International Journal of Multilingualism
2012, 1–23, iFirst article

Multilingual Stroop performance: Effects of trilingualism and proficiency on inhibitory control

Viorica Marian[a]*, Henrike K. Blumenfeld[b], Elena Mizrahi[c], Ursula Kania[d] and Anne-Kristin Cordes[e,f]

[a]*Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL, USA;* [b]*School of Speech, Language and Hearing Sciences, San Diego State University, San Diego, CA, USA;* [c]*Department of English Language and Literature, University of Haifa, Haifa, Israel;* [d]*English Department, University of Leipzig, Leipzig, Germany;* [e]*Department fuer Anglistik und Amerikanistik, Ludwig-Maximilians-Universität, Munich, Germany;* [f]*Max Planck Child Study Centre, University of Manchester, Manchester, UK*

(*Received 28 February 2012; final version received 26 June 2012*)

Previous research suggests that multilinguals' languages are constantly co-activated and that experience managing this co-activation changes inhibitory control function. The present study examined language interaction and inhibitory control using a colour-word Stroop task. Multilingual participants were tested in their three most proficient languages. The classic Stroop effect was detected in all three languages, with participants performing more accurately on congruent than on incongruent trials. Multilinguals were faster and more accurate in the within-language-competition condition than in the between-language-competition condition, indicating that additional processing costs are required when stimulus and response languages differ. Language proficiency influenced speed, accuracy and error patterns in multilingual Stroop task performance. These findings augment our understanding of language processing and inhibitory control in multilingual populations and suggest that experience using multiple languages changes demands on cognitive function.

Keywords: multilingualism; Stroop task; cognitive control; proficiency; within- and between-language interference

Introduction

Understanding and producing language entails narrowing down a number of competing linguistic options (e.g. Aitchison, 2003; Clahsen & Felser, 2006; Swinney, 1979). Hearing *can-*, for instance, activates a number of potential candidates including *candle*, *candy*, *canon* and many more. As the word unfolds, more and more options are excluded and finally the target word is selected. During the process, the listener must ignore incorrect words and focus on the correct one. This task becomes even more demanding for speakers who know more than one language and hence select words from multiple languages during comprehension (e.g. Marian & Spivey, 2003a, 2003b) and production (e.g. Costa & Santesteban, 2004). Because at least two languages are activated in parallel during comprehension (Blumenfeld & Marian, 2007; Marian & Spivey, 2003a, 2003b) and production (Christoffels, De

*Corresponding author. Email: v-marian@northwestern.edu

ISSN 1479-0718 print/ISSN 1747-7530 online

http://dx.doi.org/10.1080/14790718.2012.708037
http://www.tandfonline.com

Groot, & Kroll, 2006; Costa & Santesteban, 2004; Kaushanskaya & Marian, 2007; Kroll, Bobb, & Wodniecka, 2006), suppression of the non-target language(s) becomes necessary. Not only do multilinguals need to suppress irrelevant information within the language in use, but they also have to suppress the non-target language(s) not in use (e.g. Kroll, 2008; Levy, McVeigh, Marful, & Anderson, 2007; Linck, Kroll, & Sunderman, 2009; Meuter, 1994).

Constant inhibition of the irrelevant language has been linked to changes in bilinguals' cognitive control abilities (e.g. Bialystok, 2006; Bialystok, Craik, Klein, & Viswanathan, 2004; Blumenfeld & Marian, 2011; Colzato, Bajo, van den Wildenberg, & Paolieri, 2008; Costa, Hernández, & Sebastian-Gallés, 2008). Both bilingual children and bilingual adults have been found to outperform monolinguals on cognitive control tasks (e.g. Bialystok et al., 2004; Bialystok, Craik, & Luk, 2008). Moreover, early evidence suggests that the more extensive the experience with linguistic inhibition, the greater the cognitive benefits may be. For example, Kavé, Eyal, Shorek, and Cohen-Mansfield (2008) compared bilingual, trilingual and multilingual older adults (aged 83–91 years) and found that multilinguals outperformed trilinguals, and trilinguals outperformed bilinguals on tests of general cognitive ability (Katzman et al. 1983 cognitive-screening test and Folstein, Folstein and McHugh's 1975 MMSE). Linck, Schwieter, and Sunderman (2012) also studied speakers of more than two languages and identified a correlation between language switching and cognitive abilities in young adult trilinguals, with individuals who showed smaller switching costs from L2 or L3 into L1 also showing more efficient inhibition in a nonlinguistic inhibitory control task. The majority of studies to date, however, have focused on inhibitory control function in bilinguals. Because of the limited number of studies that have studied trilinguals or multilinguals (e.g. Abunuwara, 1992; Kavé et al., 2008; Linck et al., 2012; van Heuven, Conklin, Coderre, Guo, & Dijkstra, 2011), the consequences of speaking more than two languages on inhibitory control and executive function remain poorly understood. The present study aims to extend research in this area by examining inhibitory control in speakers of more than two languages and by investigating how trilinguals manage competition within and across their three languages.

In order to measure inhibitory control in trilinguals, we employed a classic Stroop task (Figure 1). The Stroop task has been widely used with monolinguals (e.g. Boyden & Gilpin, 1978; Cohen, Dunbar, & McClelland 1990; Warren & Marsh, 1978) and bilinguals (e.g. Costa, Albareda, & Santesteban, 2008; Dyer, 1971; Goldfarb & Tzelgov, 2007; Preston & Lambert, 1969; Zied et al., 2004). In the original version of the task (Stroop, 1935), participants were asked to name the ink colour of colour words, while the print was either congruent or incongruent with the colour that the word spelled out (e.g. the word *red* printed in red or in blue ink). Results showed that colour-naming was significantly slower and less accurate in the incongruent colour-word condition (e.g. naming the red ink 'red' when the colour-word 'green' was printed) than in the congruent condition where the ink and colour-word matched. The term 'Stroop effect' has since been used to describe differences in naming latency and accuracy between incongruent and congruent trials, and reflects inhibitory control and language processing abilities (for a review, see MacLeod, 1991).

The processing abilities underlying Stroop performance include both automaticity of reading and the ability to name colours. The Stroop effect emerges as recognition of the written word becomes automatic and the written word begins to

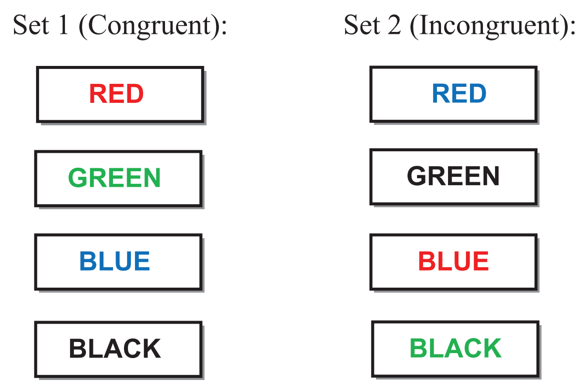

Figure 1. Sample stimulus sets from the Stroop task. The Stroop effect refers to the fact that naming the colour of the first set of words is easier and quicker than naming the colour of the second set of words. (In the within-language condition, participants name the ink colour in English; in the between-language condition, they name the ink colour in a language other than English.)

interfere with the naming of the ink colour, at which point cognitive control processes must be deployed to inhibit this interference (Kornblum, 1992; MacLeod, 1991). In the congruent condition, naming is faster either because participants can rely on the inadvertent reading response or because both reading and colour-naming yield the same correct response (Bialystok et al., 2008). Based on a set of Stroop experiments with bilinguals, Roelofs (2010) proposed that it is the converging evidence from reading and colour-naming that is responsible for better performance on congruent trials. Responding correctly in the incongruent condition, on the other hand, requires participants to resolve the conflict between the well-learned automatised reading response and the colour-naming response (Bugg, Jacoby, & Toth, 2008; Kornblum, 1992).

