



# Kent Academic Repository

Qiu, Yicheng and van Heuven, Walter J. B. (2024) *Distinct Components of Stroop Interference and Facilitation: The role of phonology and response modality*. Quarterly Journal of Experimental Psychology . ISSN 1747-0226.

## Downloaded from

<https://kar.kent.ac.uk/107889/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.1177/17470218241302490>

## This document version

Publisher pdf

## DOI for this version

## Licence for this version

CC BY-NC (Attribution-NonCommercial)

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

### Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

# Distinct components of Stroop interference and facilitation: The role of phonology and response modality



Yicheng Qiu<sup>1,2</sup>  and Walter JB van Heuven<sup>1</sup> 

## Abstract

Multi-stage accounts of Stroop effects suggest that Stroop effects result from different conflict and facilitation components. Consistent with these accounts, Augustinova et al. reported evidence for task, semantic, and response components in Stroop effects. They also investigated how vocal and manual responses impacted the magnitude of each of the conflict and facilitation components. However, the role of phonological components in Stroop effects was not investigated in their study. The impact of phonology on Stroop effects has been observed in several studies. However, these studies did not investigate the role of different conflict/facilitation components in Stroop effects. To investigate the impact of phonological components as well as task, semantic, and response components on Stroop effects, a vocal and manual Stroop task was for the first time conducted with Chinese speakers using a design similar to that of Augustinova et al. The data revealed only in the vocal Stroop task phonological conflict and facilitation, whereas semantic and response conflicts were found with vocal and manual responses. Implications of the findings for response modality effects and the measures of facilitation/conflict components are discussed.

## Keywords

Stroop interference and facilitation; response modality; phonological conflict; response conflict; semantic conflict; task conflict

Received: 18 December 2023; revised: 19 September 2024; accepted: 23 September 2024

In Stroop's (1935) original experiments, participants had to name aloud the ink colour of incongruent colour words (e.g., red in green colour) and neutral signs in different colours. The neutral signs were used as a control condition. The difference in reaction times (RTs) between those two conditions in this vocal Stroop task has been referred to as Stroop interference. The study by Dalrymple-Alford and Budayr (1966) was the first to include congruent colour words in the Stroop task. The difference in response times between congruent colour word trials and neutral trials is referred to as Stroop facilitation. Rather than responding in the Stroop task by naming the ink colour (vocal Stroop task), participants can also respond to the ink colour by pressing the corresponding keys (e.g., pressing the B key for the colour *RED*; manual Stroop task). Interestingly, it has been shown that the way you respond in the Stroop task (vocal or manual) modifies Stroop effects (White, 1969). This is known as the response modality effect.

The response modality effect in terms of larger Stroop interference or facilitation effects in vocal responses than in manual responses has been reported in many studies (Augustinova et al., 2019; Fennell & Ratcliff, 2019; Neill, 1977; Redding & Gerjets, 1977; Sharma & McKenna, 1998; Zahedi et al., 2019). To explain this response modality effect, Kinoshita et al. (2017) argued that vocal and manual response modes in 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 also be understood in terms of the

Quarterly Journal of Experimental Psychology  
1–19

© Experimental Psychology Society 2024



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/17470218241302490  
qjep.sagepub.com



<sup>1</sup>School of Psychology, University of Nottingham, Nottingham, UK

<sup>2</sup>School of Psychology, University of Kent, Canterbury, UK

## Corresponding author:

Yicheng Qiu, School of Psychology, University of Kent, Canterbury CT27F7, UK.

Email: y.qiu@kent.ac.uk

broader discussion of stimulus-response (S-R) compatibility (Kornblum, 1992). In a vocal Stroop task, the stimulus and the response are directly connected (e.g., *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 S-R mappings that associate the colour of words to the corresponding response keys.

The response modality effect is crucial for understanding the mechanisms behind Stroop effects. Single-stage accounts of Stroop effects (Klein, 1964; Roelofs, 2003) explain the competition as exclusively taken place at the response stage. 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 are not influencing the task. Furthermore, according to Roelofs’s interpretation, 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. When the semantic relationship between words and colours becomes closer, it is more difficult to name the colour words (i.e., stronger interference effect). For example, colour-associated words like *SKY* that are associated with a particular colour word *BLUE* took longer to process than neutral words that are not associated with a specific colour like *SEAT*. Klein argued that the semantic gradient determines the amount of response competition (or response conflict) that he viewed as the unique driving force behind the Stroop effect. However, Schmidt and Cheesman’s (2005) data suggest that incongruent colour-associated words do not result in response conflict. Furthermore, although the meaning of *SKY* can trigger a vocal response other than colour blue, this does not belong to the response set (e.g., blue, green, yellow) in a colour-naming Stroop task. Based on the work of Klein (1964) and Neely and Kahan (2001), Augustinova et al. (2018) used the *semantic Stroop paradigm* to investigate Stroop interference effect as well as response, semantic, and task conflicts.

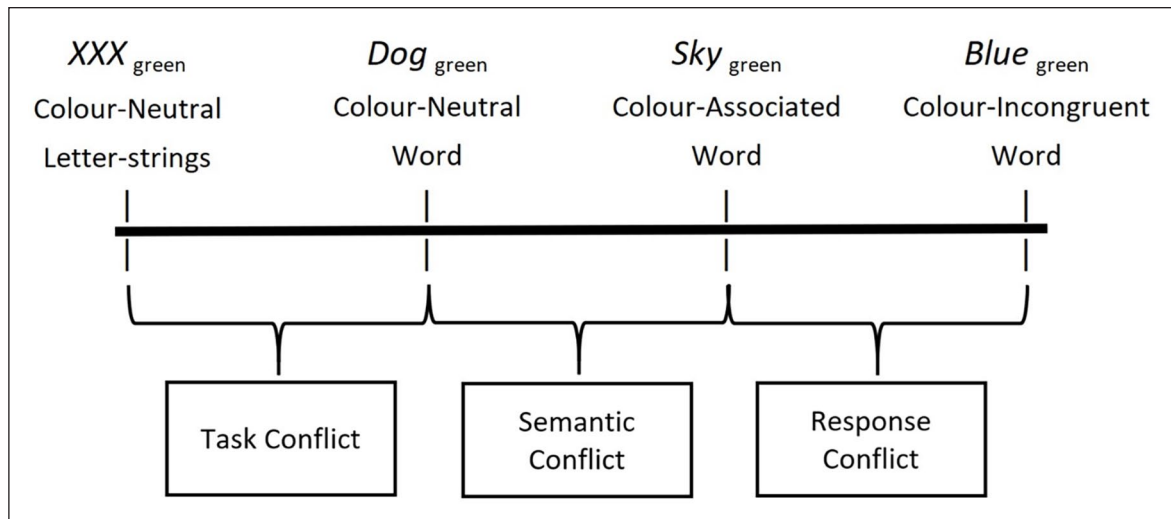
Task conflict originated from the view that word reading is automatic, whereas ink colour naming is not (see MacLeod, 1991 for a detailed review of the automaticity

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 produces task conflict (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 (MacLeod & Dunbar, 1988). However, it has also been argued that word reading may not be automatic (Besner et al., 1997). Kinoshita et al. (2018) argued that the automaticity of word reading cannot be manipulated by endogenous control. They manipulated the proportion of non-readable neutral trials (a row of #’s) to change the level of attentional control. However, this manipulation did not affect the magnitude of the semantic Stroop effect (i.e., comparing incongruent colour-associated words with neutral trials).

To measure the magnitude of task conflict, Augustinova et al. (2018) contrasted colour-neutral stimuli with a row of X’s (e.g., XXXX). Neutral words were expected to trigger word reading, whereas a row of X’s is unpronounceable and meaningless and would therefore not trigger word reading (see Figure 1). Task conflict can also be explained as the competition between whole task sets, which is in addition to any specific competition between stimuli and responses (S-R associations, Monsell et al., 2001; Parris et al., 2023). Thus, task sets refer to the whole set of S-R associations.

As illustrated in Figure 1, semantic conflict can be measured by subtracting the RTs of incongruent colour-associated words (e.g., *SKY*<sub>green</sub>, word *SKY* written in green ink) from neutral words (e.g., *DOG*<sub>green</sub>). Semantic conflict is due to two competing incompatible semantic representations activated by word and the ink. Augustinova et al. (2019) argued that colour-associated words do not activate (pre-)motor responses linked to the associated colour because the word *SKY* is not part of the response set, whereas the word *BLUE* activates (pre-)motor responses to *BLUE*. Therefore, the difference between incongruent colour words and incongruent colour-associated words can be used to measure response conflict, which refers to competition between two possible motor responses.

Augustinova et al. (2018) observed strong semantic and response conflicts in both the manual and vocal Stroop tasks, whereas task conflict occurred only in the vocal Stroop task. However, response modality was a between-subject factor in their study. To compare those conflict components in the manual and vocal Stroop tasks, Augustinova et al. (2019) used a within-subject design in their second experiment, which revealed that response and semantic conflicts were present in both the vocal and manual Stroop tasks. Task conflict, 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



**Figure 1.** The subtractive logic of Stroop interference effects (adapted from the study by Augustinova et al., 2018). The colour word in subscript that follows the stimulus written in italics indicates the ink colour of the stimulus (e.g., *DOG*<sub>green</sub> refers to the word *DOG* written in green ink). The task is to identify the ink colour either orally or manually.

