, phonological conflict is due to phonological cues provided by homophones condition, which are not provided by words in the colour-associated conditions. This definition of phonological conflict was used in Experiment 4. The data from Experiment 4 revealed stronger phonological conflicts in vocal Stroop than in the manual Stroop. This effect of response modality is to be expected when verbal responses are required. The phonological cues were activated in the vocal Stroop, and this led to interference; when responses in the Stroop task were manual, phonological cues were not activated or not fully activated.

Besner and Stolz (1998) reported phonological conflicts in a manual Stroop using pseudohomophones of colour words in English. Pseudohomophones in English have some orthographic overlap with the base words (e.g. bloo vs. blue) (Parris et al., 2022), whereas in Chinese, such orthographic overlap can be avoided completely. Therefore, what Besner and Stolz measured as phonological conflicts may involve both phonological and orthographic conflicts. With Chinese homophones, it is possible to measure purely phonological conflicts.

When phonological conflicts were calculated using the data from Spinks et al. (2000) and Wang et al. (2010), a numerically reversed phonological conflict was observed. Thus, incongruent colour-associated words were processed slower than incongruent homophones. However, whether the numerical difference was significant could not be established because the raw data were not available.

Phonological facilitation was defined in this thesis as the RT difference between congruent homophones and congruent colour-associated words. Strong phonological facilitation for vocal but not for manual responses was observed in Experiment 4. This is consistent with the vocal Stroop task data of Spinks et al.'s (2000), but not with Wang et al.'s (2010) manual Stroop task because they reported numerically a 36 ms phonological facilitation. Parris et al. (2019) found phonological facilitation in both vocal and manual responses. Phonological facilitation was measured by phonological overlap between the irrelevant word and the colour word in English. It was found that the overlap at the initial letter would facilitate the processing speed compared to control words with no overlap in phonology.

6.2.3 Semantic conflict/facilitation

The measure of semantic conflict/facilitation with Chinese stimuli in the Stroop task used in this thesis is consistent with the Stroop literature using alphabetic languages (Augustinova & Ferrand, 2014; Augustinova et al., 2019; Augustinova et al., 2018; Sharma & McKenna, 1998). This is because the only difference between those two conditions is the colour-related information provided by colour-associated words. In Experiment 4, both vocal and manual responses observed strong semantic conflicts using colour-associated words in Chinese, which is consistent with what has been found in Augustinova et al.'s study (2018, 2019). The data of Spinks et al.'s (2000) and Wang et al.'s (2010) showed a numerical semantic conflict of 30 ms (vocal Stroop) and a 36 ms (manual Stroop). However, Sharma and McKenna (1998) did not find semantic conflicts with manual responses but only with vocal responses. Therefore, they argued that the locus of the lexical effect is the vocal output system. The current results suggest that this lexical effect is not exclusive to vocal responses. Semantic information can also be activated automatically in the manual Stroop task.

Augustinova et al. (2019) investigated semantic facilitation using French words in the Stroop task. The RT difference between neutral words and congruent colour-associated words was used as a measure of semantic facilitation. They found semantic facilitation in both vocal and manual responses. In contrast, Experiment 4 did not find semantic facilitation in either response modality. Instead, a reversed semantic facilitation was found in the vocal responses; that is, congruent colour-associated words were processed slower than neutral words. Whether this reversed semantic facilitation is specific to Chinese is unclear. Further research is needed to investigate Stroop facilitation components with Chinese characters.

In sum, the presence of semantic conflict is universal in the Stroop task regardless of which language is used, because semantic effects are not language-specific. Moreover, semantic conflicts can be observed in both vocal and manual responses.

6.2.4 Task conflicts

The definition of task conflicts in the current thesis was based on Augustinova et al.'s (2018) who argued that task conflicts are due to attention drawn to the irrelevant task (i.e. word reading) rather than the relevant task (i.e. colour naming). More specifically, in a Stroop task, task conflicts are defined as the RT difference between neutral words and neutral signs (e.g. a string of Xs) because the former involves word reading, the latter not. Early literature also referred to this difference as the "lexical effect" (Levin & Tzelgov, 2016) or "lexicality cost" (Brown, 2011).

In Experiment 4 of this thesis, no task conflict was observed with vocal and manual responses. Thus, responses in the Stroop task for neutral words in Chinese characters did not differ from neutral signs (%). Task conflicts were also absent in the vocal responses with Hebrew speakers (Levin & Tzelgov, 2016) and in the manual responses with Russian speakers (Goldfarb & Henik, 2007). However, Augustinova et al. (2018) and Augustinova et al. (2019) found strong task conflicts in French speakers with vocal responses but not with manual responses. Goldfarb and Henik (2007) observed task conflicts with Hebrew speakers in the manual Stroop task. The data of Spinks et al.'s (2000) showed numerically a 33 ms task conflict, although task conflicts here were measured by the RT difference between neutral words and colour patches, which difference is defined as "orthographic conflict" by Levin and Tzelgov (2016).

Task conflicts were found in the vocal Stroop task with alphabetic languages (i.e. French) and in the manual Stroop task with non-alphabetic languages (i.e. Hebrew). However, no task conflicts were found with Chinese characters. Whether task conflicts are dependent on the type of response modalities and the type of writing systems requires more evidence from different languages. Further discussion over task conflicts will be carried out in Section 6.3.3 when talking about task conflicts in pinyin words.

6.2.5 Response modality effects

Experiment 4 of this thesis revealed strong Stroop interference effects in both vocal and manual Stroop tasks. Vocal responses were found to be slower than manual responses, which is consistent with the literature (Augustinova et al., 2019; Fennell & Ratcliff, 2019; Neill, 1977; Redding & Gerjets, 1977; Sharma & McKenna, 1998; Zahedi et al., 2019). Stroop facilitation effects were observed only in the vocal Stroop task but not in the manual Stroop task. In contrast, Augustinova et al. (2019) found strong Stroop facilitation effects in both response modalities.

The response modality effects observed in this thesis indicated that vocal and manual responses are qualitatively different, consistent with Kinoshita et al. (2017) who suggested that this qualitative difference is due to task differences. The vocal Stroop task requires colour naming, whereas the manual Stroop task requires colour classification. It is also argued in this thesis that language impacts the locus of Stroop effects. Different response modalities can also impact the locus of Stroop effects. Augustinova et al.'s (2019) study involved French stimuli and the authors argued that Stroop interference effects were due to reduced response and task conflicts in the manual responses and Stroop facilitation effects were attributed to the reduction in response facilitation. Data from Experiment 4 suggests that with Chinese characters, Stroop interference effects are due to reduced phonological and semantic conflicts in the manual responses, whereas Stroop facilitation effects are solely due to the reduction in phonological facilitation.

Roelofs's (2003) WEAVER++ model assumes that there is no qualitative difference between vocal and manual responses. Parris et al. (2019) found no qualitative difference in facilitation effects of phoneme overlap. Parris et al. (2022) argued response conflict might be indirectly measured in semantic conflict; thus, it is hard to claim a qualitative difference between the two response modalities. The current results, however, argued that vocal and manual responses are qualitatively different from each other, because there is no response, semantic, and phonological facilitation in the manual responses, but strong phonological facilitation in the vocal responses only. Even if there is an indirect measure of response conflict/facilitation in semantic conflict/facilitation or phonological conflict/facilitation, it is clear that no facilitation was found in the manual responses, and vocal responses are qualitatively different from it because of the presence of strong phonological facilitation.

Sharma and McKenna (1998) argued that manual responses do not have "privileged" access to the lexical system. Thus, no semantic conflicts were obtained in the manual responses. The results of Experiment 4 revealed that both vocal and manual responses can produce strong semantic conflicts, and more importantly, they are qualitatively different from each other.

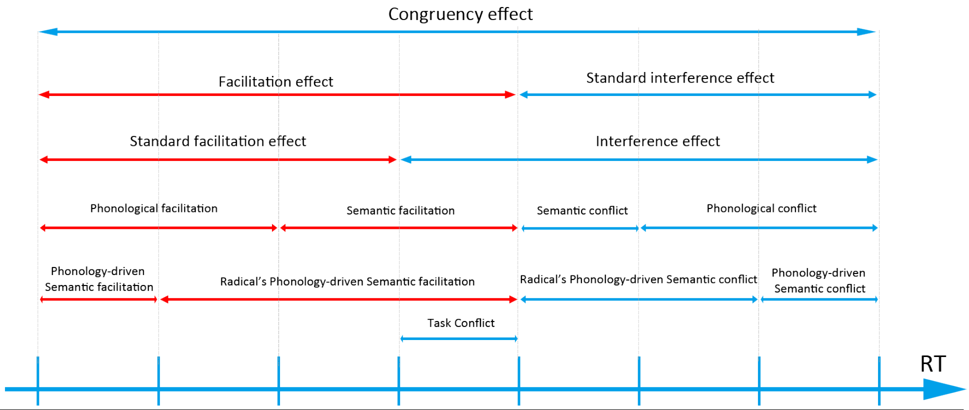
To summarize, Stroop interference effects with Chinese characters were affected by response, phonological and semantic conflicts in vocal responses, whereas phonological conflicts were absent in manual responses because verbal responses were not required. Stroop facilitation effects observed with Chinese characters were solely attributed to phonological facilitation in the vocal Stroop task.

6.3 Stroop effects with pinyin stimuli

In Experiments 5 to 8 of this thesis, the impact of pinyin stimuli on Stroop effects was investigated. Phonological and semantic conflict/facilitation, as well as task conflict with pinyin stimuli were discussed, as well as semantic conflict/facilitation from the radical level of Chinese characters written in pinyin.

6.3.1 Phonological conflict/facilitation

Pinyin is the Romanization transcription of Chinese characters, referring to the pronunciation of a Chinese character (see also Section 1.1). The phonology activated by pinyin does not refer to a single Chinese character due to the huge number of homophones in Chinese. Because pinyin refers to the pronunciation of a Chinese word, words in the homophone condition and colour word condition are written identical in pinyin. In this thesis, it is argued that colour words written in pinyin (homophones) do not activate response conflict because they do not refer to a single Chinese word but to all Chinese words with the same phonology. For example, lǜ can refer to the colour word “green”, but it can also refer to the word “consider” and “law”. Therefore, phonological conflict with pinyin stimuli is calculated in terms of the RT difference between incongruent homophones and colour-associated words written in pinyin (see Figure 6.2).



Conditions	Homophone congruent	Invalid-Radical congruent	Semantic-associative congruent	Neutral Sign	Neutral Word	Semantic-associative incongruent	Invalid-Radical incongruent	Homophone incongruent
Examples in Pinyin	lán	cāi	tiān	xxxxx	jiǎng	cǎo	cāi	lǜ
Chinese Character	蓝	猜	天	xxxxx	奖	草	猜	绿
English Translation	blue	guess	sky	xxxxx	prize	grass	guess	green
Properties	+ Word Reading + Phonological (+ Semantic)	+ Word Reading + Radical’s Phonological (+ Radical’s Semantic)	+ Word Reading + Semantic		+ Word Reading	+ Word Reading + Semantic	+ Word Reading + Radical’s Phonological (+ Radical’s Semantic)	+ Word Reading + Phonological (+ Semantic)

Figure 6.2. Decomposed Stroop conflict/facilitation components in pinyin.

In Experiment 6 of this thesis, phonological conflicts and phonological facilitation were not observed when the words were written in pinyin. Thus, homophones written in pinyin do not differ in response time from colour-associated words written in pinyin. Phonological information provided by colour words in pinyin does not have privileged access to colour characters compared to semantic information provided by colour-associated words in pinyin. Potentially, this could be attributed to the manual responses used in Experiments 5 to 8. In Experiment 4, when homophones and colour-associated words were presented using Chinese characters, phonological conflicts and facilitation only occurred in the vocal responses but not in the manual responses. Further research is needed to investigate whether phonological information can be activated when presenting Chinese words in pinyin in the vocal Stroop task.

Interestingly, in alphabetic languages like English, phonological conflicts have been found using pseudohomophones in the manual Stroop task (Besner & Stolz, 1998). Phonological facilitation was also found in terms of using phonetic overlap with the initial phoneme that corresponds to the colour words in both vocal and manual Stroop tasks (Parris et al., 2019).

