Journal of Memory and Language 105 (2019) 76–92

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

On the encapsulation of bilingual language control

Kenneth R. Paap*, Regina Anders-Jefferson, Roman Mikulinsky, Shigeaki Masuda, Lauren Mason

San Francisco State University, United States

ARTICLE INFO

Keywords:
Bilingualism
Inhibitory control
Stimulus–response competition
Congruency effects, executive functions

ABSTRACT

One purpose of this study was to test the hypothesis that prevalent patterns of bilingual language control lead to greater enhancement of the ability to resolve Stimulus-Stimulus conflict compared to Stimulus-Response conflict. To that end 104 bilinguals and 62 monolinguals completed four commonly used nonverbal interference tasks with varied S-S and S-R incompatibilities. No bilingual advantages were observed in any of the tasks. A second purpose was to further investigate whether a general inhibitory-control ability exists by examining inter-task correlations in the current and previous studies. We conclude that there may be a shared mechanism for interference control across spatial Stroop tasks and Simon tasks, but that this mechanism is apparently not recruited during bilingual language control to the extent that such practice would enhance a general ability. Rather, inhibition in language processing may be encapsulated within a lexical-identification system as assumed by Dijkstra and van Heuven's bilingual-interactive-activation plus model and its update Multilink.

Introduction

Bilinguals have been claimed to perform better than monolinguals in nonverbal interference tasks because they constantly practice inhibiting the language currently not in use. There is no doubt that when bilinguals intend to produce a sentence in a target language, the translation equivalents in the other language become coactivated and create competition (see Paap, 2018 for a review). For example, the intention to say "bull" may coactivate "toro" in a Spanish-English bilingual. Likewise, during bilingual language comprehension similar or identical word forms are coactivated in both mental lexicons. For example, even in an English context /lief/ activates the word for "sweet" (a false cognate) in Dutch-English bilinguals. There are several possible adaptations to this competition: (1) the non-target lexical competitors are inhibited via a general inhibitory control mechanism, (2) the non-target lexical competitors are inhibited via a specialized mechanism within a word identification system, (3) the target candidates are up regulated, or (4) no conflict resolution mechanism is employed at the level of lexical representations. If the last is true, then bilinguals (5) may live with the competition and the occasional unintended intrusion from the other language, (6) employ domain-general response suppression, or (7) rely on a specialized mechanism for articulatory suppression. Some of these options could occur in combination, but logically only when (1) or (6) are involved could bilingual language-control transfer to and lead to bilingual advantages in nonverbal interference tasks requiring manual responses.

Two models of bilingual language control

There are influential models of bilingual language-control that either assume or eschew the recruitment of general inhibitory control. Green's (1998) inhibitory control model (ICM) presumes the need for higher-level task schemas controlled by an even higher level supervisory attentional system (SAS). For example, when presented with a printed word a bilingual can be given the task of reading it silently, reading it aloud, translating it, generating an associate, classifying it as animate or inanimate, and so on. Each requires a different task schema. The ICM emphasizes that task schemas compete with each other for controlling action. Thus, depending on changes in topic or conversational partners, bilinguals may decide to switch from the "speak-English" schema to a "speak-Spanish" schema that reactively inhibits the English lexical representations via their language tags. The crucial point is that bilingual language control is governed by a SAS that selects and schedules specific task schemas (e.g., "speak English") in the same way as it does in novel situations and tasks (e.g., "name the color" in the standard Stroop task). If the ICM assumptions are correct, then bilingual advantages should frequently occur in nonverbal tasks where conflict between stimulus representations needs to be resolved.

A contrasting view is presented by Dijkstra and van Heuven (2002) bilingual interactive-activation-plus model (BIA+). The BIA+ makes a clear distinction between the encapsulated *word identification system* (WIS) and domain-general cognitive control (*the task schema/decision system*). Because the WIS is encapsulated, the activation of the word-

* Corresponding author at: Department of Psychology, San Francisco State University, San Francisco, CA 94132, United States.
E-mail address: kenp@sfsu.edu (K.R. Paap).

https://doi.org/10.1016/j.jml.2018.12.001
Received 21 January 2018; Received in revised form 30 November 2018

form nodes cannot be affected by top-down control. The WIS is assumed to be part of a larger *language processing system*. The BIA+ further assumes that the syntactic and semantic constraints of this language processor can affect the activation of the lexical nodes. In the BIA+ the activation of word-form units is affected only by processing within the language processing system. Linguistic information from the visual input (*c-a-t*) or from the preceding sentence context (e.g., *The mouse was chased by the…*) can and will affect the activation of word form units, but nonlinguistic context such as the participant's expectations and strategies will not. Rather, these nonlinguistic factors influence performance via adjustments to parameter settings in the task schema/decision system. The decision mechanism is part of a task schema and reads out the activation of nodes within the WIS.