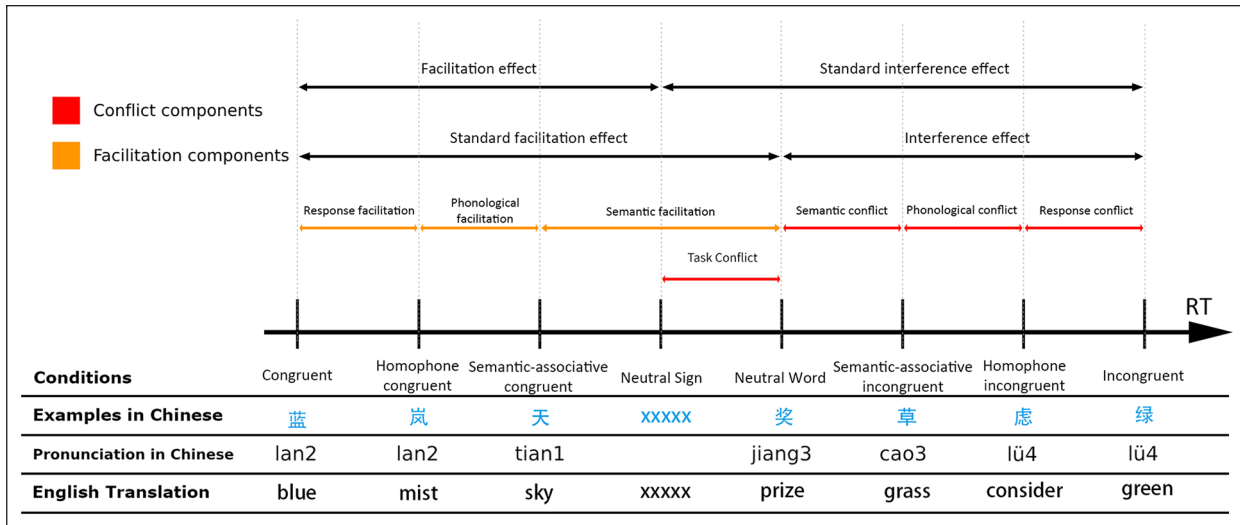
response and task conflicts. In contrast, semantic conflict was similar in both response modalities, so they did not contribute to the response modality effect. These findings suggested that Stroop effects are not solely due to response competition as suggested by Roelofs (2003).

In addition to semantic, response, and task conflicts, the role of phonological conflict in the Stroop task has also been studied. Besner and Stolz (1998) included pseudo-homophones of colour words: *WRED*, *BLOO*, *YELEO*, *GRENE* in a manual Stroop task. The results revealed significant Stroop interference effects for pseudo-homophones relative to the baseline (xxxx) and neutral words, indicating that the 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 positions 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, the study by Parris et al. (2019) did not investigate the role of phonology in Stroop interference and how it is potentially modulated by response modality. Parris et al. (2023) found phonological facilitation with manual responses. Even though simple-onset words (e.g., *RED/PURPLE*, words with a single consonant followed by a vowel) and complex-onset words (e.g., *GREEN/BLUE*, words with two or more consonants) share the same level of orthographic similarity, stronger facilitation effects were observed with complex-onset words than simple ones. This shows the effect of onset complexity on Stroop facilitation.

Parris et al. (2022) argued that measuring phonological conflicts using pseudo-homophones 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 conflict. However, there are only a few heterographic homophones or pseudo-homophones in English and other alphabetic languages. In contrast, there are abundant homophones in other languages. For example, Chinese has many homophones of the colour word green “绿” (pronounced as /lǜ/), which have completely different meanings and orthography: “虑” (consider), “律” (restrain), “率” (rate), “滤” (filter), and so on. Thus, in a language with plenty of homophones, such as Chinese, phonological conflict in the Stroop task can be studied without the confound of orthography.

The Stroop paradigm used by Augustinova et al. (2018) can be extended with additional conditions to also investigate effects of phonology. Figure 2 illustrates Stroop conditions based on Chinese stimuli that can be utilised 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 conflict can be measured by the RT difference between neutral words and neutral signs because 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 (see Dalrymple-Alford, 1972, who first showed facilitation with congruent colour-associated words). The RT difference between colour-associated words and homophones in Chinese is assumed to measure phonological activation resulting for incongruent conditions in phonological conflict and for congruent conditions in phonological facilitation. In Chinese,



**Figure 2.** Decomposed Stroop conflict/facilitation components with Chinese characters. The red arrows refer to conflict components, and the orange arrows refer to facilitation components.

homophones of colour words activate the phonology of colour words. However, this phonological activation is assumed to also activate the semantics of colour words because otherwise there would be no impact on Stroop performance. Therefore, this is indicated as properties of the conditions (see Figure 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 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. (2018, 2019). However, unlike Augustinova et al., Spinks et al. did not distinguish between task, response, semantic, and phonological conflicts. The impact of phonology in the manual Stroop task with Chinese characters was also investigated in an Event-Related Potentials (ERPs) study by Wang et al. (2010). The stimuli in this study consisted of colour characters, colour-associated characters, homophones, and neutral characters. The size of each Stroop component was estimated using the means provided in the studies by Spinks et al. (2000) and Wang et al. (2010) because the data in these studies were not analysed in terms of Stroop conflict and facilitation components. The numerical differences suggest phonological facilitation with both vocal and manual responses (12 and 36 ms) with Chinese homophones, but no phonological conflict (vocal Stroop:  $-1$  ms, manual Stroop:  $-7$  ms).

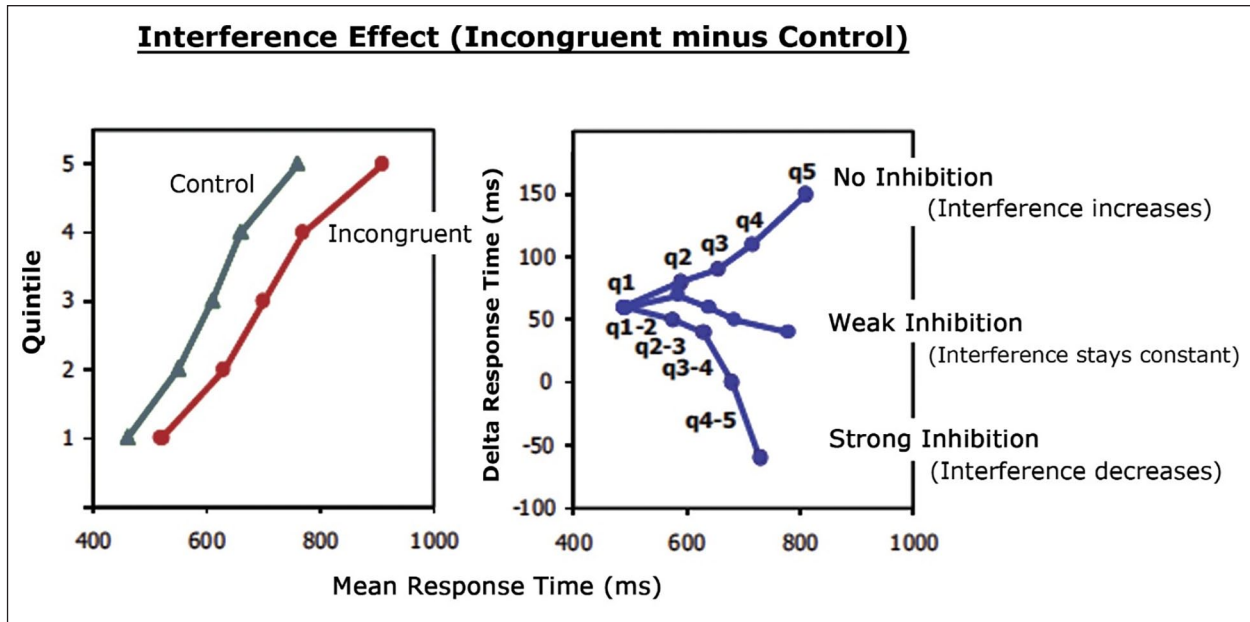
Stroop studies with Chinese stimuli have explored phonological components using homophones (Spinks et al., 2000; Wang et al., 2010). However, these studies did not

decompose Stroop effects into distinct components; therefore, the magnitude of those phonological components remains unknown. Most importantly, as far as we are aware, no Stroop studies with Chinese have used a within-subject design to study vocal and manual responses in the Stroop task.

## Present study

This study investigates for the first time the role of phonological conflict and phonological facilitation in a vocal and manual Stroop task with Chinese stimuli. The findings will be compared with those of Augustinova et al. (2019) who investigated task, semantic, and response conflicts in an alphabetic language (French). Semantic conflict as measured by subtracting responses to incongruent colour associates from colour-neutral words was not affected by response modality in Augustinova et al. (2019). However, task conflict and response conflict measured in the same study resulted in a smaller effect for manual responses than for vocal responses (i.e., response modality effect). Task and response conflicts in the current study are expected to be similar to those in the study by Augustinova et al. (2019) because these are not expected to be language specific. Phonological conflict as measured by the difference between responses to incongruent homophones and those to incongruent colour associates are expected to be larger in vocal than in manual responses because vocal responses require the explicit activation of phonological representations. Augustinova et al. (2019) did not find that semantic facilitation was affected by response modality, but response facilitation was affected; therefore, a similar pattern is expected in the present study.

Data will be analysed using a mean RT analysis and a distributional analysis. The mean RT analysis will be



**Figure 3.** Quantile plot (left) and delta plot (right). The term “q1” relates to Quintile 1 and so forth; q1–2 is the segment connecting quintiles 1 and 2, etc.

Source. Adapted from Roelofs et al. (2011).

conducted using generalised linear mixed-effect models (GLMMs). An advantage of using GLMMs rather than linear mixed-effect models (LMMs) is that it allows assessing raw RTs without transforming the data (Lo & Andrews, 2015) and selecting a distribution (e.g., Gamma or Inverse Gaussian) that fits raw RTs better than the normal (Gaussian) distribution used in LMMs.

Distributional analysis can capture the dynamics of attentional control that are likely to be lost when analysing only the mean RTs. 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). Delta plots will be created, which are useful for investigating the impact of experimental manipulations across the entire RT distribution (Ridderinkhof et al., 2005; Roelofs et al., 2011) and reflect “the effect of an experimental factor tends to increase as a function of RT” (Roelofs et al., 2011, p. 2). The effects are stronger for long RTs than for short RTs. They can be directly derived from cumulative density functions (i.e., quantile plots). Figure 3 shows the possible interpretation of reaction distribution effects in the Stroop task by Roelofs et al. (2011). Inhibition refers to active/willed inhibition as defined by Aron (2007), which suppresses the irrelevant response, stimulus, or memory. In the vocal Stroop task, inhibition takes place when word information is 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. When this rationale is applied to the Stroop interference effect, an upwards

trend in the delta plot shows that interference increases with slower RTs, 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, while a downwards trend shows strong inhibition in cognitive control leads to decrease in interference.

