The influence of cross-language similarity on within- and between-language Stroop effects in trilinguals

Walter J. B. van Heuven[1]*, Kathy Conklin[2], Emily L. Coderre[1], Taomei Guo[3] and Ton Dijkstra[4]

[1] School of Psychology, University of Nottingham, Nottingham, UK
[2] School of English Studies, University of Nottingham, Nottingham, UK
[3] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China
[4] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands

This study investigated effects of cross-language similarity on within- and between-language Stroop interference and facilitation in three groups of trilinguals. Trilinguals were either proficient in three languages that use the same-script (alphabetic in German–English–Dutch trilinguals), two similar scripts and one different script (Chinese and alphabetic scripts in Chinese–English–Malay trilinguals), or three completely different scripts (Arabic, Chinese, and alphabetic in Uyghur–Chinese–English trilinguals). The results revealed a similar magnitude of within-language Stroop interference for the three groups, whereas between-language interference was modulated by cross-language similarity. For the same-script trilinguals, the within- and between-language interference was similar, whereas the between-language Stroop interference was reduced for trilinguals with languages written in different scripts. The magnitude of within-language Stroop facilitation was similar across the three groups of trilinguals, but smaller than within-language Stroop interference. Between-language Stroop facilitation was also modulated by cross-language similarity such that these effects became negative for trilinguals with languages written in different scripts. The overall pattern of Stroop interference and facilitation effects can be explained in terms of diverging and converging color and word information across languages.

Keywords: trilinguals, Stroop, interference, facilitation, script

INTRODUCTION

Proficient bilinguals are able to communicate in both of their languages without much difficulty. This is true whether the languages they speak are highly similar in terms of orthography and phonology (e.g., German and Dutch) or dissimilar (e.g., Chinese and English). The ease of bilingual communication is surprising in light of a large body of research that has demonstrated that lexical access is non-selective with respect to both language comprehension (e.g., Dijkstra et al., 1998; van Hell and de Groot, 1998; van Heuven et al., 1998; de Groot et al., 2000; Jared and Kroll, 2001; for a review, see Dijkstra and van Heuven, 2002) and language production (e.g., Hermans et al., 1998; Colomé, 2001; Guo and Peng, 2006; Costa et al., 2008; Hoshino and Kroll, 2008). Language non-selective access implies that word representations from both languages are active during processing, even when only one language is relevant to the situation or task at hand. Because an irrelevant language is often coactivated during processing, bilinguals must rely on cognitive control to respond in the appropriate language.

An important issue on the bilingual research agenda is how the interaction between languages is affected by their similarity or dissimilarity. It is important to explore, for instance, how cross-language similarity, in terms of phonological and/or orthographic/script overlap, influences the bilingual/multilingual language processing system and whether potential differences between languages, such as script, can be exploited to reduce the amount of cross-language interference, thereby influencing how much cognitive control is required to speak exclusively in the target language.

A task that is well-suited to investigate issues of cognitive control and cross-language similarity in bilingual processing is the Stroop task (Stroop, 1935). In a color naming Stroop task, color words are presented in colored ink and participants are asked to ignore the printed word and instead name the color of the ink. To avoid reading the printed word aloud, this task requires averting the highly practiced reading process. In incongruent conditions (where word and ink color do not match; e.g., "red" printed in green ink), the conflicting word and color information requires cognitive control and conflict resolution processes to be engaged, leading to a delay in response times (RTs) compared to control conditions (typically a non-linguistic or non-response set stimulus printed in colored ink; e.g., "XXXX" printed in blue). This delay is referred to as *Stroop Interference*. In contrast, *Stroop Facilitation* refers to the faster RTs in congruent conditions (where word and color match; e.g., "blue" printed in blue ink) than in control conditions. In a multilingual version of the Stroop task, input and output languages can be manipulated so that within- and between-language interference and facilitation effects can be investigated. In what follows, we will first discuss within- and between-language interference and then facilitation effects in the Stroop task.

STROOP INTERFERENCE: WITHIN- AND BETWEEN-LANGUAGE

In a traditional monolingual Stroop task, interference is generally thought to be due to conflicting color and word information (Roelofs, 2010). Thus, seeing "red" printed in green ink leads to long RTs for naming the ink color due to the diverging available information from the color and word. An important question is how this interference is modulated within- and between-languages. Does seeing "red" in blue ink when /blu/ is the required response in an English task, produce a similar amount of interference for German–English bilinguals as seeing "rot" ("red" in German)? In terms of semantics, both *red* and *rot* provide information that diverges from that of the ink color (blue), so one might expect similar degrees of within-language (intralingual) and between-language (interlingual) interference. However, based on his survey of the bilingual Stroop literature, MacLeod concludes that "*Interference between the two languages of a bilingual, although not as great as that within either one of the languages, is very robust: Between-language interference typically is about* 75% *of within-language interference*... There may also be differences in the processing of orthographic and idiographic languages..." (1991, p. 187).

MacLeod's view is supported by early research from Preston and Lambert (1969), who found that between-language interference was only 68% of the within-language interference for English–Hungarian bilinguals, but 95% for French–English bilinguals. Similarly, a study by Dyer (1971) with Spanish–English bilinguals showed that between-language effects were 63% of within-language ones. In a study with Chinese–English, Spanish–English, and Japanese–English bilinguals, Fang et al. (1981) also found greater within- than between-language interference. Interestingly, the between-language effect was modulated by the orthographic similarity of the two languages, such that more overlap lead to stronger Stroop interference in the between-language condition. Crucially, if orthographic similarity underpins the modulation of between-language interference, there should be a larger amount of Stroop interference when the two languages of bilinguals have similar scripts (e.g., alphabetic) than when the languages are written in different scripts (e.g., logographic and alphabetic).

The finding of larger within- than between-language Stroop interference in bilinguals (e.g., Fang et al., 1981; Mägiste, 1984; Tzelgov et al., 1990; Lee et al., 1992; Brauer, 1998) and trilinguals (Abunuwara, 1992) has been termed the within-language Stroop superiority effect (WLSSE; Goldfarb and Tzelgov, 2007). Research has demonstrated that the WLSSE is modulated by cross-language similarity and the proficiency of the participants (e.g., Preston and Lambert, 1969; Fang et al., 1981; Mägiste, 1984; Chen and Ho, 1986; Brauer, 1998; Sumiya and Healy, 2004, 2008). For example, Chen and Ho (1986) conducted a Stroop task with Chinese–English bilinguals in five different age groups. When responses were in the first language (L1) Chinese, all age groups showed greater within- than between-language interference. When responses were in the second language (L2) English, there was a shift from greater between-language interference for the youngest group to greater within-language interference for the oldest three groups.

Similarly, Brauer (1998) conducted two Stroop studies with high and low proficiency bilinguals in languages with high (German, English) and low (English–Greek or English–Chinese) overlap. He found that the low proficiency bilinguals, regardless of how much the languages overlapped, showed more within- than between-language interference when they were required to respond in their L1, and the opposite pattern when they responded in their L2. In the case of high proficiency participants speaking languages with no overlap, there was greater within- than between-language interference when they responded both in the L1 and in the L2. Finally, in high proficiency participants of languages with high overlap, there was no difference between within- and between-language interference effects.

In Sumiya and Healy (2004), Japanese–English bilinguals engaged in a Stroop task in Japanese and English. Color words were used that were phonologically similar across the two languages (/bru:/ and /blu/, with Katakana and English scripts, respectively) or different (/ao/ and /blu/, with Hiragana and English scripts, respectively). Even though script provided a strong cue about the task-relevant language, a significant between-language Stroop effect arose that was larger for phonologically similar words. In a similar study with English–Japanese bilinguals, Sumiya and Healy (2008) found a between-language interference effect that was larger for phonologically similar words, in particular when responses were in L2 Japanese. Additionally, the size of the phonological effect increased with proficiency in Japanese, which was taken as an indication of increased phonological processing when speakers were more proficient in their L2. Such results suggest that not only the degree of form overlap (orthographic and phonological) may modulate interference effects in trilinguals, but proficiency and response language (L1 or L2) as well.

With respect to the WLSSE, it must be considered that when the language of the written word is different from the response language, the influence of the irrelevant language might be minimized through inhibitory control (Green, 1998) or decision criteria (Dijkstra and van Heuven, 2002). Alternatively, response set competition might be involved (Roelofs, 2003; Goldfarb and Tzelgov, 2007). Goldfarb and Tzelgov (2007) examined the cause of the WLSSE by having Hebrew–English bilinguals name an ink color when the distractor was either a color word (*red*, *green*, *blue*) or a color associated word (*tomato*, *grass*, *sky*). In the between-language condition, both the color and associated words were in the irrelevant language. However, color words, but not color associated words, demonstrated larger within- than between-language effects. It was proposed that in the case of color words in the between-language condition, activation at the conceptual level provides activation for items in the response set, while activation at the lexical level does not. This would then induce less interference than in the within-language condition, where the color word activates items in the response set at both the semantic and lexical levels, thereby increasing competition. In the case of color associated words, neither the words in the within-language condition nor in the between-language condition are part of the response set; therefore the WLSSE was not observed.

STROOP FACILITATION: WITHIN- AND BETWEEN-LANGUAGE

Let us now focus on Stroop facilitation, which arises from the difference between responses in the congruent condition (e.g., the word "red" written in red ink) and the control condition (e.g., row of X's in red ink). Typically, congruent trials are processed faster than control trials. There is disagreement in the Stroop literature about the locus of Stroop facilitation. According to the *converging information hypothesis*, in congruent trials the information available from the ink color and the word "converge": they support the correct response, which leads to faster RTs (e.g., Cohen et al., 1990; Melara and Algom, 2003; Roelofs, 2003, 2010). According to the *inadvertent reading hypothesis*, there are attentional lapses on some trials that result in the color word being read out instead of the ink color being named (MacLeod and MacDonald, 2000; Kane and Engle, 2003). In incongruent trials, this inadvertent reading yields an incorrect response. However, in congruent trials such errors are undetectable. They therefore contaminate RTs with incorrect responses to the color word and may lead to apparent but invalid facilitation effects. Research with bilinguals offers a way of testing the two hypotheses (Roelofs, 2010), because a between-language version of the Stroop task allows such reading errors to be detected (e.g., reading the German "blau" printed in blue ink when producing /blu/ is the required response). If previously observed Stroop facilitation effects arise from undetected covert reading errors, they can be eliminated in a between-language task where such errors are apparent. Thus, any facilitation in the between-language condition is not underpinned by invalid facilitation and would be due to converging information (Roelofs, 2010).