In addition to the classic within-language Stroop effect (where the stimulus and response languages coincide), the Stroop task can also be performed across different languages (where the stimulus and response languages do not coincide). Table 1 illustrates both a monolingual version (within-language condition) and a bilingual

Table 1. The Stroop task within- and between-languages. Participants are asked to name the colour of the ink (as opposed to reading the word). In the within-language condition, both the written stimulus word and the colour-naming response are in the same language (in this example, English). In the between-language condition, the written stimulus word is in one language (in this example, English), but the colour-naming response is in a different language (in this example, Spanish).

Ink colours: *blue, red* Printed Words: BLUE, RED				
Presented stimuli	BLUE in *blue*	BLUE in *red*	RED in *blue*	RED in *red*
Trial	Congruent	Incongruent	Incongruent	Congruent
Correct Response in English (within-language condition)	blue	red	blue	red
Correct Response in Spanish (between-language condition)	azul	rojo	azul	rojo

version (between-language condition) of the Stroop task. The within-language condition represents the classic Stroop task where the stimulus and the response are in the same language (the difference in performance on the congruent and incongruent trials constitutes the within-language Stroop effect). The between-language condition represents the cross-linguistic Stroop task where the stimuli are presented in a language different than the response language (the difference between congruent and incongruent trials constitutes the between-language Stroop effect). In other words, in addition to manipulating the congruency of the colour terms and the ink colour, in the bilingual version of the Stroop task, the stimulus and response languages coincide in one condition and differ in another, so that interference can be compared both within and between languages (Dyer, 1971; Preston & Lambert, 1969). For example, in an incongruent between-language trial, a Spanish–English bilingual might see the word BLUE in red ink and has to respond with Spanish *rojo* (red). In manipulating stimulus and response languages, additional inhibitory processes become necessary in order to suppress the language of the stimuli, which is irrelevant for the response (e.g. Costa, Albareda, & Santesteban, 2008; Dyer, 1971; Goldfarb & Tzelgov, 2007; Preston & Lambert, 1969). That is, suppression of the automatic reading of the word may be influenced by inhibitory mechanisms under a within-language or a between-language condition.

For bilinguals, the within-language Stroop effect has been found to be larger than the between-language Stroop effect. The size of the between-language Stroop effect is estimated to be approximately 75% of the size of the within-language Stroop effect (e.g. Dyer, 1971; MacLeod, 1991; Preston & Lambert, 1969). This difference in within-language and between-language Stroop effects is likely due to strong activation of the language in which naming occurs (resulting in more interference for written words in the within-language Stroop task), as well as bilinguals' ability to partially suppress non-target languages (Green, 1998, resulting in less interference for written words in the between-language Stroop task). The magnitude of the effects in multilinguals thus provides information about the accessibility of linguistic terms in the response language as well as the ease of suppression of the stimulus language. It is influenced by language proficiency (e.g. Mägiste, 1985), by whether or not writing systems (e.g. alphabetic, syllabic, logographic) are shared across languages (van Heuven et al., 2011), and by the similarity between languages (e.g. less similar languages show weaker between-language interference compared to within-language interference, Chen & Ho, 1986; Fang, Tzeng, & Alva, 1981; Tzelgov, Henik, & Leisner, 1990; but see van Heuven et al., 2011).

For example, Mägiste (1985) found that within-language interference is greater in unbalanced bilinguals when they use their dominant language as the response language, whereas between-language interference is stronger when bilinguals respond in their weaker language (see Meuter & Allport, 1999, for similar asymmetries in language suppression for unbalanced bilinguals, as indexed by language switch costs on a number-naming task). The dominant language is thus more difficult to suppress than the weaker language. Differences between within-language and between-language interference are less marked in more balanced bilinguals. In *trilinguals*, the availability of linguistic items in L3 can therefore be expected to be influenced by L3 proficiency and similarity to the L1 or L2, which in turn may yield a Stroop effect that differs from that of bilinguals. Trilinguals have been found to be slower in their responses than bilinguals, but suppression of a weaker language has also been shown to require less effort than suppression of a stronger language (Mägiste, 1985). This finding is consistent with data suggesting that trilinguals' highly proficient languages may be activated simultaneously, while less proficient languages may not necessarily become significantly co-activated (van Hell & Dijkstra, 2002).

Proficiency effects during word recognition and priming can further inform patterns of multilingual Stroop performance. In visual word recognition tasks, performance becomes increasingly automatic with higher L2 proficiency (Favreau & Segalowitz, 1983; Oren & Breznitz, 2005; Segalowitz, Segalowitz, & Wood, 1998). In a priming study with written words, Favreau and Segalowitz (1983) found that bilinguals who were highly proficient in their L2 showed priming effects within their L2, while less proficient bilinguals did not. Similarly, in a cross-linguistic study, Silverberg and Samuel (2004) showed that bilinguals who were highly proficient in their L2 showed between-language priming from L2 to L1 while bilinguals who were less proficient in their L2 did not. Thus, language proficiency plays an important role in the automatic activation of linguistic information, both within and between languages. According to the Bilingual Interactive Activation model (Dijkstra & van Heuven, 1998), proficiency effects can be explained by different levels of resting-level activation, where words that have higher resting-level activation are easier to access and harder to suppress. Compared to monolinguals, highly proficient bilinguals in particular are expected to face two types of inhibition during performance on the Stroop task: inhibition of a dominant reading response (which monolinguals face as well) and suppression of the language not used in the task (which is more demanding for high-proficiency languages).

The competition among languages and the task demands may be even greater in speakers of more than two languages. In a study with trilinguals, Abunuwara (1992) observed interference effects both within and across all three languages, with reaction times during Stroop performance susceptible to language proficiency effects. Reaction times were faster in L1 than in L2 and L3. Consistent with previous research, within-language Stroop effects were larger than between-language Stroop effects, with the between-language effect amounting to 43% of the within-language effect when all languages (L1, L2 and L3) were considered, and to 51% of the within-language effect when only L1 and L2 were taken into account. Compared to the finding that the between-language Stroop effect approximates 75% of the within-language effect in bilinguals (Dyer, 1971; MacLeod, 1991; Preston & Lambert, 1969), the between-language effect in trilinguals appears smaller than for bilinguals. Yet, in a recent study, van Heuven et al. (2011) found that within-language Stroop effects were equivalent to between-language Stroop effects when the trilinguals' languages were highly similar and shared writing systems (in German–English–Dutch trilinguals). Conversely, when the trilinguals' languages differed and did not share writing systems, the between-language Stroop effects were reduced to 72% and 73% of the within-language Stroop effects (in Chinese–English–Malay and in Uyghur–Chinese–English trilinguals, respectively). Thus, the limited research on trilingual processing to date has yielded a variable range of cross-linguistic Stroop effects.

The present study examined trilinguals' performance on a within-language and between-language Stroop task in order to explore the effects of trilingualism and language proficiency on inhibitory control. We were particularly interested in how the two types of inhibitory processes would interact – in other words, what would be the interplay between within-language inhibition of automatic reading responses and between-language inhibition of irrelevant languages? We used a colour-word Stroop task in which we: (1) presented congruent and incongruent items in an intermixed rather than a blocked fashion so as to increase task demands on inhibition and (2) used only one stimulus language (English), since stimulus language has not been found to significantly influence Stroop performance (Abunuwara, 1992). Finally, we introduced an additional manipulation by including both focal (red, blue, green, yellow and black) and non-focal (brown, pink, purple, orange and grey) colour terms (based on Berlin & Kay, 1969). We hypothesised that colour effects would be akin to language proficiency effects because of similarities in processing – focal colours are more frequently used (e.g. Hays, Margolis, Naroll, & Perkins, 1972) and are more universal across languages (e.g. Lindsey & Brown, 2006; Regier, Kay, & Cook, 2005), and are therefore likely to be more deeply entrenched in the lexicon.

Hypotheses

Cumulatively, the present study on trilingual inhibitory control tested the following five hypotheses.

First, in line with previous research on trilingual Stroop performance (e.g. Abunuwara, 1992), we predicted that trilinguals would show the classic within-language Stroop effect by performing more accurately on congruent trials than on incongruent trials in all of their languages, due to the need to suppress the irrelevant reading response in incongruent trials.