6.3.2 Semantic conflict/facilitation

The measure of semantic conflict/facilitation with Chinese words written in pinyin is the same as when the words are written using Chinese characters (see Figure 6.2). In Experiment 6, semantic conflict was not observed when Chinese words were presented in pinyin; however, strong semantic facilitation was found. This is consistent with what Sharma and McKenna (1998) found with English stimuli. Augustinova et al. (2019), on the other hand, found strong semantic conflict and semantic facilitation in both response modalities using French stimuli. Experiment 4 involved Chinese characters and the results revealed strong semantic conflicts in both response modalities but no semantic facilitation in either response modality. In sum, the data of the Stroop experiments with Chinese words written in pinyin in terms of semantic conflict/facilitation is close to what was found with alphabetic languages.

6.3.3 Task conflicts

In Section 6.2.4, task conflicts were discussed in terms of the difference between alphabetic and non-alphabetic languages. Task conflicts tend to occur in the vocal Stroop task with alphabetic languages (i.e. French) and in the manual Stroop task with non-alphabetic languages (i.e. Hebrew). However, Experiment 4 which involved Chinese characters revealed no task conflicts in both response modalities. As discussed in Section 4.4, Experiment 4 used the percent sign “%” as a neutral sign, in contrast to a row of Xs in Augustinova et al.'s (2019) study. Comparing the results of Experiments 5 to 8, a different explanation can be put forward (see Figure 6.3 for task conflicts in Experiments 4 to 8).

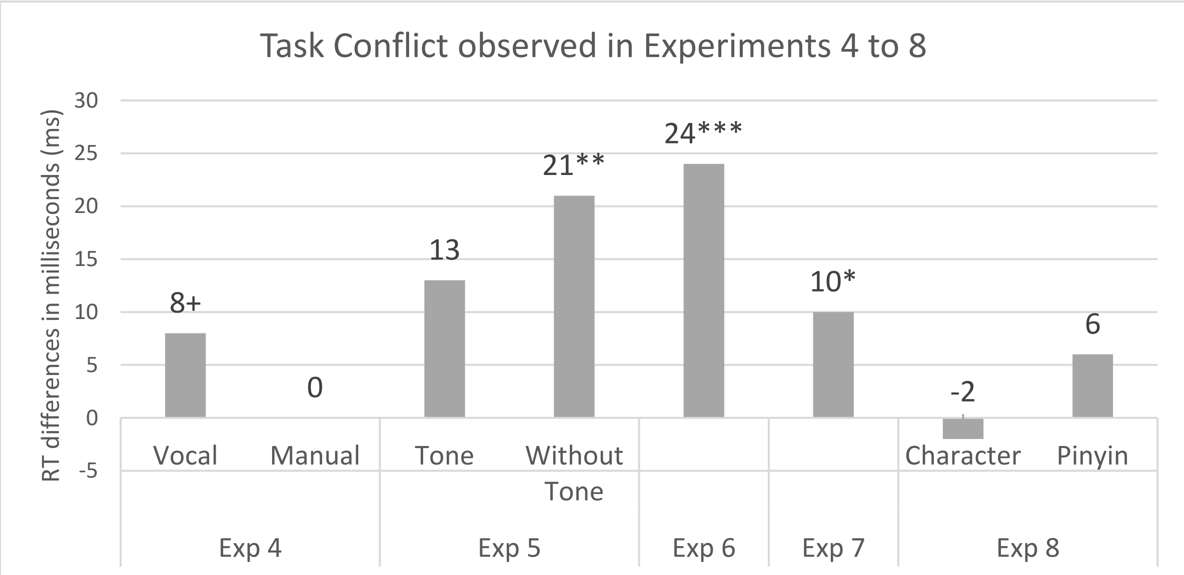


*Figure 6.3. Task conflicts observed in Experiments 4 to 8. ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < ⁺p < 0.1.*

As can be seen in Figure 6.3, there are no task conflicts when comparing neutral signs (i.e. "%" sign) with Chinese characters (Experiment 4). However, there are strong task conflicts observed when comparing to neutral signs (i.e. "xxxxx") in the experiments with Chinese words written in pinyin (Experiments 5 to 7). Most importantly, using a row of Xs as neutral signs does not guarantee the presence of task conflicts in the manual responses for alphabetic languages (Augustinova et al., 2019; Augustinova et al., 2018; Sharma & McKenna, 1998). This means that neutral words written in pinyin trigger word reading in the manual Stroop task and thus, responses are slower than neutral signs that activate no word reading, whereas neutral words in Chinese characters, English, and French do not elicit strong interference from word reading compared to neutral signs. Moreover, task conflicts in pinyin were found to be greatly affected by the presence of Chinese characters. In Experiment 8, the mixed presentation of both character and pinyin stimuli resulted in no task conflicts. When presenting Chinese words written in pinyin, strong task conflicts were found; when presenting the words using Chinese characters, no task conflicts were observed. The magnitude of task conflicts in Experiment 8 (character: -2 ms, pinyin: 6 ms) also suggests that pinyin processing was affected by the presence of Chinese characters. This could be due to the mixed presentation of different scripts that affected the visual recognition process. A constant shift between the colour naming task and word reading task in different scripts would make participants less sensitive to different types of stimuli; and thus, smaller task conflicts observed (Stroop interference and facilitation effects were also affected by this).

6.3.4 Radical's semantic conflict/facilitation

Experiments 1 to 3 revealed that phonetic radicals can activate semantic information. This finding is consistent with other studies using the Stroop task and studies using other cognitive tasks (Lee et al., 2006; Yeh et al., 2017; Zhou & Marslen-Wilson, 1999b). In Experiment 7, the Invalid-Radical condition was included again in a Stroop task but in this experiment the Chinese words were written in pinyin. This means that its original orthographic link to the colour character that contains it as a phonetic radical is lost. The only way the effect could appear is when the phonology activates the Chinese logographic script (Chinese character), which then subsequently activates the radicals that it contains, and these then need to activate their semantics (see Figure 6.4). The results of Experiment 7 revealed that the Invalid-Radical condition cannot activate phonology-driven semantic conflict/facilitation and radical's phonological-driven semantic conflict/facilitation. Such indirect activation of phonetic radicals in pinyin was not supported by the current study.

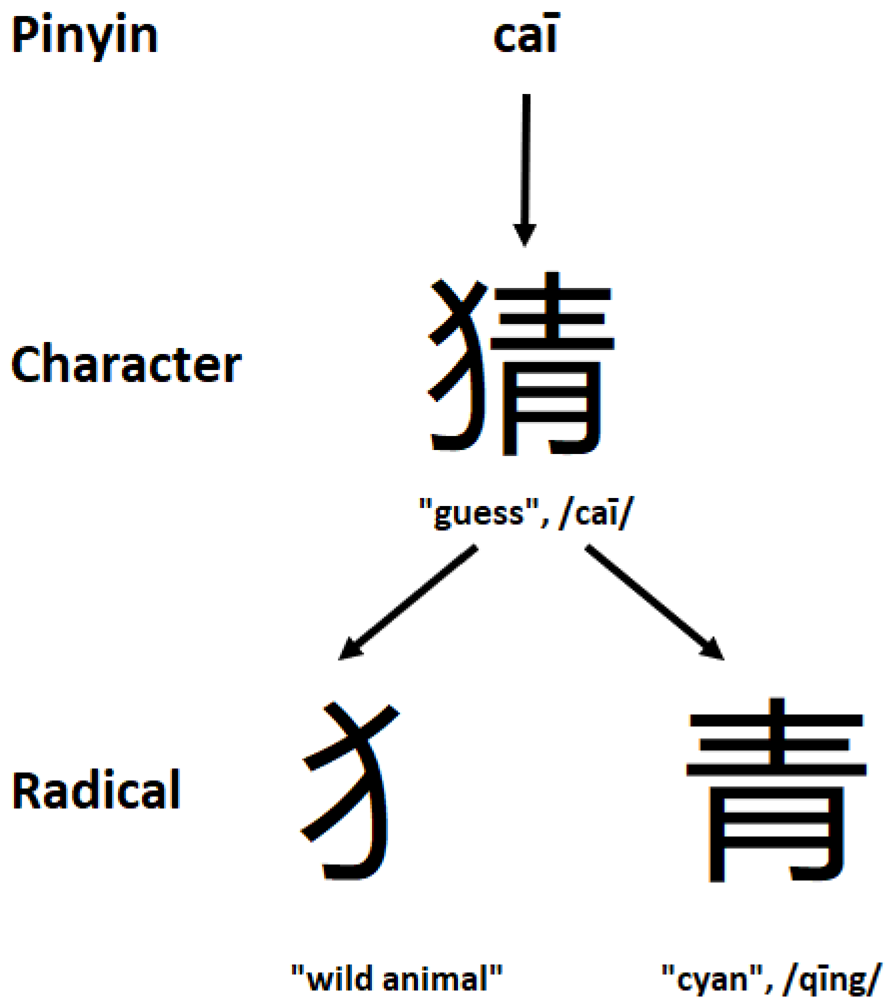


Figure 6.4. An example of the Invalid-Radical condition character written in pinyin.

6.4 Implications from the distributional analyses

In the mean RT analysis, the effects presented were averaged by subject and by condition; therefore, it was not possible to see how the effects change depending on the speed of the responses. By plotting the RTs into different bins, a cumulative distribution enables investigation of how the conditions change across the whole RT range. By subtracting one condition from the other, a delta plot can be obtained that illustrates the changes of RT differences between those two conditions (Ridderinkhof et al., 2005; Roelofs et al., 2011). An upward trend in delta plots indicated the effects grow larger with the increase of RTs, which suggested no inhibition was applied to resolve certain conflicts. This pattern was often reported in the Stroop task (Pratte et al., 2010; Roelofs et al., 2011). A flat pattern in delta plots means the effects remained constant across the quantiles, indicating that weak inhibition was applied. A downward trend indicates strong inhibition, which is often seen in the Simon effect (Burle et al., 2005).

Most of the Stroop interference effects reported in this thesis show upward trends, indicating that no inhibition was applied to resolve the conflict between colour and word information.

The significant effects observed in the mean RT analysis could be due to different levels of suppression over time. For example, in Experiment 4, the Stroop facilitation effect was significant; however, the trend analysis showed that there was no linear trend. This means weak enhancement was applied to converge the colour and word information. In contrast, Stroop facilitation effects found using Chinese characters written in pinyin showed linear trends, which suggested strong enhancement applied. Such information cannot be revealed by a mean RT analysis. Thus, the current thesis argues for using a combination of mean RT and distributional analyses when investigating response time-related studies because the significant effects observed in the mean RT analysis could be due to different distribution patterns, namely upward, downward, flat, or even quadratic. Such pattern may not be discernible when averaging the RTs.

6.5 Implications for theories of Stroop effects

6.5.1 Theories on the Stroop effects

Section 2.2 of this thesis introduced different theories of Stroop effects. The relative speed of processing theory assumes that the processing speed of colour and word information is different (Cattell, 1886; Fraisse, 1969; Klein, 1964; Morton & Chambers, 1973). Both potential responses compete before the actual response is produced. However, this theory cannot be used to explain the results of SOAs (stimulus onset asynchrony), where the presentation of colour and word information is separated. Glaser and Glaser (1982) found strong Stroop interference effects in the colour naming task when colour information is presented before the word information. In contrast, no Stroop interference was observed in the word reading task when word information is presented before the colour information. According to the relative speed of processing theory, there should be strong Stroop interference effects in the word reading task when word information is presented first.

Another theory of Stroop effects is the automaticity view that word reading is more automatic and requires less attention than colour naming (Logan, 1978; Shiffrin & Schneider, 1977). The automaticity view cannot explain why there are reduced Stroop interference effects when shifting the presentation of colour and word information from integrated to separated. Comparing the results of Experiment 1 (integrated mode) and Experiment 3 (separated mode), the Stroop interference effects were reduced almost by half, even though both experiments used the same stimuli and the same procedure.