Neither the inadvertent reading hypothesis nor the converging information hypothesis makes explicit predictions about how cross-language similarity might affect Stroop facilitation effects (however, see Roelofs, 2010, for an account of how diverging information at the word form level affects Stroop facilitation). It could be the case that inadvertent reading would not occur when script provides a strong cue. However, even if inadvertent reading occurs less when scripts are different, the claim that there will be more within- than between-facilitation still stands. Because facilitation is underpinned by inadvertent reading, these trials are removed in the between-language condition regardless of script or language overlap (unless the color words are absolutely identical in pronunciation) and thus the invalid facilitation is removed. To summarize, according to the inadvertent reading hypothesis, bilinguals and multilinguals should show within- but not between-language facilitation. In contrast, if facilitation stems from converging information, it should occur whenever word and color information converge. However, the question remains as to whether the degree of cross-language convergence modulates the facilitation effect.

Even though bilingual and multilingual research can shed light on whether Stroop facilitation is caused by inadvertent reading or converging information, not many studies in the Stroop literature have focused on facilitation effects in bilinguals and multilinguals. The limited research indicates that between-language Stroop facilitation is modulated by cross-language similarity, such that a negative or interference effect is apparent when languages are more dissimilar. Thus, when the color and the meaning of the word are congruent and the input and output languages differ (i.e., the presented word is a translation equivalent of the word that has to be produced), responses are delayed relative to a control condition when the languages are different, whereas responses are faster when they are similar. For example, Abunuwara (1992) conducted a Stroop task with Arabic–Hebrew–English trilinguals. Although not reported or analyzed as such, congruent trials in the within-language condition yielded a 45-ms facilitation effect relative to the control condition. In contrast, congruent trials in the between-language condition lead to an interference effect of 58 ms. The presence of interference supports the view that the irrelevant language is activated and slows RTs to the ink color. Furthermore, in Experiment 4 of Roelofs (2010) with Dutch–English bilinguals, the between-language facilitation effect appears to be absent at a stimulus onset asynchrony (SOA) of 0 ms. In contrast, MacLeod and MacDonald (2000) reported interference in French–English bilinguals in the between-language congruent condition.

In sum, the review of previous Stroop-research in bilinguals and multilinguals suggests that within-language facilitation should arise in all languages of trilinguals, irrespective of the script involved. However, the picture is less clear for the between-language congruent condition, which may or may not elicit faster RTs compared to the appropriate control condition and might even yield slower RTs. There is, as far as we know, only one study in the literature that has looked at Stroop effects in trilinguals (Abunuwara, 1992). However, this study only focused on Stroop interference and involved only a group of different script trilinguals. Thus, the current research is the first to consider the nature of between-language facilitation in trilinguals whose languages overlap to varying degrees in terms of their script and the orthographic/phonological similarity of their color word translations.

THE PRESENT STUDY

The above literature review suggests that within-language Stroop interference should be apparent in all three languages of trilinguals irrespective of the script involved. Between-language interference should overall be less than within-language interference. In addition, it may be modulated by factors such as script similarity and/or the form overlap (orthographic and phonological) of the color word translations, such that increased similarity may lead to more between-language interference. In terms of Stroop facilitation, there should be evidence of within-language facilitation that is unaffected by script in all languages of the trilinguals. However, previous research is equivocal on whether faster naming responses would be expected in the between-language congruent condition. If between-language Stroop facilitation arises, it might be modulated by language similarity, such that there is more Stroop facilitation when the languages have greater overlap.

In the present study, three groups of trilinguals performed a Stroop color naming task that involved three colors (red, green, and blue). The response language was blocked and the stimulus language was manipulated within each block. Two control conditions were included in each block: a color patch and a control stimulus (e.g., %). The results were analyzed in terms of within- and between-language Stroop interference (incongruent–control stimulus) and facilitation (control stimulus–congruent) to investigate whether cross-language similarity modulates between-language interactions.

Experiment 1 was conducted with German–English–Dutch (GED) trilinguals. In German, English, and Dutch, the color word translations (e.g., *rot–red–rood*) overlapped not only in terms of semantics but also in script (all alphabetic), orthography, and phonology (same orthographic/phonological onset). Experiment 2 involved Chinese–English–Malay (CEM) trilinguals. In Chinese, English, and Malay, the script is shared in English and Malay (both alphabetic) but differs from Chinese. Furthermore, orthography and phonology of the color word translations differ across all three languages [e.g., 红 (hong)–*red–merah*]. Finally, Experiment 3 was conducted with Uyghur–Chinese–English (UCE) trilinguals. The color translations between Uyghur, Chinese, and English are completely different in terms of script, orthography, and phonology [e.g., قىزىل (gizil)–红 (hong)–*red*].

EXPERIMENT 1: GERMAN–ENGLISH–DUTCH TRILINGUALS
METHOD
Participants
Thirty GED trilinguals (eight males) participated in the experiment. All participants studied at the Radboud University in Nijmegen, the Netherlands. They were first language German speakers proficient in English and Dutch. Furthermore, most of them were also proficient in one or more other languages (e.g., French, Spanish, Italian). **Table 1** provides an overview of their mean age and their subjective proficiency scores for each language (scale: from 1 = very poor to 7 = fluent), as well as the year of the first contact with each language and the number of years of experience with each language. In this and the following experiments, the order of the year of first contact with each of the three languages was used to determine the first (L1), second (L2), and third (L3) language of the trilinguals.

MATERIALS AND DESIGN
The stimuli used in the Stroop task were the English color words *red*, *green*, and *blue*, the corresponding color words in Dutch (*rood*, *groen*, *blauw*) and German (*rot*, *grün*, *blau*), control stimuli (row of percent signs), and color patches of red, green, and blue. Colored rectangles about 10 cm × 5 cm (248 × 142 pixels) were used to present the colors. The center of each colored rectangle contained a small black rectangle (142 × 42 pixels) with a color word

or control stimulus presented in a white lowercase Courier font (32-point). For each language, a control stimulus was constructed that matched the length of the color word in the languages (e.g., %%% for *red* and *rot*; %%%% for *rood*). The color patch controls were fully colored rectangles (248 × 142 pixels). In total 39 stimuli were created (3 color patches, 9 control stimuli, 9 congruent stimuli, and 18 incongruent stimuli) that were repeated a number of times in such a way that for each output language there were 108 trials: 36 congruent, 36 incongruent, 18 control stimuli, and 18 color patches.

PROCEDURE AND APPARATUS
A Sennheiser headset (PC 161) was connected to a PC and DMDX (Forster and Forster, 2003) was used to present the stimuli, to measure the voice onset latency and to record the vocal response. Each trial started with a fixation sign (+) presented for 500 ms at the center of the 17″ monitor (1024 × 768 pixels, 85 Hz). Next, a blank screen appeared for 300 ms and then the stimulus was presented until the participant responded vocally or for 2000 ms. After 1000 ms the next trial started. Participants were instructed to ignore the letter strings (color words and control stimuli) and to overtly name the color of the rectangle as fast as possible without making any errors. Participants performed the Stroop task in each of their three languages separately. Thus, output language was blocked. At the beginning of each block the required output language was indicated. The order of the output language was counterbalanced across participants. Within blocks all stimuli were randomized differently for each participant so that in contrast to the output language the input language was randomized within blocks. After the experiment participants filled out a language background questionnaire.

ANALYSIS
CheckVocal (Protopapas, 2007) was used to check whether vocal responses were correct and to find and correct voice key errors. Responses outside ±2.5 SD of each subject mean across all trial types were excluded. For the RT analysis, erroneous responses were removed and the mean RTs were calculated. In all ANOVAs a Greenhouse–Geisser correction was applied when the assumption of sphericity was violated, and all reported *p*-values from *post*

Table 1 | Subjective proficiency scores (scale: 1 = very poor to 7 = fluent) and subject demographics for the trilinguals in Experiments 1–3.

Trilinguals	n	Age	Language	Subjective proficiency scores					First contact	Years of experience
				Speaking	Listening	Reading	Writing	Overall		
Experiment 1:	30	23.2	German	6.9	7.0	7.0	6.7	6.9	0.0	22.9
German–English–Dutch (GED)			English	4.4	5.5	5.8	4.4	5.0	9.4	12.4
			Dutch	4.8	5.7	6.1	4.7	5.3	19.1	3.7
Experiment 2:	24	21.8	Chinese	6.6	6.6	5.8	5.0	6.0	1.6	18.4
Chinese–English–Malay (CEM)			English	5.3	5.7	5.8	5.4	5.5	3.7	17.0
			Malay	3.8	5.0	5.1	4.0	4.5	5.6	14.3
Experiment 3:	32	22.4	Uyghur	6.3	6.6	6.1	5.8	6.2	0.2	22.1
Uyghur–Chinese–English (UCE)			Chinese	5.2	6.3	5.7	5.0	5.5	8.7	14.1
			English	3.7	4.5	4.7	3.8	4.2	15.2	7.3

Table 2 | Mean RTs and SE of the congruent, incongruent, and control conditions for each input and output language combination in Experiments 1–3.

	Input	Output language					
		German		**English**		**Dutch**	
		Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Experiment 1, GED	German	544 (12)	653 (17)	569 (14)	651 (14)	599 (12)	695 (13)
	English	571 (13)	625 (11)	559 (14)	648 (14)	644 (17)	687 (15)
	Dutch	556 (13)	669 (18)	572 (11)	653 (16)	592 (13)	685 (15)
	%%%%	574 (12)		577 (13)		618 (12)	
	Patch	558 (10)		577 (13)		596 (11)	
		Chinese		**English**		**Malay**	
		Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Experiment, 2, CEM	Chinese	545 (18)	618 (20)	564 (16)	603 (22)	596 (20)	596 (18)
	English	584 (20)	609 (23)	537 (19)	623 (20)	593 (22)	644 (24)
	Malay	573 (19)	617 (15)	571 (20)	640 (25)	549 (18)	665 (21)
	%%%%	573 (21)		563 (19)		577 (15)	
	Patch	577 (20)		552 (18)		566 (14)	
		Uyghur		**Chinese**		**English**	
		Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Experiment 3, UCE	Uyghur	637 (20)	724 (27)	711 (20)	761 (27)	761 (24)	789 (30)
	Chinese	675 (20)	704 (27)	690 (19)	769 (23)	734 (22)	777 (26)
	English	683 (21)	705 (24)	760 (26)	766 (22)	695 (20)	822 (29)
	%%%%	642 (19)		716 (23)		731 (19)	
	Patch	651 (21)		692 (20)		742 (24)	

GED, German–English–Dutch; CEM, Chinese–English–Malay; UCE, Uyghur–Chinese–English.

hoc t-tests were Bonferroni-corrected. We first calculated Stroop interference and facilitation effects based on the raw means and then analyzed whether the magnitude of interference and facilitation effects were modulated by input and output languages. Because error rates were very low ($<1.7\%$) no error analyses were conducted.