Second, we predicted that our study would replicate previously-found language proficiency effects in Stroop performance and would extend such findings to a heterogeneous trilingual group. Specifically, we predicted that trilinguals' performance on the Stroop task would be faster and more accurate in their most proficient language and would decline with decreased proficiency in the second and third languages.

Third, based on previous research (Abunuwara, 1992; Dyer, 1971; Preston & Lambert, 1969), we expected the within-language Stroop effect to be larger than the between-language effect, due to stronger activation of the stimulus/input language and suppression of non-target languages.

Fourth, we expected the between-language Stroop effect (where participants had to name the ink colour in a different language than the stimulus language) to be larger when the stimulus language was the dominant language than when the stimulus language was not the dominant language. This expectation was based on previous research that suggested stronger co-activation of more proficient languages (Dijkstra & van Heuven, 1998; Mägiste, 1985) and on the findings that suppression of interference from a more proficient language requires use of more extensive cognitive resources (Meuter & Allport, 1999). Such cross-linguistic interference from a more dominant language has two possible outcomes. On the one hand, general cross-linguistic interference (of either translation equivalents in the congruent condition or incorrect colour names in the incongruent condition) could result in overall lower performance across trials in the between-language condition. On the other hand, translation equivalents from the stimulus language could facilitate naming (Costa, Miozzo, & Caramazza, 1999, but see Hermans, 2004) while only unrelated colour words would interfere, effectively increasing the magnitude of the between-language Stroop effect.

Our fifth and final hypothesis was that performance would be faster and more accurate on focal colour items than on non-focal colour items, with better retrieval of more familiar and more frequent focal colour terms than of less familiar and less frequent similar colour terms. We hypothesised that colour term effects would be similar to language proficiency effects because focal colours are likely to be easier to access than non-focal colours.

Method

Participants

A total of 26 participants were tested, 9 males and 17 females. Participants' mean age at the time of testing was 33.1 years ($SD = 9.0$). All participants were fluent in English and in at least two other languages and were enrolled in an international summer course at the time of testing. Table 2 summarises the demographic data including participants' age at testing, age of acquisition and self-reported proficiency levels in each language. English was the native language for 11 participants, the second language for 12 participants and the third language for three participants. The participants were otherwise highly heterogeneous with respect to language backgrounds and included speakers of German, French and Spanish as other frequently spoken languages, as well as Baatonum, Czech, Danish, Hungarian, Italian, Japanese, Malay, Maya, Norwegian, Romanian, Russian, Swedish, Tagalog and Ukrainian, yielding a truly multilingual testing sample.

Design

The study employed a 2×3 repeated-measures design, with the two within-group independent variables being (1) congruence of the written words and the ink they were printed in (congruent, incongruent) and (2) the language in which participants responded (L1, L2, L3). The two dependent variables were (1) speed of participants' responses, operationally defined as the total number of correct cards completed within a 30 second response window and (2) error rates, operationally defined as the number of errors produced in each of the pre-determined error types.

The four error types consisted of Stroop interference, disfluencies, language switch and skipped items. *Stroop interference* referred to errors in which participants read-out the words instead of naming the colour of the ink; *disfluencies* referred to errors in which participants hesitated before answering; *language switches* referred to errors in which participants answered in an inappropriate language; *skipped items* referred to errors in which participants intentionally skipped items due to inability to think of the correct response in a speedy manner. Other error types observed included

Table 2. Participants' demographic information and self-reported proficiency in their three most proficient languages ($N = 26$).

		Mean (SD)
Age at testing		33.1 (9.0)
Age at onset of acquisition	L1	from birth
	L2	10.2 (6.7)
	L3	15.0 (7.4)
Proficiency (on a scale from 0 – none to 10 – perfect)	L1	9.6 (0.8)
	L2	7.5 (1.9)
	L3	5.3 (2.1)

corrected responses, naming of an incorrect colour and false starts, but those were infrequent and did not provide sufficient data to make analysis possible.

Materials and procedure

Each participant completed a language questionnaire, and was then administered the Stroop test. The language questionnaire consisted of two parts. In the first part, participants were asked to provide information about their three most proficient languages and self-assess their proficiency in understanding, speaking, reading and writing each language. In the second part of the questionnaire, participants were asked to provide names for 10 colours in each of their three primary languages, in order to ensure that they were familiar with all colour terms necessary to perform the Stroop task. After completing the language questionnaire, each participant was individually tested using the Stroop task. Participants were asked to name the colour of the ink in which the word was written. At the beginning of the session, the task was explained to the participant and practice stimuli were presented in English. Each participant performed the task three times, once in each language.

The stimuli used in the Stroop task were 40 paper cards with 10 different colour names written out in English using coloured ink. The colour-words and ink colours were selected based on the theory of universal colour categorisation (Berlin & Kay, 1969; Kay & McDaniel, 1978) and consisted of the following 11 colours (the most frequently-found colour-words across the majority of world languages): white, black, yellow, red, blue, green, brown, orange, pink, grey and purple. All colours except white were used as colour-words during the test trials; white was used as the colour-word in the practice trials and was not used as a target colour because it was a background colour for the ink – i.e. all cards were white. The 10 colours were divided into two groups, five focal colours (black, yellow, red, blue and green, as in e.g. Lindsey & Brown, 2006; Regier, Kay, & Cook, 2005) and five non-focal colours (brown, orange, pink, grey and purple). The written words denoted colour terms that were either the same as (congruent condition) or different than (incongruent condition) the colour of the ink. Each colour was presented four times, i.e. twice in the congruent condition with the matching colour-word and twice in the incongruent condition with two different mismatching colour terms. Non-focal colour terms were never mixed with the focal colours and vice versa. Table 3 shows all colour-word combinations that were used in the study. The order of the cards was pseudo-randomised so that congruent or incongruent cards would not appear more than twice consecutively. The colour of the ink was never repeated in a consecutive card and the written word was never repeated more than twice.

Participants had 30 seconds to complete the task in each language and were told to go through as many cards as they could as quickly and as accurately as possible. Language order was counterbalanced across the three response languages. All sessions were videotaped and the data were coded in two ways: (1) on-line during the experiment by two independent judges and (2) off-line from the videotaped recording by two coders. Disagreements were discussed until a consensus was reached for 100% agreement. The total number of cards completed as well as different error types were coded and analysed across languages and across congruent and incongruent conditions.

Table 3. The stimulus colour-word combinations in congruent and incongruent conditions and the number of times each was presented.

Colour words:		BLACK	GREEN	YELLOW	RED	BLUE	BROWN	ORANGE	PINK	GREY	PURPLE
Focal colours:	Black	2	1	1							
	Green	1	2		1						
	Yellow	1		2		1					
	Red		1		2	1					
	Blue			1	1	2					
Non-focal colours:	Brown						2	1	1		
	Orange						1	2			1
	Pink						1		2	1	
	Grey								1	2	1
	Purple							1		1	2

Results

Congruency-incongruency classic Stroop effects in trilinguals

Our first hypothesis, that trilinguals would show the classic within-language Stroop effect and perform more accurately on congruent trials than on incongruent trials, was confirmed. Stroop effects indexing inhibitory control ability (as marked by lower accuracy when ink colour and text did not match) were found across all three languages (see Figure 2). To examine whether participants showed differences across languages in resolving competition from colour-word incongruities, a 2 × 3 ANOVA was conducted, with colour-word congruence (congruent, incongruent) and language (L1, L2, L3) as within-subjects factors and percentage of correct cards completed as the dependent variable (i.e. correct cards completed divided by total cards completed). Results yielded a main effect of colour-word congruence, F_1 (1, 25) = 13.6, $p = 0.001$, $\eta^2 = 0.4$, F_2 (1, 19) = 12.7, $p < 0.01$, $\eta^2 = 0.4$ and no interaction between colour-word congruence and language, F_1 (1, 25) = 0.01, $p > 0.5$, $\eta^2 = 0.001$, F_2 (1, 19) = 0.4, $p > 0.5$, $\eta^2 = 0.02$. These findings suggest that participants were equally efficient at inhibiting incongruent information across their three languages.[1]