What is missing in these two theories is a more detailed description of the strength of that processing. The Parallel Distributed Processing (PDP) model proposed by Cohen et al. (1990) was developed to provide a computational model simulating the Stroop effects. In this model, the processing of different information depends on the strength of connections. When two pathways have a similar strength but different activation, the Stroop interference is produced; when two pathways share the same activation, the Stroop facilitation is produced. Attention in this model has no privileged status. It modulates how information is processed in a pathway. This explains why different levels of suppression are observed in one study with different types of stimuli. For example, the Colour-Character and Valid-Radical conditions did not show evidence of the application of inhibition to resolve the conflict between colour and word information, whereas weak inhibition was applied in the Invalid-Radical condition. If attentional control has a privileged status in monitoring the processing of words, then this effect should be constant throughout the experiment. However, as has been observed in the distributional analysis, the execution of attentional control can vary for each individual stimulus, which depends on the strength of connections that activate the information contained in each stimulus.

6.5.2 Stroop facilitation effects

In Section 2.1, the Stroop facilitation effects were discussed in terms of the selection of neutral conditions. This could be one of the reasons why the Stroop facilitation effects are less stable and often referred to as a "fragile" effect (Logan & Zbrodoff, 1998; Macleod, 1991; MacLeod & MacDonald, 2000a). Studies have reported Stroop facilitation effects when using a row of Xs as the baseline (Dyer, 1973b; Regan, 1978) but there were cases when negative facilitation effects happened (i.e. Xs were responded to faster than congruent stimuli) (Logan & Zbrodoff, 1998; Nealis, 1973; Schulz, 1979; Sichel & Chandler, 1969; Vanayan, 1993). However, many studies reported stable Stroop facilitation effects when neutral words were the baseline (Augustinova & Ferrand, 2012; Brown, 2011; Dalrymple-Alford, 1972; Duncanjohnson & Kopell, 1980; Redding & Gerjets, 1977; Spinks et al., 2000; Yeh et al., 2017). The reason why neutral words are better markers for stable Stroop facilitation effects than neutral signs were that there exists task conflict between neutral words and neutral signs. Neutral words involve word reading, whereas neutral signs do not involve word reading. This point can now be supported by the results of Experiments 4 to 8, as shown in Figure 6.5.

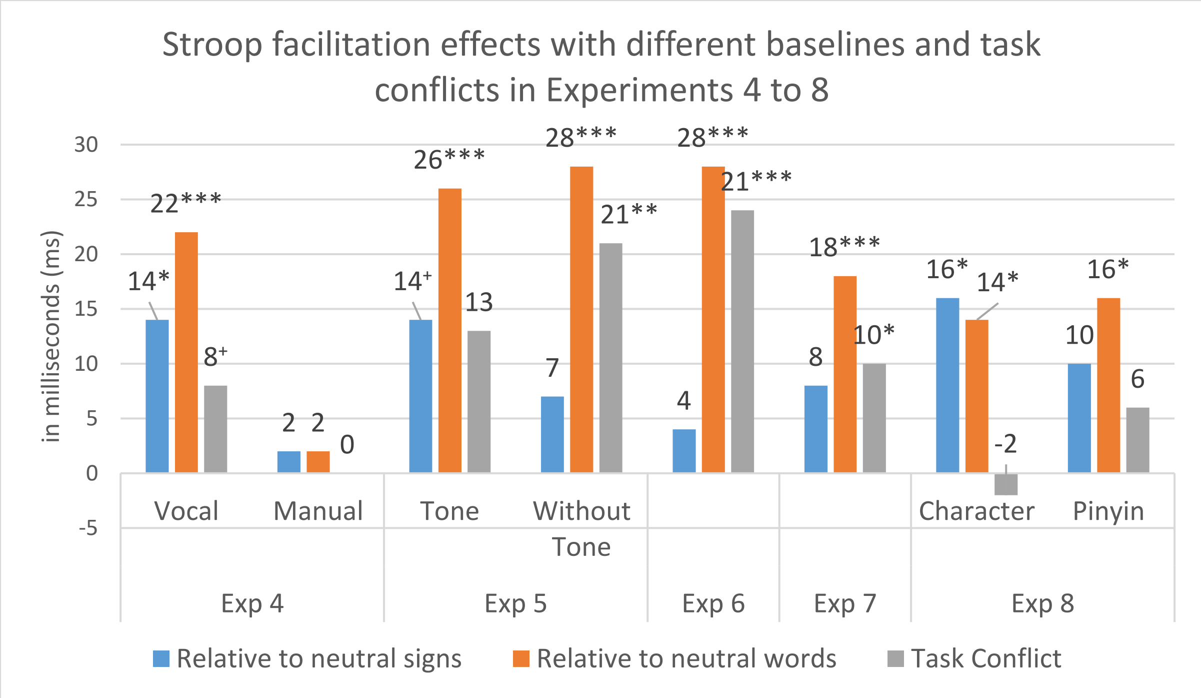


*Figure 6.5. Stroop facilitation effects with different baselines and task conflicts in Experiments 4 to 8. ***p < 0.001; **p < 0.01; *p < 0.05; ; 0.05 < $^{+}$p < 0.10.*

It is obvious that Stroop facilitation relative to neutral words is larger than Stroop facilitation relative to neutral signs. This difference is called task conflict, reflecting the involvement of the word reading task that is irrelevant to the colour naming task in the Stroop task. Strong task conflicts were often seen in studies with pinyin stimuli (Experiments 5 to 7), while it is absent in characters (Experiment 4) or a mixed presentation of both character and pinyin (Experiment 8).

Stroop facilitation relative to neutral words is stronger and more stable than Stroop facilitation relative to neutral signs because the Stroop facilitation relative to neutral words contains task conflict and Stroop facilitation relative to neutral signs does not.

In Section 2.2, there are two hypotheses on the Stroop facilitation effects: the converging information hypothesis (Cohen et al., 1990; Roelofs, 2003) and the inadvertent reading hypothesis (Kane & Engle, 2003; MacLeod & MacDonald, 2000a). From the distributional analyses of Experiments 1 to 8 in this thesis, it is clear that Stroop facilitation occurs across the quantiles, which supports the converging information hypothesis. This disagrees with the inadvertent reading hypothesis because reading error should not be present in all congruent trials.

6.5.3 Multi-stage accounts of the Stroop effects

Multi-stage accounts of Stroop effects have been developed in many phases. Klein (1964) investigated the Stroop interference effects of words with varying relationships to the colour words. They found a semantic gradience effect that the Stroop interference effects decrease as the relationship to the colour words becomes more distant. This phenomenon implied that the Stroop interference effects may have different components. Based on Klein's (1964) work, Sharma and McKenna (1998) further proposed different components in the Stroop task, including lexical, semantic relatedness, semantic relevant, and response set membership. Most importantly, they argued that different response modalities contain different components, indicating a qualitative difference between different response modalities. They found that the vocal responses have four components: lexical, semantic relatedness, semantic relevance, and response set membership, whereas manual responses contain response set membership only. Augustinova et al. (2018) proposed a semantic Stroop paradigm that interprets the Stroop interference effects as composed of response, semantic, and task conflicts. Later, Augustinova et al. (2019) extended this paradigm to Stroop facilitation effects and argued that there are two components: semantic facilitation and response facilitation. They also found that different response modalities are composed of different components as in Sharma and McKenna's (1998) study. The differences in response time between vocal and manual responses are attributed to the reduced response conflict and task conflict in the manual responses for the Stroop interference effects and attributed to the reduced response facilitation for the Stroop facilitation effects.

Experiments 4 to 8 extended Augustinova et al.'s (2019) semantic Stroop paradigm and adapted it for the Chinese language in character and pinyin. A key difference between the two paradigms was the phonological component. This is because there are abundant homophones in the Chinese language, and the same phonology can link to multiple characters with different semantics and orthography, whereas the phonological and orthographic information cannot be easily separated in many alphabetic languages (see Parris et al., 2022, for the interpretation of orthographic overlap using pseudohomophones).

When processing Chinese characters, response modality effects were attributed to the reduced phonological and semantic conflict in the manual responses for the Stroop interference effects and attributed to the reduced phonological facilitation for the Stroop facilitation effects.

The manual Stroop task conducted with pinyin stimuli revealed a contribution of task conflict in the Stroop interference effects and semantic facilitation in the Stroop facilitation effects.

To summarize, the current thesis proposed an adapted paradigm for the Chinese language. The results of the current thesis supported multi-stage account of the Stroop interference and facilitation effects. In addition, the contribution of different conflict/facilitation components differs in terms of the response modality, which suggests a qualitative difference between vocal and manual responses.

6.6 Implications

This thesis provides data to evaluate the theories and models of Chinese word recognition.

The results of Experiments 1 to 3 support the hypothesis that radicals are first activated when recognizing Chinese characters (Chen & Yeh, 2015; Feldman & Siok, 1997, 1999; Taft & Zhu, 1997; Yeh et al., 2017). This also supports the lexical constituency model (Perfetti et al., 2005). The semantic processing of radicals in Chinese characters is automatic. The radical is the basic unit of processing in Chinese, just as a letter is the basic level in alphabetic languages. However, the results of Experiments 1 to 3 also suggest that radicals can directly activate the semantics of Chinese characters. In the lexical constituency model, radical input has indirect access to semantics and phonology through the mediation of orthography. The results of Experiment 4 favoured the indirect access hypothesis (Perfetti et al., 2005; Perfetti & Tan, 1998; Perfetti & Zhang, 1995; Tan & Perfetti, 1997), as the role of phonology is confirmed with character stimuli. The results of Experiments 5 to 8 suggest that Chinese characters written in pinyin can still elicit Stroop interference and facilitation effects. As far as I am aware, no models of Chinese word recognition have considered pinyin as an input unit, but it is often considered as the representation of phonology. Further research is needed to establish a model suitable for explaining the results of pinyin words.

In addition to the implications for Chinese word recognition models and theories, this thesis provides support for the theories and models of the Stroop task.

The results of Experiment 4 support multi-stage accounts of the Stroop task and introduce the phonological components as one of the distinct components in Stroop effects when studying Chinese characters.

The distributional analyses of all experiments in this thesis favour the converging information hypothesis (Cohen et al., 1990; Roelofs, 2003) for Stroop facilitation effects and disagree with the inadvertent reading hypothesis (Kane & Engle, 2003; MacLeod & MacDonald, 2000a) because the Stroop facilitation effects are present across the quantiles rather than for certain trials.

The use of delta plots in the distributional analysis also provides details on the trend of the observed effects from the mean RT analysis. Two effects can be significant in the mean RT analysis; however, the distributional analysis may show that they have experienced different levels of suppression/enhancement. This information is hardly discernible in the mean RT analysis, but the pattern is very clear when plotting the delta plots for those effects.

6.7 Limitations

In Experiments 1 to 4, it was found that different conflict/facilitation components have different contributions to the vocal or manual responses. Experiments 5 to 8 examined the processing of pinyin words in manual Stroop tasks only. As suggested by the results of Experiments 1 to 4, the Stroop effects produced in vocal responses are generally larger than in manual responses. Certain effects that have not been observed in manual responses using pinyin words may have stronger effects when using vocal responses. In addition, a more complete picture of pinyin word processing can be achieved with both vocal and manual responses.

All experiments conducted in this thesis were behavioural. Augustinova et al. (2019) suggested that a neuropsychological approach would add value to the observed conflict/facilitation components in the behavioural study. When observing the Stroop effects from a neuropsychological perspective, it is possible to see the difference between alphabetic and non-alphabetic languages or different types of scripts in one language (Coderre et al., 2008). The differences between certain critical conditions (e.g. phonological activation of phonetic radicals) may not be conclusive in the behavioural data but may have a different story when using neuropsychological approaches.

6.8 Conclusions

The research presented in this thesis investigated Stroop interference and facilitation effects with Chinese characters and pinyin.

Strong Stroop interference and facilitation effects were observed in characters and pinyin. The distributional analysis showed that no inhibition was applied in Stroop interference to resolve the competition between the colour and word information using characters, radicals, and pinyin. However, the distributional analysis showed different levels of enhancement were applied in Stroop facilitation to converge the colour and word information: character stimuli tended to show weak enhancement applied, while radicals and pinyin stimuli had strong enhancement applied. The distributional analysis favoured the converging information hypothesis of Stroop facilitation effects because a constant growth of effects can be observed with the increase of RT, whereas the inadvertent reading hypothesis would not predict reading errors occur in all congruent trials.