RESULTS AND DISCUSSION
The total percentage of errors was 1.66%, and the percentage of outliers was 0.79%. The mean RTs for all conditions are presented in **Table 2**. In all subsequent analyses, we treated the control character as the control condition when calculating Stroop interference and facilitation effects. Separate analysis of the control conditions (control characters vs. color patch) by means of a two-way ANOVA on the mean RTs across all language outputs revealed a significant effect of control type, $F(1,29) = 12.27$, $p < 0.01$, $\eta^2 = 0.01$. Analyses using the control patch yielded the same main effects and interactions as the analyses using the control character.[1]

---
[1]A 3 (input) × 3 (output) repeated-measures ANOVA of the Stroop interference comparison showed there was an interaction of input and output language, $F(4,116) = 2.52$, $p < 0.05$. A significant effect was found of language input on German output, $F(2,58) = 6.95$, $p < 0.01$, but not in English or Dutch output.

STROOP INTERFERENCE (INCONGRUENT VS. CONTROL CONDITION)
The means of Stroop interference for all input and output language combinations are presented in **Table 3**. Bonferroni-corrected paired-sample t-tests revealed significant Stroop interference, with all ps < 0.0001, for each input and output language combination.

To investigate whether the magnitude of interference varied across input and output languages, a 3 (input language) × 3 (output language) repeated-measures ANOVA was conducted. This analysis revealed a significant interaction between input and output language, $F(4,116) = 2.52$, $p < 0.05$, $\eta^2 = 0.03$. For the German output, there was a significant main effect of input language, $F(2,58) = 6.95$, $p < 0.01$, $\eta^2 = 0.09$. *Post hoc* paired-sample t-tests showed a smaller Stroop interference for English input (51 ms, SE = 7 ms) than for German (79 ms, SE = 12 ms), $t(29) = 2.54$,

---
A similar Stroop facilitation analysis revealed an interaction between input and output, $F(4,116) = 10.77$, $p < 0.0001$, due to a significant effect of language input for German output, $F(2,58) = 6.12$, $p < 0.01$, and Dutch output, $F(2,58) = 27.42$, $p < 0.0001$, but not English output. A 2 (type: within or between) × 3 (output) repeated-measures ANOVA showed a significant effect of type on the magnitude of the Stroop facilitation effect, $F(1,29) = 40.49$, $p < 0.0001$, but not on the magnitude of the Stroop interference effect. As mentioned, all these effects do not differ from those found with the control character.

Table 3 | Magnitude of the Stroop interference and facilitation effects in Experiments 1–3.

	Input language	Interference (incongruent–control)			Facilitation (control–congruent)		
		Output language			Output language		
		German	English	Dutch	German	English	Dutch
Experiment 1, GED	German	79 (12)***	74 (9)***	77 (7)***	30 (8)*	7 (7)	20 (6)¥
	English	51 (7)***	72 (7)***	69 (8)***	3 (9)	18 (7)	−26 (10)
	Dutch	95 (13)***	77 (8)***	67 (10)***	18 (8)	5 (7)	26 (8)¥
		75	74	71	17	10	7
		Chinese	English	Malay	Chinese	English	Malay
Experiment 2, CEM	Chinese	45 (9)**	40 (11)¥	19 (12)	28 (10)§	−1 (7)	−19 (11)
	English	36 (9)*	60 (10)***	67 (15)*	−11 (12)	26 (8)¥	−16 (12)
	Malay	44 (11)*	77 (12)***	88 (15)***	1 (14)	−9 (10)	28 (11)
		42	59	58	6	5	−2
		Uyghur	Chinese	English	Uyghur	Chinese	English
Experiment 3, UCE	Uyghur	81 (14)***	45 (12)*	57 (18)¥	6 (7)	5 (8)	−29 (12)
	Chinese	62 (16)*	52 (11)**	45 (12)*	−32 (10)¥	26 (12)	−4 (11)
	English	63 (13)**	50 (13)*	91 (16)***	−40 (10)*	−44 (13)¥	36 (8)*
		69	49	64	−22	−4	1

Magnitudes are shown in milliseconds with SE in parentheses. Significant effects after Bonferroni corrections are indicated: $^§p < 0.10$; $^¥p < 0.05$; $^*p < 0.01$; $^{**}p < 0.001$; $^{***}p < 0.0001$. GED, German–English–Dutch; CEM, Chinese–English–Malay; UCE, Uyghur–Chinese–English. A negative value for Stroop facilitation indicates that congruent condition was slower than the control condition.

$p = 0.05$ corrected, $\eta^2 = 0.18$, and for Dutch input (95 vs. 51 ms), $t(29) = 3.38$, $p < 0.01$, $\eta^2 = 0.28$. In contrast, no effect of language input was found on Stroop interference for English output, $F(2,58) < 1$, and Dutch output, $F(2,58) < 1$.

To compare within- vs. between-language effects, we performed a 2 (type: within or between) × 3 (output language) repeated-measures ANOVA. This analysis revealed no main effect of type, $F(1,29) < 1$, indicating an equal amount of Stroop interference within-languages (73 ms, SE = 10 ms) and between-languages (74 ms, SE = 9 ms). No effect of language output or interaction between type and output language was found (all ps > 0.58). Equal within- and between-language interference has previously been reported with Dutch–English bilinguals (Roelofs, 2010). This can be explained in terms of the high cross-language similarity between the color word translations in terms of orthography and phonology (*red–rood–rot*, *green–groen–grun*, and *blue–blauw–blau* in English, Dutch, and German respectively). In fact, most of these translations can be considered to be cognates (same meaning and very similar orthography and phonology across languages). There is a wide literature that suggests that cognates have a special status in the multilingual lexicon because their processing differs from matched non-cognates (e.g., Dijkstra et al., 1998, 1999; Costa et al., 2000; van Hell and Dijkstra, 2002; Lemhöfer and Dijkstra, 2004; Lemhöfer et al., 2004; Hoshino and Kroll, 2008). If there is a special status for these cognates and they are activated in parallel across the three languages, within- and between-language interference should be similar.

STROOP FACILITATION (CONTROL VS. CONGRUENT CONDITION)

The size of the Stroop facilitation across input and output languages is also shown in **Table 3**. Significant Stroop facilitation was found for German output when the input language was German, $t(29) = 3.75$, $p < 0.01$, $\eta^2 = 0.33$. Furthermore, facilitation was observed for Dutch output when Dutch was the input language, $t(29) = 3.49$, $p < 0.05$, $\eta^2 = 0.30$, and German, $t(29) = 3.20$, $p < 0.05$, $\eta^2 = 0.26$.

The 3 (input language) × 3 (output language) repeated-measures ANOVA on the magnitude of Stroop facilitation effects revealed an interaction between input and output languages, $F(4,116) = 10.77$, $p < 0.0001$, $\eta^2 = 0.07$. For German output, a significant effect of language input was found, $F(2,58) = 6.12$, $p < 0.01$, $\eta^2 = 0.06$. Paired-sample t-tests revealed only a significant difference in facilitation magnitude between German and English input (30 ms, SE = 8 vs. 3 ms, SE = 9 ms), $t(29) = 3.14$, $p < 0.05$, $\eta^2 = 0.25$. For English output, no effect of language input was found, $F(2,58) = 1.22$, $p = 0.30$. A significant effect of language input was found for Dutch output, $F(1.62,47.0) = 27.42$, $p < 0.0001$, $\eta^2 = 0.22$. Paired-sample t-tests showed a significant difference in the magnitude of facilitation effects for German and English input (20 ms, SE = 6 vs. −26 ms, SE = 10 ms), $t(29) = 5.14$, $p < 0.0001$, $\eta^2 = 0.48$, and for Dutch and English input (26 ms, SE = 8 vs. −26 ms, SE = 10 ms), $t(29) = 6.38$, $p < 0.0001$, $\eta^2 = 0.58$.

In contrast to the analyses with respect to Stroop interference, the 2 (type: within or between) × 3 (output language) repeated-measures ANOVA revealed a significant main effect of type,

$F(1,29) = 40.49$, $p < 0.0001$, $\eta^2 = 0.05$, such that larger Stroop facilitation occurred within-languages (25 ms, SE = 7 ms) than between-languages (4 ms, SE = 8 ms). No effect of language output or interaction between type and output language was found (all $ps > 0.21$). Thus, Stroop facilitation was absent between-languages. Roelofs (2010) found an equal amount of within- and between-language Stroop facilitation in Dutch–English bilinguals in a Stroop task that separated color and word information in time (SOA manipulation). However, the experiment in Roelofs's study that was most comparable to the current one (color words and control conditions fully crossed), Stroop facilitation within- and between-languages was not significant at the 0-ms SOA (Experiment 4). Unfortunately, the analysis of that experiment was collapsed across within- and between-languages; thus it is unclear whether at the 0-ms SOA, within- and between-facilitation differed. Close inspection of the graphs (Figure 8 in Roelofs, 2010) shows that the within-language facilitation was in fact larger than the between-language facilitation. This suggests that the current results with GED trilinguals are very comparable to those of the Dutch–English bilinguals.

To summarize, in Experiment 1 with GED trilinguals, we found an equal amount of Stroop interference within- and between-languages, but Stroop facilitation was stronger within- than between-languages. To investigate whether similarity between the involved languages modulates these effects, a second experiment was conducted with trilinguals for whom the cross-language similarity between the color word translations was much less: two were alphabetic languages, Malay and English, and one was a logographic language, Chinese.

EXPERIMENT 2: CHINESE–ENGLISH–MALAY TRILINGUALS
METHOD
Participants
In this experiment 24 CEM trilinguals (10 males) participated. Participants were born in Malaysia and had received formal education in Mandarin, Malay, and English. They arrived in the UK between the age of 15 and 23 ($M = 21.8$ years) and were studying at the University of Nottingham (United Kingdom) at the time of testing. Most of the trilinguals could speak both Mandarin and Cantonese and rated their Cantonese proficiency higher than their Mandarin proficiency (see **Table 1**). However, seven trilinguals considered themselves more proficient in Mandarin than Cantonese. Several participants could also understand and speak other spoken Chinese dialects (e.g., Hakka). **Table 1** provides an overview of their mean age and their subjective proficiency scores for each language, as well as their first contact and years of experience with each language.