Positive slopes in delta plots are often reported in the Stroop task (Bub et al., 2006; Pratte et al., 2010; Roelofs et al., 2011; Scaltritti et al., 2022), which are interpreted in terms of the absence of inhibition to resolve the conflicts between colour and word information. Several studies (Labuschagne & Besner, 2015; Scaltritti et al., 2022; Sulpizio et al., 2022, 2024) reported positive trends in a semantic Stroop task (comparing colour-associated words to neutral words, which measures semantic conflict in the same way as Augustinova & Ferrand, 2014). Hasshim et al. (2019) examined both response and semantic conflicts using distributional analyses and found that response conflict takes place earlier than semantic conflict. Martinon et al. (2024) also observed an absence of semantic conflict in faster responses. Here we investigate the contribution of phonological components, measured using Chinese homophones, along with response, semantic, and task components in both Stroop interference and facilitation. By separating the effects of each component in the delta plot, it is possible to determine which component has largest influence on the overall Stroop effect. Based on previous literature, we predict that semantic conflict would show a positive linear trend in both response modalities. Phonological conflict on the

**Table 1.** Stimuli used in this study.

Condition	Character	Meaning	Pronunciation
Colour words	绿	green	/lǜ/
	黄	yellow	/huáng/
	蓝	blue	/lán/
Associated Colour words	草	grass	/cǎo/
	金	golden	/jīn/
	天	sky	/tiān/
Homophone words	虑	ponder	/lǜ/
	皇	emperor	/huáng/
	岚	mist	/lán/
Colour-Neutral words	贯	pass through	/guàn/
	奖	prize	/jiǎng/
	炭	charcoal	/tàn/

other hand is expected to contribute to Stroop interference only in vocal responses.

## Methods

### 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.

### Stimuli and design

The stimuli were adapted from the study by Spinks et al. (2000). 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. Although the number of colours used reduced from four to three, it falls within the recommended number of words/colours (between three to five) by MacLeod (2005). The control character 华 (/huá/, magnificent) was changed to 炭 (/tàn/, 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,” /cǎo/, “grass” is associated with the colour green). The homophones condition consisted of characters with the same pronunciation as colour words (e.g., “虑,” /lǜ/, “consider” is pronounced the same as “绿,” /lǜ/, “green” but has a completely different meaning). Table 1 presents the stimuli used in this study. Detailed linguistic properties of these stimuli can be found in the online Supplementary Material A.

Each character type was either congruent (e.g., *BLUE* surrounded by blue rectangle) or incongruent (e.g., *BLUE* surrounded by green rectangle). The reason for choosing separate presentation of word and colour information instead of integrated (i.e., *BLUE* presented in blue colour) was to control the amount of colour information in each stimulus, so that the number of strokes of each character would not impact the region being coloured (e.g., 草 [green] has more strokes than 天 [sky]; thus it would receive more colour information). 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)  $\times$  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  $\times$  3 colours  $\times$  2 repetitions=48 trials).

As suggested by Brysbaert and Stevens (2018), a minimum of 1,600 observations per condition is recommended for a well-powered repeated-measure study. The present experiment exceeds this recommendation because there are 1,920 observations in each condition and each response modality (40 participants  $\times$  8 blocks  $\times$  6 trials).

### 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 (楷体). 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 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 programme (Protopapas, 2007). In manual response modality, each colour (green, yellow, and blue) was associated with keys 1, 2, and 3 on the numeric keypad of the keyboard which had stickers indicating the colour. Participants were instructed to use the index finger to press the green key (1) when the ink was green, the middle finger to press the yellow key (2) when the ink was yellow, and to use the ring finger to press the blue key (3) when the ink colour of the stimulus was blue. The order of conducting the task with vocal or manual responses was counter-balanced across participants. 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 3,000 ms or until the participant responded. Characters disappeared as soon as the participant responded. There was an intertrial interval of 1,000 ms. For the vocal Stroop task, participants first conducted 32 practice trials. For the manual Stroop task, 128 neutral signs were used as key-matching practice trials to train participants in the key-colour correspondences (Augustinova et al., 2019; MacLeod, 2005).

## Results

Incorrect responses were discarded (3.27%) for the response time analyses. RTs below 200 ms and above 1,500 ms were also discarded before the analyses (0.73%).

Correct responses were then analysed using GLMMs analyses in R Statistical Software (v4.3.2; R Core Team, 2021) using the packages lme4 (v1.1-35.1; Bates et al., 2015), emmeans (v1.8.9; Lenth, 2023), and lmerTest (v3.1-3; Kuznetsova et al., 2017). The fixed-factor stimulus type was coded using simple coding with the neutral trials as the baseline. The fixed-factor response modality (manual vs. vocal) was coded using sum coding (-0.5, 0.5). Colour was entered as a covariate to the model, to account for addition variance, and not as a random factor because the number of levels (3: green, yellow, blue) is below 5 (see Bolker, 2015; Bolker et al., 2009). Colour was coded using sum coding.

First, three models with no fixed structure were constructed using three different types of distributions (Gaussian, Inverse Gaussian, and Gamma). The models were compared in *R* (v4.3.2; R Core Team, 2021) using the compare\_performance function provided by the Performance package (v0.10.8; Lüdtke 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), and BIC (Bayesian Information Criterion, Akaike, 1978a, 1978b; Schwarz, 1978) indices. All indices are normalised, and the mean value of all indices in each model was calculated for the performance score. Results showed that a model based on an Inverse Gaussian distribution had the highest performance scores among the three.

Next, a full GLMM was constructed using inverse gaussian distribution:

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

However, this full model did not converge, and the random structure was simplified until the model successfully converged. The final model:

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

A summary of the final model can be found in the online Supplementary Material B. Main effects of stimulus type and response modality were obtained (stimulus type:  $\chi^2(7)=913.57, p < .001$ ; response modality:  $\chi^2(1)=29.139, p < .001$ ), as well as the interaction between the two ( $\chi^2(7)=123.72, p < .001$ ). Next, planned comparisons were conducted using the emmeans package (v1.8.9; Lenth, 2023) to investigate Stroop interference and Stroop facilitation for each of the components. *p*-Values were corrected using Holm-Bonferroni correction (Holm, 1979).

Table 2 provides a summary of the descriptive statistics and the decomposition of each interference and facilitation components presented in Table 3.

### Mean RT analysis

**Stroop interference.** Strong Stroop interference effects (relative to neutral signs) occurred in both vocal and manual responses (vocal:  $M=95$  ms,  $SE=4.85$  ms,  $z=19.568, p < .001$ ; manual:  $M=51$  ms,  $SE=4.39$  ms,  $z=11.623, p < .001$ ). The distinct conflict components were all significant in both response modalities except for phonological conflict. Response conflicts were significant in both vocal ( $M=39$  ms,  $SE=4.42$  ms,  $z=8.749, p < .001$ ) and manual responses ( $M=29$  ms,  $SE=4.25$  ms,  $z=6.799, p < .001$ ). Phonological conflict was significant in vocal ( $M=16$  ms,  $SE=4.30$  ms,  $z=3.80, p < .01$ ) but not in manual responses ( $M=-1$  ms,  $SE=4.04$  ms,  $z=-0.227, p=1$ ). Semantic conflicts were significant in both vocal



**Table 2.** Mean reaction times (RT, in milliseconds), error rates (ER, %), and standard errors (SE, in parentheses) of each stimulus type in the vocal and manual tasks.

Stimulus type (examples)		Vocal	Manual
Colour-Incongruent words	RT (SE)	680 (6.57)	636 (6.54)
蓝 <sub>绿</sub> (BLUEgreen)	ER (SE)	8.28 (0.63)	5.36 (0.51)
Associated Colour-Incongruent words	RT (SE)	625 (6.41)	608 (6.62)
天 <sub>绿</sub> (SKYgreen)	ER (SE)	1.77 (0.30)	3.85 (0.44)
Homophone-Incongruent words	RT (SE)	641 (6.33)	607 (6.33)
岚 <sub>绿</sub> (MISTgreen)	ER (SE)	6.25 (0.55)	4.22 (0.46)
Colour-Neutral words	RT (SE)	593 (5.95)	585 (5.92)
奖 <sub>绿</sub> (PRIZEgreen)	ER (SE)	1.25 (0.25)	3.18 (0.40)
Colour-Neutral signs	RT (SE)	585 (6.63)	585 (6.56)
% <sub>绿</sub> (%green)	ER (SE)	1.20 (0.25)	4.06 (0.45)
Colour-Congruent words	RT (SE)	571 (6.29)	583 (6.48)
蓝 <sub>蓝</sub> (BLUEblue)	ER (SE)	0.78 (0.20)	2.81 (0.38)
Associated Colour-Congruent words	RT (SE)	602 (6.47)	586 (6.34)
天 <sub>蓝</sub> (SKYblue)	ER (SE)	1.77 (0.30)	2.92 (0.38)
Homophone-Congruent words	RT (SE)	570 (6.47)	583 (6.22)
岚 <sub>蓝</sub> (MISTblue)	ER (SE)	0.52 (0.16)	4.06 (0.45)

( $M=32$  ms,  $SE=3.55$  ms,  $z=8.934$ ,  $p<.001$ ) and manual responses ( $M=23$  ms,  $SE=3.36$  ms,  $z=6.817$ ,  $p<.001$ ). Task conflicts were not significant either in vocal or manual responses (vocal:  $M=8$  ms,  $SE=3.82$  ms,  $z=2.126$ ,  $p=.436$ ; manual:  $M=0$  ms,  $SE=3.48$  ms,  $z=0.028$ ,  $p=1$ ).

The overall Stroop Interference (relative to neutral signs) was larger in vocal than in the manual Stroop tasks ( $M=44$  ms,  $SE=6.07$  ms,  $z=7.219$ ,  $p<.001$ ). Further analyses revealed that phonological conflict was larger in vocal than in manual responses ( $M=17$  ms,  $SE=5.32$  ms,  $z=3.244$ ,  $p<.05$ ), but not in other conflict components (response conflict:  $M=10$  ms,  $SE=5.39$  ms,  $z=1.806$ ,  $p=.284$ ; semantic conflict:  $M=9$  ms,  $SE=3.99$  ms,  $z=2.217$ ,  $p=.15$ ; task conflict:  $M=8$  ms,  $SE=4.52$  ms,  $z=1.773$ ,  $p=.284$ ).

