Recognizing two dialects in one written form: a stroop study

Wu, J.; Heuven, V.J.J.P. van; Schiller, N.O.; Chen, Y.

Bilingualism: Language and Cognition

cambridge.org/bil

Research Article

Cite this article: Wu, J., van Heuven, V.J., Schiller, N.O., & Chen, Y. (2024). Recognizing two dialects in one written form: A Stroop study. *Bilingualism: Language and Cognition* **27**, 845–861. https://doi.org/10.1017/S1366728924000142

Received: 5 April 2023
Revised: 28 January 2024
Accepted: 5 February 2024
First published online: 18 March 2024

Keywords:
Chinese; bi-dialectism; logographic visual word recognition; Stroop effect

Author for correspondence:
Junru Wu;
Email: jrwu@zhwx.ecnu.edu.cn

Recognizing two dialects in one written form: A Stroop study

Junru Wu[1,2,3], Vincent J. van Heuven[2,4], Niels O. Schiller[2,3,5] and Yiya Chen[2,3]

[1]Lab of Language Cognition and Evolution, Dept. Chinese Language and Literature, East China Normal University, Shanghai, China; [2]Leiden University Centre for Linguistics, Leiden, Netherlands; [3]Leiden Institute for Brain and Cognition, Leiden, Netherlands; [4]Dept. Applied Linguistics, University of Pannonia, Veszprém, Hungary and [5]Department of Linguistics and Translation (LT), School of Liberal Arts and Social Sciences (CLASS), City University of Hong Kong, Kowloon, Hong Kong

Abstract

This study aims to examine the influence of dialectal experience on logographic visual word recognition. Two groups of Chinese monolectals and three groups of Chinese bi-dialectals performed Stroop color-naming in Standard Chinese (SC), and two of the bi-dialectal groups also in their regional dialects. The participant groups differed in dialectal experiences. The ink-character relation was manipulated in semantics, segments, and tones separately, as congruent, competing, or different, yielding ten Stroop conditions for comparison. All the groups showed Stroop interference for the conditions of segmental competition, as well as evidence for semantic activation by the characters. Bi-dialectal experience, even receptive, could benefit conflict resolution in the Stroop task. Chinese characters can automatically activate words in both dialects. Comparing naming in Standard Chinese and naming in the bi-dialectals' regional dialects, still, a regional-dialect disadvantage suggests that the activation is biased with literacy and lexico-specific inter-dialectal relations.

1. Introduction

1.1 Bilingual Stroop effects

Visual word recognition in bilingual individuals has been extensively studied to test theoretical perspectives on bilingual visual word processing. It is well-established that bilinguals activate relevant lexical items from both languages, even when the languages employ different orthographic systems. This cross-linguistic activation occurs regardless of attention to written forms (Dijkstra et al., 1998; Van Heuven et al., 2008). The Stroop paradigm is a paradigm for the investigation of speech production that relies on the unintentional recognition of visual words (Kinoshita et al., 2017; Risko et al., 2005). Many studies have consistently documented cross-linguistic Stroop effects among bilingual and trilingual individuals (Chen & Ho, 1986; Dyer, 1971; Fang et al., 1981; Kiyak, 1982; MacLeod, 1991; Preston & Lambert, 1969; Van Heuven et al., 2011). All these studies revealed cross-linguistic interference, although it generally appears weaker in comparison to within-language interference. Moreover, a trilingual study demonstrated that the extent of interference and facilitation between languages is contingent on script similarity (Van Heuven et al., 2011).

The Stroop paradigm's conflict resolution element has been employed to investigate bilingual cognitive control. Bilinguals exhibit slower lexical retrieval (Bialystok, 2009; Gollan et al., 2005; Ransdell & Fischler, 1987; Rogers et al., 2006), while demonstrating superior performance in resolving task conflicts (Bialystok, 2009; Carlson & Meltzoff, 2008; Hilchey & Klein, 2011; Prior, 2012). One prominent and robust conflict observed in the Stroop task is the ink-word incongruence (Bialystok et al., 2008). These findings suggest that bilinguals possess enhanced abilities in managing conflicting information.

However, recent research has casted doubt on the notion of a universal "bilingual cognitive advantage" (Antón et al., 2014; Costa et al., 2009; Duñabeitia et al., 2014; Hernández et al., 2013; Paap & Greenberg, 2013). The current understanding suggests that if bilinguals do possess a cognitive advantage over monolinguals, it is specific to certain cognitive tasks and particular groups of bilingual individuals (for a review see Paap, 2015; Paap et al., 2015). Thus, it is crucial to reassess the role of bilingual linguistic experience in Stroop production and visual word recognition and delve deeper into understanding the well-established Stroop executive advantage.

This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

Despite the considerable research conducted on cross-linguistic Stroop effects and its relationship with cognitive control, there remains limited understanding of the cognitive processes by bilingual individuals or bi-dialectals who utilize the same logographic writing system, also known as ideographic script, for two different languages or dialects. Investigating the word processing of bi-dialectals using the same logographic writing system has the potential to contribute to our understanding of cognitive control in bilingual individuals and shed light on the dynamic nature of visual word recognition processes.

1.2 Bi-dialectal visual word recognition in Chinese

Chinese speakers of different dialects can use the same written form to represent inter-dialectal cognates, which are pronounced differently in different dialects. For instance, the Chinese character 蓝 '*blue*' can represent cross-dialectal cognates in all Chinese dialects, despite the differences in pronunciation. The extensive cross-linguistic utilization of Chinese characters has played a pivotal role in facilitating cultural exchange and transmission across a considerable span of Chinese and Asian history.

In spite of this special property of the Chinese writing system, previous cognitive models of the processing of Chinese characters have mainly focused on Standard Chinese (e.g., Li & Pollatsek, 2020; Zhou & Marslen-Wilson, 1999), and have not taken the participants' dialectal background into consideration. Actually, a large portion of Standard Chinese speakers also use at least one regional dialect. Many Chinese bi-dialectals learn to read and write Chinese characters mainly in Standard Chinese, but in passing have developed the ability to visually process Chinese characters to support speech production in their regional dialects. What remains largely unexplored is how such ability, which developed through incidental learning, differs from the visual word recognition skills acquired through systematic instruction. Related to this issue is the lack of knowledge both in terms of theory and practical applications concerning Chinese bi-dialectals' reading process of Chinese characters.

It is possible that Chinese bi-dialectals process Chinese characters differently than monolectals who only know Standard Chinese. If so, many questions remain open, such as (1) whether Chinese bi-dialectals differ from monolectals in their automatic visual word recognition, (2) when they read Chinese characters, how phonological and semantic representations are retrieved in their mental lexicon, (3) whether and how the relation between the readers' two dialects influence the recognition process, and (4) whether and how bi-dialectal experience contributes to a similar Stroop conflict-resolution advantage as reported in previous bilingual studies.

The current study was thus designed to employ the Stroop paradigm to examine the automatic semantic and phonological processing of Chinese characters by five groups of readers, who have different dialectal backgrounds in addition to Standard Chinese. We are interested in how dialectal backgrounds may shape the visual word recognition process. Furthermore, by investigating the effects of bi-dialectism on conflict resolution within the Stroop paradigm, we will gain further insights into the possible differences between bi-dialectal and bilingual cognitive control.

1.3 Stroop paradigm applied to Chinese characters

In comparison to the classical Stroop design, this study adapted and restructured the experimental variables to account for the logographic and homophone-rich characteristics of Chinese orthography. In Chinese, characters typically represent morphemes - the smallest units of language that carry meaning. In most (if not all) Chinese dialects, there are plenty of homophonic monosyllabic morphemes, which leads to a wide selection of homophonic characters that have different meanings.

Stroop-related paradigms have long been used to investigate visual word recognition (MacLeod, 1991). The classical Stroop effect (Stroop, 1935) emerges when participants are shown color words printed in different colors. When the word and the ink color are *congruent* (e.g., 'RED' in red ink), the word facilitates the color naming relative to an *irrelevant control word* (Dalrymple-Alford, 1972). When the word and the ink color are *incongruent* (e.g., 'RED' in green ink), the word interferes with the color naming relative to different types of controls, such as *color patches*, 'X's, *irrelevant words*, and *non-words* (MacLeod, 1991).

The Stroop effect is sensitive to the phonological relationship between the printed word and the ink name. Earlier studies on alphabetic writing systems verified that when the printed word shares phonological features with a competing color word (e.g., 'GREAN' in red, or 'YELLXX' in red, as compared to controls such as GREEF, XXLLOW), the interference increases (Dennis & Newstead, 1981; Singer et al., 1975). Furthermore, when the printed word shares phonological features with the ink name (e.g., 'GREET' in green), the interference is reduced, or the color naming is facilitated, depending on whether color patches or irrelevant words are used as the control condition (Effler, 1978). Also, the effect size increases with the amount of phonological overlap (Dennis & Newstead, 1981). Moreover, comparing pseudo-homophones of color words (e.g., *bloo* for *blue*) to non-words matched for visual similarity and initial phonological overlap with color words (e.g., *blir* for *blue*), Dennis and Newstead (1981) showed that the phonological effects in alphabetic writing systems are not just a result of orthographic similarity but involve phonological and lexical effects.

Comparatively, Chinese stands out with two distinct features: (1) its use of lexical tones, and (2) its logographic writing system. The Stroop interference obtained with Chinese characters is at least of the same strength as was found for alphabetic scripts (Lee & Chan, 2000; Smith & Kirsner, 1982) if not greater (Biederman & Tsao, 1979; Saalbach & Stern, 2004; Tsao et al., 1981). It has also been shown that Chinese Stroop tasks may involve more right hemisphere interference (Tsao et al., 1981).

Phonological Stroop effects observed in Chinese are consistent with the broader findings, yet with a greater reliance on segmental information than on tonal information. Both segmental and tonal overlap between the ink-color and the color character (i.e., the orthographic unit that represents a syllable with a specific meaning, segmental make-up, and tone) showed phonological effects in segmental homophones of color characters (S+T–, e.g., 轰 /xuŋ (Hl)/ 'boom', for 红 /xuŋ(Hr)/ 'red'[1]), facilitated the naming of congruent colors and interfered with the naming of incongruent colors (Li et al., 2013; Spinks et al., 2000). Additional tonal overlap (S + T+, e.g., 洪 /xuŋ(Hr)/ 'flood', for红 /xuŋ(Hr)/ 'red') slightly increased the facilitation effect (Li et al., 2013; Spinks et al., 2000, Experiment 1). Li et al. (2013) but not Spinks et al. (2000) also found phonological facilitation based on tonal overlap (S–T+, e.g., 涂 /t^hu(Hr)/ 'smear', for 红 /xuŋ(Hr)/ 'red') alone. In a study utilizing a related picture-word interference paradigm, it has been found that when there is both segmental and tonal overlap between distractor and target words, it can facilitate picture naming as compared with segmental-alone overlapping. This effect holds true whether the tonal categories underlying the words match or if they have similar tonal pronunciations (Nixon et al., 2015).

Moreover, as Chinese color words and their homophones can be distinguished by their entirely distinct Chinese characters (e.g., 洪 'flood', is an exact homophone of the Chinese color character 红 'red', but the two characters involve no orthographic overlap[2]), by comparing the Chinese characters for color words to their homophones, these studies (Li et al., 2013; Spinks et al., 2000) also verified that the meanings of the Chinese characters were automatically activated in the Stroop task.

Thus, with orthographic overlap controlled, three distinct aspects – semantics, segments, and tones of Chinese ink-character relations – can be manipulated in the Stroop paradigm. Furthermore, taking into consideration the character's relation with both the color name and a potentially competing color word (e.g., see Dalrymple-Alford, 1972), the ink-character relation of "incongruency" can be broken down into "difference" and "competition". By "competition", we refer to the implicit relationship between the printed character (e.g., 红 'red') and any other potentially competing color words (e.g., /ly(F)/ 'green'). In short, a printed Chinese character's relation with the ink color could ideally be manipulated in a threefold manner: *segmentally congruent/competing/different*, *tonally congruent/competing/different*, *semantically congruent/competing/different*. After excluding inapplicable combinations of conditions, this manipulation is illustrated in Table 1.