MATERIALS AND DESIGN
The design was identical to Experiment 1. The word stimuli were the English color words *red*, *green*, and *blue* (same as in Experiment 1), and their Malay (*merah*, *hijau*, *biru*) and Chinese translations. For 18 of the participants, the Chinese words were presented in Cantonese (traditional Chinese script): 紅 (hung), 綠 (luk), 藍 (laam), whereas for the six participants who rated themselves more proficient in Mandarin, the Chinese words were presented in Mandarin (simplified Chinese script): 红 (hong), 绿 (lu), 蓝 (lan).

English and Malay words were presented in 32-point lowercase Courier font, and Chinese characters were presented in 32-point STHeiti font.

PROCEDURE AND APPARATUS
Same as Experiment 1.

ANALYSIS
Same as Experiment 1.

RESULTS AND DISCUSSION
The total percentage of outliers (0.71%) and the total percentage of errors (1.83%) were again very low. The mean RTs for all conditions are shown in **Table 2**. A two-way (control type: character or patch) ANOVA on the mean RTs across all language outputs showed no significant effect of control type, $F(1,23) = 1.84$, $p = 0.19$; therefore the control character was used in subsequent analyses.

STROOP INTERFERENCE (INCONGRUENT VS. CONTROL CONDITION)
Significant Stroop interference across input and output language combinations were found, $ps < 0.05$ for Bonferroni-corrected paired-sample t-tests, except for Chinese input and Malay output (see **Table 3**).

A 3 (input language) × (output language) repeated-measures ANOVA revealed an interaction between input and output language, $F(4,92) = 4.26$, $p < 0.01$, $\eta^2 = 0.04$. There was no effect of input language for Chinese output, $F(2,46) < 1$. However, for English output an effect of language input, $F(2,46) = 4.67$, $p < 0.05$, $\eta^2 = 0.07$, was found, with significant differences in Stroop interference between the Chinese and Malay input (40 ms, SE = 11 vs. 77 ms, SE = 12 ms), $t(23) = 2.90$, $p < 0.05$, $\eta^2 = 0.27$. Malay output also revealed a significant effect of language input, $F(2,46) = 11.21$, $p < 0.001$, $\eta^2 = 0.16$, with significant differences between the Chinese and English input (19 vs. 67 ms), $t(23) = 2.93$, $p < 0.05$, $\eta^2 = 0.27$, and between the Chinese and Malay input (19 ms, SE = 12 vs. 88 ms, SE = 15 ms), $t(23) = 4.44$, $p < 0.001$, $\eta^2 = 0.46$.

The 2 (type: within or between) × 3 (language output) repeated-measures ANOVA showed a significant effect of type, $F(1,23) = 6.50$, $p < 0.05$, $\eta^2 = 0.02$, such that the magnitude of interference was larger within-languages (64 ms, SE = 12 ms) than between-languages (47 ms, SE = 12 ms). Remarkably, the between-language interference is 73% of within-language interference. This percentage is very similar to the percentage of 74% reported by Francis (1999), and of 75% reported by MacLeod (1991), which were based on a review of studies with bilinguals. No main effect of output arose, $F(2,46) = 2.27$, $p = 0.11$, but there was an interesting interaction between type and output, $F(2,46) = 5.53$, $p < 0.01$, $\eta^2 = 0.02$. To follow up on this interaction, we ran a two-way (type) ANOVA for each language output. Interestingly, this analysis revealed that the Chinese and English output showed no effect of type, $F(1,23) < 1$, all $ps > 0.64$, whereas a significant effect of type was found for the Malay output, $F(1,23) = 15.20$, $p < 0.001$, $\eta^2 = 0.08$, such that the magnitude of within-language interference for Malay output was larger than between-language (Malay: 88 ms, SE = 15 ms vs. English and Chinese: 43 ms, SE = 14 ms).

Because the CEM trilinguals have two languages that share the same-script, we looked at the within- and between-language interference for same-script language pairs (Malay–English) and different script language pairs (Malay–Chinese and English–Chinese). The data revealed that for same-script languages the within- and between-language interference was similar (within: 74 ms, between: 72 ms), whereas for different script languages the between-language interference was reduced to 72% of the within-language interference for Chinese–English (within: 53 ms, between: 38 ms) and to 48% for Chinese–Malay (within: 67 ms, between: 32 ms). Thus, for the Malay–English language combination an equal amount of within- and between-language interference was found, which is consistent with the results of the GED trilinguals. Remarkably, the size of the within- and between-language interference was similar as well (GED within: 73 ms, between: 74 ms vs. CEM same-script within: 74 ms, between: 72 ms). Importantly, the color words in Malay and English do not overlap in terms of orthography and phonology, except for the color word *blue* and Malay *biru*. To investigate whether the orthographic and phonological overlap of the color word blue affected the interference effects in the CEM, we analyzed the data after excluding the *biru* trials. This analysis again revealed similar within- and between-language interference (within: 65 ms, between: 66 ms) for the same-script languages (Malay and English). Thus, the similarly sized within- and between-language interference suggests that between-language interference is not stronger because the color word translations are orthographically/phonologically similar (cognates) but because they are written in the same-script. Therefore, the reduction of between-language interference in different script languages might be due to the use of script as a cue to reduce interference. This was tested in Experiment 3, involving trilinguals with three languages that differ in orthography/script and phonology.

STROOP FACILITATION (CONTROL CHARACTER VS. CONGRUENT CONDITION)

Significant facilitation was only found for English input and output, $t(23) = 3.10$, $p < 0.05$, $\eta^2 = 0.29$ (see **Table 3**). Again, an interaction was found between input and output languages in the 3 (input language) × 3 (output language) repeated-measures ANOVA on the magnitude of Stroop facilitation effects, $F(3.0, 68.4) = 9.19$, $p < 0.0001$, $\eta^2 = 0.10$. For Chinese output, an effect of language input was found, $F(2, 46) = 7.96$, $p < 0.01$, $\eta^2 = 0.07$. Paired-sample t-tests identified the difference in Stroop facilitation between Chinese and English input (28 ms, SE = 10 vs. −11 ms, SE = 12 ms), $t(23) = 3.89$, $p < 0.01$, $\eta^2 = 0.40$, and between Chinese and Malay input (28 ms, SE = 10 vs. 1 ms, SE = 14 ms), $t(23) = 2.65$, $p < 0.05$, $\eta^2 = 0.23$. The English output group also showed a significant effect of language input, $F(2, 46) = 6.91$, $p < 0.01$, $\eta^2 = 0.12$, with significant differences between the Chinese and English input (−1 ms, SE = 7 vs. 26 ms, SE = 8 ms), $t(23) = 3.46$, $p < 0.01$, $\eta^2 = 0.34$, and between the Malay and English input (−9 ms, SE = 10 vs. 26 ms, SE = 8 ms), $t(23) = 3.19$, $p < 0.05$, $\eta^2 = 0.31$. For Malay output a significant effect of language input was also found, $F(2, 46) = 6.75$, $p < 0.01$, $\eta^2 = 0.13$. This effect was due to differences in Stroop facilitation between the English and Malay input (−16 ms, SE = 12 vs.

28 ms, SE = 11 ms), $t(23) = 3.13$, $p < 0.05$, $\eta^2 = 0.30$, and between Chinese and Malay (−19 ms, SE = 11 vs. 28 ms, SE = 11 ms), $t(23) = 3.77$, $p < 0.01$, $\eta^2 = 0.38$.

The 2 (type: within or between) × 3 (output language) repeated-measures ANOVA revealed only a significant effect of type, $F(1, 23) = 62.11$, $p < 0.0001$, $\eta^2 = 0.10$, such that the magnitude of facilitation was larger within-languages (27 ms, SE = 9 ms) than between-languages (−9 ms, SE = 11 ms). There was no main effect of output, $F(1.6, 35.7) < 1$, and no interaction, $F(2, 46) < 1$.

Interestingly, the between-language Stroop facilitation effect became negative. Thus, responses to the congruent Stroop condition (e.g., for English output: blue colored rectangle with the Chinese translation of blue written in the center) were slower than to the control condition. However, this negative Stroop facilitation effect was not significant across the different combinations of input and output languages. As discussed in the Introduction, between-language interference for Stroop facilitation has been observed before (Dalrymple-Alford, 1968; Abunuwara, 1992; MacLeod and MacDonald, 2000). Only numerically is there interference for same and different script language combinations, so it remains unclear whether script similarity modulates between-language Stroop facilitation. If script similarity does modulate the between-language facilitation, the interference effect should become larger and thus become significant when all three languages of the trilinguals differ in script. This was investigated in the next experiment, which involved trilinguals who were proficient in three languages that are written in different scripts.

EXPERIMENT 3: UYGHUR–CHINESE–ENGLISH TRILINGUALS

METHOD

Participants

Thirty-two UCE trilinguals (12 males) participated in the study. All participants were native Uyghur speakers who were born in Xinjiang, China. They received formal education in Uyghur and Mandarin and at the time of testing studied at Beijing Normal University, Beijing, China. **Table 1** presents an overview of their mean age and their subjective proficiency scores for each language, as well as their first contact and years of experience with each language.

MATERIALS AND DESIGN

The design was identical to that of the previous experiments. The only difference was that in addition to the English color words *red*, *green*, and *blue*, the word stimuli consisted of their translations in Chinese (Mandarin, simplified Chinese script): 红 (hong), 红 (lu), 绿 (lan), and Uyghur (Arabic script): قىزىل (gizil), شىلدىي (yéshil), كۆك (kök). English words were presented in 32-point lowercase Courier font, Chinese characters in 32-point STHeiti font and Uyghur words in 32-point Geeza Pro font.

PROCEDURE AND APPARATUS

Same as in the previous experiments.

ANALYSIS

Same as in the previous experiments.

RESULTS AND DISCUSSION

The total percentage of errors was 1.33%, and the total percentage of outliers was 0.74%. The mean RTs for all conditions are shown

in **Table 2**. An analysis on the two control conditions (character and color patch) using a two-way (control type: character or patch) repeated-measures ANOVA revealed, as in Experiment 2, no main effect of control stimulus type, $F(1,31) < 1$. Therefore, the control character was used in subsequent analyses.

STROOP INTERFERENCE (INCONGRUENT VS. CONTROL CONDITION)

Across all input and output language combinations significant Stroop interference effects were found, $ps < 0.05$ in all Bonferroni-corrected paired-sample t-tests (see **Table 3**).