We also performed error analyses aimed at determining error patterns across congruent and incongruent conditions in the three languages. Because disfluencies were the only error type that yielded enough instances to make analyses possible, their occurrence was examined with a 2 × 3 ANOVA, with colour-word congruence (congruent, incongruent) and response language (L1, L2, L3) as within-subjects independent variables. Results yielded a main effect of colour-word congruence in disfluencies, F_1 (1, 25) = 11.4, $p < 0.01$, $\eta^2 = 0.3$, F_2 (1, 19) = 6.8, $p < 0.05$, $\eta^2 = 0.3$, showing that disfluencies were more common in the incongruent condition, likely

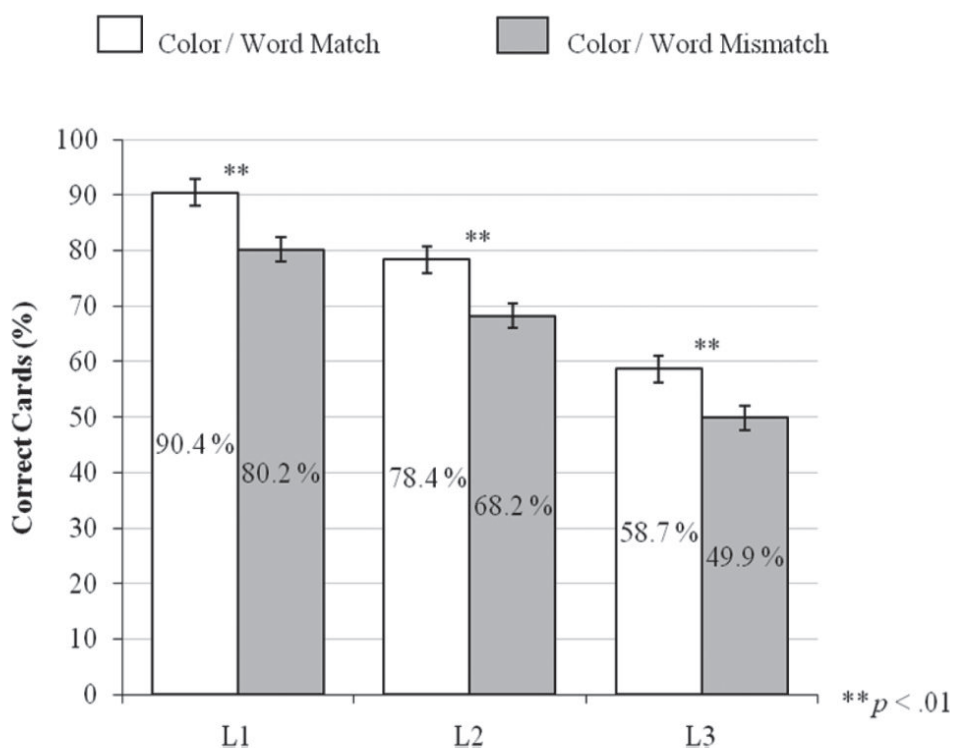

Figure 2. Performance on colour-word congruent vs. incongruent trials in participants' first language (L1), second language (L2) and third language (L3).

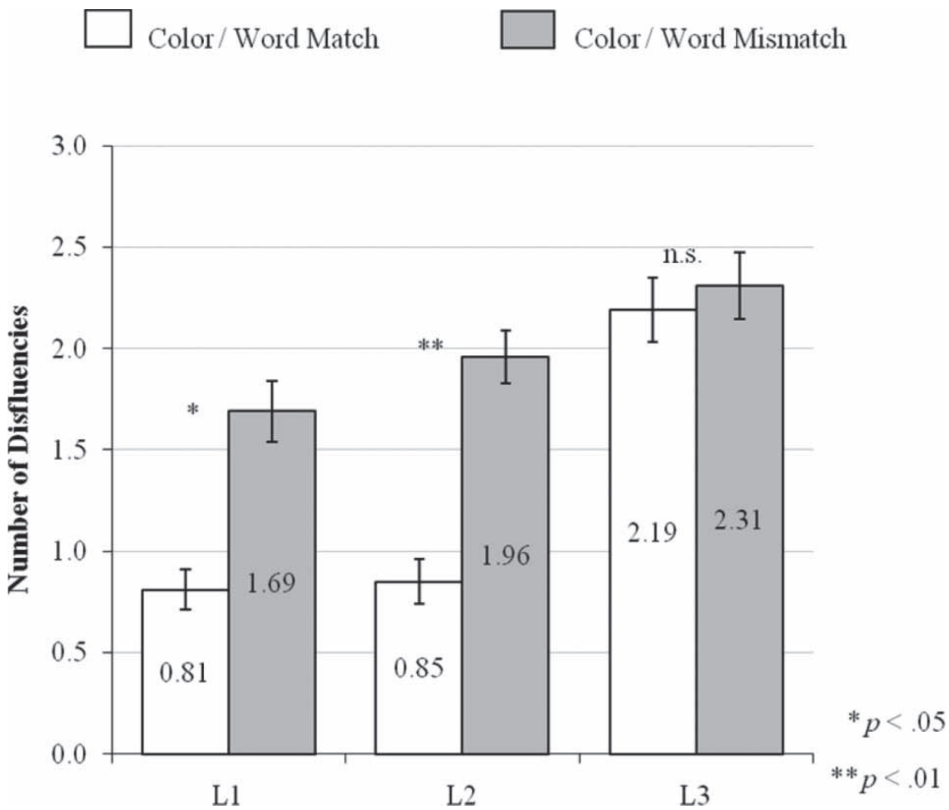

Figure 3. Number of disfluencies on colour-word congruent vs. incongruent trials in participants' first language (L1), second language (L2) and third language (L3).

due to additional task demands because of interference from irrelevant responses (see Figure 3).

Although the interaction between language and colour-word congruence did not reach significance, F_1 (1, 25) = 2.3, $p = 0.1$, $\eta^2 = 0.1$, F_2 (1, 19) = 3.5, $p = 0.08$, $\eta^2 = 0.2$, planned comparisons showed that disfluencies were more frequent in the incongruent condition than in the congruent condition for L1, t_1 (25) = 2.5, $p < 0.05$, t_2 (19) = 2.5, $p < 0.05$ and L2, t_1 (25) = 3.7, $p = 0.001$, t_2 (19) = 3.0, $p < 0.01$. However, for L3, disfluencies were equally frequent in the congruent and incongruent conditions, t_1 (25) = 0.3, $p > 0.5$, t_2 (19) = 0.3, $p > 0.5$.

Language proficiency effects in trilingual Stroop performance

Confirming our second hypothesis, analyses across participants' three languages revealed effects of language proficiency on both speed and accuracy of performance. For response speed, a 1-way ANOVA was performed with response language (L1, L2, L3) as a within-subjects variable, and the total number of items completed within 30 seconds as the dependent variable. Results yielded a main effect of language, F_1 (1, 25) = 63.5, $p < 0.001$, $\eta^2 = 0.7$, F_2 (1, 39) = 122.3, $p < 0.001$, $\eta^2 = 0.8$, with more cards completed in L1 ($M = 26.8$, $SE = 1.4$) than in L2 ($M = 21.8$, $SE = 1.5$), t_1 (25) = 2.9, $p < 0.01$, t_2 (39) = 6.0, $p < 0.001$, or L3, t_1 (25) = 8.0, $p < 0.001$, t_2 (39) = 11.1, $p < 0.001$ and more cards completed in L2 than in L3 ($M = 18.0$, $SE = 1.0$), t_1 (25) = 2.6, $p < 0.05$, t_2 (39) = 10.1, $p < 0.001$. These findings suggest that participants performed faster in their L1 than in their L2 or their L3.

Corresponding accuracy data (percentage of correct cards completed) were entered into a 1-way ANOVA, with language (L1, L2, L3) as a within-subjects variable. Results yielded a main effect of language, $F_1(1, 25) = 64.4$, $p < 0.001$, $\eta^2 = 0.7$, F_2 $(1, 39) = 84.0$, $p < 0.001$, $\eta^2 = 0.7$ and follow-up t-tests revealed that participants completed a higher percentage of correct cards in L1 than in L2, t_1 $(25) = 2.7$, $p < 0.05$, t_2 $(39) = 3.3$, $p < 0.01$, or L3, t_1 $(25) = 8.0$, $p < 0.001$, t_2 $(39) = 9.2$, $p < 0.001$, and a higher percentage of correct cards in L2 than in L3, t_1 $(25) = 3.7$, $p = 0.001$, t_2 $(39) = 7.2$, $p < 0.001$. Participants' performance accuracy and error patterns are summarised in Table 4.