**Stroop facilitation.** Stroop facilitation effects (relative to neutral words) were significant in the vocal Stroop task ( $M=22$  ms,  $SE=3.16$  ms,  $z=6.925$ ,  $p<.001$ ), 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,  $z=8.614$ ,  $p<.001$ ). Semantic facilitation showed a trend for a negative effect in the vocal Stroop task,  $M=-9$  ms,  $SE=3.23$  ms,  $z=-2.736$ ,  $p=.087$ .

Overall, Stroop facilitation (relative to neutral words) was larger in vocal than in manual Stroop tasks ( $M=20$  ms,  $SE=3.71$  ms,  $z=5.46$ ,  $p<.001$ ). Further analysis of each facilitation component revealed that phonological

facilitation was significantly larger in vocal than in manual responses ( $M=30$  ms,  $SE=4.81$  ms,  $z=6.246$ ,  $p<.001$ ).

**Error analysis.** The error analysis started with the full model, which failed to converge:

```
glmer(Error ~ stimulus_type * response_modality + colour
+ (1 + stimulus_type * response_modality|subject), family
= binomial, control=glmerControl(optimizer="bobyq", opt
Ctrl=list(maxfun=2e5)))
```

The random structure of the full model was reduced until it successfully converged:

```
glmer(Error ~ stimulus_type * response_modality + colour +
(1 |subject), family = binomial, control=glmerControl(optimizer="bobyq", optCtrl=list(maxfun=2e5)))
```

Main effects of stimulus type and response modality were obtained (stimulus type:  $\chi^2(7)=241.01$ ,  $p<.001$ ; response modality:  $\chi^2(1)=29.296$ ,  $p<.001$ ), as well as the interaction between the two ( $\chi^2(7)=149.3$ ,  $p<.001$ ). Next, planned comparisons were conducted to investigate Stroop interference and Stroop facilitation for each of the components.

**Stroop interference.** For vocal responses, error rates were significant for Stroop interference effects and phonological conflicts but not for other conflict components (Stroop interference:  $M=2.04\%$ ,  $SE=0.22\%$ ,  $z=9.09$ ,  $p<.001$ ; phonological conflict:  $M=1.33\%$ ,  $SE=0.20\%$ ,  $z=6.76$ ,  $p<.001$ ; response conflict:  $M=0.31\%$ ,  $SE=0.13\%$ ,  $z=2.47$ ,  $p=.50$ ; semantic conflict:  $M=0.36\%$ ,  $SE=0.27\%$ ,  $z=1.34$ ,  $p=1$ ; task conflict:  $M=0.04\%$ ,  $SE=0.29\%$ ,  $z=0.15$ ,  $p=1$ ). For manual responses, none of error rates were significant for Stroop interference and its conflict components (Stroop interference:  $M=0.30\%$ ,  $SE=0.15\%$ ,  $z=1.93$ ,  $p=1$ ; phonological conflict:  $M=0.10\%$ ,  $SE=0.16\%$ ,  $z=0.58$ ,  $p=1$ ; response conflict:  $M=0.26\%$ ,  $SE=0.15\%$ ,  $z=1.69$ ,  $p=1$ ; semantic conflict:  $M=0.20\%$ ,  $SE=0.18\%$ ,  $z=1.16$ ,  $p=1$ ; task conflict:  $M=0.26\%$ ,  $SE=0.17\%$ ,  $z=1.49$ ,  $p=1$ ).

**Stroop facilitation.** For vocal responses, error rates were significant for phonological facilitation but not for other facilitation components (phonological facilitation:  $M=1.24\%$ ,  $SE=0.36\%$ ,  $z=3.46$ ,  $p=.02$ ; Stroop facilitation:  $M=0.48\%$ ,  $SE=0.33\%$ ,  $z=1.46$ ,  $p=1$ ; response facilitation:  $M=0.41\%$ ,  $SE=0.41\%$ ,  $z=1.01$ ,  $p=1$ ; semantic facilitation:  $M=0.37\%$ ,  $SE=0.27\%$ ,  $z=1.34$ ,  $p=1$ ).

For manual responses, none of the error rates were significant for Stroop facilitation and facilitation components (Stroop facilitation:  $M=0.13\%$ ,  $SE=0.19\%$ ,  $z=0.67$ ,  $p=1$ ; phonological facilitation:  $M=0.35\%$ ,  $SE=0.18\%$ ,  $z=1.96$ ,  $p=1$ ; response facilitation:  $M=0.39\%$ ,  $SE=0.18\%$ ,  $z=2.15$ ,  $p=.98$ ; semantic facilitation:  $M=0.09\%$ ,  $SE=0.19\%$ ,  $z=0.48$ ,  $p=1$ ).

**Table 3.** Stroop effects (in milliseconds) observed with vocal and manual responses.

Stroop-like effects		Vocal		Manual	RME
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	17*
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***
Response Facilitation (Homophone-Congruent—Colour-Congruent)	RT diff.	-2		0	2
Phonological Facilitation (Associated-Congruent—Homophone-Congruent)	RT diff.	32***	>	2	30***
Semantic Facilitation (Colour-Neutral Words—Associated-Congruent)	RT diff.	-9 <sup>+</sup>		-1	-8

RME: response modality effect; RT diff.: reaction time differences.

\*\*\* $p < .001$ ; \*\* $p < .01$ ; \* $p < .05$ ; +  $p < .10$ .

### Distributional analysis

The RTs of each subject for each condition were rank ordered and divided into five quintiles. For each quintile in each condition, the mean RT was calculated. The delta plots for each conflict and facilitation components were obtained by computing the RT differences between each condition as described in the mean RT analysis (e.g., phonological conflict was calculated by subtracting RTs to incongruent homophones from those to incongruent colour-associated words).

**Stroop interference.** Distributional analyses were conducted to investigate the RT distribution of each conflict/facilitation component and how they contribute to the overall Stroop interference/facilitation effects across different quantiles. Delta plots and the contribution of each component towards overall Stroop interference (%) based on untrimmed data for Stroop interference effects in vocal and manual responses are presented in Figures 4 and 5.

Analyses were separated for vocal and manual responses. Then, a series of one-way analyses of variance (ANOVAs) were conducted for each conflict component in vocal and manual responses, where the dependent variable is the delta value calculating each component in each quantile (e.g., the delta value for semantic conflict is the RT difference between neutral word and incongruent colour-associated words). The independent variable is the five quantiles. To assess the trend of each component within five quantiles, the quantile variable is fitted using orthogonal polynomial contrast, which can then reflect the linear and quadratic trends for each condition.

For vocal responses, the delta plots for response, phonological, and semantic conflicts revealed positive linear trends (response conflict:  $F(1, 195) = 18.365$ ,  $p < .001$ ,  $d = 0.68$ ; phonological conflict:  $F(1, 195) = 22.228$ ,  $p < .001$ ,  $d = 0.75$ ; semantic conflict:  $F(1, 195) = 50.332$ ,  $p < .001$ ,  $d = 1.12$ ), and no quadratic trends (response conflict:  $F(1, 195) = 1.44$ ,  $p = .232$ ,  $d = 0.19$ ; phonological conflict:  $F(1, 195) = 0.01$ ,  $p = .922$ ,  $d = 0.02$ ; semantic conflict:

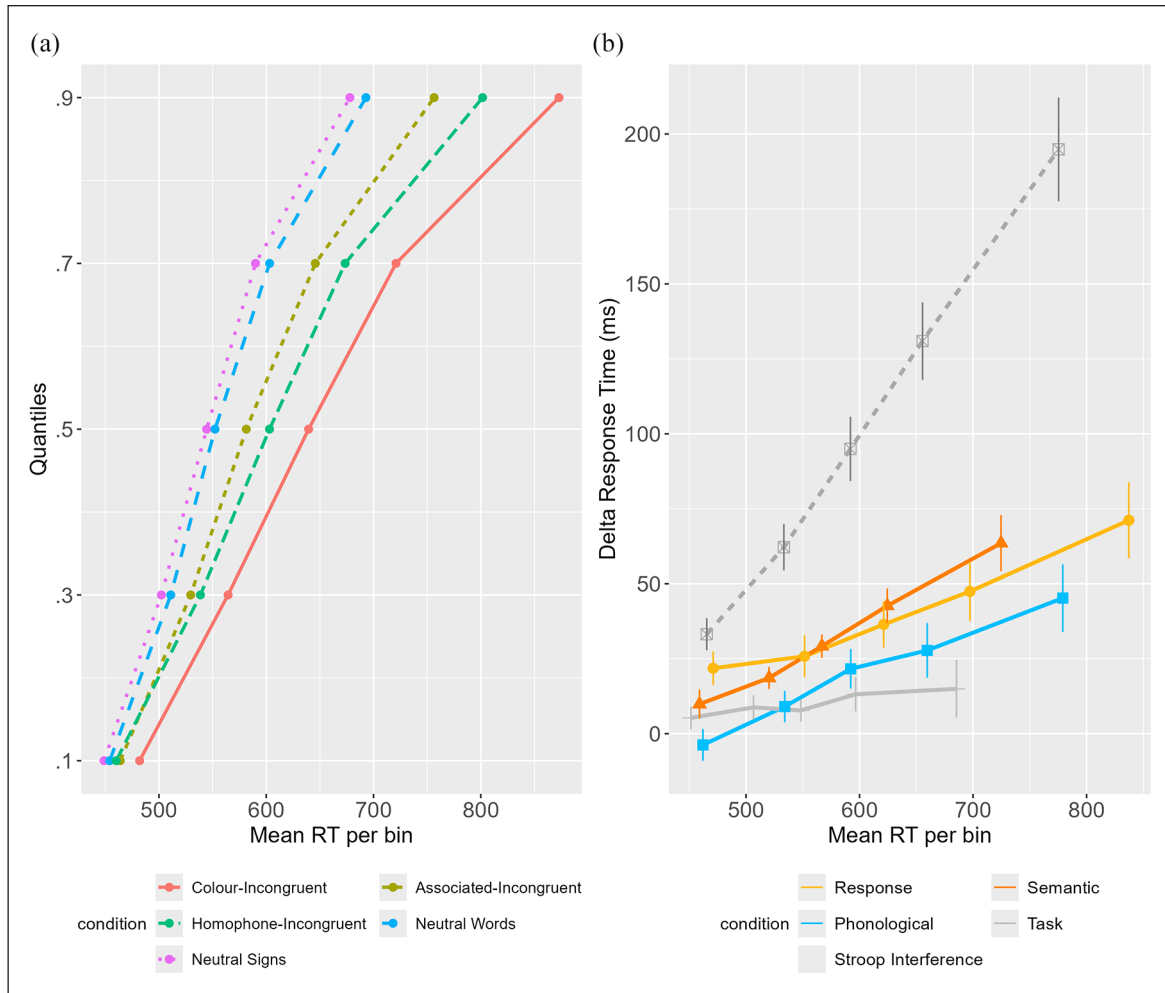
$F(1, 195) = 1.539$ ,  $p = .216$ ,  $d = 0.20$ ). The delta plot for task conflict revealed no linear component ( $F(1, 195) = 1.682$ ,  $p = .196$ ,  $d = 0.21$ ) nor a quadratic component ( $F(1, 195) = 0.02$ ,  $p = .887$ ,  $d = 0.02$ ).