Firstly, by the comparison between complete congruency (c2 in Table 1) and homophone congruency (c3 in Table 1), we may replicate the investigations conducted by Li et al. (2013) and Spinks et al. (2000) to verify that, given both segmental and tonal congruency with the ink color name, adding direct semantic information of the target ink color name via a logographic character would facilitate lexical retrieval. In addition, by the comparison of color-word competition (c5 and c8 in Table 1) with color-homophone competition (c6 and c9 in Table 1), we investigate whether, given phonological competition with the ink color name, adding direct semantic activation of a competing color word (via a logographic character, c5, c8) would introduce additional interference through lexical competition.

Secondly, further investigation may be warranted to elucidate the segmental and tonal effects of "*competition*" and "*difference*" and to better understand their individual contributions. For instance, regarding the condition of tone-alone congruency, we are comparing the previously tested condition of segment-only difference (c1 in Table 1) with two untested segmentally competing conditions: c8 and c9 in Table 1. In these segmentally competing conditions, printed characters segmentally compete but tonally align with the ink color name. We expect segmentally mediated semantic competition to yield greater interference. Also, we seek to better understand the tonal effect, as previous findings have been conflicting.

Similarly, regarding the condition of complete phonological incongruency, we compared the previously tested condition of complete irrelevance (c0 in Table 1) with three conditions of competition – namely, competing color characters with no phonological overlapping (c5 in Table 1) and homophones of these competing color characters (c6 and c7 in Table 1). We expect competing color characters and their homophones, due to stronger lexical competition, to introduce greater interference than completely irrelevant characters.

Note that, regarding the two conditions of homophone competition, we further distinguished tonal *competition* (c6 in Table 1) as an exact homophone of the implicitly competing color word from tonal *difference* (c7 in Table 1) as a segment-alone homophone for the implicitly competing color word. With this comparison, we examine whether additional tonal congruency with a potentially interfering color word would influence the interference introduced with lexical competition.

Moreover, a few previously tested comparisons were included for verification in various dialectal contexts. For instance, by replicating the comparison of completely congruent homophones (c3 in Table 1) and segment-alone homophones (c4 in Table 1), we aim to verify Li et al.'s (2013) finding that additional tonal congruency would increase the facilitation introduced by segmental congruency.

Table 1. Design for the relationship between the ink name and the character name. Congruent parts in the examples are marked **same** (bold roman); competing parts in the examples are marked ***comp*** (bold italics).

	Domain			Example				
Code	Segm.	Tone	Semantic	Ink name			Character name[a]	
c0_irrlvt_control	diff	diff	diff	xuŋ(Hr)	'red'	贯	kuan(F)	'penetrate'
c1_sdiff_tcong_ff	diff	**same**	diff	xuŋ**(Hr)**	'red'	涂	t^hu**(Hr)**	'smear'
c2_scong_tcong_cw[b]	**same**	**same**	**same**	**xuŋ(Hr)**	**'red'**	红	**xuŋ(Hr)**	**'red'**
c3_scong_tcong_ff [c]	**same**	**same**	diff	**xuŋ(Hr)**	'red'	洪	**xuŋ(Hr)**	'flood'
c4_scong_tdiff_ff	**same**	diff	diff	**xuŋ**(Hr)	'red'	轰	**xuŋ**(Hl)	'boom'
c5_scomp_tcomp_cw	***comp***	***comp***	***comp***	***xuŋ(Hr)***	***'red'***	绿	***ly(F)***	***'green'***
c6_scomp_tcomp_ff	***comp***	***comp***	diff	***xuŋ(Hr)***	'red'	虑	***ly(F)***	'consider'
c7_scomp_tdiff_ff	***comp***	diff	diff	***xuŋ***(Hr)	'red'	旅	***ly***(Lr)	'travel'
c8_scomp_tcong_cw	***comp***	**same**	***comp***	***xuŋ*****(Hr)**	***'red'***	蓝	***lan*****(Hr)**	***'blue'***
c9_scomp_tcong_ff	***comp***	**same**	diff	***xuŋ*****(Hr)**	'red'	栏	***lan*****(Hr)**	'fence'

[a]See Table 3 for more meta information of the Chinese characters.
[b]cw: character directly represents (the meaning and pronunciation of) a color word.
[c]ff: character represents a false friend (i.e., homophone of different meaning) of a color word

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

Similarly, by comparison of tone-alone congruency (c1 in Table 1) and complete irrelevance (c0 in Table 1) Li et al. (2013) and Spinks et al. (2000) reported conflicting findings. By replicating this condition with Chinese participants with different dialectal backgrounds, we may be able to better understand the confounding factors behind this inconsistency.

In addition, we further tested tone-alone congruency effect under segmental competition. By comparing competing ink-character pairs with no phonological overlap (c5 in Table 1) and competing ink-character pairs with tone-alone overlap (c8 in Table 1), as well as by comparing the homophones of these characters, in c6 and c9, we investigate the influence of tonal congruency on segmentally mediated lexical competition.

Some combinations of conditions remain inapplicable because in Chinese the tonal inventory is structured differently than the segmental inventories. For instance, SC has only four lexical tones, three of which are used in common color words, and tone alone is not enough for the identification of a word. Thus, when the character is tonally congruent with an ink color name or a competing color name, but not segmentally congruent (e.g., c1 in Table 1), it is not possible to make a distinction between semantic competition and semantic difference based on tones alone; tonal difference and tonal competition cannot be implied from this character either. Similarly, when the character segmentally conflicts with the ink color name (whether they are competing or different) and has no orthographic congruency with any other color words (c0, c5, and c6 in Table 1), it is not possible to differentiate between semantic or tonal competition and difference either. These conditions are treated as semantically and tonally different in the current study.

Moreover, since there are no segmental homophone color words in SC when the character is segmentally congruent but orthographically incongruent with the ink color name (e.g., c3 in Table 1), semantic competition is impossible. When the character is segmentally or tonally incongruent (whether competing or different) with the ink color name (e.g., c5-6-7-8-9 in Table 1), due to orthographic incongruency, they cannot be semantically congruent with it either.

To briefly summarize, to comprehensively investigate the interplay between semantic, segmental, and tonal relations of ink and characters, the current study has identified ten viable experimental conditions, excluding any inapplicable combinations, as shown in Table 1 with a detailed overview.

1.4 Bi-dialectal influences on the Stroop effect

Extensive experience in resolving linguistic conflicts has been proposed as a key factor contributing to the cognitive control advantages observed in bilingual individuals (Bialystok et al., 2008). To gain a deeper understanding of the intricate nuances of linguistic experiences and their impact on cognitive processing, this study focuses on examining the visual word processing of Chinese bi-dialectals during the Stroop task.

It is important to note that the commonly accepted dichotomy between "bilinguals" and "monolinguals" may gain from a more comprehensive analysis that examines the nuances of their respective linguistic backgrounds. Specifically, it remains uncertain whether previous findings regarding bilinguals of non-tonal languages are applicable to tonal bi-dialectals. Hence, the current study aims to explore the impact of different dialectal experiences in automatic Chinese visual world processing and conflict resolution. To achieve this, two groups of Chinese monolectals and three groups of Chinese bi-dialectals were included in the study, as outlined in Table 2.

Firstly, two groups of monolectals of Standard Chinese were included to test for the effect of receptive dialectal experiences. The first monolectal group (BJ) included monolectal speakers of Beijing Mandarin (the basis of Standard Chinese, SC) who were born, brought up, and tested in Beijing. It was confirmed that their parents and other caregivers also only spoke Beijing Mandarin or Standard Chinese. The dialectal experiences of the BJ monolectal group were minimal or absent so that they have limited experience with comprehending other Chinese dialects.

In contrast, the second monolectal group (SC) included speakers of Standard Chinese who were living in Shanghai. It was also confirmed that their parents and other caregivers only spoke Beijing Mandarin or Standard Chinese to them, and they reported that they could only speak Standard Chinese. However, most of them were not raised in Beijing, but in communities where other local Chinese dialects coexisted with Standard Chinese. Moreover, they had lived in Shanghai for at least three months before they took the test, where Shanghai Wu Chinese (Shanghainese, SH) is used beside Standard Chinese (SC). Thus, the SC group of monolectals had some receptive experiences with regional Chinese dialects.

Secondly, three groups of proficient Chinese bi-dialectals were included to assess the impact of inter-dialectal experiences. The first bi-dialectal group (JN) was recruited from the city of Jinan, speaking both Jinan Mandarin (JM) and Standard Chinese (SC). They were also living in Jinan during the test. The second bi-dialectal group (SH) was recruited from Shanghai, who spoke Shanghai Wu Chinese (SH) beside Standard Chinese (SC), living in Shanghai during the test. The third bi-dialectal group (OD) came from bi-dialectal areas other than Shanghai, spoke their own regional dialect beside Standard Chinese (SC), but had lived in Shanghai for at least three months before they were recruited in Shanghai to take the test[3].

The JN and SH groups were included to evaluate the influence of variations in the relationship between their regional dialects

Table 2. Participant groups and their dialectal backgrounds.

Group	N dialects mastered	Dialect mastered beside Standard Chinese	City of residence	Receptive dialect experience
BJ	1 (monolectal)	none*	Beijing	little
SC	1 (monolectal)	none	Shanghai	complex
JN	2 (bi-dialectal)	Jinan Mandarin (JM)	Jinan	local
SH	2 (bi-dialectal)	Shanghainese Wu (SH)	Shanghai	local
OD	2 (bi-dialectal)	other non-Wu dialect	Shanghai	complex

*Beijing Mandarin ≈ Standard Chinese

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

and SC. Regarding the former group, JM and Beijing Mandarin are both Mandarin Chinese dialects. The two dialects are highly similar and are also highly mutually intelligible (Cheng, 1997; Gooskens & Van Heuven, 2021; Tang & Van Heuven, 2009). Moreover, the majority of etymologically related translation equivalents (ETEs) between JM and SC are identical in meaning, morphology, and segmental combinations, and may only differ in the tonal aspect. For instance, the JM pronunciation /ly(Lf)/ 'green' is very similar to its SC counterpart /ly(F)/, since they have the same segments and both carry falling tones. In contrast, the JM pronunciation /lan(Hf)/ 'blue' is segmentally identical to its SC counterpart /lan(Hr)/, but they carry different tonal contours. It is important to note that, even when realized with different tonal contours, SC-JM tonal relations on ETE words are nevertheless highly regular. A previous study (Wu et al., 2016) has shown that systematic correspondence can explain more than 70% of SC-JM ETE pairs' tonal correspondence (e.g., 76% of JM monosyllabic words with (Hf) tones have ETEs in SC with (Hr) tones).

Regarding the latter group, SH is from a different family of Chinese dialects. As a Wu dialect, SH only partially overlaps with SC in sound inventory and phonotactics. Although no direct data are available about the mutual intelligibility between SH and SC, it has been verified that a very close dialect of SH, Suzhou Wu Chinese, is hardly intelligible to Beijing Mandarin monolectals (26% in isolated words, Tang & Van Heuven, 2009). Most of the translation equivalents between SH and SC are etymologically related, but some are not. Also, the range of both segmental and tonal similarities between monosyllabic ETEs in SH and SC can be quite large. For instance, the SH pronunciation of 'green' /loʔ(Lr)/ differs from its SC counterpart /ly(Lf)/ in both segmental and tonal make-up. In contrast, although the SH pronunciation of 'blue' /lᴇ(Lr)/ is segmentally distinct from its SC counterpart /lan(Hr)/, both pronunciations carry similar rising tones. A difference in tone alone is also possible. The SH pronunciation of 'purple' /tsɿ(Hl)/ is segmentally identical to its SC counterpart /tsɿ(Lr)/, but they carry different tonal contours. Moreover, although there is some overlap in the systematic correspondence between SH and SC ETEs, the relation is significantly less regular than between JM and SC. As Chinese characters represent related morphemes across dialects, we expect the SH group to differ from the JN group in their visual Chinese recognition.