For example, in general lexical decision tasks a word response can be made when a word unit from either language crosses an activation threshold. As the task unfolds, the activation thresholds and temporal deadlines may be adapted based on the frequency of different types of items. These adaptations involve dynamic adjustments of response criteria, but do not affect the activation level of individual word form units or the language nodes. For present purposes, the crucial aspect of BIA+ is that inhibitory control is domain specific and encapsulated within the word-identification and language processing system. It is a specific instantiation of Adaptation 2 from above that the non-target lexical competitors are inhibited by a specialized mechanism. If the BIA+ correctly characterizes bilingual language control, then there should be no differences between bilinguals and monolinguals in nonverbal tasks requiring conflict resolution between stimulus representations because task-specific mechanisms, by definition, do not generalize.

The computational version of the BIA+ was limited to visual word recognition, but as a verbal model included both phonological and semantic representations. Dijkstra et al. (2018) extended the computational version of BIA+ to production. This new model, dubbed Multilink, successfully simulated many classic results obtained with picture naming and translation tasks. Critical to the present discussion, bilingual language control remains encapsulated within the WIS and thus the model predicts no transfer to tasks using general executive functioning (EF). Critical to an issue raised later, the WIS eschews inhibitory connections both within and between representational levels. This is an instantiation of Adaptation 4 that no conflict resolution mechanism is employed at the level of lexical representations. To state the obvious, if there are no inhibitory connections, then there is no inhibitory control ability that can be enhanced by practice.

Conflict between stimulus representations or between stimulus and response

Although the ICM and BIA+/Multilink make clear and opposing assumptions regarding the direct application of general inhibitory control on lexical units, the models do not explicitly consider whether or how conflict might be resolved at the response level. To return to the earlier example, if the orthographic word forms for both BULL and TORO are coactivated when the intention is to read aloud in English, but the lexical competitor TORO is not suppressed, then the articulatory plans associated with the corresponding phonological word forms will compete. Inhibition may be applied in order to induce fluent speech and avoid intrusions from the unintended language. If this recruits a general inhibitory control mechanism, then that mechanism may be strengthened through ubiquitous practice.

The possibility of inhibitory control at either the stimulus or response levels has been extensively studied in cognitive psychology. Consider the most influential taxonomy for analyzing task differences, Kornblum (1994) Dimensional Overlap Model. The model distinguishes between tasks with stimulus-response (S-R) or stimulus-stimulus (S-S) incompatibility. The incompatibilities of the four tasks used in the present study are illustrated in Fig. 1. For each panel the S-R rule is at the top, a display representing a correct response on an incongruent trial is in the middle, and the Venn diagrams at the bottom represent, by their intersections, where conflict can be generated and resolved. The first (leftmost) panel is a pure S-R task that is often referred to as a Simon task. A single arrow (pointing either up or down) is presented either to the left or right of fixation and the rule is to press the left key if the arrow points up and the right key if it points down. Given the natural tendency to react toward the source of stimulation (Simon & Small, 1969), competition can occur when the physical location and the rule are incongruent as illustrated by overlap between S_I (the irrelevant stimulus) and R (the correct response). Note that there is no overlap between the task relevant and task irrelevant stimulus representations because the arrow's form varies on an up-down dimension whereas its location varies on a left-right axis. More generally, on the incongruent trials of an S-R task the response specified by the task rule is incompatible with the response of a prepotent but task irrelevant stimulus.

If the vertical arrows are displaced either above or below fixation as in third panel, the task transforms into a pure S-S task. Because the upward pointing arrow appears below the fixation, the task-relevant dimension (up arrow) is opposite its location (below) and causes S-S incompatibility. There is no S-R incompatibility because the layout of the response keys (horizontal) is orthogonal to the up-down direction of the arrow. This pure S-S task will be referred to as the vertical Stroop task.