For manual responses, the delta plots for semantic conflict showed positive linear trends ( $F(1, 195) = 5.537$ ,  $p = .02$ ,  $d = 0.37$ ) and no quadratic trend ( $F(1, 195) = 1.517$ ,  $p = .22$ ,  $d = -0.19$ ). The delta plot for response conflict revealed a positive linear trend ( $F(1, 195) = 26.232$ ,  $p < .001$ ,  $d = 0.81$ ), as well as a quadratic trend ( $F(1, 195) = 4.757$ ,  $p = .03$ ,  $d = 0.34$ ). The delta plot for phonological conflict revealed no linear component ( $F(1, 195) = 1.627$ ,  $p = .204$ ,  $d = 0.20$ ) and no quadratic component ( $F(1, 195) = 0.038$ ,  $p = .845$ ,  $d = 0.03$ ). The delta plot for task conflict revealed a trend towards a linear component ( $F(1, 195) = 3.237$ ,  $p = .074$ ,  $d = 0.28$ ) but no quadratic component ( $F(1, 195) = 1.185$ ,  $p = .278$ ,  $d = 0.17$ ).

**Stroop facilitation.** Delta plots and the contribution of each component towards overall Stroop facilitation (%) based on untrimmed data for Stroop facilitation effects in vocal responses are presented in Figure 6.

A series of one-way ANOVAs were performed for each facilitation component in vocal and manual responses separately. For vocal responses, the delta plot for phonological facilitation revealed a positive linear trend ( $F(1, 195) = 36.731$ ,  $p < .001$ ,  $d = 0.96$ ) and a trend towards a quadratic trend ( $F(1, 195) = 2.879$ ,  $p = .091$ ,  $d = 0.27$ ), whereas the delta plot for semantic facilitation showed a negative linear trend ( $F(1, 195) = 17.457$ ,  $p < .001$ ,  $d = -0.67$ ) and a trend towards a quadratic trend ( $F(1, 195) = 2.791$ ,  $p = .096$ ,  $d = -0.26$ ). The delta plot for response facilitation showed no linear trends ( $F(1, 195) = 2.921$ ,  $p = .089$ ,  $d = -0.27$ ) and no quadratic trends ( $F(1, 195) = 0.433$ ,  $p = .511$ ,  $d = -0.10$ ).

For manual responses, the delta plots for all facilitation components revealed no linear trends (semantic facilitation:  $F(1, 195) = 1.003$ ,  $p = .318$ ,  $d = 0.16$ ; phonological facilitation:  $F(1, 195) = 0.464$ ,  $p = .496$ ,  $d = 0.11$ ; response facilitation:  $F(1, 195) = 0.169$ ,  $p = .682$ ,  $d = -0.06$ ), and no



**Figure 4.** Quantile plots (a) and delta plots (b) for Stroop interference effects in vocal responses. Error bar stands for standard error (SE) for each quantile.

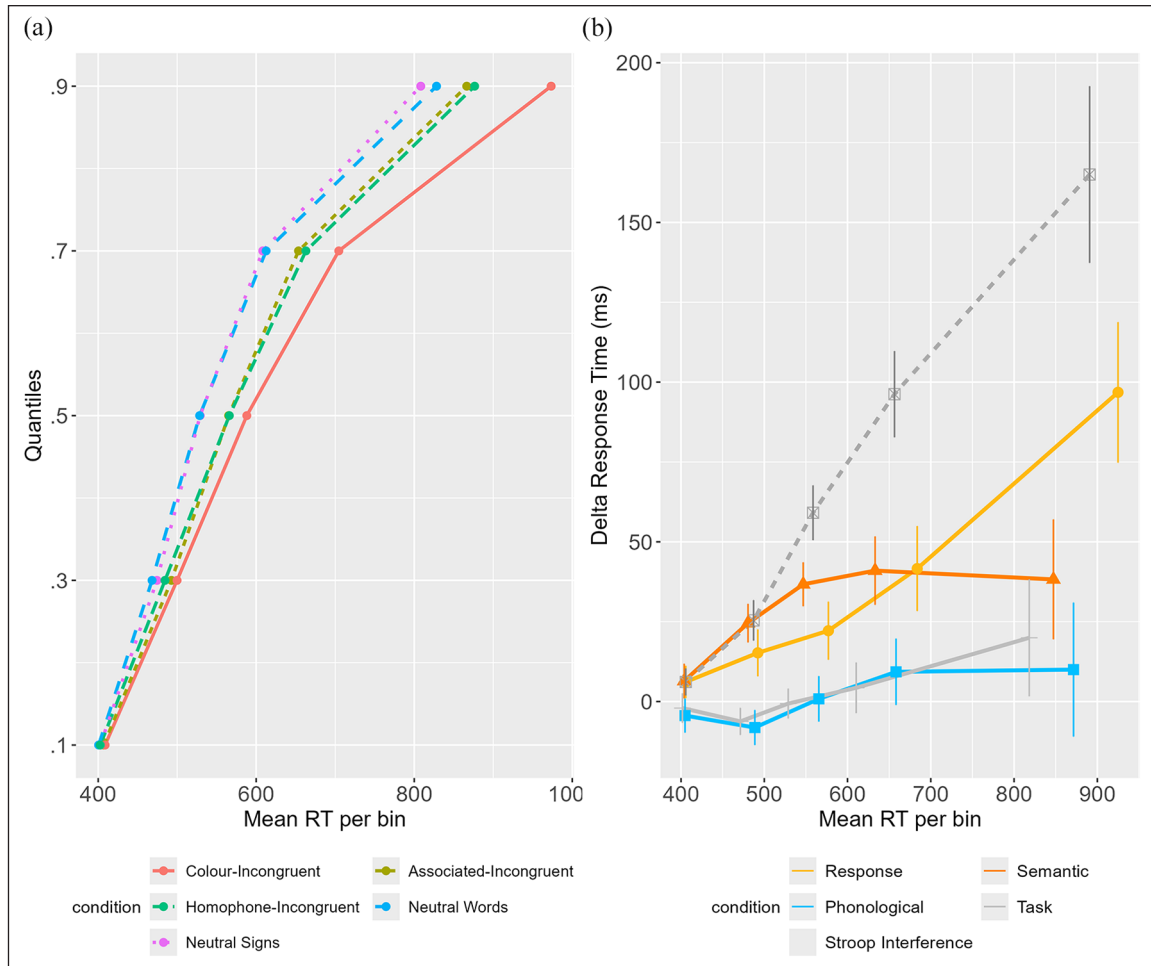
quadratic trends (semantic facilitation:  $F(1,195)=1.556$ ,  $p=.214$ ,  $d=0.20$ ; phonological facilitation:  $F(1,195)=0.01$ ,  $p=.909$ ,  $d=0.02$ ; response facilitation:  $F(1,195)=0.002$ ,  $p=.963$ ,  $d=-0.01$ ).

## Discussion

The present study investigated the distinct components of Stroop interference and facilitation using Chinese characters with both vocal and manual responses. Overall, vocal responses resulted in stronger Stroop interference than manual responses. This response modality effect is consistent with what has been reported in the literature (e.g., Augustinova et al., 2019; Fennell & Ratcliff, 2019; Neill, 1977; Redding & Gerjets, 1977; Sharma & McKenna, 1998; Zahedi et al., 2019). In contrast, Stroop facilitation was found in vocal responses but not in manual responses, which is inconsistent with the findings in French reported by Augustinova et al. (2019). The absence of Stroop facilitation effects in the current study could be due to the

separate presentation of the word and colour information compared to an integrated presentation (word written using congruent or incongruent ink colour), as MacLeod (1998) reported smaller Stroop interference and facilitation when using separate presentation instead of an integrated presentation.

In terms of the distinct components of Stroop interference, the data revealed semantic and response conflicts in both response modalities, whereas phonological conflict was only found with vocal responses. Surprisingly, task conflict, as measured by a response time difference between neutral signs (triggering only a single task: colour identification) and neutral words (triggering two possible tasks: word reading and colour identification), was not found in both the vocal and manual Stroop task. There was also no evidence of semantic and response facilitation in both response modalities. Importantly, phonological facilitation was found with vocal responses but not with manual responses. Thus, the present study found evidence for distinct components of Stroop interference and facilitation



**Figure 5.** Quantile plots (a) and delta plots (b) for Stroop interference effects in manual responses. Error bar stands for standard error (SE) for each quantile.