In addition, the inclusion of the OD bi-dialectal group aimed to validate the applicability of the findings acquired from the previously mentioned well-controlled bi-dialectal groups, and to investigate the impact of intricate receptive dialectal experiences. The OD group did not reside in their native dialectal areas, but instead lived in Shanghai, where they use SC but had receptive exposure to the Shanghainese dialect. To differentiate them from the SH bi-dialectal group and to rule out potentially high mutual intelligibility between SH and any of the native dialects of the OD participants, it was stipulated that the candidate participants whose regional dialects belong to the Wu family of Chinese dialects were excluded. Moreover, examining the dialectal effects of the OD group can reveal commonalities and differences among Chinese dialects in terms of cross-dialectal visual word recognition and explore the generalizability and variance of these effects.

2. Materials and methods

2.1 Participants

Five groups of paid participants took part in the experiment, whose meta information is shown in Table 3.

To assess their language and dialectal backgrounds in SC (and their regional dialects), the participants completed a questionnaire prior to the experiment. All participants were right-handed, had received a literacy education in Simplified Chinese, reported having attained a level of education equivalent to college level or higher, and had some prior exposure to English in school. Fifteen participants from the BJ group, five from the SC group, four from the JN group, eight from the SH group, and six from the OD group also have some knowledge of other non-tonal languages, including French, German, Russian, Japanese (although pitch accent), Korean, and Uygur.

The bi-dialectal participants were interviewed during recruitment to confirm acquisition of both Standard Chinese (SC) and their regional dialect before reaching school age. Furthermore, the questionnaire included self-reported information on their general proficiency and frequency of usage in both Standard Chinese and their regional dialect. Although the bi-dialectal participants vary in their dominant dialects, they all reported high proficiency in both SC and their regional dialects, which was verified with a questionnaire on their specific lexical and sentential knowledge, as well as with face-to-face interviews. The reader may visit the data availability link for further participant information that was collected in the questionnaire.

2.2 Design and stimuli

An unbalanced mixed design was adopted. The stimuli were presented in five different ink colors (blue, green, purple, red, and yellow) including 230 critical trials (23 characters × 5 colors × 2 repetitions) and 10 neutral training trials. Of the 23 characters, four were related to the color character of 'blue', four to 'green', three to 'purple', three to 'red', four to 'yellow', and the remaining five were neutral characters each matched to one specific color character, with comparable character frequency and number of strokes. The four or five characters corresponding to the same color character include the color character itself, its S + T+ exact homophone, its S + T− (segmental only) homophone(s), and one of its irrelevant neutral characters (see Table 4 for the stimulus characteristics).

Each non-neutral character was segmentally *congruent* with one color word and *competing* with the other four color words. The neutral characters had a *different* segmental structure compared to the five color words. The semantic and tonal relationships between the characters and the color words varied, in each aspect ranging from *congruency* to *competition* and *difference*. Accordingly, the trials were recoded into the ten Stroop conditions as listed above in Table 1.

The selection of competition colors was based on exhaustive testing of all possible color combinations, while the inclusion of random intercepts and color-related predictors in later analyses aimed to account for potential variations in competition levels and ensure a comprehensive understanding of the interaction between color perception and word recognition.

Moreover, we acknowledge the concern regarding the wide range of word frequencies in Table 4 and the potential impact on word identification and the Stroop effect and have addressed this by including by-stimuli random intercepts in later analyses to account for the influence of character frequency variations.

Also, we have thoroughly documented and marked the special aspects within each regional dialect and the differences between SC and these dialects in the uploaded 'datacoding.xlsx' file on the Open Science framework, which we have taken into

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

Table 3. Meta data of the participant groups

		Group[a]				
		Monolectal		Bi-dialectal		
		BJ	SC	JN	SH	OD
N	all	48	30	54	41	40
	male	7	4	16	15	6
	female	41	26	38	26	34
Age (years)	range	19-30	17-34	19-36	18-38	18-30
	mean	22.73	21.30	22.70	22.90	22.40
	SD	2.95	3.93	3.85	5.26	2.33
N	SC dominant	48	30	44	28	37
	balanced	0	0	0	10	1
	dialect dominant	0	0	10	3	2
Proficiency in SC[b]	range		5-10	4-10	6-10	6-10
	mean		8.9	7.98	8.76	8.66
	SD		1.21	1.37	1.14	0.99
Proficiency in dialect[b]	range			3-10	5-10	5-10
	mean			7.66	8.1	8.11
	SD			1.68	1.34	1.33
Frequency of SC use[b]	range		7-10	4-10	3-10	7-10
	mean		9.8	8.18	9.4	9.53
	SD		0.66	1.90	1.22	0.80
Frequency of dialect use[b]	range		[c]	1-10	1-10	0-10
	mean			5.93	7.30	5.37
	SD			2.76	2.22	2.28

[a]BJ = Beijing, SC = Standard Chinese, JN = Jinan, SH = Shanghai, OD = Other dialects. The sample sizes were determined to ensure sufficient data points for each Stroop condition. We aimed to recruit as many participants as possible within our resources, considering statistical power and feasibility. Then candidates with special linguistic backgrounds were excluded to maintain group homogeneity.

[b]Proficiency and frequency of use by self-rating on a scale from 0 to 10 (= best/most). Participants' dialectal dominance was determined based on a comparison between their reported frequency of Standard Chinese usage and their reported frequency of regional dialect usage. If their reported frequency of Standard Chinese usage was higher, equal, or lower than their reported frequency of regional dialect usage, they were categorized as SC dominant, balanced, or regional dialect dominant, respectively.

[c]The SC group comprises participants from Northeast China, where the local dialect shares similar pronunciation with Beijing Mandarin and is fully mutually intelligible with Standard Chinese (SC). Five Northeast participants mentioned in the questionnaire that they frequently use the "local dialect." However, through interviews, we confirmed that these participants do not speak Standard Chinese with a Northeast accent in their daily lives. Instead, their responses indicate that they have considerable exposure to Northeast accented speech, allowing them to mimic the accent to some extent when asked to.

consideration; and later we specifically discussed the cross-dialectal effects and effects specific to certain dialects that hold theoretical value in our study.

2.3 Procedure

The experiment was implemented using the E-Prime software (Schneider et al., 2002). Participants named the color of the characters shown on the screen as quickly and accurately as possible. Each trial started with the presentation of a fixation cross that appeared in the center of the screen for 1,000 ms, followed by the target character printed in 48-point SimSun font, which disappeared after 2,000 ms. Then the following trial started. A sound recording was made from the appearance of each target character until the appearance of the next character. Critical trials were preceded by ten neutral training trials. The two repetitions of the critical trials were split into two blocks and the trials within each block appeared in different randomized orders for each participant. The order of the trials in the whole experiment was recorded for further analysis.

Half of the JN and SH groups of bi-dialectals were first-tested in SC and then in their regional dialect and the other half were first-tested in their regional dialect and then in SC. The dialect of naming was prompted by auditory instructions and five auditory examples of color naming in the target dialect. The bi-dialectals had a short break and auditory lexical identification and production tasks (first in the previously tested dialect and then in the coming dialect) between the two Stroop experiments in different dialects to avoid abrupt switching. The monolectals and the OD bi-dialectals were tested using the same procedure (with auditory instruction and other tasks before); however, only in SC.

The naming latencies of the recordings were automatically measured with a Praat (Boersma & Weenink, 2017) script

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

Table 4. Character stimuli.

			S + T− [b]		
	Color	S + T+[a]	1st	2nd	S−T− (neutral)
Character	蓝	栏	烂	览	抱
Translation	blue	Fence	rotten	View	hold
Frequency[c]	103	32	49	53	131
N strokes	13	9	9	9	8
Pronunciation SC	lan(Hr)	lan(Hr)	lan(F)	lan(Lr)	pɑu(F)
Pronunciation JM	lan(Hf)	lan(Hf)	lan(Lf)	lan(Hl)	pɑu(Lf)
Prounuciation SH	lᴇ(Lr)	lᴇ(Lr)	lᴇ(Lr)	lᴇ(Hf)	bɔ(Lr)
SC-JM common pinyin	lɑn2	lɑn2	lɑn4	lɑn3	bɑo4
Character	绿	虑	旅	驴	涂
Translation	green	consider	travel	donkey	smear
Frequency	133	180	154	26	5
N strokes	11	10	10	6	10
Pronunciation SC	ly(F)	ly(F)	ly(Lr)	ly(Hr)	t^hu(Hr)
Pronunciation JM	ly(Lf)	ly(Lf)	ly(Hl)	ly(Hf)	t^hu(Hf)
Pronunciation SH	loʔ(Lr)	ly(Lr)	ly(Lr)	ly(Lr)	du(Lr)
SC-JM common pinyin	lü4	lü4	lü3	lü2	tu2
Character	紫	子	自		鼻
Translation	purple	child	self		nose
Frequency	100	4001	3113		84
N strokes	12	3	6		13
Pronunciation SC	ʦɿ(Lr)	ʦɿ(Lr)	ʦɿ(F)		pi(Hr)
Pronunciation JM	ʦɿ(Hl)	ʦɿ(Hl)	ʦɿ(Lf)		pi(Hf)
Pronunciation SH	ʦɿ(Hf)	ʦɿ(Hf)	zɿ(Lr)		bieʔ(Lr)/bəʔ(Lr)
SC-JM common pinyin	zi3	zi3	zi4		bi2
Character	红	洪	轰		贯
Translation	red	flood	boom		penetrate
Frequency	419	122	79		89
N strokes	6	9	8		8
Pronunciation SC	xuŋ(Hr)	xuŋ(Hr)	xuŋ(Hl)		kuan(F)
Pronunciation JM	xuŋ(Hf)	xuŋ(Hf)	xuŋ(R)		kuan(Lf)
Pronunciation SH	ɦoŋ(Lr)	ɦoŋ(Lr)	hoŋ(Hf)		kuᴇ(Hr)
SC-JM common pinyin	hong2	hong2	hong1		guan4
Character	黄	皇	荒	谎	岸
Translation	yellow	emperor	shortage	lie	shore
Frequency	478	438	106	28	16
N strokes	11	9	9	11	8
Pronunciation SC	xuɑŋ(Hr)	xuɑŋ(Hr)	xuɑŋ(Hl)	xuɑŋ(Lr)	an(F)
Pronunciation JM	xuɑŋ(Hf)	xuɑŋ(Hf)	xuɑŋ(R)	xuɑŋ(Hl)	an(Lf)
Pronunciation SH	ɦuʌŋ(Lr)	ɦuʌŋ(Lr)	huʌŋ(Hf)	ɦuʌŋ(Hf)	ŋø(Lr)
SC-JM common pinyin	huɑng2	huɑng2	huɑng1	huɑng3	ɑn4

[a]in SC
[b]also arranged in SC. JM High-falling is undergoing merging with JM High-level (Wu et al., 2018). SH tonal system and rhyme system vary across generations, generally with high tones merging with high tones, and low tones merging with low tones. The notation here is based on the youngest urban generation's pronunciation (You 2013). Abbreviation for tones: Hl = High-level, Hr = High-rising, Lr = Low-rising (dip tone), F = Falling, R = Rising, Hf = High-falling, Hl = High-level, Lf = Low-falling.
[c]Character frequencies per million on the Chinese Text Computing website (Da, 2004; http://lingua.mtsu.edu/chinese-computing/).
[d]The missing column for the target colors "紫" and "红" in Table 4 is due to the unavailability of suitable characters for a second S + T− SC stimulus. Potential candidate characters either had complex orthography (a high number of strokes) and low frequency or shared sound radicals with other characters in the set, making them unsuitable for inclusion in the table.