As illustrated in the second panel tasks have both S-S and S-R incompatibilities when arrows pointing left or right are displaced either to the left or right of fixation. On an incongruent trial the task-relevant direction (e.g., left) is incompatible with task-irrelevant location (e.g., right) causing S-S incompatibility. Furthermore, the task-relevant direction (e.g., left) is also incompatible with the predisposition to react toward the source of stimulation (e.g., right) causing S-R incompatibility. This task will be referred to as the spatial Stroop task and in the literature is sometimes referred to as the Simon arrows task.

The rightmost panel illustrates a variant of the flanker task. If the central arrow points left press the left key, if it points right then press the right key. The flanking arrows are irrelevant but when they point in the opposite direction there is both S-S and S-R incompatibility because the relevant and irrelevant stimuli share the same dimension (viz., left-right) as do the relevant stimulus and the response.

The Kornblum taxonomy and this set of tasks provides a testbed for how S-S or S-R inhibitory control is moderated by any individual difference or group variable such as aging, video gaming, or bilingualism. Given bilingualism as our focus, here are some possibilities. If S-S conflict is resolved by language-specific mechanisms (as assumed in the BIA+/Multilink), then no bilingual advantage will accrue due to S-S incompatibilities and no advantage will occur in the vertical Stroop task as its only source of conflict is S-S. In contrast, if domain-general inhibitory control is routinely recruited (as assumed in the ICM), then bilingual advantages would be expected in the three tasks that have S-S incompatibilities. To take a third example, if a general response suppression mechanism is routinely recruited, during bilingual language control, then a bilingual advantage should appear in the three S-R tasks. Finally, if bilingual language control does not involve general inhibitory control at either the lexical or response levels, then there should be no performance differences between bilinguals and monolinguals on any of the tasks. The consequences of other permutations can be deduced from inspection of Fig. 1.

Blumenfeld and Marian (2014) developed a specific hypothesis based on the Kornblum taxonomy and the following analysis of bilingual language control. They reasoned that S-S competition may be the most common type of bilingual competition because competition between the lexicons occurs during both comprehension and production, whereas S-R inhibition may be limited to production contexts where both languages remain active until the response stage. Because cross-linguistic co-activation results in S-S competition more frequently than S-R competition, they predicted that bilingual advantages would be

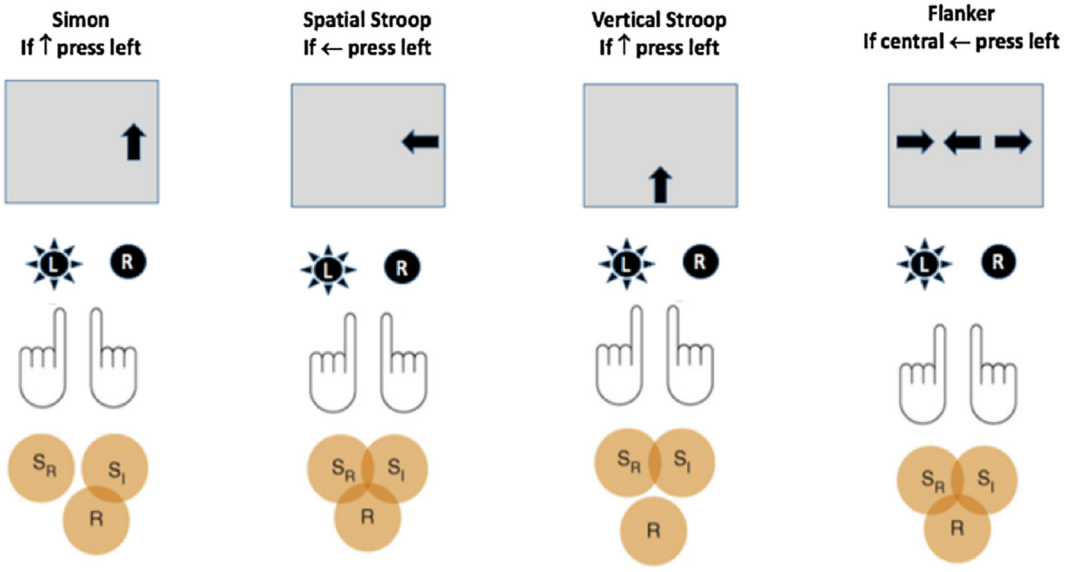


Fig. 1. The four nonverbal interference tasks used in the present study. The representational scheme is based on Fig. 1 from Egner (2008). From top to bottom for each task is the name of the task, the response rule, a screen illustrating an incongruent trial, response keys with the correct response radiated, and finally Venn diagrams showing potential conflict between the irrelevant stimulus (S_I), relevant stimulus (S_R), and response (R).