A multivariate analysis of variance (MANOVA) was used to examine participants' error patterns[2] across the three languages. Language (L1, L2, L3) was entered as the within-subjects independent variable, and the four error types were entered as dependent variables (disfluencies, language interference, skipped items and read words). Results revealed a main effect of language for errors that involved production of disfluent responses, $F_1(2, 75) = 6.7$, $p < 0.01$, $\eta^2 = 0.2$, with fewer disfluencies in L1 than L3, t_1 $(25) = 3.5$, $p < 0.01$, and fewer disfluencies in L2 than L3, t_1 $(25) = 3.3$, $p < 0.01$.

A main effect of language was also found for errors of interference from other languages, $F_1(2, 75) = 4.6$, $p = 0.01$, $\eta^2 = 0.1$, with fewer errors in L1 than L3, t_1 $(25) = 2.9$, $p < 0.01$. A trend towards a main effect for skipped items was also observed, $F_1(2, 75) = 2.5$, $p = 0.09$, $\eta^2 = 0.1$, with fewer skips in L1 than L2, t_1 $(25) = 2.3$, $p < 0.05$, and fewer skips in L1 than L3, t_1 $(25) = 2.1$, $p < 0.05$. No main effect of language was observed for read word errors, but planned comparisons revealed that participants were more likely to read words instead of naming the ink colour in their L3 than in their L1, t_1 $(25) = 2.1$, $p < 0.05$.

Within-language and between-language Stroop effects

Our third hypothesis was that the within-language Stroop effect would be larger than the between-language Stroop effect, due to stronger activation of the stimulus/input language and suppression of non-target languages. To test this hypothesis, we probed for within-language Stroop effects (colour-word incongruencies during language match) and between-language Stroop effects (colour-word incongruencies during language mismatch) simultaneously, when the response language was L1, L2 or L3. A 2×2 ANOVA was conducted for each response language (L1 and L2) separately,

Table 4. Accuracy and error rates in the first, second and third languages.

	First language (L1)	Second language (L2)	Third language (L3)
	Mean (SE)	*Mean (SE)*	*Mean (SE)*
Percentage correct cards completed	85.3 (4.5)	73.3 (4.2)	54.3 (4.4)
Number of disfluencies	2.5 (0.4)	2.8 (0.4)	4.5 (0.5)
Number of word naming errors	0.2 (0.1)	0.3 (0.1)	0.7 (0.3)
Number of language interference errors	0.0 (0.0)	0.2 (0.1)	0.5 (0.2)
Number of skipped items	0.0 (0.0)	0.4 (0.2)	0.5 (0.2)

Table 5. Accuracy rates on language match (within-language effect) and language–mismatch (between-language effect) trials and colour-word congruent and incongruent trials in participants' first language (L1) and second language (L2).

	Stimulus-Response Language Match[a]	Stimulus-Response Language Mismatch[a]
Response in L1		
Colour-Word Congruent	95.1 (3.1)	91.0 (2.9)
Colour-Word Incongruent	84.8 (6.7)	77.8 (6.4)
Stroop Effect (%)	10.3*	13.2*
Response in L2		
Colour-Word Congruent	93.9 (6.6)	68.4 (6.9)
Colour-Word Incongruent	79.0 (4.9)	68.1 (5.2)
Stroop Effect (%)	14.9*	0.3[ns]
Response in L3		
Colour-Word Congruent		54.1 (6.1)
Colour-Word Incongruent		45.4 (6.7)
Stroop Effect (%)		8.7 [ns]

[a]Percentage correct cards completed by each participant.
*Marks significant Stroop effects ($p < 0.05$).
[ns]Marks non-significant Stroop effects ($p > 0.1$).

with colour-word congruence (congruent, incongruent) as a within-subjects factor, and stimulus-response language relationship (match, mismatch) as a between-subjects factor. For a summary of findings, see Table 5. Results yielded no interactions between colour-word congruence and stimulus-response language (match or mismatch) when the participants' response language was their L1, F_1 (1, 22) = 0.09, $p > 0.5$, $\eta^2 = 0.04$, F_2 (1, 38) = 0.07, $p > 0.5$, $\eta^2 = 0.002$. However, when the participants' response language was their L2, an interaction between colour-word congruence and stimulus-response language emerged, F_1 (1, 22) = 5.0, $p < 0.05$, $\eta^2 = 0.2$, F_2 (1, 38) = 3.4, $p = 0.07$, $\eta^2 = 0.1$. Follow-up t-tests suggested a colour-word Stroop effect for participants whose response language (L2) was English, creating a language match, t_1 (11) = 5.0, $p < 0.001$, t_2 (19) = 3.8, $p = 0.001$, but not for participants whose response language (L2) was another language, creating a language mismatch, t_1 (10) = 0.06, $p > 0.5$, t_2 (19) = 0.3, $p > 0.5$. These findings suggest that participants showed comparable Stroop interference effects when they responded in L1, regardless of whether stimulus and response languages matched or mismatched. In other words, both *within-language* and *between-language* Stroop effects were found when the response language was L1. However, when the response language was the participants' L2, colour-word interference effects emerged in participants whose L2 was English (creating a language match), but not in participants whose L2 was a different language (creating a language mismatch). In other words, only *within-language* Stroop effects were found when the response language was L2.

In order to examine colour-word congruency effects in the language *mismatch* condition when participants responded in their L3, performance of the 23 participants who experienced a language mismatch when responding in L3 (i.e. whose L3 was *not* English) was compared on colour-word congruent vs. incongruent conditions, but no significant differences were found, t (22) = 1.3, $p > 0.1$. Note that a *within-language* Stroop effect for L3 could not be examined because only 3 participants had English as their L3. Together, these findings suggest that *within-language* Stroop effects were found for both L1 and L2, but *between-language* Stroop effects were only found when the response language was L1. These results suggest that the overall *within-language* Stroop effect is larger than the *between-language* Stroop effect (when considering L1 and L2), lending support to our third hypothesis.

Inhibitory control effects in language mismatch

Our fourth hypothesis focused on the between-language Stroop effect, where participants had to name the ink colour in a language different from the stimulus language. To examine whether the between-language Stroop effect would be larger when the stimulus language was the dominant language than when the stimulus language was not the dominant language, analyses compared native English speakers and non-native English speakers when responding in English vs. another language. The written stimulus words always appeared in English, and accuracy was examined when participants responded in English (language match) vs. when participants responded in another language (language mismatch). English was the L1 for 11 participants and the L2 for 12 participants.[3] *T*-tests with the stimulus-response language relationship (match, mismatch) as a between-subjects independent variable were conducted to examine language match/mismatch effects when the response language was the participants' L2 and when the response language was the participants' L1. When the response language was L2, participants completed *more* correct items if their L2 was English (language match), than if their L2 was another language (language mismatch), $t_1\,(22) = 2.3$, $p < 0.05$, $t_2\,(78) = 2.2$, $p < 0.05$. When the response language was L1, participants performed *similarly* if their L1 was