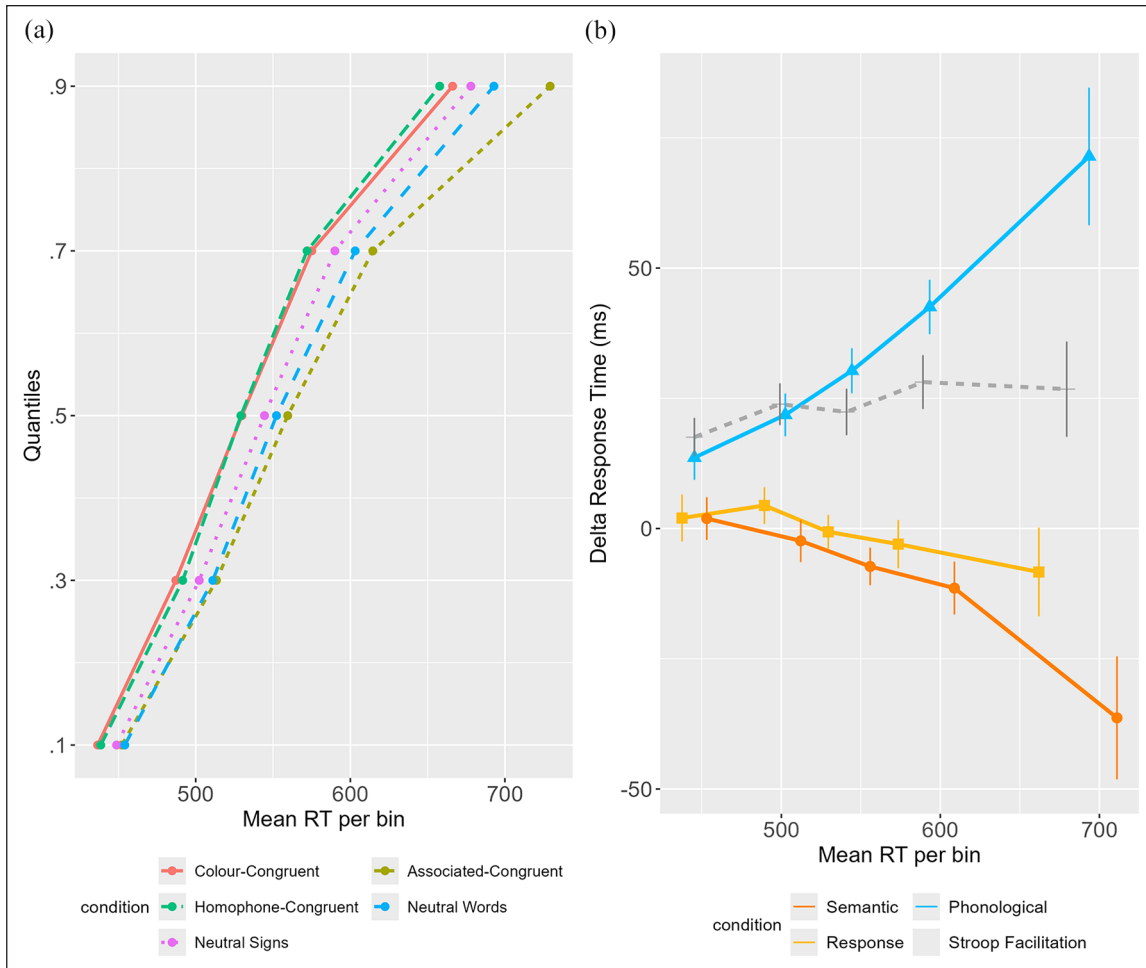
with Chinese stimuli, consistent with the findings of Augustinova et al. (2019) with French stimuli. There are, however, some important differences between the two studies that will be discussed below in detail when discussing each of the conflict and facilitation components.

**Response conflict**

Response conflict refers to the competition between two different incompatible (pre-)motor responses triggered by a stimulus, e.g., a colour-incongruent word. However, a colour-incongruent word also triggers semantic conflict (incompatible semantic information from the word and from the ink colour). To disentangle response conflict from semantic conflict, different methods have been used in the literature. For example, studies have measured response conflict by subtracting RTs to colour-incongruent words from those to incongruent colour-associated words because the latter words only trigger semantic conflict and not response conflict (Schmidt & Cheesman, 2005), whereas the latter

triggers both. This subtractive approach isolates response-based conflict from semantic conflict (Augustinova et al., 2019; Augustinova & Ferrand, 2014; Ferrand & Augustinova, 2014).

Alternatively, studies have used an experimental paradigm (2:1 paradigm) in which colour information of two different colours is mapped to either the same response or to two different responses so that response conflict is only present in the different-response condition (Schmidt & Cheesman, 2005; Shichel & Tzelgov, 2018). Response conflict is then measured by subtracting RTs to different-response trials from those to same-response trials. Studies that used the 2:1 paradigm (Burca et al., 2021, 2022; Martinon et al., 2024) found direct support for semantic conflict that was independent from response conflict. This is because semantic conflict is measured as the RT difference between same-response trials and colour-neutral trials. Such a measure of semantic conflict is independent of response conflict because response conflict is measured as the RT difference between same- and different-response trials.



**Figure 6.** Quantile plots (a) and delta plots (b) for Stroop facilitation effects in vocal responses. Error bar stands for standard error (SE) for each quantile.

The present study included Chinese colour word homophones that do not overlap in orthography with colour words (unlike homophones in alphabetic language). Homophones made it possible to measure response conflict without phonological conflict by subtracting response times of incongruent colour words (incompatible phonology and response) from those of incongruent homophones. The latter trials activate incompatible phonology but not incompatible responses (no response conflict) because the orthography of the homophone and associated meaning is not in the response set.

The present study revealed response conflict in the Stroop task with vocal and manual responses. Furthermore, the amount of conflict was not significantly different (no response modality effect). In contrast, Augustinova et al. (2019) reported a response modality effect for response conflict because the conflict with vocal responses was significantly larger than that with manual responses. However, as mentioned earlier, Augustinova et al. used a different definition of response conflict compared to that in the

present study. In Augustinova's study, response conflict was defined as the RT difference between incongruent colour words and incongruent colour-associated words, whereas in the current study, response conflict was defined as the RT difference between incongruent colour words and incongruent homophones. When response conflict was calculated in the present study in the same way as in the study by Augustinova et al., a modality effect was found, and just as in the study by Augustinova et al., the amount of response conflict in the vocal Stroop task was about twice that of the response conflict in the manual Stroop task (see Table 4).

The findings of the present study therefore suggest that the modality effect in response conflict observed by Augustinova et al. (2019) was potentially driven in part by phonological conflict in the vocal Stroop task. Alternatively, it could be argued that the response conflict observed in the current study is in fact reflecting reduced semantic activation. Thus, it reduced response conflict in the incongruent homophone condition relative to the incongruent

**Table 4.** Conflict components of Stroop interference comparing the current study and Augustinova et al. (2019) in both vocal and manual tasks.

Conflict components	Chinese (present study)		French (Augustinova et al., 2019)	
	Vocal	Manual	Vocal	Manual
Semantic Conflict (Associated-Incongruent—Colour-Neutral Words)	32***	23***	23***	18***
Phonological Conflict (Homophone-Incongruent—Associated-Incongruent)	16** >	-1		
Response Conflict <sup>a</sup> (Colour-Incongruent—Homophone-Incongruent)	39***	29***		
Response Conflict <sup>b</sup> (Colour-Incongruent—Associated-Incongruent)	55*** >	28***	101*** >	50***
Task Conflict (Neutral Words—Neutral Signs)	8	0	44*** >	5

<sup>a</sup>Response conflict as defined in the present study.

<sup>b</sup>Response conflict as defined by Augustinova et al. (2019).

\*\*\* $p < .001$ ; \*\* $p < .01$ .

condition. Both conditions activate the phonology and semantics of the colour word; however, in the homophone condition, the activation of semantics is due to phonological mediation and therefore weaker. This would explain the absence of a modality effect in the response conflict.

Augustinova et al. (2019) used the comparison between incongruent colour trials (e.g., *BLUE* in green) and incongruent colour-associated trials (e.g., *SKY* in green) as their measure of response conflict. The associated trials have indirect access to the semantic and phonological components of the related response option (e.g., *SKY* indirectly activates the concept blue and the phonology of blue). In contrast, incongruent colour trials directly activate these components. Thus, the difference between these two trial types is the amount of semantic information, and semantically-mediated phonological information of the incongruent response is activated. The definition of response conflict used in the present study and that in the study by Augustinova et al. are different as they measure different contributors to response conflict.

### Response facilitation

Response facilitation was defined in the present study as the RT difference between congruent colour words and congruent homophones. No evidence for response facilitation was found in the vocal and manual Stroop task. In contrast, with French stimuli, Augustinova et al. (2019) found strong response facilitation with vocal but not with manual responses. However, response facilitation was defined differently in Augustinova's study (the RT difference between congruent colour words and congruent colour-associated words). Using this definition, the present study also revealed strong response facilitation with vocal responses and no response facilitation with manual responses. In fact, the difference observed match those reported by Augustinova et al. closely (31 vs. 2 ms in the present study, and 39 vs. 7 ms in the study by Augustinova et al., see Table 5). Thus, the present study suggests that

response facilitation in the vocal Stroop as reported by Augustinova's study could in part have been driven by phonological facilitation (i.e., overlap in the pronunciation between the colour word and the coloured rectangle).

A numerical 36-ms response facilitation was observed with vocal responses in the study by Spinks et al. (2000), but not with manual responses in the study by Wang et al. (2010). In both studies, response facilitation was defined in the same way as in the study by Augustinova et al. (2019). Thus, the effect in the vocal Stroop task could also have been driven by phonological facilitation in the vocal Stroop task and not in the manual task. Therefore, the modality effect observed with response facilitation might also be due to phonological facilitation with vocal responses.

### Semantic conflict/facilitation

Semantic conflict was measured in the present study by subtracting the RTs to incongruent colour-associated words from those to neutral words because the only difference between these two conditions is the colour-related information activated by colour-associated words. This measurement of conflict and our measurement of semantic facilitation are the same as that in other studies in the literature (Augustinova et al., 2018, 2019; Augustinova & Ferrand, 2014; Sharma & McKenna, 1998).