larger in an S-S task compared to a pure S-R task. The S-S task tested by Blumenfeld and Marian was like the spatial Stroop task illustrated in the third panel of Fig. 1 and their S-R task was like the Simon task shown in the first panel.

Across Blumenfeld and Marian's two experiments there was some evidence for bilingual advantages in the spatial Stroop task. Experiment 1 recruited young adults from the Chicago area and included 38 English-Spanish bilinguals and 30 English monolinguals. Experiment 2 recruited young adults from San Diego with 60 participants in each group. Overall accuracy (about 90%) across experiments and groups was somewhat lower than typically observed for young adults in nonverbal interference tasks (see Table 1). To protect against a speed-accuracy trade-off Blumenfeld and Marian used efficiency scores (ES) as a composite measure of speed and accuracy. An ES is calculated as the mean correct RT divided by the proportion correct (PC). In Experiment 1, the Task × Group × Congruency interaction was significant for the ES and PC measures but not for RT. For the ES and PC measures, the pattern of interaction was as predicted: larger bilingual advantages in the S-S task and larger task differences for bilinguals. Despite the larger sample size and greater power, neither the Group × Congruency nor the three-way interaction with Task was significant for any of the three measures in Experiment 2 and consequently there was no statistical support in the second experiment for the hypothesis that the interference effects would be smaller in bilinguals compared to monolinguals in the S-S (spatial Stroop) task. Blumenfeld and Marian's reconciliation of the inconsistencies between their two experiments is considered in our general discussion.

Previous research testing for bilingual advantages in S-S and S-R tasks

Blumenfeld and Marian reviewed 21 prior tests and showed that bilingual advantages tend to occur more often in S-S tasks compared to S-R tasks. Only one pair of tests in their review involved the same sample of participants. Although their review supported the hypothesis that bilingual advantages occur more often in S-S tasks it is risky to compare the results of studies using an S-S task to different studies using an S-R task. A chronic problem in testing for bilingual advantages is that bilinguals and monolinguals often differ in terms of ethnicity, culture, education, immigrant status, SES and other factors that may influence measures of inhibitory control (see Paap, Johnson, & Sawi, 2015, for a review). For that reason, it is better to focus on studies that test the same language groups with both an S-S and an S-R task. In these studies, any confound that favors one language group over the other should apply a comparable bias to each task.

Do bilingual advantages occur more often in S-S tasks when the same participants perform both types of tasks? Table 1 shows the results of each such study and a summary on the bottom row. The first important observation is that statistically significant advantages in inhibitory control are infrequent. There was only 1 bilingual advantage and that occurred in a pure S-R Simon task. There were also 4 monolingual advantages, and 19 non-significant differences. When the magnitude of the bilingual advantage is averaged across studies, neither task type shows a bilingual advantage, and the magnitude of the monolingual advantage (−6.5 ms) is actually greater for the S-S tasks than for the S-R tasks (−3.9 ms). In summary, when examining only those studies in which the same participants completed both an S-S and S-R task the evidence supporting the hypothesis that bilingual-language control employs a general S-S inhibitory control mechanism is absent.

Meta-analyses of bilingual-advantage hypothesis across all nonverbal interference tasks

To this point the emphasis has been on whether S-S tasks may yield more consistent bilingual advantages than S-R tasks. Blumenfeld and Marian's review of mostly between-subject comparisons appeared more promising in that regard than our review of the 12 studies that used within-subject designs. Given that the S-S versus S-R distinction may not be a good predictor of bilingual advantages, it is also instructive to look at all of the data derived from nonverbal interference tasks that include both congruent and incongruent trials. Three recent meta-analyses converge on the conclusion that significant bilingual advantages in inhibitory control are relatively rare (15% of all comparisons), that the average effect sizes are very small, and that there is evidence for publication bias, which when taken into account, appears to completely eliminate the effect. In the meta-analysis by Paap (2019) the mean bilingual advantage across all 146 comparisons was +4.4 ms. If the 146 effect sizes are treated as a single sample the Bayes Factor (using the JZS prior and Rouder's calculator) favoring the alternative is 2.9, an odds ratio that according to Jeffreys (1961) guidelines is barely worth mentioning. The meta-analyses by Lehtonen et al. (2018) examined bilingual advantages across six domains of EF, but their analysis of inhibitory control is most central to our focus. Their meta-analysis