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

(Pacilly, 2010) based on the intensity of the sound pressure. Then a trained phonetician (the first author) listened to each recording, inspected the waveform and spectrogram, and manually corrected any errors in the marking of the response onset. The naming accuracy of each response was also manually marked in this process. We chose to record participants' responses and extract the naming latencies and accuracies from the recordings instead of using the voice keys provided by E-prime to ensure measurement accuracy and minimize data loss rate due to false triggers and potential truncation of recordings.

3. Analyses and results

We analyzed, by-participant and by-stimuli, the naming accuracy rates (ACC rates) and naming latencies (RTs) of the correct responses only from the Stroop ink-color naming task with Gaussian Linear Mixed Effect (LME) models (Bates et al., 2013).[4] To normalize the distribution of the data, RT data were log-transformed. Naming latency outliers were excluded for each participant using a distribution-based approach (Van der Loo, 2010, method I, limits = 0.01-0.99) on the log transformed naming latency. *Trial order in the same color* and *trial distance from the same color* were calculated for each trial using the trial order data.

Two sets of analyses were performed (1) to investigate the influence of dialectal background on Stroop effects (to answer research questions 1 and 2), and (2) to compare within-dialectal Stroop effects and cross-dialectal Stroop effects (to answer research question 3), respectively.

The first set of analyses was performed on all the five groups' Stroop naming data in SC (N = 48,760). The models included the ten rearranged *Stroop conditions* (as listed in Table 1, baseline = c0, character and ink color completely irrelevant) and the participants' *dialectal backgrounds* (BJ, SC, JN, SH, and OD, baseline = BJ) as fixed predictors and included them in all the candidate-models. In addition to these essential predictors, the potential interactions between them, as well as a few color-related predictors (i.e., the scaled *trial order in the same color*, the scaled *trial distance from the same color*, and the scaled *color distance* between ink color and character-related color in a color wheel) and their multi-way interactions were also included as candidates of fixed predictors. By-*participant* random intercepts (nested under *dialectal backgrounds* or not), by-*stimuli* random intercepts (nested under *Stroop conditions* or not), by-*participant* random slopes of the scaled *trial order in the same color*, and random intercepts of *color pairs* of ink color and character related color were included as the candidates for the random terms.

The selective inclusion of terms reported in the manuscript below was selected via model comparison based on Akaike's Information Criteria (Sakamoto & Ishiguro, 1986). In the process of model-comparison, main effects of the crucial fix predictors were always included, while the other fixed and random terms were allowed to be excluded according to the result of model comparison. The model estimates were calculated with the *emmeans* function and *pair* method from the *emmeans* R package, without adjustment (Lenth, 2019).[5]

The second set of analyses was performed on two groups of bi-dialectals' Stroop naming data in both SC and their regional dialects (N = 43,470). The models also included the ten *Stroop conditions* (as listed in Table 1, baseline = c0) and the participants' *dialectal backgrounds* (JN and SH, baseline = JN), but further included the dialect used for color naming (SC = Standard Chinese, RD = regional dialect; baseline = RD), which were specified as fixed predictors and included them in all the candidate-models. The rest of the model settings were arranged in a similar way as the first set of analyses. Alongside the forthcoming modelling results discussed in subsequent sections, Table 5 presents the average accuracy rates and reaction times for each group in every condition.

3.1 Stroop naming in SC

3.1.1 Accuracy rates of Stroop naming in SC

The accuracy data in SC were collapsed by participant and by stimuli, respectively, and analyzed twice. Statistics of LME models are shown in Appendices 2 and 3 with Satterthwaite approximation (Kuznetsova et al., 2013), and the model estimates are shown in Figure 1.

The main effect of Stroop conditions was significant in both models, $F_{Stroop\_by\text{-}participant}(9) = 30.03$, $p < 0.001$, $F_{Stroop\_by\text{-}stimuli}(9) = 14.39$, $p < 0.001$. Specifically regarding the main effects of Stroop conditions in t-statistics, three conditions of segmental competition – namely, c5, c8, and c9 – significantly reduced accuracies as compared with the completely irrelevant baseline (c0) in both models, $t_{c5\_by\text{-}participant}(1863.01) = -4.02$, $p < 0.001$, $t_{c5\_by\text{-}stimuli}(457.24) = -4.20$, $p < 0.001$, $t_{c8\_by\text{-}participant}(1863.01) = -6.36$, $p < 0.001$, $t_{c8\_by\text{-}stimuli}(457.24) = -5.07$, $p < 0.001$, $t_{c9\_by\text{-}participant}(1863.01) = -2.04$, $p < 0.05$, $t_{c9\_by\text{-}stimuli}(457.24) = -2.12$, $p < 0.05$. See **Appendix 1** for the description of the random terms, similarly hereinafter

The main effect of dialectal backgrounds was significant in the by-stimuli model, but was only marginally significant in the by-participant model, $F_{background\_by\text{-}participant}(4) = 2.29$, $p = 0.07$, $F_{background\_by\text{-}stimuli}(4) = 6.75$, $p < 0.001$. Nevertheless, none of the terms were significant in t-statistics.

The interaction term of Stroop conditions and dialectal backgrounds was insignificant in F-statistics. However, t-statistics showed that the SH group of bi-dialectals' accuracy rates when presented with tonally congruent characters that competed segmentally (c8) were significantly higher, $t_{c8:SH\_by\text{-}participant}(1863.01) = 2.55$, $p < 0.05$, $t_{c8:SH\_by\text{-}stimuli}(420) = 2.36$, $p < 0.05$.

Please find detailed descriptions for post hoc contrasts in **Appendix1** (similarly hereinafter).

3.1.2 Stroop naming latencies in SC

Here only correctly named trials were taken into consideration. Statistics of LME models are shown in Appendix 4, and the model estimates are shown in Figure 2.

The main effects of Stroop conditions, dialectal backgrounds, and their interaction were all significant, $F_{Stroop}(9) = 39.21$, $p < 0.001$, $F_{background}(4) = 2.79$, $p < 0.05$, $F_{Stroop:background}(36) = 3.14$, $p < 0.001$.

Specifically, regarding the main effects of Stroop conditions, naming latencies of tone-alone congruent condition (c1) and the two color character competing conditions (c5 & c8) were significantly longer than the baseline (c0), $t_{c1}(184.89) = 2.94$, $p < 0.001$, $t_{c5}(29.83) = 4.50$, $p < 0.001$, $t_{c8}(33.43) = 4.69$, $p < 0.001$.

Regarding the main effect of dialectal background, the OD bi-dialectal group showed significantly longer naming latencies than the Beijing SC monolectals (SC) in general, $t_{OD}(230.48) = 3.36$, $p < 0.001$.

Regarding the interaction terms, the three bi-dialectal groups (JN, SH, OD), as compared with the control group of BJ monolectals, exhibited significantly distinct response to Stroop

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

Table 5. Mean Accuracy Rates (ACC, %) and Naming Latencies (RT, ms) in the 10 Stroop conditions (c1-c10) for the 5 participant groups (BJ, SC, JN, SH, and OD). ACC represents the percentage (%), while RT is given in milliseconds (ms). Stroop conditions follow the coding from Table 1. For the JN and SH groups, "SC" refers to naming in Standard Chinese, while "reg" indicates naming in the speaker's regional dialect.

	BJ	SC	JN	SH	OD
c0_irrlvt_control	ACC: 99.28% RT: 695.5(168)	ACC: 99.68% RT: 727.9(193)	ACC: SC:99.65%; reg:99.54% RT: SC:726.2(189); reg:722.9(186)	ACC: SC:99.85%; reg:99.53% RT: SC:716.1(174); reg:749.4(185)	ACC: 99.38% RT: 764.9(184)
c1_sdiff_tcong_ff	ACC: 99.77% RT: 711.7(184)	ACC: 100% RT: 729.4(208)	ACC: SC:99.18%; reg:99.59% RT: SC:728.5(187); reg:724.9(191)	ACC: SC:99.73%; reg:99.58% RT: SC:713.5(183); reg:744.1(187)	ACC: 99.03% RT: 762.4(187)
c2_scong_tcong_cw	ACC: 100% RT: 702.1(192)	ACC: 99.66% RT: 724.4(199)	ACC: SC:99.63%; reg:99.81% RT: SC:690.3(181); reg:707.3(194)	ACC: SC:99.51%; reg:100% RT: SC:725.2(192); reg:733.5(190)	ACC: 99.75% RT: 769.6(226)
c3_scong_tcong_ff	ACC: 99.79% RT: 684.4(173)	ACC: 100% RT: 713.6(216)	ACC: SC:99.44%; reg:100% RT: SC:693.6(176); reg:686.2(176)	ACC: SC:100%; reg:100% RT: SC:686.8(186); reg:708(175)	ACC: 99.25% RT: 734.2(188)
c4_scong_tdiff_ff	ACC: 99.48% RT: 687.6(174)	ACC: 100% RT: 710.5(200)	ACC: SC:99.88%; reg:99.54% RT: SC:700.7(171); reg:697.3(183)	ACC: SC:99.24%; reg:99.38% RT: SC:700.1(181); reg:721.9(172)	ACC: 99.69% RT: 752.6(189)
c5_scomp_tcomp_cw	ACC: 97.32% RT: 843.5(224)	ACC: 97.29% RT: 876.4(239)	ACC: SC:97.35%; reg:96.89% RT: SC:825.8(220); reg:837.3(223)	ACC: SC:98.43%; reg:97.41% RT: SC:837.3(223); reg:857.8(210)	ACC: 96.25% RT: 904.5(225)
c6_scomp_tcomp_ff	ACC: 98.81% RT: 752.1(189)	ACC: 99.14% RT: 791.9(220)	ACC: SC:99.14%; reg:98.28% RT: SC:765.4(186); reg:772.6(196)	ACC: SC:99.65%; reg:99.2% RT: SC:758.6(187); reg:789.8(192)	ACC: 98.21% RT: 822.4(212)
c7_scomp_tdiff_ff	ACC: 99.22% RT: 726.4(186)	ACC: 99.28% RT: 755.1(214)	ACC: SC:99.42%; reg:98.96% RT: SC:737(189); reg:746(196)	ACC: SC:99.39%; reg:99.38% RT: SC:732.9(186); reg:763.1(190)	ACC: 98.85% RT: 784.4(201)
c8_scomp_tcong_cw	ACC: 96.18% RT: 818.3(234)	ACC: 97.7% RT: 832.2(241)	ACC: SC:97.84%; reg:95.99% RT: SC:781.8(217); reg:795.3(239)	ACC: SC:98.58%; reg:98.54% RT: SC:798.6(209); reg:836.5(222)	ACC: 96.88% RT: 862.1(247)
c9_scomp_tcong_ff	ACC: 98.29% RT: 736.2(189)	ACC: 98.65% RT: 779.4(213)	ACC: SC:98.41%; reg:98.61% RT: SC:747.9(183); reg:752(203)	ACC: SC:99.04%; reg:98.57% RT: SC:743.7(181); reg:776.5(193)	ACC: 97.86% RT: 814.8(212)

conditions, with reduced naming latencies observed across the following conditions:

(1) the OD and SH groups when presented with the condition of tone-alone congruency (c1), $t_{c1:OD}(47112.52) = -2.16$, $p < 0.05$, $t_{c1:SH}(47112.94) = -2.15$, $p < 0.05$,
(2) the JN group when presented with congruent colors and characters (c2), $t_{c2:JN}(47108.93) = -3.87$, $p < 0.001$,
(3) the JN and SH groups when presented with complete competition between colors and characters (c5), $t_{c5:JN}(47114.52) = -5.88$, $p < 0.001$, $t_{c5:SH}(47116.15) = -3.13$, $p < 0.01$,
(4) the JN group when presented with competition from another color character's homophone (c6), $t_{c6:JN}(47112.17) = -2.04$, $p < 0.05$,
(5) the JN and SH groups when presented with a competing color character's tonally-different homophone (c7), $t_{c7:JN}(47114.36) = -2.86$, $p < 0.01$, $t_{c7:SH}(47112.82) = -2.06$, $p < 0.05$,
(6) the JN, SH, and OD groups when presented with competition from a tonally congruent color character (c8), $t_{c8:JN}(47111.21) = -6.47$, $p < 0.001$, $t_{c8:OD}(47122.08) = -2.95$, $p < 0.01$, $t_{c8:SH}(47111.15) = -3.31$, $p < 0.01$,
(7) the JN group when presented with competition from exact homophones of competing color characters (c9), $t_{c9:JN}(47116.14) = -2.34$, $p < 0.05$.