The 3 (input language) × 3 (output language) repeated-measures ANOVA on the magnitude of Stroop interference effects revealed a trend toward a significant interaction between input and output languages, $F(4,124) = 2.18$, $p = 0.08$, $\eta^2 = 0.02$. No effect of language input was found for Uyghur output, $F(2,62) = 1.10$, $p = 0.34$, or Chinese output, $F(2,62) < 1$. In contrast, for English output a significant effect of input language was found, $F(2,62) = 4.55$, $p < 0.05$, $\eta^2 = 0.05$, with a significant difference in Stroop interference between Chinese (45 ms, SE = 12 ms) and English (91 ms, SE = 16 ms) inputs, $t(31) = 3.10$, $p < 0.05$, $\eta^2 = 0.24$.

The 2 (type: within or between) × 3 (output language) repeated-measures ANOVA revealed a main effect of type, $F(1,31) = 8.92$, $p < 0.01$, $\eta^2 = 0.02$, such that there was stronger interference for within (75 ms, SE = 14 ms) than between-languages (54 ms, SE = 14 ms), but no effect of output, $F(2,62) = 1.57$, $p = 0.22$, and no interaction, $F(12,62) = 1.96$, $p = 0.15$.

Thus, the reduction of the magnitude from within- to between-languages was 72%, which is very similar to what has been reported in the literature (MacLeod, 1991; Francis, 1999). Furthermore, the percentage is identical to the different script languages of the CEM trilinguals in Experiment 2 (72% for Chinese–English).

STROOP FACILITATION (CONTROL CHARACTER VS. CONGRUENT CONDITION)

English input lead to significant Stroop facilitation across all output languages, all $ps < 0.05$ (see **Table 3**), although only the English input and output combination yielded a positive effect (36 ms), while the others produced a negative effect (−40 and −44 ms). Significant negative Stroop facilitation was also found for Chinese input and Uyghur output (−32 ms), $t(31) = 3.17$, $p < 0.05$, $\eta^2 = 0.24$.

The 3 (input language) × 3 (output language) repeated-measures ANOVA revealed an interaction between input and output languages for Stroop facilitation effects, $F(4,124) = 21.01$, $p < 0.0001$, $\eta^2 = 0.16$. Uyghur output showed an effect of input language, $F(2,62) = 13.31$, $p < 0.0001$, $\eta^2 = 0.13$, with significant differences between English and Uyghur input (−40 ms, SE = 10 vs. 6 ms, SE = 7 ms), $t(31) = 4.71$, $p < 0.001$, $\eta^2 = 0.42$, and between Chinese and Uyghur (−32 ms, SE = 10 vs. 6 ms, SE = 7 ms), $t(31) = 4.09$, $p < 0.001$, $\eta^2 = 0.35$. An effect of language input was also found with Chinese output, $F(2,62) = 16.96$, $p < 0.001$, $\eta^2 = 0.18$. Significant differences in the magnitude of Stroop facilitation were found between the Chinese and English input languages (26 ms, SE = 12 vs. −44 ms, SE = 13 ms), $t(31) = 5.64$, $p < 0.0001$, $\eta^2 = 0.51$, and the English and Uyghur

input (−44 ms, SE = 13 vs. 5 ms, SE = 8 ms), $t(31) = 3.64$, $p < 0.001$, $\eta^2 = 0.30$. Also English output revealed an effect of language input, $F(2,62) = 13.60$, $p < 0.0001$, $\eta^2 = 0.18$, with a significant difference between Chinese and English input (−4 ms, SE = 11 vs. 36 ms, SE = 8 ms), $t(31) = 3.99$, $p < 0.01$, $\eta^2 = 0.34$ and between Uyghur and English (−29 ms, SE = 12 vs. 36 ms. SE = 8 ms), $t(31) = 4.88$, $p < 0.0001$, $\eta^2 = 0.43$.

The 2 (type: within or between) × 3 (language output) repeated-measures ANOVA showed a main effect of type, $F(1,31) = 59.12$, $p < 0.0001$, $\eta^2 = 0.12$, such that there was a larger effect for within (22 ms, SE = 10 ms) than between language conflict (−24 ms, SE = 11 ms), and a trend toward a main effect of output, $F(2,62) = 3.08$, $p = 0.053$, $\eta^2 = 0.02$, but no interaction, $F(2,62) < 1$. Interestingly, the magnitude of within-language Stroop facilitation was positive (22 ms), while between-language Stroop facilitation was negative (−24 ms). The size of this negative Stroop facilitation was larger than for the CEM trilinguals reported in Experiment 2 (−9 vs. −24 ms). Thus, script appears to modulate the Stroop facilitation effect. This will be analyzed further in the next section and taken up in the Section "General Discussion."

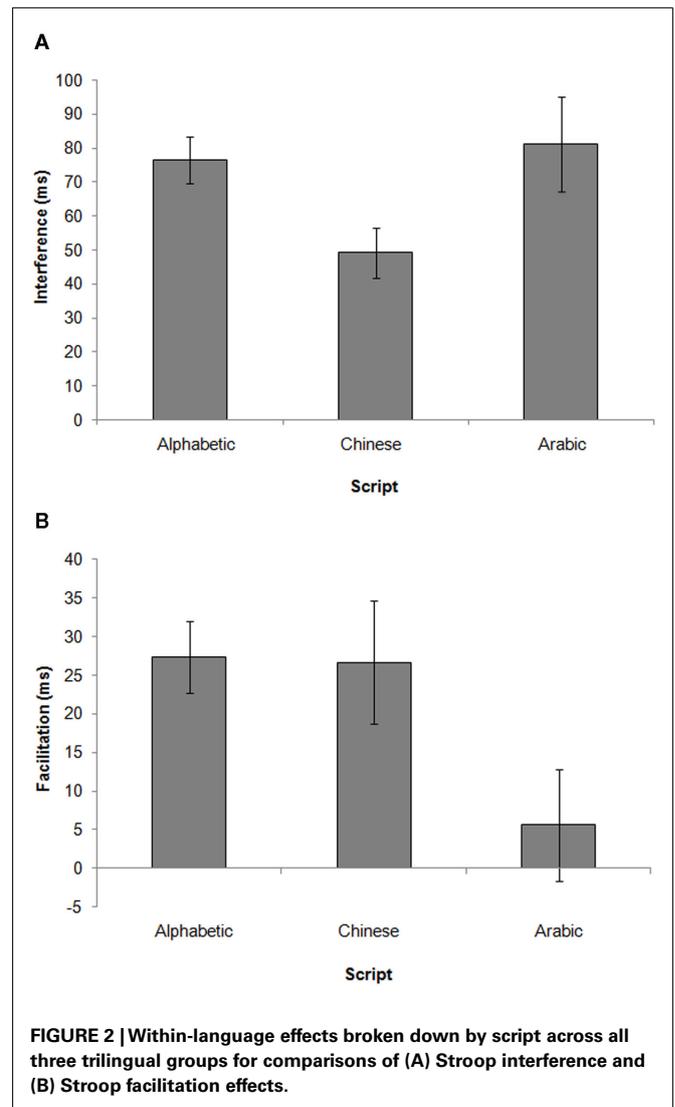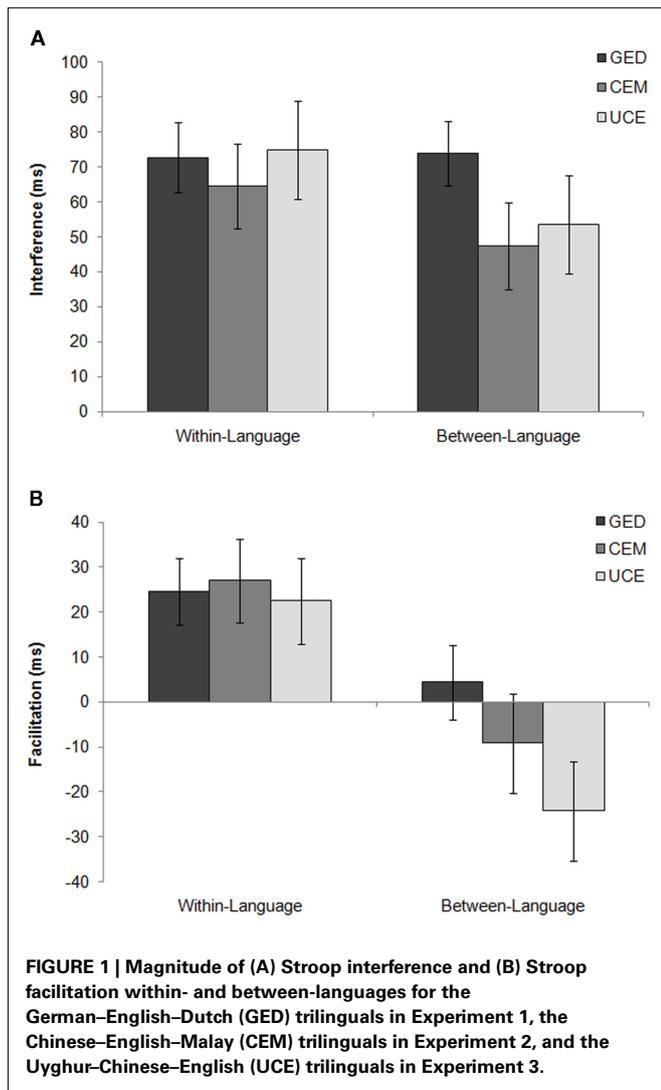## ANALYSES ACROSS EXPERIMENTS 1–3

To compare the results across the three groups of trilinguals, we analyzed first the magnitude of within- and between-language Stroop interference and facilitation effects. Next, we looked across the three groups of trilinguals at the amount of within- and between-language Stroop interference and facilitation in terms of alphabetic (German, English, Dutch, Malay), Chinese, and Arabic (Uyghur) scripts to investigate the role of script.

STROOP INTERFERENCE: WITHIN- AND BETWEEN-LANGUAGE

Figure 1A summarizes the amount of within- and between-language Stroop interference when data was collapsed across input and output languages. The analysis of the data of the three groups of trilinguals revealed a similar amount of within-language Stroop interference, $F(2,255) < 1$ (GED: 73 ms, SE = 10 ms; CEM: 64 ms, SE = 12 ms; UCE: 75 ms, SE = 14 ms). In contrast, the between-language interference varied between the three trilingual groups, $F(2,513) = 7.58$, $p < 0.001$, $\eta^2 = 0.03$ (see **Figure 1B**). As reported in Experiment 1, the amount of within- and between-language interference was similar for the GED trilinguals, whereas it was reduced for the CEM (73% of the within-language interference, see Experiment 2) and UCE trilinguals (72% of the within-language interference, see Experiment 3). After discussing the Stroop facilitation we will analyze the Stroop interference in terms of script similarity to investigate whether script similarity can explain the reduction.