Table 1
Bilingual advantages (RT congruency effects) in S-S versus S-R tasks. Below each specific task is the percentage of correct responses averaged across both the congruent and incongruent trials.

Study	S-R Task	Bi. Adv.	Decision	S-S Task	Bi. Adv.	Decision
Antón, Carreiras, and Duñabeitia (submitted for publication)	Simon 96%	−3	ns	Flanker 98%	0	ns
Bialystok (2006)	Simon 97%	0#^	ns	Spatial Stroop 98%	−10#^	ns
Blackburn (2013)	Simon 96%	−4	ns, p = .53	Flanker 98%	+11	ns, p = .48
Guido Mendes (2016)	Simon 97%	+5	ns, p = .054	Flanker 98%	+5	ns
Humphrey and Valian (2012)	Simon	−1	ns	Flanker	−14	ns
Kousaie and Phillips (2012)	Simon 95%	−6^	ns, p = .27	Flanker 93%	0^	ns, p = .94
Kousaie and Phillips (2017)	Simon 92%	−8^	ns, p = .12	Flanker 83%	−8^	ns, p = .08
Mohades et al. (2014)	Simon 96%	−13	p = .05	Numerical Stroop, 95%	−48	p = .045
Paap and Greenberg (2013), Study 3	Simon 98%	−10	p = .026	flanker 99%	−4	ns, p = .68
Paap and Sawi (2014)	Simon 98%	−13	p = .01	Flanker 98%	−9	ns, p = .09
Prior and MacWhinney (2010)	Simon	?	ns	Flanker	?	ns
Woumans, Ceuleers, Van der Linden, Szmalec, and Duyck (2015)	Simon 94%	+10	p = .004	Flanker 96%	+6	ns, p = .84
Composite	S-R 96%	−3.9	8 ns 1 B Adv 3 M Adv	S-S 96%	−6.5	11 ns 0 B Adv 1 M Adv

Averaged across both videogame players and nonplayers for the low-switch condition that showed typical congruency effects.
^ Estimated from figures.

used a wider definition of inhibitory control tasks and identified a more heterogeneous set of 212 effect sizes compared to Paap (2019). The mean effect size for inhibitory control in Lehtonen et al. was Hedge's $g = +0.11$ [+0.05, +0.18], but when corrected for bias the mean was no longer significant, $g = -0.02$ [−0.12, +0.08]. Donnelly, Brooks, and Homer (in press) reported a meta-analysis of 80 studies using a multiverse analysis approach where each research question was tested many times while making different decisions about the inclusion criteria. The bilingual-advantage effect size, corrected for publication bias, was negative, $g = -22$ [−0.35, −0.09]. The Lehtonen et al. meta-analysis was restricted to studies using participants 18 years and older, Paap to 6 years and older; and Donnelly et al. to 4.5 years and older.

If bilingual advantages in inhibitory control are indistinguishable from zero, then neutral observers, and certainly skeptics, may well question the value of another study comparing bilinguals to monolinguals. However, proponents of the bilingual advantage hypothesis continue to argue that bilingualism enhances domain-general EF, but only when sufficient dimensions of bilingual experience are combined with sensitive tests of the affected component(s) of cognitive control (e.g., Bialystok, 2017; Tabori, Mech, & Atagi, 2018; Struys, Duyck, & Woumans, 2018). For that reason it is worthwhile to conduct a study that compares four closely-matched tasks that differ with respect to the need for spatial attention (only the flanker does) and includes pure S-S (vertical Stroop) and pure S-R (Simon) tasks.

On the other hand the possibility that bilingual advantages in inhibitory control do not exist raises the question why? One explanation already introduced is that bilingual language control relies on language-specific, not domain general, control mechanisms. That is, bilingual language control may be encapsulated within the language-processing system as implemented in the architecture of Multilink. A related, but more expansive explanation is that domain-general and top-down inhibitory control may not exist, that is, inhibitory control may be task-specific. Thus, an additional important purpose of the present study is to determine the psychometric structure among a set of four tasks that are frequently used to measure "inhibition".