3.2 Comparing Stroop naming in SC and in regional dialects

3.2.1 Bi-dialectal groups' accuracy rates in SC and regional dialects

The JN and SH groups' accuracy data in SC and their regional dialects respectively were collapsed by speaker and by stimuli respectively and analyzed twice. Statistics of LME models are shown in Appendices 5 and 6, and the model estimates are shown in Figure 3.

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

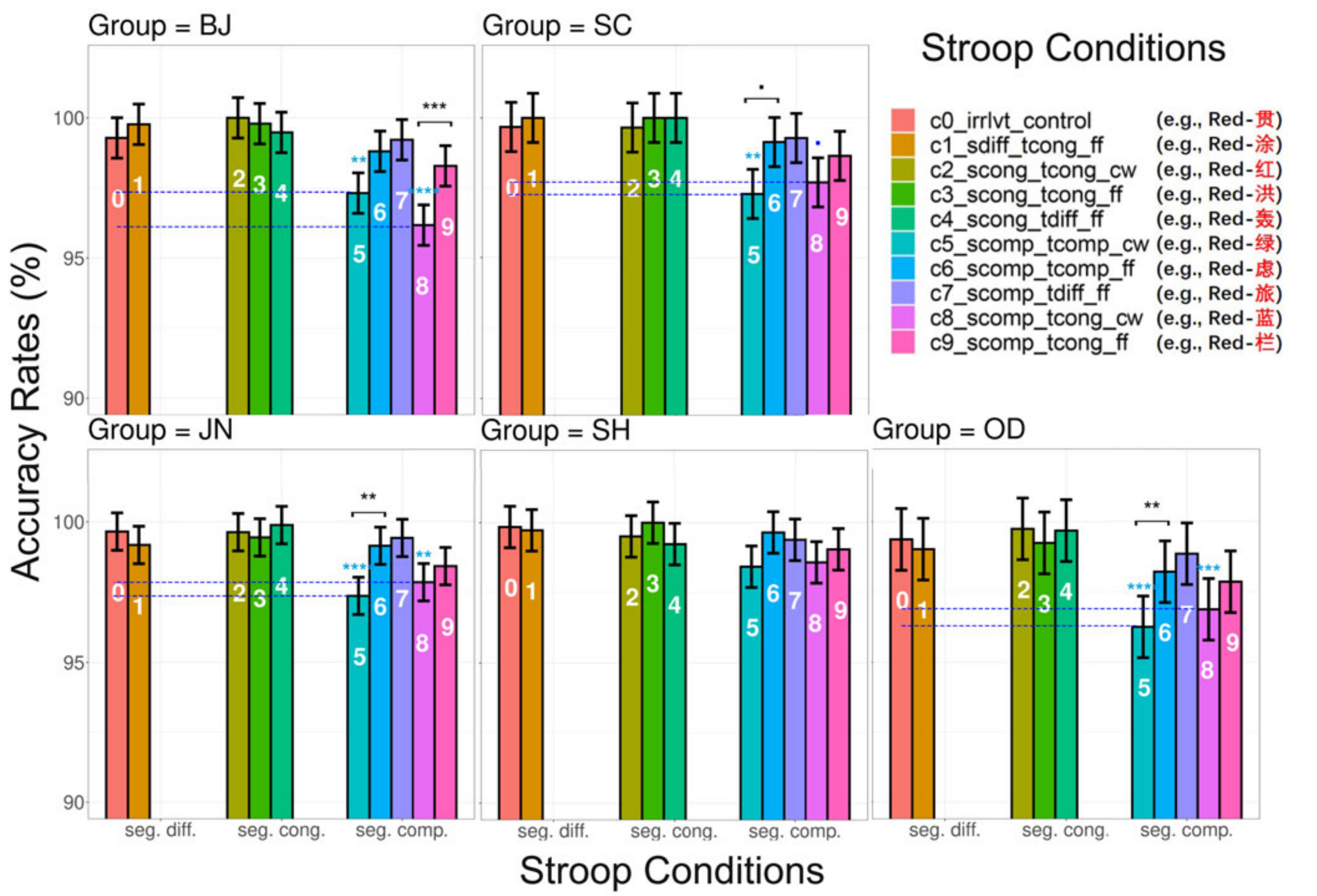

Figure 1. Estimated by-speaker mean accuracy (%) in ten different Stroop conditions (marked with colors, and clustered by segmental condition, details in Table 1. irrlvt_control = irrelevant control, sdiff = segmentally different, scong = segmentally congruent, scomp = segmentally competition, tcong = tonally congruent, tdiff = tonally different, tcomp = tonally competition, cw = color word, ff = false friend) for five groups of participants (BJ = Beijing SC mono-dialectals, SC = SC mono-dialectals living in Shanghai, JN = SC-JM bi-dialectals from Jinan, SH = SC-SH bi-dialectals from Shanghai, and OD = bi-dialectals of SC and other dialects living in Shanghai) in SC. Braces and stars serve as indicators of significant contrasts between a pair of conditions, for which only significant contrasts within participant groups and comparing two Stroop conditions that differ in a single aspect of manipulation (whether it be semantic, segmental, or tonal in nature) are marked. Single dashed line mark conditions that significantly contrast with the baseline condition (c0_irrlvt_control).

Stroop conditions showed significant main effects in both models, $F_{Stroop\_by\text{-}participant}(9) = 21.8$, $p < 0.001$, $F_{Stroop\_by\text{-}stimuli}(9) = 13.67$, $p < 0.001$. Specifically regarding the main effect of Stroop conditions in t statistics, four conditions of segmental competition – namely, c5, c6, c8, and c9 – significantly reduced accuracies as compared with the baseline of complete irrelevance in both models, $t_{c5\_by\text{-}participant}(1757.32) = -5.69$, $p < 0.001$, $t_{c5\_by\text{-}stimuli}(398.10) = -6.07$, $p < 0.001$, $t_{c6\_by\text{-}participant}(1757.32) = -2.70$, $p < 0.01$, $t_{c6\_by\text{-}stimuli}(398.10) = -2.88$, $p < 0.001$, $t_{c8\_by\text{-}participant}(1757.32) = -7.64$, $p < 0.001$, $t_{c8\_by\text{-}stimuli}(398.10) = -6.23$, $p < 0.001$, $t_{c9\_by\text{-}participant}(1757.32) = -1.99$, $p < 0.05$, $t_{c9\_by\text{-}stimuli}(398.10) = -2.12$, $p < 0.05$.

Dialect of naming showed a significant main effect in the by-participant model, but was only marginally significant in the by-stimuli model, $F_{dial\_by\text{-}participant}(1) = 4.05$, $p < 0.05$, $F_{dial\_by\text{-}stimuli}(1) = 3.82$, $p = 0.05$. The main effect of dialectal backgrounds was significant in the by-stimuli model, but was only marginally significant in the by-participant model, $F_{background\_by\text{-}participant}(1) = 3.59$, $p = 0.06$, $F_{background\_by\text{-}stimuli}(1) = 11.19$, $p < 0.001$. Nevertheless, none of the corresponding terms were significant in t statistics.

Regarding the interaction terms, Stroop conditions and dialectal backgrounds showed significant interactions in both models, $F_{Stroop:background\_by\text{-}participant}(9) = 2.44$, $p < 0.01$, $F_{Stroop:background\_by\text{-}stimuli}(9) = 1.95$, $p < 0.05$. Interactions with dialect of naming were insignificant in F-statistics, but the two models showed marked differences in t-statistics. Specifically, when presented with competition from tonally congruent color characters (c8), not only the SH group of bi-dialectals as speakers but also SC as the dialect of naming were associated with significantly greater accuracy, $t_{c8:SH\_by\text{-}participant}(1757.32) = 3.59$, $p < 0.001$, $t_{c8:SH\_by\text{-}stimuli}(315) = 3.42$, $p < 0.001$, $t_{c8:inSC\_by\text{-}participant}(1757.32) = 2.64$, $p < 0.01$, $t_{c8:inSC\_by\text{-}stimuli}(315) = 2.31$, $p < 0.05$. Nevertheless, in the by-participant model a significant three-way interaction revealed that the additive effects of the SH group and the SC naming under c8 were largely cancelled out, $t_{c8:SH:inSC\_by\text{-}participant}(1757.32) = -2.01$, $p < 0.05$.

3.2.2 Bi-dialectal groups' naming latencies in SC and regional dialects

Here only the JN and SH groups' correctly named trials were taken into consideration. Statistics of LME models are shown in Appendix 7, and the model estimates are shown in Figure 4.

The main effect of Stroop conditions was significant, $F_{Stroop}(9) = 32.13$, $p < 0.001$. As shown with t statistics, naming latencies of the two conditions of color-character competition (c5 & c8) were

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

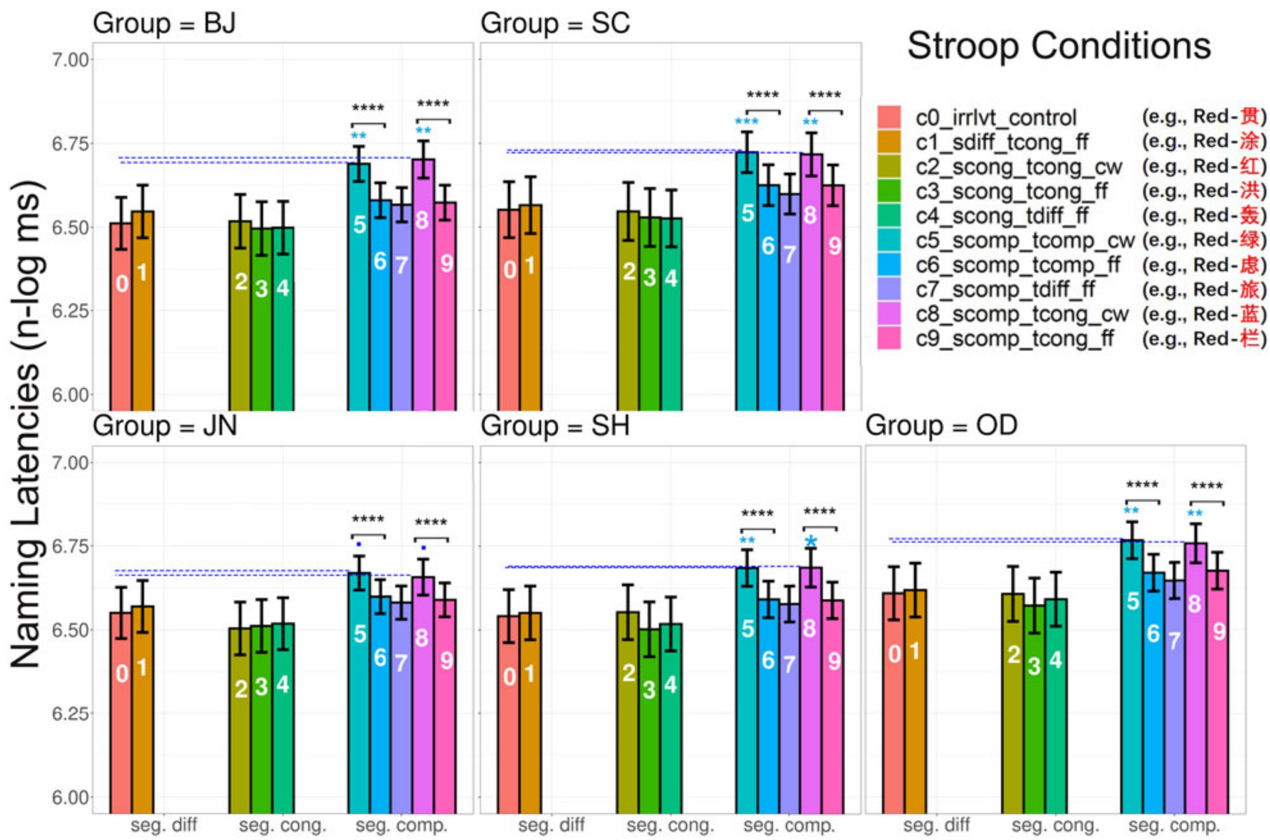

Figure 2. Estimated SC correct naming latencies (n-log ms) in ten different Stroop conditions (marked with colors, and clustered under segmental conditions, for more information see Table 1, irrlvt_control = irrelevant control, sdiff = segmentally different, scong = segmentally congruent, scomp = segmentally competition, tcong = tonally congruent, tdiff = tonally different, tcomp = tonally competition, cw = color word, ff = false friend; for details of significance markers, see Figure 1 caption)

significantly longer than the baseline (c0), $t_{c5}(30.07) = 3.85$, $p < 0.001$, $t_{c8}(34.02) = 3.32$, $p < 0.01$.