The vocal and manual Stroop tasks with Chinese stimuli revealed semantic conflicts that were numerically slightly larger than what has been found in the study by Augustinova et al. (2019) with French stimuli. The data from Spinks et al. (2000) and Wang et al. (2010), who also used Chinese stimuli, showed a similar numerical semantic conflict of 30 ms with the vocal Stroop task as the current study but a larger 36 ms with manual Stroop than in the present study (23 ms). Importantly, the difference in semantic conflict between vocal and manual responses in the present study was not significant; thus, it was not modulated by response modality, which was also

observed in the study by Kinoshita et al. (2018). In contrast, Sharma and McKenna (1998) did not find semantic conflict 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 and those of Augustinova and colleagues indicate that this lexical effect is not exclusive to vocal responses. Thus, semantic information is also 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. However, the effects of semantic facilitation observed in French were relatively small (vocal: 11 ms; manual: 14 ms), and the authors did not correct for multiple comparisons. The present study did not reveal semantic facilitation in either response modality after Holm-Bonferroni correction. Interestingly, a trend towards a reversed semantic facilitation effect was observed with vocal responses, which is actually consistent with the findings in the study by Dalrymple-Alford (1972), although with a different baseline (a row of X's) compared to that in the current study. Regarding the response modality effect, a smaller semantic facilitation was found in vocal than in manual responses. A similar pattern was also found in French (Augustinova et al., 2019). In addition, the distributional analysis in the present study revealed a significant negative linear trend was observed. These findings are unexpected considering that strong semantic conflicts were observed in both response modalities. To investigate whether the specific colours used in the current study led to the reversed semantic facilitation effect, a post hoc distributional analysis was conducted for each colour separately (Table 6). The analysis revealed that the colour blue did not contribute to the negative linear effects in semantic facilitation but contributed to positive linear effects in semantic conflict. An explanation for this could be that the colour-associated word 天 (SKY) has a higher word frequency than 草 (GRASS) and 金 (GOLD). Studies have found that the use of high-frequency distractor words leads to smaller Stroop interference than low-frequency distractor words (Burt, 1999, 2002; Navarrete et al., 2015). When presented in the congruent condition (i.e., SKY in blue colour), there is a tendency to name SKY rather than BLUE because of its high frequency; thus, we observed no semantic facilitation in colour blue.

In sum, the presence of semantic conflict in the Stroop task is universal. Moreover, semantic conflict can be observed with both vocal and manual responses. Semantic facilitation is absent in the manual Stroop, and there is a trend towards a reverse semantic facilitation effect (i.e. -9 ms) in the vocal Stroop.

The measurement of semantic conflict/facilitation assumes that both the incongruent/congruent colour-associated words and the neutral words do not trigger (pre-) motor responses. Klein (1964), however, suggested that incongruent colour-associated words generate response conflict. Critically, data from Schmidt and Cheesman (2005) provided evidence that semantic colour-associated words do not result in response conflict (see the previous section about response conflict). However, Parris et al. (2022) argued that semantic associates are not purely measuring semantic conflict/facilitation because the activated colour meanings are part of the response set. Parris argued that more sensitive techniques should be used, such as electromyography (EMG), to determine if colour associates do not trigger (pre-)motor responses. Quéard et al. (2023) used a mouse-tracking paradigm to distinguish between response and semantic conflicts. Non-response set trials (e.g., *ORANGE* that was not used as a colour response) were used to measure response conflict with incongruent colour words and to measure semantic conflict with incongruent colour-associated words. They found that there was no response conflict in non-response set trials and argued that it is better to distinguish semantic relevance from incongruent colour-associated words.

### *Phonological conflict/facilitation*

In the present study, strong phonological conflict and facilitation were observed with vocal responses and not with manual responses, which indicates a qualitative rather than quantitative difference between response modalities. Moreover, as discussed earlier, the response modality effect was driven by phonological conflict and facilitation. These results can only be explained by assuming that phonology is automatically activated when using Chinese characters in the Stroop task. Besner and Stolz (1998) reported phonological conflict in a manual Stroop using pseudo-homophones of colour words in English. Parris et al. (2022) expressed concerns that measuring phonological conflict using pseudo-homophones would be confounded by orthographic conflict. This confound is not an issue in Chinese because Chinese homophones do not overlap in orthography with colour words. Thus, it is possible to investigate pure phonological conflict/facilitation in a Chinese Stroop task.

The magnitude of phonological conflict calculated from the vocal Stroop data collected by Spinks et al. (2000) was -2 ms, and -7 ms for the manual Stroop task of Wang et al. (2010). In the present study, a significant phonological conflict (16 ms) was observed in the vocal but not in the manual responses (1 ms), and this difference was also significant between those response modalities. Spinks et al. (2000) used both same-tone and different-tone homophones in the study, whereas the current study used same-tone homophones only. The several homophone conditions

**Table 5.** Facilitation components of Stroop facilitation comparing the current study and the study by Augustinova et al. (2019) in both vocal and manual tasks.

Facilitation components	Chinese (present study)		French (Augustinova et al., 2019)	
	Vocal	Manual	Vocal	Manual
Semantic Facilitation (Colour-Neutral Words—Associated-Congruent)	-9 <sup>+</sup>	-1	11 <sup>***</sup>	14 <sup>***</sup>
Phonological Facilitation (Homophone-Congruent—Colour-Congruent)	32 <sup>***</sup>	> 2		
Response Facilitation <sup>a</sup> (Homophone-Congruent—Colour-Congruent)	-2	0		
Response Facilitation <sup>b</sup> (Associated-Congruent—Colour-Congruent)	31 <sup>**</sup>	> 2	39 <sup>***</sup>	> 7

<sup>a</sup>Response conflict as defined in the present study.

<sup>b</sup>Response conflict as defined by Augustinova et al. (2019).

\*\*\* $p < .001$ ; \*\* $p < .01$ ; \* $p < .10$ .

**Table 6.** F values of linearity for each colour in vocal responses.

Components	Vocal responses		
	Colour blue	Colour green	Colour yellow
Semantic facilitation	0.15	8.453 <sup>**</sup>	5.941 <sup>*</sup>
Phonological facilitation	2.571	26.497 <sup>***</sup>	21.498 <sup>***</sup>
Response facilitation	1.681	4.444 <sup>*</sup>	3.942 <sup>*</sup>

\*\*\* $p < .001$ ; \*\* $p < .01$ ; \* $p < .05$ .

in the study by Spinks et al. might have resulted in reduced sensitivity towards homophones, and at the same time, this reduced the proportion of experimental conditions and neutral controls. As a consequence, phonological conflict was much smaller in the study by Spinks et al. than that in the current study.

Phonological facilitation in the data of Spinks et al. (2000) and Wang et al. (2010) was 12 ms and 36 ms for vocal and manual responses, respectively. The current study found a significant phonological facilitation in the vocal (32 ms) but not in the manual responses (2 ms), and this difference was significant. Wang et al. used four colours in their study, whereas the current study removed the colour red because of a potential alliteration between the colour red (/hóng/) and yellow (/huáng/) in Chinese characters. This could explain why they found much larger phonological facilitation effects in manual responses because it is more likely to trigger the phonology of the words that are pronounced similarly.

In summary, the present data showed that phonological conflict and facilitation are affected by response modality, which suggests qualitative difference between response modalities. Phonological conflict and facilitation were stronger in the vocal than in the manual responses. However, the present study cannot rule out the possibility that the present measure of phonological conflict/facilitation (subtracting RTs to homophones from those to colour-associated words) could also have been in part due to

phonologically-mediated response conflict/facilitation, which is stronger in the homophone condition.

### Task conflict

The definition of task conflict in the current study was based on the logic of Augustinova et al. (2018) that task conflict occurs due to attention being drawn to the irrelevant task (i.e., word reading) rather than the relevant task (i.e., colour naming). Specifically, in a Stroop task, task conflict is defined as the RT difference between neutral words and neutral signs (e.g., a row of X's) because the former involves word reading, whereas the latter does not. In the literature, this difference is also referred to as the “lexical effect” (Levin & Tzelgov, 2016) or “lexicality cost” (Brown, 2011).

In the present study, no task conflict was found in either vocal or manual responses. In contrast, Augustinova et al. (2019) found significant task conflict in the vocal Stroop tasks only and absent in manual responses. This discrepancy between the current study and that of Augustinova et al. (2019) regarding task conflict could be related to the differences in the neutral sign between the present study and that of Augustinova et al. (2019) (percent sign “%” vs. a row of X’s “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 X’s, and could therefore have been considered by participants as readable. 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.

As an alternative to neutral signs, colour patches have been used in other Stroop studies (Spinks et al., 2000; Wang et al., 2010; Yeh et al., 2017); however, they often only served as fillers rather than experimental stimuli, so it is difficult to measure task conflict using colour patches because they were less repeated than other conditions.



Future research could consider colour patches as neutral stimuli and analyse task conflict relative to these. Noncharacters (i.e., combining real radicals/characters into pseudo-words) could also be used to measure task conflict, although these might still activate meaning or phonology because of the radicals.

The measurement of task conflict in the present study might have been affected by the use of the colour-neutral word 炭 (charcoal, /tan4/), which is strongly associated with black (黑色, Li et al., 2024). However, black is not considered a colour, and it is not in the response set. If, however, charcoal would cause semantic conflict in the Stroop task, then stronger task conflict should be expected, which is not supported by the findings. Thus, the colour-neutral word “charcoal” is unlikely to have affected the present findings.

### *Linear trends in conflict/facilitation components*

The distributional analysis findings are consistent with the observed mean RT analysis. With vocal responses, positive linear trends were observed for response, semantic, and phonological conflicts, which suggests that it was difficult to maintain a high level of suppression for those conflict components in a Stroop task. With manual responses, positive linear trends were observed for response and semantic conflicts. Labuschagne and Besner (2015), Scaltritti et al. (2022), and Sulpizio et al. (2022, 2024) also found a positive slope for semantic conflict. In the current study, a levelling-off trend was observed in phonological conflict, suggesting that weak inhibition was applied to suppress the competition between word and colour information. Likewise, a positive linear trend was observed for phonological facilitation with vocal responses, but not with manual responses. These findings are consistent with the findings of the mean RT analysis, which showed that phonological conflict and phonological facilitation contribute to response modality effects.