Intertask correlations between Simon, spatial Stroop, and flanker

If the interference scores derived from any two nonverbal interference scores correlate, this can be taken as evidence that they share a conflict-resolution mechanism. However, Kornblum's taxonomy implies that different mechanisms are employed to resolve S-S and S-R conflict. Thus, the intertask correlations between the interference scores for the four tasks shown in Fig. 1 should increase as pairs of tasks share neither type of conflict, one type, or both.

Only three of the six possible pairs from this set of four tasks have been previously tested using within-subjects designs. Each of the three pairings are discussed next. The interference scores for the pure S-R Simon task and the pure S-S vertical Stroop should not correlate if they rely on different inhibitory abilities. Although nine published studies had the same participants do both tasks (either as separate tasks or as an integrated task with trial types mixed), eight did not report intertask correlations. However, one study did (Li, Nan, Wang, & Liu, 2014) and the outcome was consistent with the prediction that S-S and S-R tasks employ separate mechanisms. They reported a small, negative, and non-significant correlation, $r(28) = -0.05$ between a pure S-R Simon task and a pure S-S vertical Stroop task. This finding is shown in the top block of Table 2.

Table 2
Intertask correlations between tasks with various combinations of S-S and S-R conflict based on Kornblum's taxonomy.

Study	Task Type		r	Decision
	Pure S-R	Pure S-S		
Li et al. (2014)	Simon (horizontal, up-down)	Vertical Stroop (vertical, up-down)	−0.05	ns
	Pure S-R	S-S & S-R		
Fan, Flombaum, McCandliss, Thomas, and Posner (Exp 1, 2003)	Simon (horizontal, lion-dog)	Flanker (arrow, left-right)	−0.02	ns
Fan, McCandliss, Sommer, Raz, and Posner (Exp. 2, 2002)	Simon (horizontal, lion-dog)	ANT flanker	+0.23	ns
Humphrey and Valian (2013)	Simon	Flanker		ns
Kousaie and Phillips (2012)	Simon (horizontal, blue, red	Arrow flanker		ns
Kousaie and Phillips (2017)	Simon (horizontal, blue-red)	Arrow flanker		ns
Paap and Greenberg (Exp 3, 2013)	Simon (horizontal, z-/)	ANT flanker	−0.01	ns
Paap and Sawi (2014)	Simon (horizontal, z-/)	ANT flanker	+0.14	ns
Rey-Mermet et al. (2018)	Simon (horizontal, square-circle)	Arrow flanker	+0.05	ns
Rey-Mermet et al. (2018)	Simon (horizontal, square-circle)	Letter flanker	+0.14	ns
Stins, Polderman, Boomsma, and de Geus (2007)	Simon (horizontal, green-red)	Arrow flanker	< +0.2	ns
Keye, Wilhelm, Oberauer, and van Ravenzwaaij (2009)	Simon (vertical, diamond-square)	Arrow flanker	+0.14	ns
Wilhelm, Hildebrandt, and Oberauer (2013)	Simon (vertical, diamond-square)	Arrow flanker	+0.01	ns
	S-S & S-R	S-S & S-R		
de Bruin and Della Sala (2017)	Spatial Stroop (horizontal, left-right)	Arrow flanker	+0.14	ns
Pettigrew and Martin (2014)	Spatial Stroop (horizontal, left-right)	Arrow flanker	−0.07	ns
Wöstmann et al. (2013)	Spatial Stroop (horizontal, left-right)	Arrow flanker	+0.24	p < .001

Note. Within parentheses for each task is an ordered pair showing the displacement relevant to fixation (horizontal or vertical) and either the directions of the arrow targets (up-down or left right) or the two targets in a Simon task.

As shown in the second block of Table 2, 12 experiments have correlated a pure S-R Simon task with a flanker task that includes both S-S and S-R incompatibilities. The intertask correlations between the interferences scores for all 12 were nonsignificant with Pearson *r*'s ranging −0.02 to +0.23. A parsimonious interpretation is that the pure S-R Simon task and the flanker task do not share a conflict resolution mechanism. Thus, the potential S-R conflict in a flanker task does not appear to be sufficient to generate a correlation with the interference scores in a pure S-R Simon task.