Dialect of naming also showed a significant main effect, $F_{dial}(1) = 69.05$, $p < 0.001$. However, the difference was very small and did not reach significance in t statistics. Also, the main effect of dialectal background was insignificant in both F- and t-statistics.

In addition, scaled *trial distance from the same color* was kept after model comparison, yielding a significant main effect, $F_{trial.from.same.col.}(1) = 389.96$, $p < 0.001$, $t_{trial.from.same.col.}(42343.75) = 19.75$, $p < 0.001$. The further the previous appearance of the same color was away, the longer it took to name this color.

Regarding the interaction terms, the interaction of Stroop conditions and dialectal backgrounds was significant in F-statistics, $F_{Stroop:background}(9) = 3.30$, $p < 0.001$. However, the difference did not reach significance in t statistics. The interaction of dialectal backgrounds and dialect of naming was also significant, $F_{background:dial}(1) = 40.07$, $p < 0.001$. The SH group were significantly faster in SC, $t_{SH:inSC}(42301.68) = -4.23$, $p < 0.001$. However, this advantage was largely cancelled in three-way interactions, when the SH group was presented with congruent color characters (c2) or competition from color characters (c5), $t_{c2:SH:inSC}(42302.96) = 2.63$, $p < 0.01$, $t_{c5:SH:inSC}(42303.4) = 2.44$, $p < 0.05$. The other terms of interaction did not reach significance.

4. Discussion

This study employed the Stroop paradigm to investigate how dialectal background affects semantic and phonological processing in Chinese word recognition. The results uncovered classical Stroop interference and automatic activation of semantic information, alongside a beneficial effect of bi-dialectism.

4.1 Typical and atypical Stroop effects

All the five participant groups exhibited classical Stroop interference when presented with competing color characters, as indicated by longer correct naming latencies. These findings are consistent with many previous Chinese Stroop studies (Biederman & Tsao, 1979; Lee & Chan, 2000; Li et al., 2013; Saalbach & Stern, 2004; Smith & Kirsner, 1982; Spinks et al., 2000; Tsao et al., 1981). Therefore, the experimental process and results of the current study are reliable, and classical Stroop effects are also observed in Chinese dialects.

Also, regarding the contrasts between color characters and their exact homophones (c5 vs. c6 and c8 vs. c9), all the five participant groups across the tested dialects showed longer naming latencies when presented with competition from color characters (c5 & c8). This finding replicated Li et al.'s (2013) and Spinks et al.'s (2000) findings. Interestingly, the contrasts between color characters and their non-exact homophones yielded similar results, regardless of whether the tone of the character was consistent (c9), different (c7), or in competition (c6) with the naming target. These findings confirmed that the semantic information of the Chinese characters is automatically activated during the Stroop task, whether the naming is made in SC or a regional dialect.

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

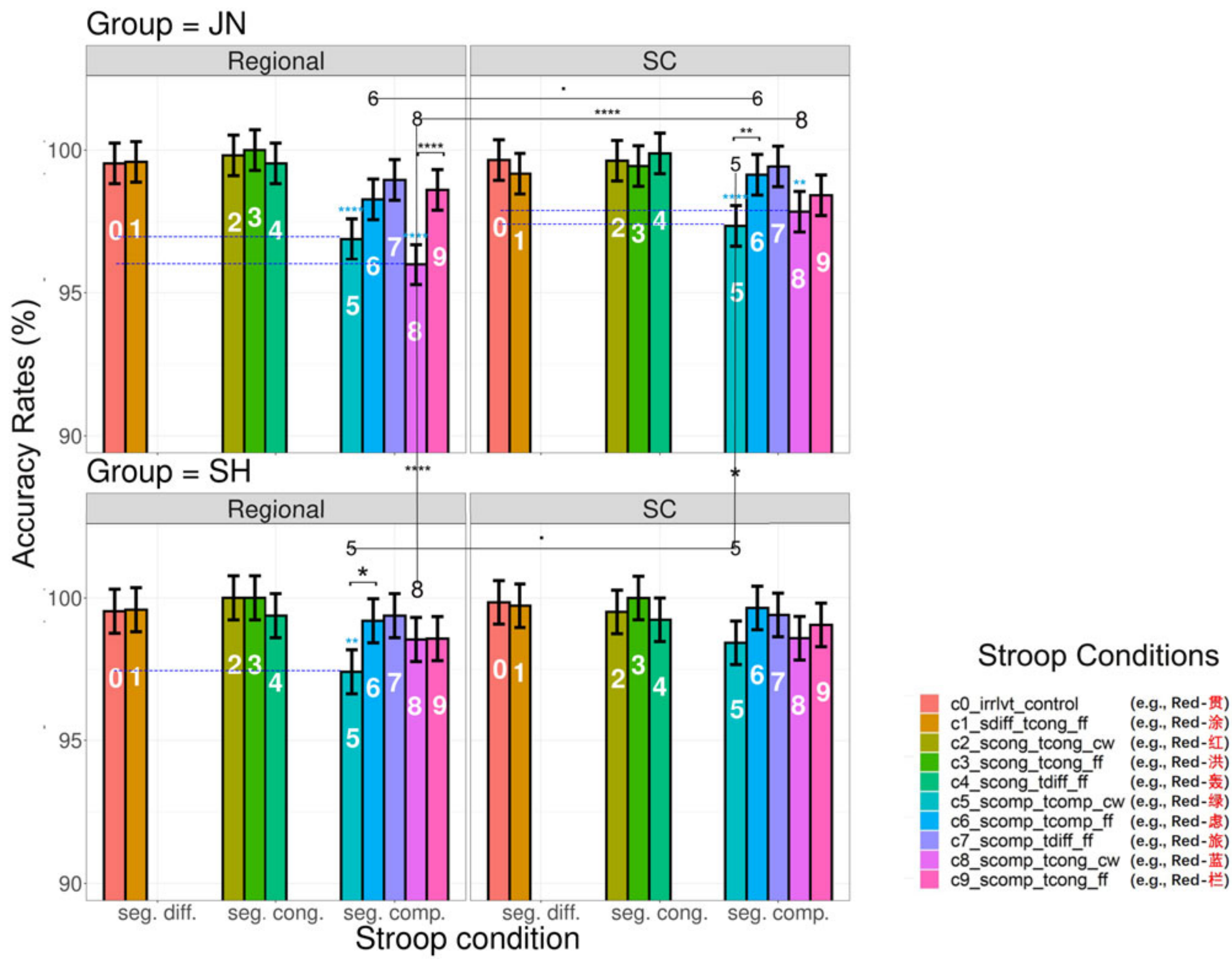

Figure 3. Estimated by-speaker mean accuracy (%) in ten different Stroop conditions by two bi-dialectal groups of participants (JN = SC-JM bi-dialectals from Jinan, SH = SC-SH bi-dialectals from Shanghai) in the participants' regional dialect and SC. (marked with colors, and clustered under segmental conditions, for more information see Table 1, irrlvt_control = irrelevant control, sdiff = segmentally different, scong = segmentally congruent, scomp = segmentally competition, tcong = tonally congruent, tdiff = tonally different, tcomp = tonally competition, cw = color word, ff = false friend; for details of significance markers, see Figure 1 caption; additionally significant contrasts of the same Stroop conditions between the regional dialect and SC, as well between participant groups are also marked with line segments and stars)

Most of the minimal phonological contrasts tested (i.e., comparisons involving tonal congruence and competition: c3 vs. c4, c8 vs. c5, and c9 vs. c6; comparisons involving tonal congruence and difference: c9 vs. c7; and comparisons involving tonal competition and difference: c6 vs. c7) did not yield statistically significant results. Given these null effects, it is important to approach the interpretation with caution. It would be premature to claim the absence of unintended phonological activation in automatic Chinese visual word recognition during Stroop production, as more sensitive paradigms or techniques may detect more subtle effects. Continued exploration using refined approaches might provide further insights into this topic.

However, one phonological effect was verified consistently across participant groups, who all exhibited longer naming latencies when presented with the tone-alone congruent condition (c1) than with the irrelevant baseline (c0). This manipulation is equivalent to the S−T− vs. S−T+ manipulation in Li et al.'s (2013) and Spinks et al.'s (2000) previous works. However, whereas Li et al. (2013) found tone-related facilitation, Spinks et al. (2000) found no significant difference, and we found tone-related interference. Note that we included more color-character combinations for these two conditions (sixteen for c0 and nine for c1). Also note that the interference, although small, was found consistently across participants of different dialectal backgrounds. Previous studies have found evidence of interference based on tone-alone similarity in other tasks (e.g., J.-Y. Chen et al., 2002), suggesting the reliability of our finding.

The current finding may indicate that additional tonal congruency without the precondition of segmental congruency can introduce more similarity-based lexical competition and compromise the selection of the proper word for production. This elucidates the intricate phonological mechanisms underlying visual word identification and sheds light on the potential impact of tonal congruency in shaping the production process. By expanding our knowledge of these processes, this finding contributes to a deeper understanding of how phonological factors influence word-level processing.