Similar to previous studies that used distributional analyses to investigate Stroop effects (Bub et al., 2006; Pratte et al., 2010; Roelofs et al., 2011; Scaltritti et al., 2022), overall positive trends were observed for Stroop interference effect with both vocal and manual responses in the current study. In particular, semantic conflict found with manual responses revealed a positive trend as in the study by Scaltritti et al. (2022).

### *Response modality effects*

The current findings indicate 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. In contrast, Parris et al.

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

Response conflict as measured by Augustinova et al. (2018, 2019) could also be phonologically-mediated, therefore resulting in the conclusion that there is a quantitative difference in the activation of phonology between the response modalities. Crucially, the data are consistent with the idea that phonology does play an essential role in reading (Frost, 1998). The difference between response modalities may be attributed to the overt vocal response in a vocal Stroop task compared to the covert subvocal response in a manual Stroop task because the overt vocal response encourages greater phonological processing of the irrelevant word (Parris et al., 2019; Zahedi et al., 2019).

## **Conclusion**

The present experiment revealed that phonology is automatically activated in a vocal Stroop task with Chinese characters. Unlike alphabetic languages, the observed phonological conflict/facilitation was not affected by orthographic overlap because the Chinese homophones used to measure phonological conflict were completely orthographically distinct from the colour words (unlike homophones in alphabetic languages).

The current findings support multi-stage accounts of Stroop effects and specifically provide evidence of phonological conflict and phonological facilitation. The results support the notion that multi-stage accounts are independent of the language used, as distinct components were identified not only with Chinese in the present study but also with French stimuli in a study conducted by Augustinova et al. (2019). In addition, our findings are consistent with those reported with English stimuli in studies conducted by Klein (1964), Sharma and McKenna (1998), and Risko et al. (2006).

The current findings suggest that the modality effects observed with response conflict/facilitation in the study by Augustinova et al. (2019) could in part be driven by phonological conflict/facilitation in the vocal Stroop. When the same definition of response conflict/facilitation as Augustinova et al. was used, our results also showed stronger response conflict/facilitation with vocal than with manual responses. However, when our definition of

response conflict/facilitation was used, the response modality effect disappeared. As discussed earlier, the two methods of measuring response conflict are both useful, and one is not necessarily better than the other. Crucially, the present Stroop study with Chinese stimuli provided a new way to measure phonological conflict/facilitation, as well as response conflict/facilitation.

### Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Yicheng Qiu  <https://orcid.org/0000-0001-7336-1542>

Walter JB van Heuven  <https://orcid.org/0000-0003-3183-4449>

### Data Accessibility Statement



The data from the present experiment are publicly available at the Open Science Framework website: <https://osf.io/gu35n/>

### Supplemental Material

The supplementary material is available at [qjep.sagepub.com](http://qjep.sagepub.com)

### 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>
- 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. *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>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- 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>
- Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, 4(2), 221–225. <https://doi.org/10.3758/BF03209396>
- 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.
- 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>
- 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. *Psychological Science*, 17(4), 351–357. <https://doi.org/10.1111/j.1467-9280.2006.01710.x>
- Burca, M., Beaucousin, V., Chausse, P., Ferrand, L., Parris, B. A., & Augustinova, M. (2021). Is there semantic conflict in the Stroop task?: Further evidence from a modified two-to-one Stroop paradigm combined with single-letter coloring and cueing. *Experimental Psychology*, 68(5), 274–283. <https://doi.org/10.1027/1618-3169/a000530>
- Burca, M., Chausse, P., Ferrand, L., Parris, B. A., & Augustinova, M. (2022). Some further clarifications on age-related differences in the Stroop task: New evidence from the two-to-one Stroop paradigm. *Psychonomic Bulletin & Review*, 29(2), 492–500. <https://doi.org/10.3758/s13423-021-02011-x>
- Burt, J. S. (1999). Associative priming in color naming: Interference and facilitation. *Memory & Cognition*, 27(3), 454–464.
- Burt, J. S. (2002). Why do non-colour words interfere with colour naming? *Journal of Experimental Psychology-Human Perception and Performance*, 28(5), 1019–1038.
- Dalrymple-Alford, E. C. (1972). Associative Facilitation and Interference in Stroop Color-Word Task. *Perception & Psychophysics*, 11(4), 274–276. <https://doi.org/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, 1211–1214.
- 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>
- Ferrand, L., & Augustinova, M. (2014). Differential effects of viewing positions on standard versus semantic Stroop interference. *Psychonomic Bulletin & Review*, 21(2), 425–431. <https://doi.org/10.3758/s13423-013-0507-z>

- 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>
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, 123(1), 71–99.
- 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>
- Hasshim, N., Downes, M., Bate, S., & Parris, B. A. (2019). Response time distribution analysis of semantic and response interference in a manual response Stroop task. *Experimental Psychology*, 66(3), 231–238. <https://doi.org/10.1027/1618-3169/a000445>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297–307. <https://doi.org/10.1093/biomet/76.2.297>
- 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>
- 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>
- Kinoshita, S., Mills, L., & Norris, D. (2018). The semantic Stroop effect is controlled by endogenous attention. *Journal of Experimental Psychology-Learning Memory and Cognition*, 44(11), 1730–1742. <https://doi.org/10.1037/xlm0000552>
- 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.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Labuschagne, E. M., & Besner, D. (2015). Automaticity revisited: When print doesn't activate semantics. *Frontiers in Psychology*, 6, 117. <https://doi.org/10.3389/fpsyg.2015.00117>
- Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means* (Version 1.8.9). <https://CRAN.R-project.org/package=emmeans>
- 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>
- Li, B., Ding, Z., De Deyne, S., & Cai, Q. (2024). *A large-scale database of Mandarin Chinese word associations from the Small World of Words project* [Under review].
- 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, 1171. <https://doi.org/10.3389/fpsyg.2015.01171>
- 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>
- 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/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. (2005). The Stroop task in cognitive research. In A. Wenzel & D. C. Rubin (Eds.), *Cognitive methods and their application to clinical research* (pp. 17–40). American Psychological Association.
- 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>
- Martinon, L. M., Ferrand, L., Burca, M., Hasshim, N., Lakhzoum, D., Parris, B. A., Silvert, L., & Augustinova, M. (2024). Distributional analyses reveal the polymorphic nature of the Stroop interference effect: It's about (response) time. *Memory & Cognition*, 52, 1229–1245. <https://doi.org/10.3758/s13421-024-01538-3>
- Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition*, 29(1), 137–151. <https://doi.org/10.3758/BF03195748>
- Navarrete, E., Sessa, P., Peressotti, F., & Dell'Acqua, R. (2015). The distractor frequency effect in the colour-naming Stroop task: An overt naming event-related potential study. *Journal of Cognitive Psychology*, 27(3), 277–289. <https://doi.org/10.1080/20445911.2014.1002786>
- Neely, J., & Kahan, T. (2001). Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of R. G. Crowder* (pp. 69–93). American Psychological Association. <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/10.1037/0096-1523.3.3.444>
- Parris, B. A., Hasshim, N., Ferrand, L., & Augustinova, M. (2023). Onset complexity and task conflict in the Stroop task. *Quarterly Journal of Experimental Psychology*, 0(0). <https://doi.org/10.1177/17470218231214515>
- 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), 361–367. <https://doi.org/10.1027/1618-3169/a000459>
- 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>
- Quétaud, B., Spatola, N., Parris, B. A., Ferrand, L., & Augustinova, M. (2023). A mouse-tracking study of the composite nature of the Stroop effect at the level of response execution. *PLoS ONE*, 18(1), Article e0279036. <https://doi.org/10.1371/journal.pone.0279036>
- R Core Team. (2021). R: A language and environment for statistical computing [Computer software]. *R Foundation for Statistical Computing*. <https://www.R-project.org/>
- 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/10.2466/pms.1977.45.1.11>
- 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>
- Risko, E. F., Schmidt, J. R., & Besner, D. (2006). Filling a gap in the semantic gradient: Color associates and response set effects in the Stroop task. *Psychon Bull Rev*, 13(2), 310–315. <https://doi.org/10.3758/bf03193849>
- 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., Piai, V., & Rodriguez, G. G. (2011). Attentional inhibition in bilingual naming performance: Evidence from delta-plot analyses. *Frontiers in Psychology*, 2, 184. <https://doi.org/10.3389/fpsyg.2011.00184>
- Scaltritti, M., Job, R., & Sulpizio, S. (2022). Different types of semantic interference, same lapses of attention: Evidence from Stroop tasks. *Memory & Cognition*, 50(5), 898–910. <https://doi.org/10.3758/s13421-021-01256-0>
- 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>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- 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>
- Shichel, I., & Tzelgov, J. (2018). Modulation of conflicts in the Stroop effect. *Acta Psychologica*, 189, 93–102. <https://doi.org/10.1016/j.actpsy.2017.10.007>
- 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/10.1016/S0010-0277\(00\)00072-X](https://doi.org/10.1016/S0010-0277(00)00072-X)
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. <https://doi.org/10.1037/0096-3445.121.1.15>
- Sulpizio, S., Job, R., Leoni, P., & Scaltritti, M. (2022). Prepotent task-irrelevant semantic information is dampened by domain-specific control mechanisms during visual word recognition. *Quarterly Journal of Experimental Psychology*, 75(3), 390–405. <https://doi.org/10.1177/17470218211030863>
- Sulpizio, S., Scaltritti, M., & Spinelli, G. (2024). Fast habituation to semantic interference generated by taboo connotation in reading aloud. *Cognition and Emotion*, 1–16. <https://doi.org/10.1080/02699931.2024.2307367>
- 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/10.1016/j.neuroscience.2009.09.070>
- White, B. W. (1969). Interference in Identifying Attributes and Attribute Names. *Perception & Psychophysics*, 6(3), 166–168. <https://doi.org/10.3758/Bf03210086>
- 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>
- 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>