Based on the Kornblum taxonomy, the cases most likely to correlate have paired the spatial Stroop task with a flanker task as both tasks include both S-S and S-R incompatibilities. However as shown in the third block of Table 2, neither de Bruin and Della Sala (2017) nor Pettigrew and Martin (2014) found a significant correlation. In contrast, for the experiment described in Wöstmann et al. (2013) there was a significant correlation (U. Ettinger, personal communication, $r(534) = +0.24$, $p < .001$). The effect size is in the small to moderate range and is highly significant, in part, because of the large degrees of freedom.

The overall pattern of intertask correlations is suggestive, at best, that nonverbal interference tasks involving S-S incompatibilities recruit a common inhibitory control mechanism. If they did (and if bilingual language control also involves the same ability to resolve S-S conflict), then the three S-S tasks should produce bilingual advantages. However, given that the intertask correlations are mostly small and nonsignificant, the outcomes reviewed in Table 1 are also consistent with the possibility that inhibitory control in S-S tasks is task specific and that no bilingual advantages should be observed.

The psychometric structure of inhibition revealed by latent-variable analyses

Exploring the construct of inhibition through individual differences often goes beyond the zero-order intertask correlations and uses latent-variable analyses such as confirmatory factor analysis (CFA) and/or structural equation modeling (SEM). This approach has not been used to examine S-S versus S-R conflict resolution as such, but in their seminal work Friedman and Miyake (2004) considered very similar constructs dubbed *inhibition of prepotent responses*[1] (i.e., the ability to suppress dominant responses) and *resistance to distractor interference* (i.e., the ability to ignore distracting information). *Inhibition of prepotent responses* is more closely aligned to S-R incompatibility and *resistance to distractor interference* to S-S incompatibility. The CFA indicated that the two types of inhibition could be considered separate but correlated factors. However, additional modeling using SEM led Friedman and Miyake to prefer a more parsimonious model that collapsed the initial two factors into a single latent variable.

The question of whether *inhibition of prepotent responses* (akin to S-R conflict) and *resistance to distractor interference* (akin to S-S conflict) should be considered separate factors was further and more extensively examined by Rey-Mermet, Gade, and Oberauer (2018). Measures from

[1] For clarity throughout this section italics will be used to designate names of latent variable (factors).

six tasks were assumed to load on *inhibition of prepotent responses* and four others on *resistance to distraction*. The results were analyzed and interpreted in two distinctively different ways. The competitive fitting of SEM models (the standard approach) yielded a best-fitting model with two positively correlated inhibition factors assumed to reflect *inhibition of prepotent responses* and *resistance to distractor interference*. This analysis reinforces a common belief that inhibition is a useful psychometric construct and that it can be differentiated into correlated, but separable latent variables.

The alternative interpretation, and the one that Rey-Mermet et al. urge us to seriously consider leads to the conclusion that there is no general ability for inhibitory control. Here are the building blocks of their argument. First, the intertask correlations for measures in their study that were intended to load on the same factor were typically low. These low intertask correlations imply that most of the variance in the inhibition measures is not accounted for by the latent factors even when the overall fit of the models by conventional standards is good. Second, each factor was dominated by a single measure. They show this trend not only in their own data, but also in a fair amount of previous work (see Rey-Mermet et al.'s Table 8). Rey-Mermet et al.'s third argument is based on Bayesian hypothesis testing. These tests used a measure of model fit called the Bayesian information criterion (BIC) approximation (Wagenmakers, 2007). For example, the BIC for a model with only a single inhibition factor was cast in the role of the null hypothesis while the BIC for a two-factor model was cast in the role of the alternative hypothesis in order to compute a Bayes factor in favor of the null hypothesis (BF_{01}) and in favor of the alternative (BF_{10}). The advantage of using Bayesian hypothesis testing is that, in theory, one can not only reject the null hypothesis, but also accept it given that the BF is strong. However, the results were ambiguous. The BF did not distinguish between the single-factor and two-factor models. Neither did it distinguish between the two-factor model with correlated factors and a two-factor model with completely unrelated factors. Putting this together the BFs do not permit us to conclude whether there is one inhibition factor or two; or, if two, whether they are correlated or orthogonal. Furthermore, given that the factors tend to be dominated by a single measure and that the simple intertask correlations are very small, it is plausible that inhibition is task specific and not a general ability. Rey-Mermet et al. conclude their paper by suggesting that *"we should perhaps stop thinking about inhibition as a general cognitive construct"* p. 516.