STROOP FACILITATION: WITHIN- AND BETWEEN-LANGUAGE

A three-way (trilingual group) ANOVA revealed that the magnitude of within-language Stroop facilitation was similar across the three groups of trilinguals, $F(2,255) < 1$ (GED: 25 ms, SE = 7 ms; CEM: 27 ms, SE = 9 ms; UCE: 22 ms, SE = 10 ms), whereas the magnitude of between-language facilitation was modulated by type of trilingual, $F(2,513) = 12.78$, $p < 0.0001$, $\eta^2 = 0.05$. No between-language Stroop facilitation was observed for the GED trilinguals (4 ms, SE = 8 ms), whereas responses to the Stroop

FIGURE 1 | Magnitude of (A) Stroop interference and (B) Stroop facilitation within- and between-languages for the German–English–Dutch (GED) trilinguals in Experiment 1, the Chinese–English–Malay (CEM) trilinguals in Experiment 2, and the Uyghur–Chinese–English (UCE) trilinguals in Experiment 3.

FIGURE 2 | Within-language effects broken down by script across all three trilingual groups for comparisons of (A) Stroop interference and (B) Stroop facilitation effects.

congruent condition were slower than to the control condition (negative Stroop facilitation) for the CEM (−9 ms, SE = 11 ms) and UCE trilinguals (−24 ms, SE = 11 ms). For CEM and UCE trilinguals in the congruent condition, the written word and color information matched at the conceptual level (converging conceptual information), but at the word form level a mismatch occurred between the spoken and written word (divergent phonological information) that resulted into a negative Stroop facilitation effect (cf. Roelofs, 2010). In the next section, we examine further the influence of script on within- and between-language Stroop interference and facilitation by analyzing the data across the three groups of trilinguals. This is particularly relevant in the case of the CEM trilinguals, as two of their languages share the same-script (alphabetic).
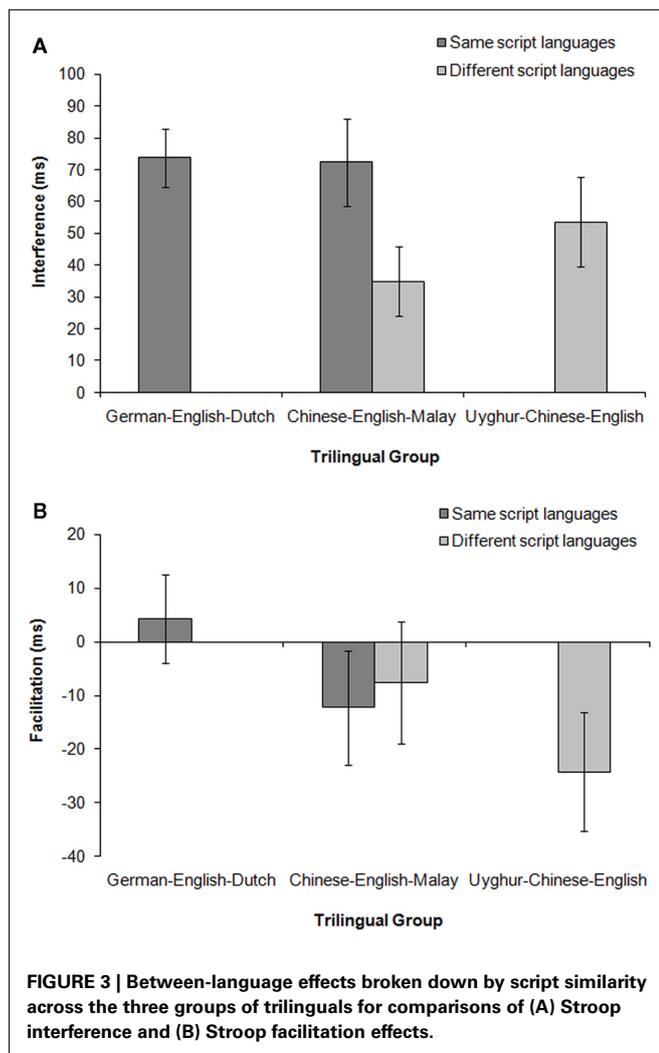
THE ROLE OF SCRIPT SIMILARITY ACROSS LANGUAGES

We explored the influence of script similarity on Stroop interference and facilitation across the languages of the three groups of trilinguals. First, the magnitude of within-language Stroop interference and facilitation across the three groups of trilinguals

for alphabetic (German, English, Dutch, Malay), logographic (Chinese), and Arabic (Uyghur) scripts was calculated (see **Figure 2**). To compare the between-language effects, the cross-language similarity of each combination of languages was coded as either same or different script (same = alphabetic scripts; different = combination of alphabetic, Chinese, and Arabic scripts). Next, the between-language Stroop effects were calculated (see **Figure 3**). We will report first the analyses of the within- and between-language Stroop interference and then the analyses of the within- and between-language Stroop facilitation.

STROOP INTERFERENCE: WITHIN- AND BETWEEN-LANGUAGE

A three-way (script) ANOVA for within-language Stroop interference showed an effect of script, $F(2,255) = 4.16$, $p < 0.05$, $\eta^2 = 0.03$. Independent-sample t-tests identified a difference in the magnitude of within-language interference between the alphabetic (77 ms, SE = 7 ms) and Chinese (49 ms, SE = 6 ms) scripts, $t(108.6) = 3.05$, $p < 0.01$, $\eta^2 = 0.08$, and a trend between Chinese (49 ms, SE = 6 ms) and Arabic (81 ms, SE = 8 ms), $t(48.8) = 2.02$,

FIGURE 3 | Between-language effects broken down by script similarity across the three groups of trilinguals for comparisons of (A) Stroop interference and (B) Stroop facilitation effects.

$p < 0.05$, $\eta^2 = 0.08$, but not between alphabetic (77 ms, SE = 7 ms) and Arabic (81 ms, SE = 8 ms), t (39.3) = 0.32, $p = 0.75$ (see **Figure 2A**). This reduction of the magnitude of Stroop interference for the Chinese script relative to the other scripts (64% of the Alphabetic and 60% of the Arabic script) is interesting. However, it is unclear why this reduction occurred because in the literature either stronger Stroop interference effects have been reported in Chinese than English (Biederman and Tsao, 1979) or an equal amount of interference (Smith and Kirsner, 1982; Lee and Chan, 2000).

A 2 (script similarity: same or different) × 3 (trilingual group) ANOVA for between-language Stroop interference revealed a significant main effect of similarity, $F(1,512) = 20.79$, $p < 0.0001$, $\eta^2 = 0.04$ (see **Figure 3A**), such that the magnitude of interference of same-script languages was larger than for different script languages (74 ms, SE = 6 vs. 47 ms, SE = 8 ms) and there was a trend toward a main effect of trilingual group, $F(1,512) = 2.68$, $p = 0.07$, $\eta^2 = 0.01$. This trend of trilingual group arose from a significant difference in the different script effects of the CEM (35 ms, SE = 11 ms) and UCE groups (54 ms, SE = 14 ms), $t(263.7) = 2.36$, $p < 0.05$, $\eta^2 = 0.02$. Importantly, as discussed in

the Section "Results and Discussion" of Experiment 2, there was no difference between GED same-script (German, English, Dutch: 74 ms, SE = 9 ms) and the CEM same-script languages (Malay and English: 72 ms, SE = 14 ms), $p = 0.88$ uncorrected, even though the color words translations were similar in terms of orthography and phonology for GED trilinguals and different for CEM same-script languages (Malay and English). The modulation of between-language Stroop interference by script will be discussed further in the Section "General Discussion."

STROOP FACILITATION: WITHIN- AND BETWEEN-LANGUAGE

For within-language Stroop facilitation (see **Figure 2B**) there was a trend toward an effect of script in the Stroop facilitation comparison, $F(2,252) = 2.95$, $p = 0.05$, $\eta^2 = 0.02$. Independent-sample t-tests showed a significant difference in within-language facilitation between alphabetic (27 ms, SE = 5 ms) and Arabic (6 ms, SE = 4 ms) scripts, $t(45.1) = 2.73$, $p < 0.05$, $\eta^2 = 0.14$, and a trend toward a difference between Chinese (27 ms, SE = 6 ms) and Arabic (6 ms, SE = 4 ms) scripts, $t(82.9) = 1.96$, $p = 0.053$, $\eta^2 = 0.04$, but no difference between alphabetic (27 ms, SE = 5 ms) and Chinese scripts (27 ms, SE = 6 ms), $t(75.2) = 0.08$, $p = 0.94$.

The analysis of the magnitude of between-language Stroop facilitation revealed a significant main effect of script similarity, $F(1,512) = 16.30$, $p < 0.0001$, $\eta^2 = 0.03$, because there was no Stroop facilitation effect for same-script (alphabetic) languages, whereas for different script languages the effect of between-language Stroop congruency was negative (1 ms, SE = 5 vs. −19 ms, SE = 7 ms). There was also a main effect of trilingual group, $F(2,512) = 4.73$, $p < 0.01$, $\eta^2 = 0.02$. *Post hoc* t-tests indicated that there was a trend toward a significant difference in the same-script facilitation effects between the GED and CEM groups, $t(66.9) = 2.02$, $p = 0.095$ ($p = 0.048$ uncorrected), $\eta^2 = 0.06$, such that the GED effect was more positive than the CEM effect (4 ms, SE = 8 vs. −12 ms, SE = 11 ms), and also a significant difference between the different script facilitation between CEM (−8 ms, SE = 11 ms) and UCE (−24 ms, SE = 11 ms), $t(212.7) = 2.31$, $p < 0.05$, $\eta^2 = 0.02$. The overall pattern indicates that between-language Stroop facilitation was primarily modulated by script similarity, such that increased language dissimilarity (in term of script) lead to slower responses to the between-language congruent condition relative to the control condition. However, the results also seem to indicate that cross-language similarity between the color word translations in terms of orthography/phonology seems to play a role as well (between-language facilitation difference between GED and CEM same-script is a trend), although it is clearly not as strong as the impact of script similarity. Overall, the findings could be explained by a combination of a conceptual match (converging information) at the output (e.g., blue in English) and input (blue colored rectangle and the translation of *blue* written in a different script language in the center of the rectangle) but a mismatch at the level of script (orthography) and phonology (diverging information). In the next Section "General Discussion," we will discuss this and other explanations in more detail.