4.2 Bi-dialectal experiences confer benefits on conflict resolution

All the three bidialectal groups (JN, SH and OD) showed reduced naming latencies as compared with the BJ group of monolectals,

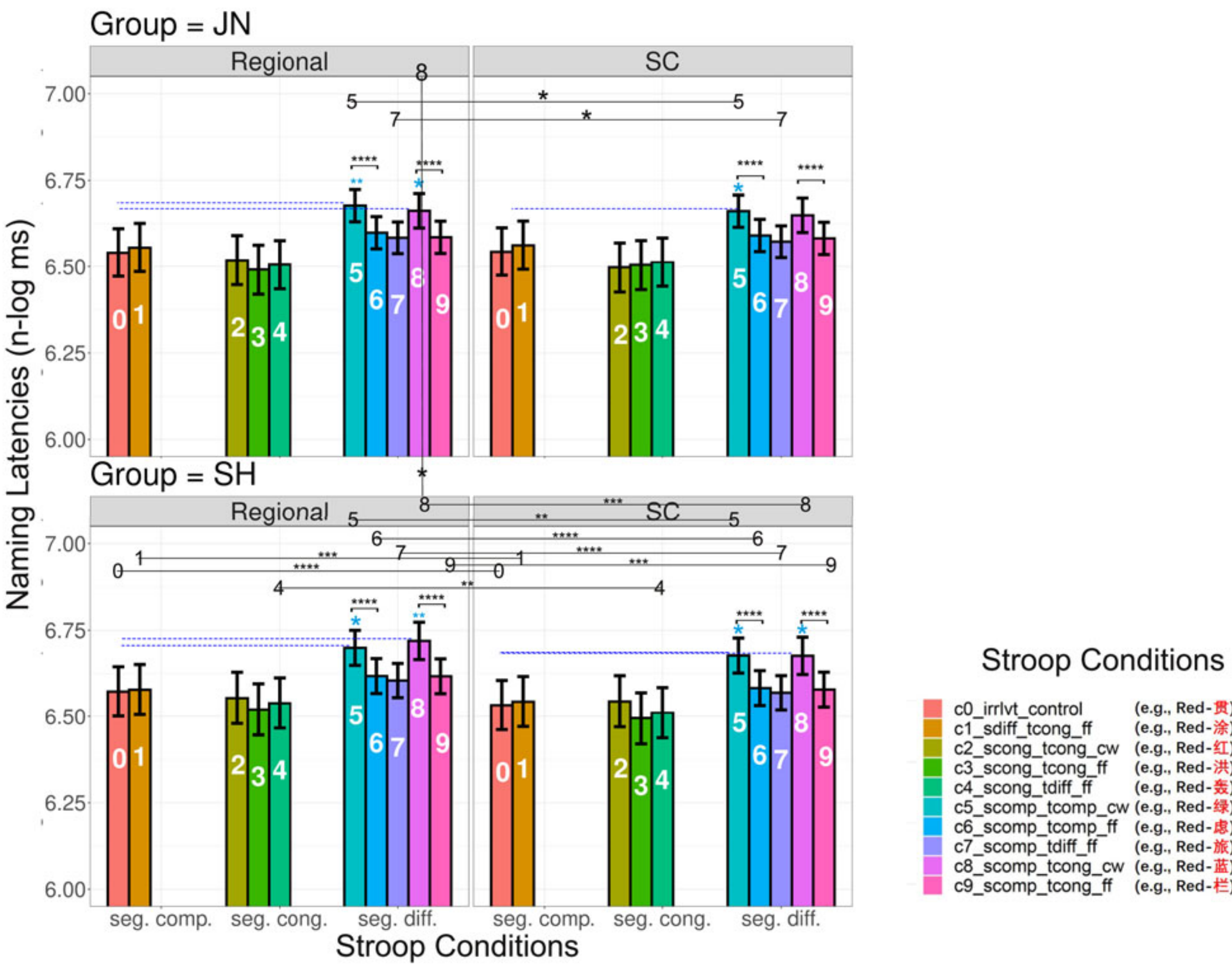

Figure 4. Estimated correct naming latencies (n-log ms) in ten different Stroop conditions by two bi-dialectal groups of participants (JN = SC-JM bi-dialectals from Jinan, SH = SC-SH bi-dialectals from Shanghai) in the participants' regional dialect and SC. (marked with colors, and clustered under segmental conditions, see Table 1 for details, irrlvt_control = irrelevant control, sdiff = segmentally different, scong = segmentally congruent, scomp = segmentally competition, tcong = tonally congruent, tdiff = tonally different, tcomp = tonally competition, cw = color word, ff = false friend; for details of significance markers, see Figure 3 caption).

under at least one of the conditions with segmental incongruency (e.g., c1, c2, c5, c6, c7, c8, or c9). This finding is in line with previous studies that bilinguals are better at solving task conflicts, especially the conflict of the Stroop task (Bialystok et al., 2008; but see Paap et al., 2015). Hence, the bilingual advantage of conflict resolution extends to Chinese bi-dialectals, including speakers of very remote and very similar dialects, indicating that bi-dialectism can also be beneficial for attentional control.

Moreover, the SC group of monolectals, who had more extensive exposure to various dialects as compared to the BJ group, did not exhibit a significant decrease in naming latencies. However, they did demonstrate a lower competition-related reduction in accuracy rates than the BJ group, particularly when c8 was compared to c9. This finding may suggest that monolectals' cognitive control may also benefit from additional exposure to multiple dialects. Alternatively, this finding may also indicate that monolectals develop an altered strategy in adapting to the complex linguistic ecology of multiple dialects. These two explanations are not mutually exclusive, and both contribute to our understanding of the results.

In the control condition (c0), when comparing the OD bi-dialectal group to both the JN and SH bi-dialectal groups, as well as the monolectal groups, it is evident that the OD group exhibited slightly slower naming latencies while maintaining the same level of accuracy. This comparison highlights that the OD bi-dialectals, in particular, take a more cautious approach to SC naming without indicating any lack of competence. In contrast, both the JN and SH bi-dialectal groups exhibited similar levels of accuracy and naming latencies to the monolectal groups in this condition. Thus, overall, there is no evidence of a proficiency difference between the bi-dialectal and monolectal groups in SC. Furthermore, the observed advantage in the bi-dialectal groups specifically applies to Stroop conditions involving segmental competition. These findings support the hypothesis that the advantage in the bi-dialectal groups can be attributed to executive control processes.

4.3 Advantages related to more challenging inter-dialectal experiences

The current study comprised three cohorts of highly skilled bi-dialectals to assess the generalizability of the impact of inter-dialectal experience.

(1) Regarding the comparison between the JN and SH bi-dialectal groups, a general finding was that the SH bi-dialectals showed an advantage when presented with segmental

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

competitions (e.g., c5, c6, c8), by responding with greater accuracy than the JN bi-dialectals. Considering that most JM and SC translation equivalents are segmentally identical, while most SH and SC words are segmentally non-identical, the SH bi-dialectals' advantages may be attributed to their richer experience and hence more fine-tuned attention control regarding the handling of challenging segmental conflicts during speech production.

Nevertheless, note that when naming in their regional dialects under the competition from a tonally congruent color character (c8), the JN bi-dialectals although yielding lower accuracy rates, demonstrated significantly faster responses than the SH bi-dialectals, indicating that they were not more effective, but instead were less cautious. This difference may be attributed to the JN bi-dialectals' experience with the tonal systematic correspondence between their two dialects: words that exhibit tonal congruency in Standard Chinese often have counterparts in Jinan Mandarin that also display tonal congruency. As the JN bi-dialectals are used to such dual-dialectal congruency, they may be more used to reduce their alertness given tonal congruency between competing lexical candidates.

One may propose that the JN and SH bi-dialectal groups developed distinct overarching strategies based on their previous linguistic experiences or during the experimental process. However, it is important to note that we thoroughly examined the two bi-dialectal groups in terms of accuracy and naming latencies in the control condition (c0) and found no significant differences. This indicates that any potential strategical disparities between the groups, if present, are likely confined to the c8 condition and influenced by the distinctive cross-dialectal characteristics of the same stimuli.

(2) The OD bi-dialectals were included for exploring the effects of complex receptive dialectal experience. As expected, with a wider range of dialectal backgrounds, this group displayed a greater degree of intra-group variance than the other groups. As compared with the BJ monolectals, the OD bi-dialectals showed greater accuracy when presented with tonally congruent segmental incongruency (c1 and c8), but longer naming latencies in general. As compared with the SH bi-dialectals who were living in the same city but as locals, OD bi-dialectals showed lower accuracy rates when presented with tonal and segmental competition (c5 and c6). Taking these findings into account, bi-dialectals who have experienced migration and have complex receptive dialectal experience may maintain their advantages in conflict resolution, but may be more cautious than monolectals, although they are still less accurate than local bi-dialectals.

4.4 Unattended character processing when naming colors in quasi-literate Chinese dialects

In previous bilingual Stroop studies, the color naming and printed words were in clearly different languages with different writing systems. In contrast, the present Chinese Stroop experiment suggests that the visually-presented Chinese characters may activate both dialects' lexical phonological representations; one aligned with and the other deviating from the dialect of naming. To investigate bi-dialectal non-attentional processing of Chinese characters, the JN and SH bi-dialectals were tested in their naming of ink colors in both SC and their regional dialects given the same Chinese characters.

A general finding was that SC as the dialect of naming was associated with greater accuracy and shorter naming latencies by both groups under most of the conditions of segmental competition (c5, c6, c7, c8, and c9). Given that this difference was rarely observed under the conditions of segmental congruency and the control condition, it is unlikely that it can be simply attributed to a general processing advantage stemming from the dominant status of SC. It seems more likely that the bi-dialectals had specifically reduced difficulties in conflict resolution during color naming in SC as compared with their regional dialect.

This finding may be explained with theories of asymmetrical cross-linguistic lexical retrieval (e.g., Kroll & De Groot, 1997). More specifically, across Chinese dialects, translation equivalents of color words are etymologically related and associated with identical written forms. It is well known that such cognates can get activated in parallel in bilingual visual word recognition (Canseco-Gonzalez et al., 2010; Lagrou et al., 2011; Marian & Spivey, 2003; Weber & Cutler, 2004), while the dominant language tends to show an advantage (e.g., Brenders et al., 2011; Van Hell & Dijkstra, 2002). SC is systematically taught and much more frequently used in Chinese literacy education than in SH and JM. Thus, it is reasonable to hypothesize that Chinese characters may activate SC lexical phonological representations during color naming in any of the dialects but activate SH and JM lexical phonological representations only when the naming was performed in these regional dialects. This additional cross-dialectal parallel activation of competing color lexical forms may explain the additional lexical interference found in the regional dialects.

Concerning the influence of dialects on this effect, the SH group exhibited a significant regional-dialect disadvantage in naming latencies under more conditions than the JN bi-dialectal group, even when segmental information was irrelevant (c0 and c1) or congruent (c4). This may be attributed to the more independent relations between SH and SC ETEs, which may have exacerbated the conflict in the regional dialect.

Moreover, further taking into consideration the impact of tonal congruency, the competition from color characters without any within-dialectal tonal overlap (c5) introduced significant reduction of accuracy in both JN and SH bi-dialectal groups, and the SH group's regional-dialect-related reduction of accuracy was restricted to this condition. In contrast, while the competition from tonally congruent color characters (c8) severely exacerbated the JN group's regional-dialect-related reduction of accuracy and semantic interference, this interference was completely absent for the SH group.

It turned out that both JNs and SHs encountered cross-dialectal tonal interference when presented with completely competing colors and characters (visit the data availability link for more details). For instance, when the ink was yellow /xuɑŋ(Hr)/ and the competing character was 绿/ly(F)/ 'green', although there was no tonal overlap within SC, the SC tone for the character (F) was close to the JM tone (Hf) for the ink name. Similarly, when the ink was green /ly(F)/ and the competing character was 红 /xuŋ(Hr) 'red', although there was no tonal overlap within SC, the SC tone for the character (Hr) shared the rising pitch with the SH tone (Lr) for the ink name. Thus, it seems that cross-dialectal tonal similarity combined with within-dialectal tonal and lexical competition may increase semantic interference and contribute to the regional-dialect disadvantage.

In contrast, under c8, there are distinct cross-dialectal ink-character tonal relations between the groups. For instance, when the ink was yellow /xuɑŋ(Hr)/ and the competing character was 蓝 /lan(Hr)/ 'blue', beside the tonal congruency of (Hr) within SC, the SC tone for the character (Hr) was also similar

https://doi.org/10.1017/S1366728924000142 Published online by Cambridge University Press

to the SH tone (Lr) for the ink name. However, it happens that under this condition no such cross-dialectal similarity was applicable between JM and SC. Thus, it seems that cross-dialectal tonal similarity combined with within-dialectal tonal congruency may reduce semantic interference and counteract the regional-dialect disadvantage.

These findings suggest that the unattended processing of characters in quasi-literate Chinese dialects is influenced by specific lexical-phonological alignments that exist across the dialects in the mind of bi-dialectal individuals.

4.5 Limitations

This study has two limitations.