In addition to the traditional Stroop interference and facilitation effects, this thesis explored whether Stroop effects are composed of different conflict/facilitation components. In addition, it is proposed that phonological components should be taken into consideration when investigating the Chinese language. Thus, the Stroop interference effects in Chinese are attributed to response, phonological and semantic conflicts, while the Stroop facilitation effects were mainly affected by phonological facilitation. The results also indicated a qualitative difference between vocal and manual responses because reduced contribution of phonological, semantic conflicts, and phonological facilitation was found in manual responses.

Results of this thesis support the decomposition view of Chinese characters, where radicals are first recognized during Chinese word recognition. This means that radicals can directly activate the semantics of Chinese characters. In addition, in contrast to Yeh et al.'s (2017) result, the phonological cues provided by phonetic radicals are confirmed by the current results. This suggested that phonetic radicals can activate both semantic and phonological information as characters do.

This thesis also investigated the processing of pinyin in the Stroop task. It was found that the presence of tonal information in pinyin does not strongly impact the magnitude of Stroop interference effects. Chinese characters associated with colour words written in pinyin can facilitate reading process, indicating that the semantic information of Chinese characters can be activated even when presented in pinyin, the phonological representation of Chinese characters. However, the semantic activation of Chinese characters that contain phonetic radicals that are colour words cannot be found in pinyin stimuli. Although the semantic activation of radicals is found using Chinese characters, pinyin cannot activate the semantics of radicals.

The decomposition view of radicals in character processing supported the lexical constituency model that radicals should be the basic unit of processing. However, as suggested by the current study, pinyin can also be recognised as an input unit rather than representing the phonology level only. Further models of Chinese word recognition can explore how pinyin activates the orthography, semantics, and phonology of Chinese characters.

References

Akaike, H. (Ed.). (1973). *Information theory and an extension of the maximum likelihood principle*. Akadémiai Kiadó.

Akaike, H. (1978a). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics, 30*(1), 9-14. https://doi.org/10.1007/BF02480194

Akaike, H. (1978b). A new look at the Bayes procedure. *Biometrika, 65*, 53-59.

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *Neuroscientist, 13*(3), 214-228. https://doi.org/10.1177/1073858407299288

Asaad, H., & Eviatar, Z. (2013). The effects of orthographic complexity and diglossia on letter naming in Arabic: A developmental study. *Writing Systems Research, 5*(2), 156-168.

Augustinova, M., & Ferrand, L. (2012). Suggestion does not de-automatize word reading: Evidence from the semantically based Stroop task. *Psychonomic Bulletin & Review, 19*(3), 521-527. https://doi.org/10.3758/s13423-012-0217-y

Augustinova, M., & Ferrand, L. (2014). Automaticity of Word Reading: Evidence From the Semantic Stroop Paradigm. *Current Directions in Psychological Science, 23*(5), 343-348. https://doi.org/10.1177/0963721414540169

Augustinova, M., Parris, B. A., & Ferrand, L. (2019). The Loci of Stroop Interference and Facilitation Effects With Manual and Vocal Responses [Original Research]. *Frontiers in Psychology, 10*(1786), 1786. https://doi.org/10.3389/fpsyg.2019.01786

Augustinova, M., Silvert, L., Spatola, N., & Ferrand, L. (2018). Further investigation of distinct components of Stroop interference and of their reduction by short response-stimulus intervals. *Acta Psychologica, 189*, 54-62. https://doi.org/10.1016/j.actpsy.2017.03.009

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/https://doi.org/10.1016/j.jml.2012.11.001

Besner, D., & Stolz, J. A. (1998). Unintentional Reading: Can Phonological Computation be Controlled? *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 52*(1), 35-43. https://doi.org/10.1037/h0087277

Bolger, D. J., Perfetti, C. A., & Schneider, W. (2005). Cross-cultural effect on the brain revisited: universal structures plus writing system variation. *Hum Brain Mapp, 25*(1), 92-104. https://doi.org/10.1002/hbm.20124

Bolker, B. M. (2015). Linear and generalized linear mixed models. *Ecological statistics: contemporary theory and application*, 309-333.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S., Poulsen, J. R., Stevens, M. H. H., & White, J. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution, 24*(3), 127-135.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences, 16*(2), 106-113. https://doi.org/10.1016/j.tics.2011.12.010

Brown, T. L. (2011). The Relationship Between Stroop Interference and Facilitation Effects: Statistical Artifacts, Baselines, and a Reassessment. *Journal of Experimental Psychology-Human Perception and Performance, 37*(1), 85-99. https://doi.org/10.1037/a0019252

Brown, T. L., Joneleit, K., Robinson, C. S., & Brown, C. R. (2002). Automaticity in reading and the Stroop task: testing the limits of involuntary word processing. *American Journal of Psychology, 115*(4), 515-543. https://www.ncbi.nlm.nih.gov/pubmed/12516527

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition, 1*(1), 9. https://doi.org/10.5334/joc.10

Bub, D. N., Masson, M. E., & Lalonde, C. E. (2006). Cognitive control in children: stroop interference and suppression of word reading. *Psychol Sci, 17*(4), 351-357. https://doi.org/10.1111/j.1467-9280.2006.01710.x

Burle, B., Spieser, L., Servant, M., & Hasbroucq, T. (2014). Distributional reaction time properties in the Eriksen task: marked differences or hidden similarities with the Simon task? *Psychonomic Bulletin & Review, 21*(4), 1003-1010. https://doi.org/10.3758/s13423-013-0561-6

Burle, B., van den Wildenberg, W., & Ridderinkhof, K. R. (2005). Dynamics of facilitation and interference in cue-priming and Simon tasks. *European Journal of Cognitive Psychology, 17*(5), 619-641. https://doi.org/10.1080/09541440540000121

Burnham, K. P. (2002). *Model selection and multimodel inference : a practical information-theoretic approach / Kenneth P. Burnham, David R. Anderson* (2nd ed.). New York: Springer.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *Plos One, 5*(6). https://doi.org/https://doi.org/10.1371/journal.pone.0010729

Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: an H215O PET study of Stroop task performance. *Neuroimage, 2*(4), 264-272. https://doi.org/10.1006/nimg.1995.1034

Cattell, J. M. (1886). The Time It Takes to See and Name Objects. *Mind, os-XI*(41), 63-65. https://doi.org/10.1093/mind/os-XI.41.63

Chang, Y. (2018). How pinyin tone formats and character orthography influence Chinese learners' tone acquisition. *Chinese as a Second Language Research, 7*(2), 195-219. https://doi.org/http://dx.doi.org/10.1515/caslar-2018-0008

Chang, Y. N., Hsu, C. H., Tsai, J. L., Chen, C. L., & Lee, C. Y. (2016). A psycholinguistic database for traditional Chinese character naming. *Behavior Research Methods, 48*(1), 112-122. https://doi.org/10.3758/s13428-014-0559-7

Chen, L., Perfetti, C. A., Fang, X., Chang, L.-Y., & Fraundorf, S. (2019). Reading Pinyin activates sublexcial character orthography for skilled Chinese readers. *Language, Cognition and Neuroscience, 34*(6), 736-746. https://doi.org/10.1080/23273798.2019.1578891

Chen, L., Perfetti, C. A., & Leng, Y. (2019). Reading Pinyin activates character orthography for highly experienced learners of Chinese. *Bilingualism: Language and Cognition, 22*(1), 103-111. https://doi.org/10.1017/S136672891700058X

Chen, L., Zhong, L., Leng, Y., & Mo, L. (2014). The Role of the Character Graphic Information in Different Pinyin Processing Tasks. *Acta Psychologica Sinica, 46*(11), 1661. https://doi.org/10.3724/SP.J.1041.2014.01661

Chen, Y. C., & Yeh, S. L. (2015). Binding radicals in Chinese character recognition: Evidence from repetition blindness. *Journal of Memory and Language, 78*, 47-63. https://doi.org/10.1016/j.jml.2014.10.002

China, P. s. R. o. (1958). *Scheme for the Chinese Phonetic Alphabet*.

Coderre, E. L., Filippi, C. G., Newhouse, P. A., & Dumas, J. A. (2008). The Stroop effect in kana and kanji scripts in native Japanese speakers: an fMRI study. *Brain and Language, 107*(2), 124-132. https://doi.org/10.1016/j.bandl.2008.01.011

Coderre, E. L., Van Heuven, W. J. B., & Conklin, K. (2013). The timing and magnitude of Stroop interference and facilitation in monolinguals and bilinguals. *Bilingualism-Language and Cognition, 16*(2), 420-441. https://doi.org/10.1017/S1366728912000405

Cohen, J. D., Dunbar, K., & Mcclelland, J. L. (1990). On the Control of Automatic Processes - a Parallel Distributed-Processing Account of the Stroop Effect. *Psychological Review, 97*(3), 332-361. https://doi.org/Doi 10.1037//0033-295x.97.3.332

Comalli, P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *The Journal of genetic psychology, 100*(1), 47-53.

Dalrymple-Alford, E. C. (1972). Associative Facilitation and Interference in Stroop Color-Word Task. *Perception & Psychophysics, 11*(4), 274-276. https://doi.org/Doi 10.3758/Bf03210377

Dalrymple-Alford, E. C., & Budayr, B. (1966). Examination of some aspects of the Stroop color-word test. *Perceptual and Motor Skills, 23*(3), 1211-1214.

De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition, 31*(3), 353-359. https://doi.org/10.3758/bf03194393

DeFrancis, J. (1989). *The Chinese language : fact and fantasy / John DeFrancis*. Honolulu : University of Hawaii Press.

Deng, Y., & Wei, R. H. (2002). The Polysemous advantage effects and Its Development in Identification of Chinese One-Characters. *Psychological Development and Education*.

Ding, C. M., & Rong, J. (2012). *A Course for Mandarin Chinese Pronunciation*. Peking University Press.

Ding, G. S., Peng, D. L., & Taft, M. (2004). The nature of the mental representation of radicals in Chinese: A priming study. *Journal of Experimental Psychology-Learning Memory and Cognition, 30*(2), 530-539. https://doi.org/10.1037/0278-7393.30.2.530

Duncanjohnson, C. C., & Kopell, B. S. (1980). The Locus of Interference in a Stroop Task - When You Read Blue, Do You See Red. *Psychophysiology, 17*(3), 308-309. <Go to ISI>://WOS:A1980JR29500090

Dunnett, C. W. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association, 50*(272), 1096-1121. https://doi.org/10.1080/01621459.1955.10501294

Dyer, F. N. (1973a). Interference and Facilitation for Color Naming with Separate Bilateral Presentations of Word and Color. *Journal of Experimental Psychology, 99*(3), 314-317. https://doi.org/DOI 10.1037/h0035245

Dyer, F. N. (1973b). Same and Different Judgments for Word-Color Pairs with Irrelevant Words or Colors - Evidence for Word-Code Comparisons. *Journal of Experimental Psychology, 98*(1), 102-108. https://doi.org/DOI 10.1037/h0034278

Entel, O., & Tzelgov, J. (2018). Focusing on task conflict in the Stroop effect. *Psychological Research-Psychologische Forschung, 82*(2), 284-295. https://doi.org/10.1007/s00426-016-0832-8

Feldman, L. B., & Siok, W. W. T. (1997). The role of component function in visual recognition of Chinese characters. *Journal of Experimental Psychology-Learning Memory and Cognition, 23*(3), 776-781. https://doi.org/Doi 10.1037//0278-7393.23.3.776

Feldman, L. B., & Siok, W. W. T. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language, 40*(4), 559-576. https://doi.org/DOI 10.1006/jmla.1998.2629

Fennell, A., & Ratcliff, R. (2019). Does response modality influence conflict? Modelling vocal and manual response Stroop interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(11), 2098-2119. https://doi.org/10.1037/xlm0000689

Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116-124. https://www.ncbi.nlm.nih.gov/pubmed/12723786

Fraisse, P. (1969). Why Is Naming Longer Than Reading. *Acta Psychologica, 30*, 96-&. https://doi.org/Doi 10.1016/0001-6918(69)90043-2

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 104-115. https://doi.org/https://doi.org/10.4324/9781315108506-4

Gao, J., Fan, K., & Fei, J. (1993). *Xiandai hanzi xue [The study of modern Chinese characters]*. Beijing: Higher Education Press.