GENERAL DISCUSSION

Our Stroop experiments involved three groups of trilinguals with languages that are highly similar (GED trilinguals), partly similar

(CEM trilinguals), or completely different (UCE trilinguals). A comparison of these trilinguals made it possible to analyze the influence of cross-language similarity on Stroop interference and facilitation in terms of orthographic/phonological overlap between the color word translations, and script. We will first consider the within- and between-language Stroop interference and facilitation effects and then focus on the WLSSE. Finally, we will discuss the implications for theories of language processing and control in bilinguals and multilinguals.

STROOP INTERFERENCE: WITHIN- AND BETWEEN-LANGUAGE

The observed magnitude of between-language Stroop interference was equal to within-language for GED trilinguals, but was reduced for CEM and UCE trilinguals (respectively 72 and 73% of the within-language interference). The size of the reduction is in line with the conclusions of Francis (1999) and MacLeod (1991). Interestingly, for CEM trilinguals, between-language interference was modulated by script similarity. For similar script (alphabetic) languages, the between-language interference effect was similar to the within-language interference effect, whereas for different script combinations between-language interference was reduced considerably. Furthermore, the size of within- and between-language interference for same-script languages (Malay and English) of CEM trilinguals was similar to that of GED trilinguals. This is theoretically important because the color word translations of the GED trilinguals are similar in terms of orthography and phonology (e.g., *rot*, *red*, and *rood*), whereas the color word translations of the same-script languages Malay and English are different in terms of their spelling and pronunciation (e.g., *merah* and *red*). Thus, unlike script similarity, the similarity between the color word translations in terms of orthography and phonology does not seem to modulate the amount of between-language Stroop interference.

Several studies in the literature have concluded that cross-language similarity modulates between-language interference (e.g., Fang et al., 1981; Brauer, 1998). However, in some studies the orthographic/phonological similarity of the color word translations was confounded with script similarity. Preston and Lambert (1969) investigated the role of color word translation similarity on Stroop interference. As discussed in the Introduction, they found that between-language Stroop interference was 68% of within-language interference in English–Hungarian bilinguals, whereas it was 95% for English–French bilinguals. In their second experiment, they manipulated cross-language similarity within German–English bilinguals using two sets of color terms (translations similar or not in terms of orthography/phonology). The results revealed an equal amount of within- and between-language interference for similar color word translations and a reduction of between-language interference for dissimilar color word translations (between-language interference was 62% of within-language interference). Thus, in contrast to our results with trilinguals, their results with bilinguals suggest that between-language interference is modulated by the orthographic/phonological similarity of the color word translations. However, note that the experimental method used by Preston and Lambert and by many others in the literature (e.g., Dyer, 1971; Fang et al., 1981; Brauer, 1998) involved cards with multiple items of the same condition (i.e., cards with 10 rows of 10 color words written in incongruent ink colors). Thus, input conditions were blocked, whereas in the current study all conditions were completely randomized. Therefore, the influence of cross-language similarity in terms of the orthography/phonology of the color word translations on between-language Stroop interference might be restricted to the specific design of Preston and Lambert.

STROOP FACILITATION: WITHIN- AND BETWEEN-LANGUAGE

The magnitude of within-language Stroop facilitation collapsed across output languages did not differ between the three groups of trilinguals. For all trilinguals, color naming latencies were faster relative to the control condition when the naming response matched the pronunciation of the written word (e.g., for English output: blue colored rectangle with the word "blue" presented in the center). In contrast, between-language Stroop facilitation was modulated by cross-language similarity. For trilinguals with languages written in the same-script (GED), responses to the congruent Stroop condition did not differ from the control condition. However, when the languages of the trilinguals were written using different scripts, the responses to the congruent Stroop condition were slower than to the control condition (negative Stroop facilitation effect). For example, when the response was in Uyghur and the stimulus consisted of a blue rectangle with the English word "blue" presented at the center, naming latencies were slower than for the control condition. Similar negative Stroop facilitation effects for between-language Stroop facilitation have been reported before with bilinguals (Dalrymple-Alford, 1968; MacLeod and MacDonald, 2000) and trilinguals (Abunuwara, 1992).

Thus, similar to between-language Stroop interference, the between-language Stroop facilitation effects are modulated by script. However, in contrast to between-language Stroop interference, orthographic and phonological similarity of the color word translations seem to have some impact on the Stroop facilitation effects in addition to script differences: for the GED trilinguals the Stroop facilitation was positive, while for the CEM and UCE trilinguals the effect was negative. In particular, for GED trilinguals and the same-script languages of the CEM trilinguals (Malay and English) the magnitude of between-language facilitation differed.

The occurrence of negative between-language Stroop facilitation effects can be explained by the *inadvertent reading hypothesis* (MacLeod and MacDonald, 2000) in terms of a covert repair processes when the congruent word is read in the incorrect language. These covert repair processes are assumed to slow down the naming process, so that naming latencies in the congruent condition are equal or slower than those in the control condition (Roelofs, 2010). However, it is unclear how the inadvertent reading hypothesis explains the modulation of between-language Stroop facilitation effects by cross-language script similarity. Furthermore, the inadvertent reading hypothesis does not explain the (positive) between-language Stroop facilitation effect observed with GED trilinguals in our study and with Dutch–English bilinguals in a Stroop task with preexposure SOAs by Roelofs (2010).

In contrast, the *converging information hypothesis* (Cohen et al., 1990; Melara and Algom, 2003; Roelofs, 2003, 2010) can account

for negative between-language Stroop facilitation, as well as for the influence of cross-language similarity on between-language facilitation. This hypothesis proposes that a combination of converging information at the conceptual and phonological levels underlies between-language facilitation. Thus, the interference in the Stroop congruent condition is explained by assuming that two different phonological forms are activated. For instance, the competing phonological form of the English color word *red* and that of the translation in Chinese (红) may slow the naming response, in spite of a match at the conceptual level (Roelofs, 2010). A modulation of between-language Stroop facilitation effects by cross-language similarity can be accounted for by this hypothesis, because the converging information at the form level is influenced by the similarity of the representations at this level.

Roelofs (2010) concluded that "…Stroop facilitation and interference have a common locus within and between-languages, supporting the converging information hypothesis of Stroop facilitation." However in a recent paper, Brown (2011) showed that there is no (or very weak and inversely) correlation between Stroop interference and facilitation after correction for spurious correlations, which argues against a common locus of both effects. As Brown argued, response conflict and resolution processes are uniquely involved in Stroop interference and therefore some processes may be shared between Stroop interference and facilitation, whereas others may be unique to Stroop interference or facilitation.

Overall, our data revealed stronger Stroop interference than Stroop facilitation in all trilinguals, a finding commonly reported in the literature (MacLeod and MacDonald, 2000). However, as Brown (2011) recently pointed out, when using X's as control condition to calculate Stroop interference and facilitation, both effects are confounded by a lexicality cost, which is supported by the finding that neutral words in fact interfer with color naming (e.g., Brown et al., 1998). Thus, the Stroop facilitation effects are underestimated, whereas Stroop interference effects are exaggerated (see Brown, 2011, p. 87). As a consequence, negative Stroop facilitation might arise due to a large lexicality cost. Furthermore, the lexicality cost could vary between-languages/scripts and potentially also depend on the familiarity/proficiency with the scripts at hand. We observed a strong trend of an effect of script in the within-language Stroop facilitation of the UCE trilinguals, such that Stroop facilitation was stronger in English than Chinese and Uyghur (36 vs. 26 vs. 9 ms). This trend could reflect a larger lexicality cost in Uyghur than in Chinese and English. At the same time, the lexicality cost would affect between-language Stroop facilitation and result in the observed negative between-language Stroop facilitation.

WITHIN-LANGUAGE STROOP SUPERIORITY EFFECT
The finding of larger within- than between-language Stroop effects has been referred to as the WLSSE. We calculated the WLSSE by collapsing the data over all within and between conditions for each language output and then calculating the difference of the within-effects minus the between-effects. **Figure 4** provides an overview of the WLSSE for Stroop interference and facilitation across all output languages of the trilinguals.

As expected, based on the analyses across the experiments the magnitude of the WLSSE in all the Stroop interference and Stroop
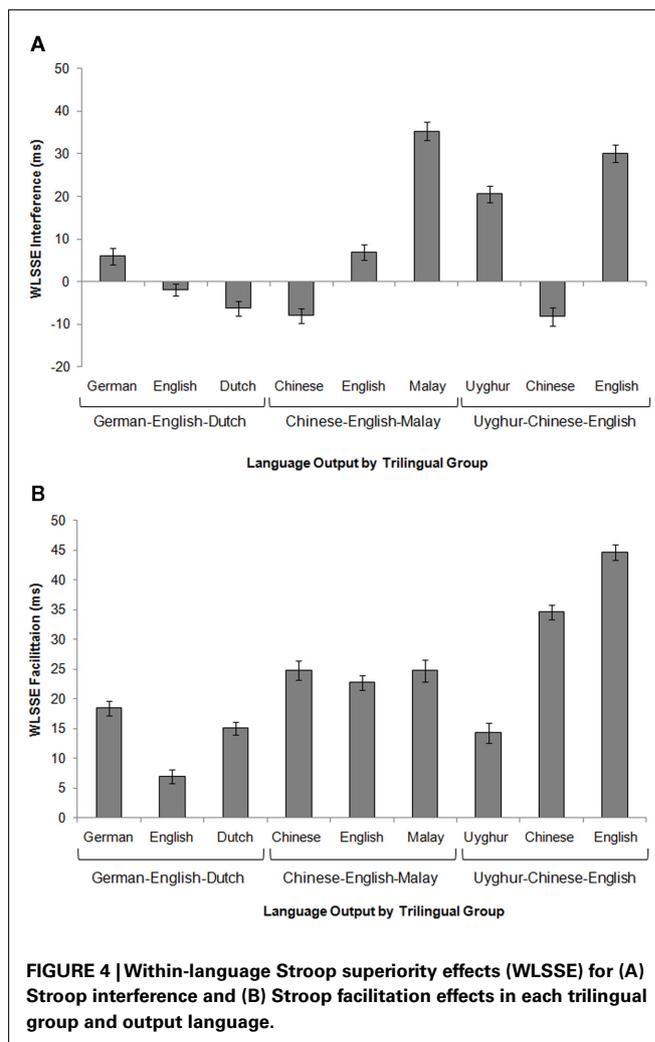
FIGURE 4 | Within-language Stroop superiority effects (WLSSE) for (A) Stroop interference and (B) Stroop facilitation effects in each trilingual group and output language.

facilitation comparisons increases with increasing language differences (in terms of script) in the trilinguals. As discussed in the Introduction, Goldfarb and Tzelgov (2007) argued that the WLSSE is driven by response set effects (see also Roelofs, 2003). According to this account, within-language Stroop interference is larger than between-language due to competition at the conceptual and lexical levels in the within-language condition, whereas in the between-language condition there is only competition at the conceptual level. There is no lexical competition in the between-language condition, because the lexical items (color words in the non-target output language) do not belong to the response set. The present results with trilinguals show that WLSSE is modulated by cross-language similarity; this cannot be explained if the WLSSE is simply due to differences in response sets given that these are the same for the three groups of trilinguals.