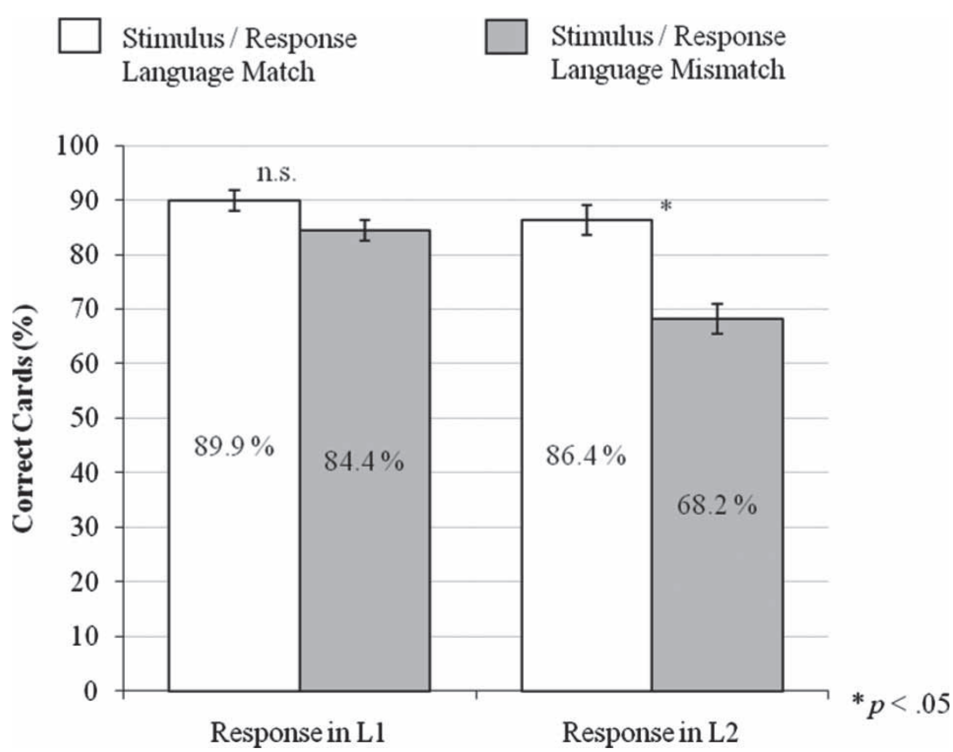

Figure 4. Performance on stimulus/response language matched vs. mismatched trials in participants' first language (L1) and second language (L2).

English (language match) or if their L1 was another language (language mismatch), t_1 (22) = 0.4, $p > 0.5$, t_2 (78) = 0.2, $p > 0.5$ (see Figure 4). These findings suggest that mismatch between stimulus and response languages resulted in greater interference when the response language was a lower-proficiency language, supporting our fourth hypothesis that the between-language Stroop effect is larger when the interfering stimuli are presented in the dominant language.

Focal and non-focal colours

In addition to Stroop and between-language effects, we also examined whether the status of focal vs. non-focal colours influenced performance, in order to test our fifth hypothesis that performance would be faster and more accurate on focal colour items. A 2×3 ANOVA was conducted, with colour (focal, non-focal) and language (L1, L2, L3) as within-subjects variables and percentage of correct cards as a dependent variable. A main effect of colour emerged, F_1 (1, 25) = 9.9, $p < 0.01$, $\eta^2 = 0.3$, F_2 (1, 19) = 11.3, $p < 0.01$, $\eta^2 = 0.4$, with participants correctly completing a higher percentage of focal colour cards (M = 79.3%, SE = 2.2) than non-focal colour cards (M = 70.7%, SE = 3.3). No interaction was found between colour and language, F_1 (1, 25) = 2.9, $p = 0.1$, $\eta^2 = 0.1$, F_2 (1, 19) = 2.1, $p > 0.1$, $\eta^2 = 0.1$. These findings support our fifth hypothesis and show that participants were more likely to accurately access focal colour-words than non-focal colour-words across their three languages.

Discussion

The present study examined trilingual Stroop performance and cognitive control across various levels of language proficiency. Results suggest that two inhibitory control processes are at work in trilinguals, one controlling the classic congruence/incongruence task effect and another controlling the languages that are irrelevant for the response. The following five findings are discussed:

(1) The trilinguals in the present study showed the classic Stroop effect and performed better when the ink colour and word were congruent than when they were incongruent in each of their three languages;
(2) Language proficiency impacted overall performance on the Stroop task, with overall performance decreasing as the level of proficiency declined;
(3) The within-language Stroop effect was more robust than the between-language Stroop effect;
(4) The magnitude of the between-language Stroop effect interacted with proficiency, with a greater between-language Stroop interference effect when the stimulus was presented in the dominant language and when task performance took place in the less-dominant language;
(5) Proficiency effects were also reflected in colour-word analyses, with better performance for focal colour-words than for non-focal colour-words.

These findings are consistent with previously-identified patterns of language co-activation and proficiency (e.g. Mägiste, 1985, van Hell & Dijkstra, 2002), as well as with the multiple (and separable) levels of cognitive control proposed by the Inhibitory Control model (Green, 1998). The model posits that a communicative goal must be established during production. In frequently-performed tasks, the relevant language task schemas are activated automatically, whereas novel tasks rely more heavily on the Supervisory Attentional System. The Supervisory Attentional System also mediates interference when several language task schemas (such as reading vs. naming of ink colour) compete. The classic Stroop effect, i.e. slower performance on incongruent vs. congruent trials, is likely due to activation of both the reading schema and the colour-naming schema on congruent trials – since both lead to the same, correct answer, there is no competition that needs to be resolved. As suggested by Roelofs (2010), responses may even be facilitated because the appropriate lexical item is activated via two different routes. Incongruent trials, however, are likely to result in competition between separate task schemas, with the Supervisory Attentional System inhibiting the reading schema and facilitating the colour-naming schema. In addition to this within-language Stroop performance, cross-linguistic Stroop performance puts further demands on the cognitive control system. Within the lexical system posited by the Inhibitory Control model, a language-specific task schema, when employed, inhibits lemmas that belong to the non-target language, with inhibition of the non-target language likely to act at the level of the lexicon. Therefore, the Inhibitory Control model predicts that, while within-language and between-language inhibition may rely on similar processes, they can act independently.

The present study confirmed our *first* hypothesis and revealed a Stroop effect in all three of the trilinguals' languages, with more accurate performance on colour-word congruent trials than colour-word incongruent trials. Finding a Stroop effect when the response language was L1, L2 or L3 suggests that incongruent written words interfere with spoken responses regardless of response language and language proficiency. However, the extent of interference from the automatic reading response – which is likely inherent in each of the three response languages since within- and between-language components are involved in all three cases, influences the magnitude of the Stroop effect. It is possible that the Stroop effect is driven more by processing costs resulting from the need to inhibit the irrelevant reading response than by successful activation of the relevant colour name.

Consistent with our *second* hypothesis and replicating previous research, performance on the trilingual Stroop task was tied to proficiency in each of the three languages. Participants' performance on the Stroop task was faster and more accurate when completing the task in their L1 than in their L2, and faster and more accurate in their L2 than in their L3. Error analyses further support language proficiency effects in multilingual Stroop performance. For example, the disfluency rates in the present study suggest that language proficiency likely influenced successful activation of relevant information during the Stroop task. Disfluencies are generally thought to signal planning difficulties during language production (e.g. Boomer, 1965; Clark & Fox Tree, 2002). Inhibitory control has been recently shown to form an integral part of the language production system that helps prevent the articulation of inappropriate words (Engelhardt, Corley, Nigg, & Ferreira, 2010). Therefore, disfluencies are to be expected when an automatic response has to be reconsidered. When participants responded in L1 and L2, significantly more disfluencies were identified in incongruent trials relative to congruent trials. This is likely due to the cognitive resources dedicated to controlling the automatic reading response during ink naming, resulting in fewer resources being available to retrieve the correct target. In the less proficient L3, an equal number of disfluencies in congruent and incongruent trials was found (in the context of overall more disfluencies in L3 than in L2 or L1). This effect might be a result of interference from stronger languages that needed to be suppressed during L3 production in both the congruent and incongruent conditions. The finding of Stroop effects on disfluencies in L1 and L2 but not L3 is also consistent with the Inhibitory Control model since the dominant language(s) are more active than the weaker languages and are therefore harder to inhibit while the successful inhibition of weaker languages is less demanding. Alternatively, the absence of differences in L3 disfluencies between congruent and incongruent conditions might reflect either a floor effect in both conditions or a proficiency effect. In other words, disfluencies in L3 may mirror lexical retrieval difficulties in the weaker language. It is possible that response times during ink colour-naming in L3 were so slow that any reading of incongruent text did not interfere. In fact, findings showed that participants responded to an average of 18 cards within the 30 second time limit in their L3 (i.e. 1.6 seconds per word stimulus) while they responded to an average of 26.8 cards in L1 (i.e. 1.1 seconds per word stimulus). With participants responding 500 milliseconds faster, on average, to each stimulus in L1 relative to L3, it is possible that automatic reading interfered strongly with early generation of a verbal response in L1 and L2 but not with later generation of a verbal response in L3.