First, while we ensured that all participants displayed high proficiency in both SC and their regional dialect, the differing distributions of SC dominant, balanced, and regional dialect dominant individuals among the bi-dialectal groups constitute a limitation that may impact our findings. Future studies could address potential variations in participant characteristics to enhance the generalizability of the results.

Second, due to the constraints of the tonal lexical systems of the dialects examined in this study, it was not feasible to implement more intensive control or manipulation of cross-dialectal phonological conditions. Nevertheless, we have carefully noted all the identifiable specifics and meticulously considered their potential impacts on the findings.

To strengthen the overall applicability of our results, future research could explore larger sample sizes and include more diverse dialectal or language groups. Additionally, investigating alternative methods to control or manipulate cross-dialectal conditions, and thus overcome the limitations inherent in tonal lexical systems, would be valuable. By addressing these limitations, future studies can greatly contribute to our understanding of the intricate interplay between dialectal variations and cognitive processes.

5. Conclusion

To sum up, the current study used the Stroop paradigm to investigate the influence of dialectal experience on visual word recognition in Chinese. Our findings corroborated the theory that semantic and phonological information contained in Chinese characters can be automatically activated during visual word recognition. These findings extend the classical bilingual executive advantage to bi-dialectals, and further found that even receptive dialectal experience can counteract Stroop interference with monolectals. Moreover, the investigation of regional-dialect disadvantage suggested that Chinese characters can automatically activate lexical phonology in both dialects, but the activation is biased with the dialect of naming and lexical-specific inter-dialectal phonological relations. Stroop interference in the bi-dialectals' regional dialect cannot only be modulated with within-dialectal conditions but also with cross-dialectal tonal similarity. These findings establish a connection between lexical prosody and bilingual executive control, offering a new understanding of how the brain processes visual information during word recognition.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728924000142

Competing interests. The author(s) declare none

Ethical statements. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Data availability statement. Data availability: The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/x5gwf/, reference number http://doi.org/10.17605/OSF.IO/X5GWF.

Notes

[1] Here, the terms Hr (High-rising), Hl (High-level), are used to refer to distinct tonal patterns, similarly hereafter.

[2] One common way to define the orthographic overlap of Chinese characters is by looking at the orthographical elements that they have in common, such as strokes, radicals, and components. Many studies factor in these various elements. As stroke overlap is inevitable and showed limited cognitive influence, previous studies usually only manipulate the number of strokes. (For more information, refer to Wu's 2022 review).

[3] This is a heterogeneous group, including fourteen participants who speak Zhongyuan Mandarin, nine who speak Xi'nan Mandarin, seven who speak Jilu Mandarin, two who speak Jianghuai Mandarin, one who speak Jin Mandarin, three who speak Yue (Cantonese), two who speak Min, one who speak Xiang, and two speak boundary dialects between Wu and Gan. For more information, please refer to allmeta_anonumous.xlsx on Open Science Framework.

[4] Naming accuracy data were also directly analyzed with logistic LME, but the models could not converge because in some combination of conditions all the responses were correct.

[5] Visit the data-availability link also for the data and scripts of the estimates.

References

Antón, E., Duñabeitia, J. A., Estévez, A., Hernãndez, J. A., Castillo, A., Fuentes, L. J., Davidson, D. J., & Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00398

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4. R package version* 1.0-4 [Computer software].

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(01), 3–11. https://doi.org/10.1017/S1366728908003477

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(4), 859–873. https://doi.org/10.1037/0278-7393.34.4.859

Biederman, I., & Tsao, Y.-C. (1979). On processing Chinese ideographs and English words: Some implications from Stroop-test results. *Cognitive Psychology*, *11*(2), 125–132. https://doi.org/10.1016/0010-0285(79)90007-0

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer [Computer program] Version* 6.0.3 *retrieved* 2017 *from* http://www.praat.org/ [Computer software].

Brenders, P., Van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology*, *109*(4), 383–396. https://doi.org/10.1016/j.jecp.2011.03.012

Canseco-Gonzalez, E., Brehm, L., Brick, C. A., Brown-Schmidt, S., Fischer, K., & Wagner, K. (2010). Carpet or Carcel: The effect of age of acquisition and language mode on bilingual lexical access. *Language and Cognitive Processes*, *25*(5), 669–705. https://doi.org/10.1080/01690960903474912

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*(2), 282–298. https://doi.org/10.1111/j.1467-7687.2008.00675.x

Chen, H.-C., & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition*, *12*(3), 397. https://doi.org/10.1037/0278-7393.12.3.397

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*(4), 751–781. https://doi.org/10.1006/jmla.2001.2825

Cheng, C.-C. (1997). Measuring relationship among dialects: DOC and related resources. *Computational Linguistics & Chinese Language Processing*, *2*.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*(2), 135–149. https://doi.org/10.1016/j.cognition.2009.08.001

Dalrymple-Alford, E. C. (1972). Associative facilitation and interference in the Stroop color-word task. *Perception & Psychophysics*, *11*(4), 274–276. https://doi.org/10.3758/BF03210377

Dennis, I., & Newstead, S. E. (1981). Is phonological recoding under strategic control? *Memory & Cognition*, *9*(5), 472–477. https://doi.org/10.3758/BF03202341

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, *1*(1), 51–66. https://doi.org/10.1017/S1366728998000121

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental Psychology*, *61*(3), 234–251. https://doi.org/10.1027/1618-3169/a000243

Dyer, F. N. (1971). Color-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 297–302. https://doi.org/10.1016/S0022-5371(71)80057-9

Effler, M. (1978). The influence of similarity in names of Stroop items. *Archiv für Psychologie*, *131*, 21–37.

Fang, S.-P., Tzeng, O. J. L., & Alva, L. (1981). Intralanguage vs. Interlanguage Stroop effects in two types of writing systems. *Memory & Cognition*, *9*(6), 609–617. https://doi.org/10.3758/BF03202355

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220–1234. https://doi.org/10.3758/BF03193224

Gooskens, C., & Van Heuven, V. J. (2021). Mutual Intelligibility. In M. Zampieri & P. Nakov (Eds.), *Similar languages, varieties, and dialects* (pp. 51–95). Cambridge University Press. https://doi.org/10.1017/9781108565080.006

Hernández, M., Martin, C. D., Barceló, F., & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language*, *69*(3), 257–276. https://doi.org/10.1016/j.jml.2013.06.004

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, *18*(4), 625–658. https://doi.org/10.3758/s13423-011-0116-7

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

Kiyak, H. A. (1982). Interlingual interference in naming color words. *Journal of Cross-Cultural Psychology*, *13*(1), 125–135. https://doi.org/10.1177/0022022182131011

Kroll, J. F., & De Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. De Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169–199). Lawrence Erlbaum.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R Package Version*, 2–0.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 952–965. https://doi.org/10.1037/a0023217

Lee, T. M. C., & Chan, C. C. H. (2000). Stroop interference in Chinese and English. *Journal of Clinical and Experimental Neuropsychology*, *22*(4), 465–471. https://doi.org/10.1076/1380-3395(200008)22:4;1-0;FT465

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means* [Computer software]. https://CRAN.R-project.org/package=emmeans

Li, C., Lin, C. Y., Wang, M., & Jiang, N. (2013). The activation of segmental and tonal information in visual word recognition. *Psychonomic Bulletin & Review*, *20*(4), 773–779. https://doi.org/10.3758/s13423-013-0395-2

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, *127*(6), 1139–1162. https://doi.org/10.1037/rev0000248

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

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97–116. https://doi.org/10.1017/S1366728903001068

Nixon, J. S., Chen, Y., & Schiller, N. O. (2015). Multi-level processing of phonetic variants in speech production and visual word processing: Evidence from Mandarin lexical tones. *Language, Cognition and Neuroscience*, *30*(5), 491–505. https://doi.org/10.1080/23273798.2014.942326

Paap, K. R. (2015). Do many hones dull the bilingual whetstone? *Bilingualism: Language and Cognition*, *18(1)*, 41–42. https://doi.org/10.1017/S1366728914000431

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*(4), 265–278. http://dx.doi.org/10.1016/j.cortex.2015.04.014

Pacilly, J. J. A. (2010). *MarkIntervals.praat [Computer program]* 24-*jan*-2010.

Preston, M. S., & Lambert, W. E. (1969). Interlingual interference in a bilingual version of the Stroop color-word task. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 295–301. https://doi.org/10.1016/S0022-5371(69)80079-4

Prior, A. (2012). Too much of a good thing: Stronger bilingual inhibition leads to larger lag-2 task repetition costs. *Cognition*, *125*(1), 1–12. https://doi.org/10.1016/j.cognition.2012.06.019

Ransdell, S. E., & Fischler, I. (1987). Memory in a monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language*, *26*(4), 392–405. https://doi.org/10.1016/0749-596X(87)90098-2

Risko, E. F., Stolz, J. A., & Besner, D. (2005). Basic processes in reading: Is visual word recognition obligatory? *Psychonomic Bulletin & Review*, *12*(1), 119–124. https://doi.org/10.3758/BF03196356

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*(03), 465–485. https://doi.org/10.1017/S014271640606036X

Saalbach, H., & Stern, E. (2004). Differences between Chinese morphosyllabic and German alphabetic readers in the Stroop interference effect. *Psychonomic Bulletin & Review*, *11*(4), 709–715. https://doi.org/10.3758/BF03196624

Sakamoto, Y., & Ishiguro, M. (1986). *Akaike information criterion statistics*. Reidel.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated.

Singer, M. H., Lappin, J. S., & Moore, L. P. (1975). The interference of various word parts on color naming in the Stroop test. *Perception* & *Psychophysics*, *18*(3), 191–193. https://doi.org/10.3758/BF03205966

Smith, M. C., & Kirsner, K. (1982). Language and orthography as irrelevant features in colour-word and picture-word Stroop interference. *The Quarterly Journal of Experimental Psychology Section A*, *34*(1), 153–170. https://doi.org/10.1080/14640748208400864

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Tang, C., & Van Heuven, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, *119*(5), 709–732. https://doi.org/10.1016/j.lingua.2008.10.001

Tsao, Y., Wu, M., & Feustel, T. (1981). Stroop interference: Hemispheric difference in Chinese speakers. *Brain and Language*, *13*(2), 372–378. https://doi.org/10.1016/0093-934X(81)90101-2

Van der Loo, M. P. (2010). Distribution based outlier detection for univariate data. *Statistics Netherlands*, 10003.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780–789. https://doi.org/10.3758/BF03196335

Van Heuven, W. J. B., Conklin, K., Coderre, E. L., Guo, T., & Dijkstra, T. (2011). The influence of cross-language similarity on within-and between-language Stroop effects in trilinguals. *Frontiers in Psychology*, *2*, 374. https://doi.org/10.3389/fpsyg.2011.00374

Van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex*, *18*(11), 2706–2716. https://doi.org/10.1093/cercor/bhn030

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Wu, J. (2022). Neurocognition in the reading of Chinese characters. *Journal of Chinese Writing Systems*, *6*(3), 167–184. https://doi.org/10.1177/25138502211061605

Wu, J., Chen, Y., Van Heuven, V. J., & Schiller, N. O. (2016). Predicting tonal realizations in one Chinese dialect from another. *Speech Communication*, *76*, 1–27. https://doi.org/10.1016/j.specom.2015.10.006

Wu, J., Chen, Y., Van Heuven, V. J., & Schiller, N. O. (2018). Applying functional partition in the investigation of lexical tonal-pattern Categories in an under-resourced Chinese dialect. In J. Tao, T. F. Zheng, C. Bao, D. Wang, & Y. Li (Eds.), *Man-machine speech communication* (Vol. 807, pp. 24–35). Springer Singapore. https://doi.org/10.1007/978-981-10-8111-8_3

Zhou, X., & Marslen-Wilson, W. (1999). Phonology, orthography, and semantic activation in reading Chinese. *Journal of Memory and Language*, *41*(4), 579–606. https://doi.org/10.1006/jmla.1999.2663