IMPLICATIONS
Overall, our results indicate that Stroop effects in trilinguals are comparable to those of bilinguals. This is evidence that the bilingual and trilingual language systems are comparable in their organization. This conclusion is consistent with Dijkstra's (2003)

theoretical analysis of language processing in bilinguals and multilinguals. An important issue is whether or not bilinguals and multilinguals can use script cues to reduce cross-language interference in the Stroop task. Our trilingual data showed that script differences between input and output languages reduced between-language Stroop interference. One potential interpretation of this finding is that trilinguals who are proficient in languages involving different scripts can use script type as a cue to reduce the impact of cross-language interference. However, this interpretation does not explain the negative between-language Stroop facilitation observed in CEM and UCE trilinguals. If script cues could be used optimally to reduce cross-language influences, between-language Stroop facilitation should be absent in different script trilinguals. Although the processes underlying Stroop interference and facilitation might be different (Tzelgov et al., 1990, 1992, 1996; Brown, 2011; see also discussion above), the different impact of script cues on Stroop interference and facilitation could be explained in terms of the level in the language processing system at which they affect processing. Script cues might affect the word production process at higher levels (thus reducing between-language Stroop interference), but might not be able to reduce interference at the lower phonological output level (thus interference occurs for between-language Stroop facilitation).

In addition to cross-language similarity, proficiency has also been shown to modulate interference effects (e.g., Preston and Lambert, 1969; Fang et al., 1981; Mägiste, 1984; Chen and Ho, 1986; Brauer, 1998; Sumiya and Healy, 2004, 2008). To date the role of proficiency and how it interacts with cross-language similarity in Stroop interference and facilitation has not been systematically investigated with trilinguals. Because the primary goal of the current study was to focus on the role of cross-language similarity in terms of script, the participants were matched as closely as possible with regards to proficiency. However, the trilinguals in the current study were not balanced across all three languages. Their L1 was clearly their native language, and their L2 and L3 were less proficient than their L1. As a consequence, the unbalanced proficiency levels across trilinguals' three languages may have modulated the observed Stroop interference and facilitation effects. Investigation of the impact of relative proficiency across languages requires further research with trilinguals having a wider variation in their language proficiency. However, taking into consideration previous research on proficiency, we suggest that the findings of the present study are mainly modulated by language similarity.

CONCLUSION

To conclude, the present study provides evidence that cross-language similarity across color words modulates both between-language Stroop interference and facilitation in trilinguals. In particular, cross-language similarity in terms of the scripts in which the languages of the trilinguals are written modulates between-language Stroop effects. Furthermore, cross-language similarity in terms of the degree of orthographic and phonological overlap between the color word translations seems to have some impact on the Stroop facilitation effects in addition to script similarity. Overall, the observed patterns of Stroop interference and facilitation can be explained by diverging and converging color and word information across languages.

ACKNOWLEDGMENTS

REFERENCES

Abunuwara, E. (1992). The structure of the trilingual lexicon. *Eur. J. Cogn. Psychol.* 4, 311–322.

Biederman, I., and Tsao, Y.-C. (1979). On processing Chinese ideographs and English words: some implications from Stroop-test results. *Cogn. Psychol.* 11, 125–132.

Brauer, M. (1998). "Stroop interference in bilinguals: the role of script similarity between two languages," in *Foreign Language Learning: Psycholinguistic Studies on Training and Retention,* eds A. F. Healy and L. E. Jr. Bourne (Mahwah, NJ: Erlbaum), 317–337.

Brown, T. L. (2011). The relationship between Stroop interference and facilitation effects: statistical artifacts, baselines, and a reassessment. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 85–99.

Brown, T. L., Gore, C. L., and Pearson, T. (1998). Visual half-field Stroop effects with spatial separation of words and color targets. *Brain. Lang.* 63, 122–142.

Chen, H.-C., and Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 12, 397–401.

Cohen, J., Dunbar, K., and McClelland, J. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* 97, 332–361.

Colomé, A. (2001). Lexical activation in Bilinguals' speech production: language-specific or language-independent? *J. Mem. Lang.* 45, 721–736.

Costa, A., Albareda, B., and Santesteban, M. (2008). Assessing the presence of lexical competition across languages: evidence from the Stroop task. *Biling. (Camb. Engl.)* 11, 121–131.

Costa, A., Caramazza, A., and Sebastian Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1283–1296.

Dalrymple-Alford, E. C. (1968). Interlingual interference in a color-naming task. *Psychon. Sci.* 10, 215–216.

de Groot, A. M. B., Delmaar, P., and Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: support for nonselective access to bilingual memory. *Q. J. Exp. Psychol. A* 53, 397–428.

Dijkstra, T. (2003). "Lexical processing in bilinguals and multilinguals: the word selection problem," in *The Multilingual Lexicon,* eds J. Cenoz, B. Hufeisen, and U. Jessner (Dordrecht: Kluwer Academic Publishers), 11–26.

Dijkstra, T., Grainger, J., and van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *J. Mem. Lang.* 41, 496–518.

Dijkstra, T., and van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Biling. (Camb. Engl.)* 5, 175–197.

Dijkstra, T., van Jaarsveld, H., and ten Brinke, S. (1998). Interlingual homograph recognition: effects of task demands and language intermixing. *Biling. (Camb. Engl.)* 1, 51–66.

Dyer, F. N. (1971). Color-naming interference in monolinguals and bilinguals. *J. Verbal Learn. Verbal Behav.* 10, 297–302.

Fang, S.-P., Tzeng, O. J. L., and Alva, L. (1981). Intralingual vs. interlingual Stroop effects in two types of writing systems. *Mem. Cognit.* 9, 609–617.

Forster, K. I., and Forster, J. C. (2003). DMDX: a Windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124.

Francis, W. S. (1999). Cognitive integration of language and memory in bilingual memory: semantic representations. *Psychol. Bull.* 125, 193–122.

Goldfarb, L., and Tzelgov, J. (2007). The cause of the within-language Stroop superiority effect and its implications. *Q. J. Exp. Psychol.* 60, 179–185.

Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Biling. (Camb. Engl.)* 1, 67–81.

Guo, T., and Peng, D. (2006). ERP evidence for parallel activation of two languages in bilingual speech production. *Neuroreport* 17, 1757–1760.

Hermans, D., Bongaerts, T., de Bot, K., and Schreuder, R. (1998). Producing words in a foreign language: can speakers prevent interference from their first language? *Biling. (Camb. Engl.)* 1, 213–229.

Hoshino, N., and Kroll, J. F. (2008). Cognate effects in picture naming: does cross-language activation survive a change of script? *Cognition* 106, 501–511.

Jared, D., and Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *J. Mem. Lang.* 44, 2–31.

Kane, M. J., and Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *J. Exp. Psychol. Gen.* 132, 47–70.

Lee, T. M. C., and Chan, C. C. H. (2000). Stroop interference in Chinese and English. *J. Clin. Exp. Neuropsychol.* 22, 465–471.

Lee, W. L., Wee, G. C., Tzeng, O. J. L., and Hung, D. L. (1992). "A study of interlingual and intralingual Stroop effect in three different scripts: logographic, syllabary, and alphabet," in *Cognitive Processing in Bilinguals,* ed. R. J. Harris (Amsterdam: Elsevier Science Publishers), 427–442.

Lemhöfer, K., and Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: effects of code similarity in language-specific and generalized lexical decision. *Mem. Cognit.* 32, 533–550.

Lemhöfer, K., Dijkstra, T., and Michel, M. C. (2004). Three languages, one ECHO: cognate effects in trilingual word recognition. *Lang. Cogn. Process.* 19, 585–611.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* 109, 163–203.

MacLeod, C. M., and MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends Cogn. Sci. (Regul. Ed.)* 4, 383–391.

Mägiste, E. (1984). Stroop task and dichotic translation: the development of interference patterns in bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 304–315.

Melara, R. D., and Algom, D. (2003). Driven by information: a tectonic theory of Stroop effects. *Psychol. Rev.* 110, 422–471.

Preston, M. S., and Lambert, W. E. (1969). Interlingual interference in a bilingual version of the Stroop color-word task. *J. Verbal Learn. Verbal Behav.* 8, 295–301.

Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav. Res. Methods* 39, 859–862.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychol. Rev.* 110, 88–125.

Roelofs, A. (2010). Attention and facilitation: converging information versus inadvertent reading in Stroop task performance. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 411–422.

Smith, M. C., and Kirsner, K. (1982). Language and orthography as irrelevant features in colour-word and picture-word Stroop interference. *Q. J. Exp. Psychol. A* 34, 153–170.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662.

Sumiya, H., and Healy, A. F. (2004). Phonology in the bilingual Stroop effect. *Mem. Cognit.* 32, 752–758.

Sumiya, H., and Healy, A. F. (2008). The Stroop effect in English-Japanese Bilinguals: the effect of phonological similarity. *Exp. Psychol.* 55, 93–101.

Tzelgov, J., Henik, A., and Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Mem. Cognit.* 20, 727–635.

Tzelgov, J., Henik, A., and Leiser, D. (1990). Controlling Stroop interference: evidence from a bilingual task. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 760–771.

Tzelgov, J., Henik, A., Sneg, R., and Baruch, O. (1996). Unintentional word reading via the phonological route: the Stroop effect with cross-script homophones. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 336–349.

van Hell, J. G., and de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: effects of concreteness and cognate status in word association. *Biling. (Camb. Engl.)* 1, 193–211.

van Hell, J. G., and Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychon. Bull. Rev.* 9, 780–789.

van Heuven, W. J. B., Dijkstra, A., and Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *J. Mem. Lang.* 39, 458–483.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.