Stronger activation and faster retrieval in more proficient languages also offer an explanation for the higher number of hesitations in both congruent and incongruent trials in L3, as well as for the patterns of skipped item errors. The percentage of skipped items was highest in L3, significantly lower in L2, and again significantly lower (no skipped items) in L1. Skipping items is likely not related to the Stroop interference effect, since skipping does not occur in the most proficient language (but the Stroop effect does). Instead, skipped items may be indicative of proficiency effects since slower retrieval makes it difficult for participants to respond in their weaker languages, which may in turn lead to strategic skipping of items to maximize performance under time pressure. Skipping of items may be interpreted as a failure to access colour names in the target language within a participant-determined time-window.

In addition to overall performance in language match and language mismatch contexts, Stroop effects were also compared across the language match and language mismatch conditions in order to test our *third* hypothesis (that Stroop effects would be smaller between-language than within-language). In addition to managing the competition between the reading schema and colour-naming schema in the classic Stroop interference effect, a multilingual needs to select the appropriate language by increasing its activation level and by inhibiting the competing inappropriate language. This is necessary because irrelevant languages are co-activated in multilinguals, even if they are not in use at a given time (Blumenfeld & Marian, 2007; Kaushanskaya & Marian, 2007; Kroll, 2008; Marian & Spivey, 2003a, 2003b). Our findings indicate that using a response language other than the stimulus language hampered performance when the response language was of lower-proficiency than the stimulus language. Findings suggested significant within-language Stroop effects in both L1 and L2, with a between-language Stroop effect found only when the response language was L1 (with L2 as the stimulus language). The findings of stronger within-language than between-language Stroop effects are consistent with previous research by Dyer (1971) and Preston and Lambert (1969) and confirm our third hypothesis. In our study, it is possible that general cross-linguistic interference from more proficient languages resulted in lower performance in weaker response languages in both congruent and incongruent conditions, effectively reducing the cross-linguistic Stroop effect when L2 and L3 were the response languages. This is in contrast to findings by Costa, Miozzo, and Caramazza (1999), who suggested that translation equivalents from the stimulus language might lead to cross-linguistic facilitation during naming. In a picture-word-interference paradigm, Costa et al. found facilitation in picture naming when translation equivalents occurred together with the picture, relative to when irrelevant words from the other non-target language occurred together with the picture. These findings of between-language facilitation rather than interference have been taken as evidence in favour of language-selective word production models, where information during naming is activated in parallel at the lemma level, with lexical form only activated and selected in the target language. Specifically, in the context of the cross-linguistic Stroop task, if naming were language-selective then facilitation of cross-linguistic colour-matches (translation equivalents) would be expected because they are conceptually shared, and interference from cross-linguistic colour-mismatches would be expected because they activate separate concepts. In contrast, if naming were language-nonselective then interference from cross-linguistic colour-matches would be expected (because two separate lexical items are activated and compete), as well as interference from cross-linguistic colour-mismatches because separate concepts and lexical items are activated and compete.

The finding that performance was slower and less accurate when the stimulus and response languages did not match, than when the stimulus and response languages matched, suggests that there may be an interaction between the inhibitory control required to suppress a non-target language and the inhibitory control required to suppress incongruent text information during the Stroop task. When participants responded in their L1, similar Stroop effects were observed within languages (response and stimulus language match) and between languages (response and stimulus language mismatch). However, when participants responded in their L2, Stroop effects were only observed within languages. Similarly, when participants responded in their L3, Stroop effects were not observed between languages. These patterns suggest that the interaction between the two sources of interference (within-language and between-language) is likely mediated by proficiency, a finding consistent with our *fourth* hypothesis.

Specifically, when participants performed the Stroop task in their L2, the participants whose L2 was English (leading to a stimulus and response language match) were faster and more accurate than the participants whose L2 was another language (leading to a stimulus and response language mismatch). In contrast, when participants performed the Stroop task in their L1, there was no difference in performance between participants whose L1 was English (creating a stimulus-response match) and participants whose L1 was another language (creating a mismatch). It is likely that this asymmetry in cross-linguistic activation was a due to differences in language proficiency. Naming in L1 was more automatic and suppression of L2 easier, resulting in no significant costs when the stimulus and response language did not match. For participants whose L2 was another language and L1 was English, naming in L2 was more effortful and suppression of L1 (the stimulus language) was more difficult, resulting in a decline in performance when stimulus language and response language did not match. This finding is consistent with previous research that cross-linguistic interference is stronger when the non-target language is more proficient than the target language (e.g. Blumenfeld & Marian, 2007).

The observed asymmetry in cross-linguistic interference during Stroop performance supports our fourth hypothesis, that the between-language Stroop effect would be larger when the stimulus language was the dominant language than when the stimulus language was not the dominant language. This asymmetry is also consistent with previous findings (Abunuwara, 1992; van Heuven et al., 2011). Re-analyses of means reported in Abunuwara (1992) suggest that, as in our study, participants performed more slowly across their three languages when stimulus and response languages mismatched (969.8 msec per response) than when they matched (927.3 msec per response). In addition, Abunuwara's participants were slower to respond in L2 when the stimulus language was L1 (1004 msec) than they were at responding in L1 when the stimulus language was L2 (850 msec). These findings are similar to the asymmetry in cross-linguistic competition between L1 and L2 that has been identified in previous studies where L1 was more proficient than L2 (e.g. Meuter, 1994).

Our findings of comparable cross-linguistic interference patterns in both colour-word congruent and incongruent trials during L2 and L3 naming are consistent with findings suggesting cross-linguistic interference rather than facilitation of translation equivalents (e.g. *red* when participants had to produce *rojo* (red) in Spanish; also see Hermans, 2004). Therefore, our findings indicate that, at least for L2 and L3, lexical selection during naming may not be language-selective. In contrast, it is possible that in L1, where a robust between-language Stroop effect was found, lexical access is at least functionally more language-selective. Specifically, during L1 production, fewer cognitive resources are likely to be dedicated to the selection of relevant information, with cross-linguistic interference from weaker languages more readily controlled. Therefore, lexical items from other languages are more likely to be fully inhibited during L1 naming. Such a scenario may result in facilitation (but not interference) during the cross-linguistic congruent condition, with the concept facilitated by the translation equivalent and without competition at the lexical level. Conversely, interference would remain during the cross-linguistic incongruent condition, with the conceptual activation of a conflicting colour term. Therefore, it can be preliminarily concluded that proficiency in both the response and stimulus languages may determine success of cross-linguistic inhibition at the lexical level (and thus the extent of functional language selectivity) during cross-linguistic Stroop naming.

Evidence from language switches (i.e. errors where an inappropriate language is used as the response language) also supports the asymmetry in interference patterns and allows us to conclude that other languages are co-activated. We found significantly more errors in L3 than in L1. The higher level of interference in the less proficient language indicates that co-activation of stronger languages is more difficult to suppress, a finding consistent with previous literature (e.g. Meuter, 2005) and with predictions of the Inhibitory Control model. In addition to previous studies of bilingual Stroop effects (e.g. Costa, Albareda, & Santesteban, 2008; Tzelgov, Henik, & Leisner, 1990), current findings of increased cross-linguistic interference errors with decreased language proficiency suggest that the extent of cross-linguistic interference may increase the weaker the target language gets and the stronger the non-target languages become.

Finally, supporting our *fifth hypothesis*, participants performed better in all three languages when the ink was a focal colour than when the ink was a non-focal colour.

This is likely because focal colour terms are more entrenched than non-focal colour terms due to their higher frequency and earlier acquisition (e.g. Regier, Kay, & Cook, 2005) leading to faster retrieval. The improved performance observed in trilinguals when presented with focal colours than when presented with non-focal colours is consistent with our hypothesis and suggests that colour labels show effects similar to language proficiency.