Glaser, M. O., & Glaser, W. R. (1982). Time Course Analysis of the Stroop Phenomenon. *Journal of Experimental Psychology-Human Perception and Performance, 8*(6), 875-894. https://doi.org/Doi 10.1037//0096-1523.8.6.875

Goldfarb, L., & Henik, A. (2006). New Data Analysis of the Stroop Matching Task Calls for a Reevaluation of Theory. *Psychological Science, 17*(2), 96-100. https://doi.org/10.1111/j.1467-9280.2006.01670.x

Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology-Human Perception and Performance, 33*(5), 1170-1176. https://doi.org/10.1037/0096-1523.33.5.1170

Grant, D. A. (1956). Analysis-of-Variance Tests in the Analysis and Comparison of Curves. *Psychological Bulletin, 53*, 141-154. https://doi.org/10.1037/h0038479

Guo, T., Peng, D., & Liu, Y. (2005). The role of phonological activation in the visual semantic retrieval of Chinese characters. *Cognition, 98*(2), B21-B34. https://doi.org/10.1016/j.cognition.2005.02.004

Guttentag, R. E., & Haith, M. M. (1978). Automatic processing as a function of age and reading ability. *Child Development, 49*, 707-716. https://doi.org/10.2307/1128239

Hasshim, N., & Parris, B. A. (2014). Two-to-one color-response mapping and the presence of semantic conflict in the Stroop task. *Frontiers in Psychology, 5*, 1157. https://doi.org/10.3389/fpsyg.2014.01157

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70. <Go to ISI>://WOS:A1979JY78700003

Hu, R. F., Cao, B., & Du, J. Y. (2013). Research on phonetic symbols of phonograms in Chinese Mandarin. *Journal of Chinese Information Processing, 27*(3), 41-47.

Huaon.com. (2021). *China's third-party input method industry development forecast and investment strategy research report from 2021-2026*. https://m.huaon.com/detail/687585.html

Hue, C.-W. (1992). Recognition Processes in Character Naming. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Advances in Psychology* (Vol. 90, pp. 93-107). North-Holland. https://doi.org/https://doi.org/10.1016/S0166-4115(08)61888-9

Hung, Y. H., Hung, D. L., Tzeng, O. J. L., & Wu, D. H. (2014). Tracking the temporal dynamics of the processing of phonetic and semantic radicals in Chinese character recognition by MEG. *Journal of Neurolinguistics, 29*, 42-65. https://doi.org/10.1016/j.jneuroling.2013.12.003

Hurvich, C. M., & Chih-Ling, T. (1989). Regression and time series model selection in small samples. *Biometrika, 76*(2), 297-307. https://doi.org/10.1093/biomet/76.2.297

Hutchison, K. A. (2011). The Interactive Effects of Listwide Control, Item-Based Control, and Working Memory Capacity on Stroop Performance. *Journal of Experimental Psychology-Learning Memory and Cognition, 37*(4), 851-860. https://doi.org/10.1037/a0023437

Iwasaki, S. i. (2002). *Japanese [electronic resource] / Shoichi Iwasaki*. Amsterdam : Philadelphia : J. Benjamins Pub. Co.

Jiang, W. (2013). 象形汉字向符号化汉字演变初探. *美术教育研究*(23), 41.

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 9*(4), 497-509. https://doi.org/https://doi.org/10.1037/0096-1523.9.4.497

Kalanthroff, E., Davelaar, E. J., Henik, A., Goldfarb, L., & Usher, M. (2018). Task Conflict and Proactive Control: A Computational Theory of the Stroop Task. *Psychological Review, 125*(1), 59-82. https://doi.org/10.1037/rev0000083

Kalanthroff, E., Goldfarb, L., Usher, M., & Henik, A. (2013). Stop interfering: Stroop task conflict independence from informational conflict and interference. *Quarterly Journal of Experimental Psychology, 66*(7), 1356-1367. https://doi.org/10.1080/17470218.2012.741606

Kalanthroff, E., & Henik, A. (2013). Individual but not fragile: Individual differences in task control predict Stroop facilitation. *Consciousness and Cognition, 22*(2), 413-419. https://doi.org/https://doi.org/10.1016/j.concog.2013.01.010

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology-General, 132*(1), 47-70. https://doi.org/10.1037/0096-3445.132.1.47

Katz, L., & Feldman, L. (1981). *Linguistic Coding in Word Recognition: Comparisons Between a Deep and a Shallow Orthography*. Routledge. https://doi.org/10.4324/9781315108506-4

Keele, S. W. (1972). Attention Demands of Memory Retrieval. *Journal of Experimental Psychology, 93*(2), 245-248. https://doi.org/DOI 10.1037/h0032460

Kinoshita, S., De Wit, B., & Norris, D. (2017). The Magic of Words Reconsidered: Investigating the Automaticity of Reading Color-Neutral Words in the Stroop Task. *Journal of Experimental Psychology-Learning Memory and Cognition, 43*(3), 369-384. https://doi.org/10.1037/xlm0000311

Klein, G. S. (1964). Semantic Power Measured through the Interference of Words with Color-Naming. *The American Journal of Psychology, 77*(4), 576-588. https://doi.org/10.2307/1420768

Kornblum, S. (1992). *Dimensional overlap and dimensional relevance in stimulus-response and stimulus-stimulus compatibility*. North-Holland.

Leck, K. J., Weekes, B. S., & Chen, M. J. (1995). Visual and Phonological Pathways to the Lexicon - Evidence from Chinese Readers. *Memory & Cognition, 23*(4), 468-476. https://doi.org/Doi 10.3758/Bf03197248

Lee, C., Tsai, J., Huang, H., Hung, D. L., & Tzeng, O. J. (2006). The temporal signatures of semantic and phonological activations for Chinese sublexical processing: an event-related potential study. *Brain Res, 1121*(1), 150-159. https://doi.org/10.1016/j.brainres.2006.08.117

Lee, C., Tsai, J., Su, E. C., Tzeng, O. J., & Hung, D. L. (2005). Consistency , Regularity , and Frequency Effects in Naming Chinese Characters. *Language and Linguistics, 6*, 75-107.

Leng, Y. L., & Wei, Y. X. (1994). *Zhonghua Zihai [Chinese Thesauruses]*. Zhonghua Shuju.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-38. https://doi.org/10.1017/S0140525X99001776

Levin, Y., & Tzelgov, J. (2016). What Klein's "Semantic Gradient" Does and Does Not Really Show: Decomposing Stroop Interference into Task and Informational Conflict Components. *Frontiers in Psychology, 7*, 249. https://www.frontiersin.org/article/10.3389/fpsyg.2016.00249

Levitt, J. S., Nakakita, M., & Katz, W. F. (2015). Role of Phonology in Reading: A Stroop Effect Case Report With Japanese Scripts. *Studies in Literature and Language, 10*(3), 1-6.

Li, C., Lin, C. Y., Wang, M., & Jiang, N. (2013). The activation of segmental and tonal information in visual word recognition. *Psychonomic Bulletin & Review, 20*(4), 773-779. https://doi.org/10.3758/s13423-013-0395-2

Li, C., Wang, M., Davis, J. A., & Guan, C. Q. (2019). The role of segmental and tonal information in visual word recognition with learners of Chinese. *Journal of Research in Reading, 42*(2), 213-238. https://doi.org/10.1111/1467-9817.12137

Lin, D., McBride-Chang, C., Shu, H., Zhang, Y., Li, H., Zhang, J., Aram, D., & Levin, I. (2010). Small Wins Big: Analytic Pinyin Skills Promote Chinese Word Reading. *Psychological Science, 21*(8), 1117-1122. https://doi.org/10.1177/0956797610375447

Liu, H., & Weng, X. (2007). Differences of Stroop Interference:Characters,Pinyin and English. *Psychological Science, 30*(2), 365-368.

Liu, I. M., Wu, J. T., & Chou, T. L. (1996). Encoding operation and transcoding as the major loci of the frequency effect. *Cognition, 59*(2), 149-168. https://doi.org/Doi 10.1016/0010-0277(95)00688-5

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01171

Locker, L., Jr., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods, 39*(4), 723-730. https://doi.org/10.3758/bf03192962

Logan, G. D. (1978). Attention in Character-Classification Tasks - Evidence for Automaticity of Component Stages. *Journal of Experimental Psychology-General, 107*(1), 32-63. https://doi.org/Doi 10.1037//0096-3445.107.1.32

Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 978-992. https://doi.org/https://doi.org/10.1037/0096-1523.24.3.978

Lüdecke, D., Ben-Shachar, S. M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

Luo, C. M., Proctor, R. W., & Weng, X. C. (2015). A Stroop effect emerges in the processing of complex Chinese characters that contain a color-related radical. *Psychological Research-Psychologische Forschung, 79*(2), 221-229. https://doi.org/10.1007/s00426-014-0553-9

Luo, C. M., Proctor, R. W., Weng, X. C., & Li, X. S. (2014). Spatial Stroop interference occurs in the processing of radicals of ideogrammic compounds. *Psychonomic Bulletin & Review, 21*(3), 715-720. https://doi.org/10.3758/s13423-013-0533-x

Macleod, C. M. (1991). Half a Century of Research on the Stroop Effect - an Integrative Review. *Psychological Bulletin, 109*(2), 163-203. https://doi.org/Doi 10.1037/0033-2909.109.2.163

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: the progression of interference and facilitation. *Memory & Cognition, 26*(2), 201-211. https://doi.org/10.3758/bf03201133

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: evidence for a continuum of automaticity. *Journal of Experimental Psychology-Learning Memory and Cognition, 14*(1), 126-135. https://doi.org/10.1037//0278-7393.14.1.126

MacLeod, C. M., & MacDonald, P. A. (2000a). Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*(10), 383-391. https://doi.org/Doi 10.1016/S1364-6613(00)01530-8

MacLeod, C. M., & MacDonald, P. A. (2000b). Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*(10), 383-391. https://doi.org/10.1016/s1364-6613(00)01530-8

Morton, J., & Chambers, S. M. (1973). Selective Attention to Words and Colors. *Quarterly Journal of Experimental Psychology, 25*(3), 387-397. https://doi.org/Doi 10.1080/14640747308400360

National Language Commission, P. s. R. o. C. (1964). *Summary of Simplified Chinese*.

Nealis, P. M. (1973). The Stroop phenomenon: some critical tests of the response competition hypothesis. *Percept Mot Skills, 37*(1), 147-153. https://doi.org/10.2466/pms.1973.37.1.147

Neely, J., & Kahan, T. (2001). Is semantic activation automatic? A critical re-evaluation. *The Nature of Remembering: Essays in Honor of R. G. Crowder*. https://doi.org/10.1037/10394-005

Neill, W. T. (1977). Inhibitory and Facilitatory Processes in Selective Attention. *Journal of Experimental Psychology-Human Perception and Performance, 3*(3), 444-450. https://doi.org/Doi 10.1037//0096-1523.3.3.444

Parris, B. A., Hasshim, N., Wadsley, M., Augustinova, M., & Ferrand, L. (2022). The loci of Stroop effects: a critical review of methods and evidence for levels of processing contributing to color-word Stroop effects and the implications for the loci of attentional selection. *Psychological Research, 86*(4), 1029-1053. https://doi.org/10.1007/s00426-021-01554-x

Parris, B. A., Sharma, D., Weekes, B. S. H., Momenian, M., Augustinova, M., & Ferrand, L. (2019). Response Modality and the Stroop Task. *Experimental Psychology, 66*(5), 361. https://doi.org/10.1027/1618-3169/a000459

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 112*(1), 43-59. https://doi.org/10.1037/0033-295x.112.1.43

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 101-118. https://doi.org/10.1037/0278-7393.24.1.101

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 24-33.

Pratte, M. S. (2021). Eriksen flanker delta plot shapes depend on the stimulus. *Attention, Perception, & Psychophysics, 83*(2), 685-699. https://doi.org/10.3758/s13414-020-02166-0

Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics, 72*(7), 2013-2025. https://doi.org/10.3758/APP.72.7.2013

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39*(4), 859-862. https://doi.org/10.3758/BF03192979

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to Deal with "The Language-as-Fixed-Effect Fallacy": Common Misconceptions and Alternative Solutions. *Journal of Memory and Language, 41*(3), 416-426. https://doi.org/https://doi.org/10.1006/jmla.1999.2650

Redding, G. M., & Gerjets, D. A. (1977). Stroop Effect - Interference and Facilitation with Verbal and Manual Responses. *Perceptual and Motor Skills, 45*(1), 11-17. https://doi.org/DOI 10.2466/pms.1977.45.1.11

Regan, J. (1978). Involuntary automatic processing in color-naming tasks. *Perception & Psychophysics, 24*(2), 130-136. https://doi.org/10.3758/bf03199539

Ridderinkhof, K. R., Scheres, A., Oosterlaan, J., & Sergeant, J. A. (2005). Delta plots in the study of individual differences: New tools reveal response inhibition deficits in AD/HD that are eliminated by methylphenidate treatment. *Journal of Abnormal Psychology, 114*(2), 197-215. https://doi.org/10.1037/0021-843x.114.2.197

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*(1), 107-142. https://doi.org/10.1016/0010-0277(92)90041-f

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*(3), 249-284. https://doi.org/10.1016/s0010-0277(97)00027-9

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modelling attentional control in the Stroop task. *Psychological Review, 110*(1), 88-125. https://doi.org/10.1037/0033-295x.110.1.88

Roelofs, A. (2010a). Attention and Facilitation: Converging Information Versus Inadvertent Reading in Stroop Task Performance. *Journal of Experimental Psychology-Learning Memory and Cognition, 36*(2), 411-422. https://doi.org/10.1037/a0018523

Roelofs, A. (2010b). Attention, temporal predictability, and the time course of context effects in naming performance. *Acta Psychologica, 133*(2), 146-153. https://doi.org/10.1016/j.actpsy.2009.11.003

Roelofs, A. (2012). Attention, spatial integration, and the tail of response time distributions in Stroop task performance. *Quarterly Journal of Experimental Psychology, 65*(1), 135-150. https://doi.org/10.1080/17470218.2011.605152

Roelofs, A., Piai, V., & Rodriguez, G. G. (2011). Attentional inhibition in bilingual naming performance: evidence from delta-plot analyses. *Frontiers in Psychology, 2*. https://doi.org/https://doi.org/10.3389/fpsyg.2011.00184

Rumelhart, D. E., McClelland, J. L., & Group, P. R. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. The MIT Press. https://doi.org/10.7551/mitpress/5236.001.0001

Saalbach, H., & Stern, E. (2004). Differences between Chinese morphosyllabic and German alphabetic readers in the Stroop interference effect. *Psychonomic Bulletin & Review, 11*(4), 709-715. https://doi.org/10.3758/BF03196624

Saito, H., Masuda, H., & Kawakami, M. (1998). Form and sound similarity effects in kanji recognition. *Reading and Writing, 10*, 323-357. https://doi.org/Doi 10.1023/A:1008093507932

Schmidt, J. R., & Besner, D. (2008). The stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology-Learning Memory and Cognition, 34*(3), 514-523. https://doi.org/10.1037/0278-7393.34.3.514

Schmidt, J. R., & Cheesman, J. (2005). Dissociating Stimulus-Stimulus and Response-Response Effects in the Stroop Task. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 59*(2), 132-138. https://doi.org/10.1037/h0087468

Schulz, T. (1979). Components of the reaction time stroop-task. *Psychological Research, 40*(4), 377-395. https://doi.org/10.1007/BF00309418

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6*(2), 461-464, 464. https://doi.org/10.1214/aos/1176344136

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*(1), 1-30. https://doi.org/https://doi.org/10.1016/0010-0277(85)90029-0

Sharma, D., Booth, R., Brown, R., & Huguet, P. (2010). Exploring the temporal dynamics of social facilitation in the Stroop task. *Psychonomic Bulletin & Review, 17*(1), 52-58. https://doi.org/10.3758/PBR.17.1.52

Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition, 26*(5), 1033-1040. https://doi.org/10.3758/BF03201181

Shen, D., & Forster, K. I. (1999). Masked phonological priming in reading Chinese words depends on the task. *Language and Cognitive Processes, 14*(5-6), 429-459.

Shichel, I., & Tzelgov, J. (2018). Modulation of conflicts in the Stroop effect. *Acta Psychologica, 189*, 93-102. https://doi.org/https://doi.org/10.1016/j.actpsy.2017.10.007

Shiffrin, R. M., & Schneider, W. (1977). Controlled and Automatic Human Information-Processing .2. Perceptual Learning, Automatic Attending, and a General Theory. *Psychological Review, 84*(2), 127-190. https://doi.org/Doi 10.1037//0033-295x.84.2.127

Shu, Hua, Anderson, Richard, C., Wu, & Ningning. (2000). Phonetic awareness: Knowledge of orthography-phonology relationships in the character acquisition of Chinese children. *Journal of Educational Psychology*.

Shu, H., & Anderson, R. C. (1999). Learning to read Chinese: The development of metalinguistic awareness. In *Reading Chinese script: A cognitive analysis.* (pp. 1-18). Lawrence Erlbaum Associates Publishers.

Sichel, J. L., & Chandler, K. A. (1969). The Color-Word Interference Test: The Effects of Varied Color-Word Combinations Upon Verbal Response Latency. *The Journal of Psychology, 72*(2), 219-231. https://doi.org/10.1080/00223980.1969.10543502

Siok, W. T., & Fletcher, P. (2001). The Role of Phonological Awareness and Visual-Orthographic Skills in Chinese Reading Acquisition. *Developmental psychology, 37*(6), 886-899. https://doi.org/10.1037/0012-1649.37.6.886

Siok, W. T., Perfetti, C. A., Jin, Z., & Tan, L. H. (2004). Biological abnormality of impaired reading is constrained by culture. *Nature, 431*(7004), 71-76. https://doi.org/10.1038/nature02865

Spinks, J. A., Liu, Y., Perfetti, C. A., & Tan, L. H. (2000). Reading Chinese characters for meaning: the role of phonological information. *Cognition, 76*(1), B1-B11. https://doi.org/Doi 10.1016/S0010-0277(00)00072-X

State Language Commission, P. s. R. o. C. (1988). *Xiandai Hanyu Changyong Zibiao [High-frequency Character in Modern Chinese]*.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662. https://doi.org/Doi 10.1037/0096-3445.121.1.15

Su, I. F., Mak, S. C. C., Cheung, L. Y. M., & Law, S. P. (2012). Taking a radical position: evidence for position-specific radical representations in Chinese character recognition using masked priming ERP. *Frontiers in Psychology, 3*. https://doi.org/https://doi.org/10.3389/fpsyg.2012.00333

Sumiya, H., & Healy, A. F. (2004). Phonology in the bilingual Stroop effect. *Memory & Cognition, 32*(5), 752-758. https://doi.org/10.3758/bf03195865

Sun, J., Rao, L., Gao, C., Zhang, L., Liang, L., & Gong, H. (2016). The Role of Phonological Processing in Semantic Access of Chinese Characters: A Near-Infrared Spectroscopy Study. Oxygen Transport to Tissue XXXVIII, Cham.

Taft, M., & Zhu, X. P. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology-Learning Memory and Cognition, 23*(3), 761-775. https://doi.org/Doi 10.1037/0278-7393.23.3.761

Taft, M., Zhu, X. P., & Peng, D. L. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory and Language, 40*(4), 498-519. https://doi.org/DOI 10.1006/jmla.1998.2625

Tan, L. H., & Perfetti, C. A. (1997). Visual Chinese Character Recognition: Does Phonological Information Mediate Access to Meaning? *Journal of Memory and Language, 37*(1), 41-57. https://doi.org/https://doi.org/10.1006/jmla.1997.2508

Tan, L. H., Spinks, J. A., Feng, C. M., Siok, W. T., Perfetti, C. A., Xiong, J., Fox, P. T., & Gao, J. H. (2003). Neural systems of second language reading are shaped by native language. *Hum Brain Mapp, 18*(3), 158-166. https://doi.org/10.1002/hbm.10089

Tzeng, O. J., Hung, D. L., & Wang, W. S. Y. (1977). Speech recoding in reading Chinese characters. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 621-630. https://doi.org/10.1037/0278-7393.3.6.621

Vanayan, M. (1993). Relating interference and facilitation in the Stroop task: An individual-differences comparison. *Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario, Canada*.

Vurpillot, E., & Ball, W. A. (1979). The concept of identity and children's selective attention. In *Attention and cognitive development* (pp. 23-42). Springer.

Wang, K., Mecklinger, A., Hofmann, J., & Weng, X. (2010). From orthography to meaning: an electrophysiological investigation of the role of phonology in accessing meaning of Chinese single-character words. *Neuroscience, 165*(1), 101-106. https://doi.org/https://doi.org/10.1016/j.neuroscience.2009.09.070

Wang, M., Anderson, A., Cheng, C., Park, Y., & Thomson, J. (2008). General auditory processing, Chinese tone processing, English phonemic processing and English reading skill: a comparison between Chinese-English and Korean-English bilingual children. *Reading and Writing, 21*(6), 627-644. https://doi.org/10.1007/s11145-007-9081-y

Wang, M., Li, C., & Lin, C. Y. (2015). The Contributions of Segmental and Suprasegmental Information in Reading Chinese Characters Aloud. *Plos One, 10*(11), e0142060. https://doi.org/10.1371/journal.pone.0142060

Wang, M. Y. (2006). Examining the bias for orthographic components using an apparent motion detection task. *Psychologia, 49*(3), 193-213. https://doi.org/DOI 10.2117/psysoc.2006.193

Wang, X. S., Pei, M., Wu, Y., & Su, Y. J. (2017). Semantic Radicals Contribute More Than Phonetic Radicals to the Recognition of Chinese Phonograms: Behavioral and ERP

Evidence in a Factorial Study. *Frontiers in Psychology, 8*. https://doi.org/https://doi.org/10.3389/fpsyg.2017.02230

White, B. W. (1969). Interference in Identifying Attributes and Attribute Names. *Perception & Psychophysics, 6*(3), 166-168. https://doi.org/Doi 10.3758/Bf03210086

Wong, K. F. E., & Chen, H.-C. (1999). Orthographic and Phonological Processing in Reading Chinese Text: Evidence From Eye Fixations. *Language and Cognitive Processes, 14*(5-6), 461-480. https://doi.org/10.1080/016909699386158

Wu, Y., Mo, D. Y., Tsang, Y. K., & Chen, H. C. (2012). ERPs reveal sub-lexical processing in Chinese character recognition. *Neuroscience Letters, 514*(2), 164-168. https://doi.org/10.1016/j.neulet.2012.02.080

Xiang, G. (2010). 文字起源考索. *信阳师范学院学报: 哲学社会科学版*.

Xu, S. (1963). *说文解字 [Shuowen Jiezi, the first Chinese dictionary]*. Zhonghua Shuju.

Xu, Y. D., Pollatsek, A., & Potter, M. C. (1999). The activation of phonology during silent Chinese word reading. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*(4), 838-857. https://doi.org/Doi 10.1037//0278-7393.25.4.838

Yeh, S. L., Chou, W. L., & Ho, P. K. (2017). Lexical processing of Chinese sub-character components: Semantic activation of phonetic radicals as revealed by the Stroop effect. *Scientific Reports, 7*(1), 15782. https://doi.org/10.1038/s41598-017-15536-w

Yeh, S. L., & Li, J. L. (2004). Sublexical processing in visual recognition of Chinese characters: Evidence from repetition blindness for subcharacter components. *Brain and Language, 88*(1), 47-53. https://doi.org/10.1016/S0093-934x(03)00146-9

Yoshihara, M., Nakayama, M., Verdonschot, R. G., Hino, Y., & Lupker, S. J. (2021). Orthographic properties of distractors do influence phonological Stroop effects: Evidence from Japanese Romaji distractors. *Memory & Cognition, 49*(3), 600-612. https://doi.org/10.3758/s13421-020-01103-8

Yu, B. L., Cao, H. Q., Feng, L., & Li, W. L. (1990). Effect of morphological and phonetic whole perception of Chinese characters on the perception of components. *Acta Psychologica Sinica, 22*(03), 10-17.