In conclusion, our study illustrates that two different types of control mechanisms are at work in multilinguals' Stroop performance – one mechanism to control the classic Stroop interference effect and a second language control mechanism to inhibit non-target languages. The need to simultaneously negotiate multiple inhibitory control demands inherent in the multilingual Stroop task is reflective of the cognitive demands experienced by multilinguals in everyday settings, where both non-target language inhibition and task-irrelevant information inhibition are required. The practice in managing such competing demands is likely to contribute to the executive control advantages observed in speakers of more than one language. Future studies may further explore the cognitive consequences of simultaneously performing multiple types of inhibitory control tasks, where one type of inhibitory control involves inhibition of an entire language system and another type of inhibitory control involves inhibition of task-irrelevant information.

Acknowledgements

Preparation of this manuscript was supported in part by grant NICHD RO1HD059858 to the first author. We thank Maria Khristina S. Manueli and Vera Denisova for their contributions to data collection and the members of the *Northwestern Bilingualism and Psycholinguistics Laboratory* for comments on earlier drafts of this paper.

Notes

1. Response speed is reported across the three languages, but not for incongruent and congruent items separately. Because incongruent and congruent trials were presented in an intermixed fashion, with an equal number of responses in the two conditions, separate response speed measures for the two conditions could not be obtained.
2. For analyses on specific error types, only by-subjects analyses (F_1, t_1) are reported, because by-item analyses were not possible due to the small number of errors.
3. Analyses on match vs. mismatch of stimulus and response languages were conducted only on L1 and L2, because not enough participants were available who showed a match between stimulus language and response language for their L3 (i.e. L3 was English for three participants only), therefore this analysis was conducted with 23 participants.

References

Abunuwara, E. (1992). The structure of the trilingual lexicon. *European Journal of Cognitive Psychology*, *4*, 311–322.

Aitchison, J. (2003). *Words in the mind*. Oxford: Blackwell.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology*, *60*(1), 68–79.

Bialystok, E., Craik, F.I.M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon Task. *Psychology and Aging*, *19*, 290–303.

Bialystok, E., Craik, F.I.M., & 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.

Blumenfeld, H.K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633–660.

Blumenfeld, H.K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition*, *118*(2), 245–257.

Boomer, D.S. (1965). Hesitation and grammatical encoding. *Language and Speech*, *8*, 148–158.

Boyden, J.G., & Gilpin, A.R. (1978). Matching familiar figures test and Stroop test performance in adults. *Perceptual and Motor Skills*, *46*, 854–858.

Bugg, J.M., Jacoby, L.L., & Toth, J.P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, *36*(8), 1484–1494.

Chen, H.-C., & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *12*, 397–401.

Christoffels, I.K., De Groot, A.M.B., & Kroll, J.F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*, 324–345.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.

Clark, H.H., & Fox Tree, J.E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111.

Cohen, J.D., Dunbar, K., & McClelland, J.L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361.

Colzato, L.S., Bajo, M.T., van den Wildenberg, W., & Paolieri, D. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 302–312.

Costa, A., Albareda, B., & Santesteban, M. (2008). Assessing the presence of lexical competition across languages: Evidence from the Stroop task. *Bilingualism: Language and Cognition*, *11*(2), 121–131.

Costa, A., Hernández, M., & Sebastian-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*, 59–86.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*, 365–397.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491–511.

Dijkstra, A., & Van Heuven, W.J.B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189–225). London: Erlbaum.

Dyer, F. (1971). Color-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 297–302.

Engelhardt, P.E., Corley, M., Nigg, J.T., & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Memory and Cognition*, *38*, 617–628.

Fang, S.-R., Tzeng, O.J., & Alva, L. (1981). Intralanguage vs. interlanguage Stroop effects in two types of writing systems. *Memory and Cognition*, *9*, 609–617.

Favreau, M., & Segalowitz, N.S. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory & Cognition*, *11*, 565–574.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Psychiatric Research*, *12*, 189–198.

Goldfarb, L., & Tzelgov, J. (2007). The cause of the within language Stroop Superiority Effect and its implications. *Quarterly Journal of Experimental Psychology*, *60*, 179–185.

Green, D.W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67–81.

Hays, D.G., Margolis, E., Naroll, R., & Perkins, D.R. (1972). Color term salience. *American Anthropologist*, *74*, 1107–1121.

Hermans, D. (2004). Between-language identity effects in picture-word interference tasks: A challenge for language-nonspecific or language-specific models of lexical access? *International Journal of Bilingualism*, *8*(2), 115–125.

Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a short orientation-memory-concentration test of cognitive impairment. *American Journal of Psychiatry*, *140*, 734–739.

Kaushanskaya, M., & Marian, V. (2007). Non-target language recognition and interference: Evidence from eye-tracking and picture naming. *Language Learning*, *57*(1), 119–163.

Kavé, G., Eyal, N., Shorek, A., & Cohen-Mansfield, J. (2008). Multilingualism and cognitive state in the oldest old. *Psychology and Aging*, *23*(1), 70–78.

Kay, P., & McDaniel, C.K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, *54*, 610–646.

Kornblum, S. (1992). Dimensional overlap and dimensional relevance in stimulus-response and stimulus-stimulus compatibility. In G. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior II* (pp. 743–777). Amsterdam: North-Holland.

Kroll, J.F. (2008). Juggling two languages in one mind. *Psychological Science Agenda, American Psychological Association*, *22*(1). Retrieved from http://www.apa.org/science/about/psa/2008/01/kroll.aspx

Kroll, J.F., Bobb, S.C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*(2), 119–135.

Levy, B.J., McVeigh, N.D., Marful, A., & Anderson, M. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science*, *18*(1), 29–34.

Linck, J.A., Kroll, J.F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second language learning. *Psychological Science*, *20*(12), 1507–1515.

Linck, J.A., Schwieter, J.W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition*, *15*, 651–662.

Lindsey, D.T., & Brown, A.M. (2006). Universality of color names. *PNAS*, *103*(44), 16608–16613.

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

Mägiste, E. (1985). Development of intra- and interlingual Interference in bilinguals. *Journal of Psycholinguistic Research*, *14*(2), 137–154.

Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*, 173–193.

Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*, 97–115.

Meuter, R.F.I. (1994). *Language switching in naming tasks* (Unpublished Ph.D. thesis). University of Oxford, UK.

Meuter, R.F.I. (2005). Language selection in bilinguals. In J.F. Kroll & A.M.B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (Chapter 17, pp. 349–370). Oxford, UK: Oxford University Press.

Meuter, R.F.I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*, 25–40.

Oren, R., & Breznitz, Z. (2005). Reading processes in L1 and L2 among dyslexic as compared to regular bilingual readers: Behavioral and electrophysiological evidence. *Journal of Neurolinguistics*, *18*(2), 127–151.

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*, 295–301.

Regier, T., Kay, P., & Cook, R.S. (2005). Focal colors are universal after all. *PNAS*, *102*(23), 8386–8391.

Roelofs, A. (2010). Attention and facilitation: Converging information versus inadvertent reading in Stroop task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 411–422.

Segalowitz, S.J., Segalowitz, N.S., & Wood, A.G. (1998). Assessing the development of automaticity in second-language word recognition. *Applied Psycholinguistics*, *19*(1), 53–67.

Silverberg, S., & Samuel, A.G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *Journal of Memory and Language*, *51*, 381–398.

Stroop, J.R. (1935). Studies in interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–661.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–660.

Tzelgov, J., Henik, A., & Leisner, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology*, *16*, 760–771.

van Hell, J.G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*, 780–789.

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*, 1–15.

Warren, L.R., & Marsh, G.R. (1978). Hemispheric asymmetry in the processing of Stroop stimuli. *Bulletin of the Psychonomic Society*, *12*, 214–216.

Zied, K.M., Phillipe, A., Pinon, K., Havet-Thomassin, V., Ghislaine, A., Roy, A.R., & Le Gall, D. (2004). Bilingualism and adult differences in inhibitory mechanisms: Evidence from a bilingual stroop task. *Brain and Cognition*, *54*(3), 254–256.