Zahedi, A., Abdel Rahman, R., Stürmer, B., & Sommer, W. (2019). Common and specific loci of Stroop effects in vocal and manual tasks, revealed by event-related brain potentials and posthypnotic suggestions. *Journal of Experimental Psychology-General, 148*(9), 1575-1594. https://doi.org/10.1037/xge0000574

Zhang, S.-Z., Georgiou, G. K., Inoue, T., Zhong, W.-w., & Shu, H. (2020). Do pinyin and character recognition help each other grow? *Early Childhood Research Quarterly, 53*, 476-483. https://doi.org/https://doi.org/10.1016/j.ecresq.2020.06.004

Zhou, X. L., & Marslen-Wilson, W. (1999a). The nature of sublexical processing in reading Chinese characters. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*(4), 819-837. https://doi.org/Doi 10.1037//0278-7393.25.4.819

Zhou, X. L., & Marslen-Wilson, W. (1999b). Phonology, Orthography, and Semantic Activation in Reading Chinese. *Journal of Memory and Language, 41*(4), 579-606. https://doi.org/https://doi.org/10.1006/jmla.1999.2663

Zhou, Y. G. (1978). Xiandai hanzizhong shengpangde biaoyin gongneng wenti [To what degree are the "phonetics" of present-day Chinese characters still phonetic?]. In *Zhongguo Yuwen* (Vol. 146, pp. 172-177).

Zhou, Y. G. (2003). The Historical Evolution of Chinese Languages and Scripts. Pathways to Advanced Skills Series. Volume 8. *Foreign Language Publications*.

Appendices

Appendix A: Stimuli used in Experiment 1, 2, and 3

Condition (Associated Colour)	Character (Phonetic radical)	Meaning (Phonetic radical)	Pronunciation	Frequency (Stroke count)	Neutral-Control (Phonetic radical)	Meaning (Phonetic Radical)	Pronunciation	Frequency (Stroke count)
Colour-Character								
Cyan	青	cyan	[qing1]	257 (8)	具	tool	[ju4]	262 (8)
Yellow	黄	yellow	[huang2]	513 (12)	曾	already	[ceng2]	543 (12)
Red	朱	red	[zhu1]	117 (6)	丢	discard	[diu1]	117 (6)
Valid-Radical								
Cyan	清（青）	clear (cyan)	[qing1]	1760 (11)	理（里）	reason (length unit)	[li3]	1666 (11)
Yellow	潢（黄）	pond (yellow)	[huang2]	2 (15)	谆（享）	iterate (enjoy)	[zhun1]	2 (15)
Red	珠（朱）	pearl (red)	[zhu1]	44 (10)	轩（干）	pavilion (dry)	[xuan1]	45 (10)
Invalid-Radical								
Cyan	猜（青）	guess (cyan)	[cai1]	130 (11)	帐（长）	tent (long)	[zhang4]	130 (11)
Yellow	横（黄）	horizontal (yellow)	[heng2]	73 (16)	榜（旁）	placard (side)	[bang3]	65 (14)
Red	殊（朱）	different (red)	[shu1]	62 (10)	勒（革）	strangle (leather)	[le4]	62 (11)

Note: The above information was provided by Yeh et al. (2017) in Supplementary Material 1.1. The original source of frequency count was not provided by Yeh et al. (2017), but we know that those values are based on frequency per million. The original frequency count was based on traditional Chinese, which may be different in simplified Chinese; therefore, we provided a new frequency count in Simplified Chinese in Appendix B.

Appendix B: Linguistic properties for stimuli used in Experiment 1, 2, and 3

Character (Phonetic radical)	Meaning (Phonetic radical)	Frequency SUBTLEX-CH	PR Frequency	PR Family Size	PR Regularity	PR Friends	PR Enemies Types	PR Friends Frequency	PR Enemies Frequency
Colour-Character									
青	cyan	17	NA	NA	NA	NA	NA	NA	NA
黄	yellow	29	NA	NA	NA	NA	NA	NA	NA
朱	red	4	NA	NA	NA	NA	NA	NA	NA
Valid-Radical									
清（青）	clear (cyan)	85	4528	13	1	124	8	591	5383
潢（黄）	pond (yellow)	/	/	/	/	/	/	/	/
珠（朱）	pearl (red)	6	142	9	1	51	1	84	35
Invalid-Radical									
猜（青）	guess (cyan)	195	4528	13	0	14	8	129	5666
横（黄）	horizontal (yellow)	8	56	6	0	8	2	3	47
殊（朱）	different (red)	1	142	9	0	9	1	32	108
Colour-Character's Neutral Control									
具	tool	51	NA	NA	NA	NA	NA	NA	NA
曾	already	151	692	7	1	58	5	556	248
丢	discard	122	NA	NA	NA	NA	NA	NA	NA
Valid-Radical's Neutral Control									
理（里）	reason (length unit)	52	927	9	1	186	4	1040	139
谆（享）	iterate (enjoy)	/	/	/	/	/	/	/	/

轩（干）	pavilion (dry)	0	520	12	0	2	9	1	651
Invalid-Radical's Neutral Control									
帐（长）	tent (long)	60	700	6	0	53	3	78	637
榜（旁）	placard (side)	3	147	10	0	15	4	77	56
勒（革）	strangle (leather)	36	306	4	0	0	3	0	304

Note: All of the data were based on Sun, Hendrix, Ma, and Baayen's (2018) Chinese lexical database (CLD). Frequency SUBTLEX-CH, a corpus of film subtitles (per million) (Cai & Brysbaert, 2010); PR, Phonetic Radical; PR Friends, occurrences of the same phonetic radical in a character with the same pronunciation. PR Enemies Types, occurrences of the same phonetic radical in a character with a different pronunciation.

Appendix C: Linguistic characteristics at the radical level[1]

Condition (Associated Colour)	Character (Phonetic radical)	Consistency[2]	Phonetic Combinability[3]	Semantic Combinability[4]	Neutral-Control (Phonetic radical)	Consistency	Phonetic Combinability	Semantic Combinability
Colour-Character								
Cyan	青	--	--	--	具	--	--	--
Yellow	黃	--	--	--	曾	--	--	--
Red	朱	--	--	--	丟	--	--	--
Valid-Radical								
Cyan	清（青）	0.50	16	226	理（里）	0.91	10	65
Yellow	潢（黃）	0.57	6	226	谆（享）	0.14	7	123
Red	珠（朱）	0.80	10	65	轩（干）	0.06	15	30
Invalid-Radical								
Cyan	猜（青）	0.06	16	35	帐（长）	0.67	6	20
Yellow	横（黄）	0.13	8	289	榜（旁）	0.54	10	161
Red	殊（朱）	0.20	10	17	勒（革）	0.50	10	31

Note: [1] The above information was provided by Yeh et al. (2017) in Supplementary Material 1.1. [2] Consistency is defined as the ratio of the number of characters sharing a phonetic radical that have the same pronunciation to the number of characters sharing that phonetic radical. Tonal differences are not taken into account. [3] Phonetic Combinability is defined as the number of characters that share a phonetic radical. [4] Semantic Combinability is defined as the number of characters that share a semantic radical. All of the data were based on Chang et al. (2016) and sinica.edu.tw database.

Appendix D: Linguistic properties for stimuli used in Experiment 4

Condition	Character	Meaning	Pronunciation	Frequency	Stroke count
Colour words					
	绿	green	/lü4/	178	11
	黄	yellow	/huang2/	281	11
	蓝	blue	/lan2/	106	13
Associated Colour words					
	草	grass	/cao3/	444	9
	金	golden	/jin1/	384	8
	天	sky	/tian1/	3090	4
Homophone words					
	虑	ponder	/lü4/	85	10
	皇	emperor	/huang2/	62	9
	岚	mist	/lan2/	??[1]	7
Colour Neutral words					
	贯	pass through	/guan4/	74	8
	奖	prize	/jiang3/	51	9
	炭	charcoal	/tan4/	47	10

[1] The frequency count for 岚 is not recorded in the Modern Chinese Frequency Dictionary (1986), but its frequency is similar to that of 炭 in Chinese text computing (2004).

Appendix E: Possible meanings of pinyin stimuli used in Experiment 5 and 6

Pinyin	Possible Meaning 1	Possible Meaning 2	Possible Meaning 3	Possible Meaning 4	Possible Meaning 5
Colour characters in pinyin (Exp 5, 6)					
/lv4/	green	law	filter	consider	
/huang2/	yellow	emperor	phoenix	jade	bright
/lan2/	blue	holdback	eupatorium	basket	railing
Neutral characters in pinyin (Exp 5, 6)					
/po1/	very	splash	slope		
/qiong2/	poor	red jade	high		
/xun4/	instruct	interrogate	modest	tame	sacrifice
Colour associated characters in pinyin (exclusive for Exp 6)					
/cao3/	grass				
/jin1/	gold	today	a unit of weight (=1/2 kilogram)		
/tian1/	sky	add			
Neutral characters in pinyin (exclusive for Exp 6)					
/xiu1/	repair	rest	shy	make a din	
/rou4/	meat				
/zhua1/	scratch				

Appendix F: Possible meanings of segment (pinyin without tonal information) used in Experiment 5 and 6

Segment	Possible Meaning 1	Possible Meaning 2	Possible Meaning 3	Possible Meaning 4	Possible Meaning 5
Colour characters in pinyin (Exp 5)					
/lv/	donkey	backbone	repeatedly	aluminium	journey
/huang/	shine	nervous	lie		
/lan/	lazy	hold	rot	float	
Neutral characters in pinyin (Exp 5)					
/po/	break	force	old woman		
/qiong/					
/xun/	fumigate/smoke	circulate	seek	patrol	
Colour associated characters in pinyin (added in Exp 6)					
/cao/	grasp	group	groove		
/jin/	careful beautiful	advance; move only	tense immerse	taboo exhaust	enter
/tian/	fill	field	lick	sweet	
Neutral characters in pinyin (added in Exp 6)					
/xiu/	decayed	pretty	sleeve	sniff	
/rou/	soft	rub			
/zhua/	claw				

Appendix G: Possible meanings of pinyin stimuli used in Experiment 7

Pinyin	Possible Meaning 1	Possible Meaning 2	Possible Meaning 3	Possible Meaning 4	Possible Meaning 5
Colour characters in pinyin					
/qing1/	cyan	clear	light/little	bend/lean	
/huang2/	yellow	emperor	phoenix	jade	bright
/zhu1/	red	pig	pearl	stub	
Invalid-Radical condition in pinyin					
/cai1/	guess				
/heng2/	horizontal	permanent	weight/measure	purlin	
/shu2/	different	book	transport/defeat	comb	uncle
Neutral characters in pinyin					
/ju4/	tool	sentence	assemble	rely on	distance
/ceng2/	already	layer/tier			
/diu1/	discard				
/zhang4/	tent	swell	rely on	hold	
/bang3/	placard	bind/tie			
/le4/	strangle				

Appendix H: Possible meanings of segment (pinyin without tonal information) used in Experiment 7

Segment	Possible Meaning 1	Possible Meaning 2	Possible Meaning 3	Possible Meaning 4	Possible Meaning 5
Colour characters in pinyin					
/qing/	request	affection	clear	celebrate	
/huang/	shine	nervous	lie		
/zhu/	stay	pray	boil	host	help
Invalid-Radical condition in pinyin					
/cai/	ability	vegetable	trample	wealth	
/heng/	groan	be prosperous			
/shu/	trees	count	harvest	careful	
Neutral characters in pinyin					
/ju/	raise	part	dwell		
/ceng/	rub	scold			
/diu/					
/zhang/	draw	a song	long	palm	
/bang/	group	stick	pound		
/le/	used after the verb or adjective				

Appendix I: Final model for Experiment 1

Random Effects				
Group	Name	Variance	SD	Correlation
subject	(Intercept)	1.619E+03	40.24027	
Number of observations: 8493, groups: subject, 41				

Fixed Effects				
	Estimates	SE	t	p
(Intercept)	713.039	13.387	53.262	< 2e-16 ***
congruency2	-24.23	4.1	-5.909	3.44e-09 ***
congruency3	75.759	3.825	19.808	< 2e-16 ***
character type2	-29.116	4.047	-7.195	6.24e-13 ***
character type3	-36.148	4.081	-8.857	< 2e-16 ***
colour1	4.903	2.285	2.146	0.0319 *
colour2	29.564	2.37	12.476	< 2e-16 ***
congruency2:character type2	-10.674	8.442	-1.264	0.2061
congruency3:character type2	-64.253	7.398	-8.685	< 2e-16 ***
congruency2:character type3	-1.744	7.801	-0.224	0.8231
congruency3:character type3	-81.064	8.334	-9.727	< 2e-16 ***

Model Fit		
AIC	BIC	Log Likelihood
109565.8	109657.4	-54769.9

Family: inverse.gaussian (identity)
Model: RT ~ congruency * character type + colour + (1|subject)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Note:
congruency2: neutral vs. congruent (i.e. Stroop facilitation effects)
congruency3: neutral vs. incongruent (i.e. Stroop interference effects)
character type2: Colour-Character vs. Valid-Radical
character type3: Colour-Character vs. Invalid-Radical
colour1: colour cyan vs. mean
colour2: colour red vs. mean
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Appendix J: Final model for Experiment 2

Random Effects					
Group	Name	Variance	SD	Correlation	
subject	(Intercept)	1.78E+03	42.24147		
	character type2	3.37E+02	18.35567	0.05	
	character type3	3.32E+02	18.20852	-0.03	0.41
item	(Intercept)	3.13E+01	5.59282		
Number of observations: 8783, groups: subject, 42; item, 18					

Fixed Effects				
	Estimates	SE	t	p
(Intercept)	639.851	11.629	55.02	< 2e-16 ***
congruency2	-15.607	5.129	-3.043	0.00234 **
congruency3	20.305	4.785	4.244	2.20e-05 ***
character type2	-4.49	7.063	-0.636	0.5249
character type3	-7.055	6.778	-1.041	0.29795
colour1	11.538	2.193	5.262	1.42e-07 ***
colour2	-5.66	2.16	-2.621	0.00878 **
congruency2:character type2	20.045	8.49	2.361	0.01823 *
congruency3:character type2	-15.442	8.873	-1.74	0.08181 .
congruency2:character type3	31.479	9.996	3.149	0.00164 **
congruency3:character type3	-10.411	10.105	-1.03	0.30285

Model Fit		
AIC	BIC	Log Likelihood
12586.5	112721.0	-56274.3

Family: inverse.gaussian (identity)
Model: RT ~ congruency * character type + colour + (1 + character type|subject) + (1 | item)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Note:
congruency2: neutral vs. congruent (i.e. Stroop facilitation effects)
congruency3: neutral vs. incongruent (i.e. Stroop interference effects)
character type2: Colour-Character vs. Valid-Radical
character type3: Colour-Character vs. Invalid-Radical
colour1: colour cyan vs. mean
colour2: colour red vs. mean
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Appendix K: Final model for Experiment 3

Random Effects				
Group	Name	Variance	SD	Correlation
subject	(Intercept)	1.065e+03	32.62851	
Number of observations: 16146, groups: subject, 39				

Fixed Effects				
	Estimates	SE	t	p
(Intercept)	652.8496	10.6249	61.445	< 2e-16 ***
congruency2	-39.8724	2.7464	-14.518	< 2e-16 ***
congruency3	39.8681	2.5644	15.547	< 2e-16 ***
character type2	-15.3739	2.7311	-5.629	1.81e-08 ***
character type3	-11.3264	2.7914	-4.058	4.96e-05 ***
colour1	-0.6149	1.5667	-0.392	0.69472
colour2	39.141	1.6733	23.391	< 2e-16 ***
congruency2:character type2	-0.8622	5.1815	-0.166	0.86784
congruency3:character type2	-22.3062	5.1895	-4.298 1	.72e-05 ***
congruency2:character type3	16.6671	5.6922	2.928	0.00341 **
congruency3:character type3	-34.1529	5.5108	-6.197	5.74e-10 ***

Model Fit		
AIC	BIC	Log Likelihood
206585.0	206685.0	-103279.5

Family: inverse.gaussian (identity)
Model: RT ~ congruency * character type + colour + (1 | subject)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Note:
congruency2: neutral vs. congruent (i.e. Stroop facilitation effects)
congruency3: neutral vs. incongruent (i.e. Stroop interference effects)
character type2: Colour-Character vs. Valid-Radical
character type3: Colour-Character vs. Invalid-Radical
colour1: colour cyan vs. mean
colour2: colour red vs. mean
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Appendix L: Final model for Experiment 4

Random Effects				
Group	Name	Variance	SD	Correlation
subject	(Intercept)	5.927e+02	24.34448	
Number of observations: 29499, groups: subject, 40				

Fixed Effects				
	Estimates	SE	t	p
(Intercept)	602.556	5.769	104.454	< 2e-16 ***
condition2	-11.777	2.568	-4.585	4.53e-06 ***
condition3	68.808	2.886	23.842	< 2e-16 ***
condition4	4.732	2.638	1.794	0.0729 .
condition5	27.333	2.825	9.674	< 2e-16 ***
condition6	-12.645	2.586	-4.89	1.01e-06 ***
condition7	35.04	2.83	12.38	< 2e-16 ***
condition8	-4.109	2.871	-1.431	0.1523
mode1	11.873	1.633	7.273	3.52e-13 ***
colour1	-12.446	1.133	-10.981	< 2e-16 ***
colour2	22.917	1.184	19.349	< 2e-16 ***
condition2:mode1	-20.239	3.707	-5.46	4.77e-08 ***
condition3:mode1	35.823	4.395	8.151	3.60e-16 ***
condition4:mode1	8.222	3.504	2.347	0.0189 *
condition5:mode1	8.846	3.991	2.217	0.0266 *
condition6:mode1	-21.817	3.835	-5.689	1.28e-08 ***
condition7:mode1	26.097	3.596	7.257	3.95e-13 ***
condition8:mode1	-8.022	4.524	-1.773	0.0762 .

Model Fit		
AIC	BIC	Log Likelihood
375557.8	375723.6	-187758.9

Family: inverse.gaussian (identity)
Model: RT ~ condition * mode + colour + (1 | subject)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Note:
condition2: neutral word vs. congruent (i.e. Stroop facilitation effects)
condition3: neutral word vs. incongruent (i.e. Stroop interference effects)
condition4: neutral word vs. colour-associated-congruent
condition5: neutral word vs. colour-associated-incongruent
condition6: neutral word vs. homophone-congruent
condition7: neutral word vs. homophone-incongruent
condition8: neutral word vs. neutral sign
mode1: manual vs. vocal
colour1: colour blue vs. mean
colour2: colour green vs. mean
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Appendix M: Final model for Experiment 5

Random Effects				
Group	Name	Variance	SD	Correlation
subject	(Intercept)	1.823e+03	42.6929	
Number of observations: 8913, groups: subject, 22				

Fixed Effects				
	Estimates	SE	t	p
(Intercept)	666.294	13.199	50.481	< 2e-16 ***
condition2	-26.4738	4.5829	-5.777	7.62e-09 ***
condition3	16.843	4.8886	3.445	0.000570 ***
condition4	-16.6506	4.6984	-3.544	0.000394 ***
tone1	0.6071	3.4716	0.175	0.861181
colour1	-33.3332	2.3804	-14.003	< 2e-16 ***
colour2	39.2669	2.6079	15.057	< 2e-16 ***
condition2:tone1	1.3498	7.2184	0.187	0.85166
condition3:tone1	14.4426	7.7017	1.875	0.060757 .
condition4:tone1	8.233	7.6387	1.078	0.281122

Model Fit		
AIC	BIC	Log Likelihood
116494.6	116579.7	-58235.3

Family: inverse.gaussian (identity)
Model: RT ~ condition * tone + colour + (1 | subject)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Note:
condition2: neutral word vs. congruent (i.e. Stroop facilitation effects)
condition3: neutral word vs. incongruent (i.e. Stroop interference effects)
condition4: neutral word vs. neutral sign
tone1: without tone vs. with tone
colour1: colour blue vs. mean
colour2: colour green vs. mean
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Appendix N: Final model for Experiment 6

Random Effects				
Group	Name	Variance	SD	Correlation
subject	(Intercept)	1.102e+03	33.20132	
Number of observations: 13318, groups: subject, 32				

Fixed Effects				
	Estimates	SE	t	p
(Intercept)	638.614	8.992	71.022	< 2e-16 ***
condition2	-27.964	4.562	-6.129	8.82e-10 ***
condition3	-2.136	4.656	-0.459	0.646
condition4	-21.052	4.267	-4.933	8.09e-07 ***
condition5	-1.361	4.591	-0.296	0.767
condition6	-24.202	3.761	-6.436	1.23e-10 ***
colour1	-19.632	1.923	-10.209	< 2e-16 ***
colour2	36.723	2.074	17.708	< 2e-16 ***

Model Fit		
AIC	BIC	Log Likelihood
172802.4	172877.3	-86391.2

Family: inverse.gaussian (identity)
Model: RT ~ condition + colour + (1 | subject)
Note:
condition2: neutral word vs. congruent (i.e. Stroop facilitation effects)
condition3: neutral word vs. incongruent (i.e. Stroop interference effects)
condition4: neutral word vs. colour-associated-congruent
condition5: neutral word vs. colour-associated-incongruent
condition6: neutral word vs. neutral sign
colour1: colour blue vs. mean
colour2: colour green vs. mean
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Appendix O: Final model for Experiment 7

Random Effects				
Group	Name	Variance	SD	Correlation
subject	(Intercept)	1.555e+03	39.43505	
Number of observations: 15205, groups: subject, 37				

Fixed Effects				
	Estimates	SE	t	p
(Intercept)	665.338	10.245	64.946	< 2e-16 ***
condition2	-17.435	4.156	-4.195	2.73e-05 ***
condition3	9.113	4.419	2.062	0.03916 *
condition4	-8.414	4.035	-2.085	0.03703 *
condition5	5.6	4.446	1.26	0.20783
condition6	-10.351	3.472	-2.981	0.00287 **
colour1	25.661	1.903	13.487	< 2e-16 ***
colour2	-14.207	1.785	-7.96	1.72e-15 ***

Model Fit		
AIC	BIC	Log Likelihood
197646.1	197722.4	-98813.0

Family: inverse.gaussian (identity)
Model: RT ~ condition + colour + (1 | subject)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Note:
condition2: neutral word vs. congruent (i.e. Stroop facilitation effects)
condition3: neutral word vs. incongruent (i.e. Stroop interference effects)
condition4: neutral word vs. invalid-radical-congruent
condition5: neutral word vs. invalid-radical-incongruent
condition6: neutral word vs. neutral sign
colour1: colour cyan vs. mean
colour2: colour red vs. mean
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Appendix P: Final model for Experiment 8

Random Effects						
Group	Name	Variance	SD	Correlation		
subject	(Intercept)	1.52E+03	38.98092			
condition2		9.88E+01	9.94014	-0.05		
condition3		3.97E+02	19.91165	0.13	-0.37	
condition4		1.09E+02	10.42547	0.07	0.43	-0.42
Number of observations: 12391, groups: subject, 30						

Fixed Effects				
	Estimates	SE	t	p
(Intercept)	677.206	10.625	63.736	< 2e-16 ***
condition2	-14.555	4.317	-3.372	0.000746 ***
condition3	33.447	6.334	5.281	1.29e-07 ***
condition4	-1.672	4.667	-0.358	0.720103
script1	-7.541	2.73	-2.762	0.005743 **
colour1	-32.247	1.91	-16.882	< 2e-16 ***
colour2	34.956	2.089	16.73	< 2e-16 ***
condition2:script1	-2.234	5.562	-0.402	0.687961
condition3:script1	-28.093	6.33	-4.438	9.07e-06 ***
condition4:script1	-7.973	5.448	-1.463	0.14339

Model Fit		
AIC	BIC	Log Likelihood
160967.3	161123.2	-80462.6

Family: inverse.gaussian (identity)
Model: RT ~ condition * script + colour + (1 + condition| subject)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Note:
condition2: neutral word vs. congruent (i.e. Stroop facilitation effects)
condition3: neutral word vs. incongruent (i.e. Stroop interference effects)
condition4: neutral word vs. neutral sign
script1: character vs. pinyin
colour1: colour blue vs. mean
colour2: colour green vs. mean
Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; 0.05 < . < 0.1

Appendix Q: The impact of COVID-19

Due to the impact of COVID-19, Experiments 5 to 8 were conducted online, and therefore data was collected for manual responses only. Effects that have not been observed in the manual responses may be present if vocal responses are available. In addition, with the inclusion of vocal responses, response modality effects and their relation to distinct conflict/facilitation components in Chinese characters written in pinyin can be investigated and compared to the results of Chinese characters in Experiment 4.