If inhibition is task specific this would explain why inhibition sometimes fails to surface as a coherent latent variable that is separable from other factors. For example, in a CFA intended to confirm *updating*, *shifting*, and *inhibition* as factors, Krumm et al. (2009) found that the three presumed measures of *inhibition* (antisaccade, stop-signal, and a Stroop task) did not form a coherent latent variable. Similarly, van der Sluis, de Jong, and van der Leij (2007) explored the same three-factor model of EF by testing 172 children on 11 tasks that included the Stroop color-word naming task and three other Stroop variants. A common factor for *inhibition* could not be distinguished from the naming control tasks. In yet another attempt to confirm a three-factor solution of EF Hull, Martin, Beier, Lane, and Hamilton (2008) concluded that the *inhibition* factor was not adequately determined. Less stark results, but still amenable to there being no separable inhibition factors, comes from Klauer, Schmitz, Tiege-Mocigemba, and Voss (2010) who reported that the two measures (antisaccade and stop-signal) intended to load on an *inhibition* factor were better incorporated into the *working memory* factor. The absence of a separable inhibition factor is also consistent with Miyake and Friedman's (2012) more recent unity and diversity model of EF. In multiple data sets the three measures selected to load on an *inhibition* factor were not separable from a basic EF ability. They explicitly reject the interpretation that the *common EF* factor is inhibition or that inhibition is the most central of all EFs. Rather, they suggest that the common EF factor reflects individual differences in the ability to maintain and manage goals and to use those goal to bias ongoing processing. In summary, the results of most latent-variable investigations could either neatly fit or stretch to fit a loss of faith in *inhibition* as a general construct.

For some readers abandoning inhibition as a general construct may seem a bridge too far, but Rey-Mermet et al. were not the first to cross it. In a highly cited essay titled "In Opposition to Inhibition" MacLeod, Dodd, Sheard, Wilson, and Bibi (2003) provide alternatives to inhibition for several classic demonstrations of "inhibition" (e.g., Stroop color-word interference, inhibition of return, negative priming, etc.). These alternative explanations describe how standard functional or neural net models might predict interference effects without having to evoke inhibitory control.

Hampshire and Sharp (2015) also concluded that top-down inhibitory control is not necessary. Using the stop-signal task as their primary example, they argued that top-down inhibition from frontal cortex to other areas (or even to specific representations) is simply not necessary because phenomena assumed to reflect top-down inhibition can be explained by the upregulation of correct responses that compete with incorrect responses through local lateral inhibition. They suggest that an illusion of top-down inhibition may have been created by overgeneralizing a real neural mechanism (local lateral inhibition) to one that has not been established (a global inhibitory mechanism acting between cortical regions). Similarly, and based primarily on cognitive neuroscience data derived during the antisaccade task, Curtis and D'Esposito (2013) proposed that cognitive control should be modeled as a process by which the best response (including not responding at all) among competing responses is selected. This led them to conclude that a mechanism specialized for inhibiting actions, per se, does not seem necessary for the behavioral expression of inhibiting an unwanted action.

Purpose of present study

The present study has two purposes. One is to determine the psychometric structure among a set of four tasks that are frequently used in cognitive psychology to measure "inhibition." All four allow the computation of interferences scores (the difference between congruent and incongruent trials) and based on Kornblum's taxonomy they include a pure S-S task (vertical Stroop), a pure S-R task (Simon), and two that involve both types of conflict (spatial Stroop and flanker). Predictions based on S-S and S-R overlap between tasks will be evaluated with intertask correlations and an exploratory factor analysis. Another important purpose is to test the hypothesis that bilingual advantages are most likely to occur on tasks that involve S-S conflict. This hypothesis is based on the assumption (e.g., Green's ICM) that bilingual language control and the control recruited on nonverbal interference tasks with S-S conflict are one and the same. In contrast, the absence of bilingual advantages across all four tasks would, of course, be consistent with the encapsulation assumption of the BIA + /Multilink models.

Method

Sequence of events

All parts of the study were conducted in a single session of at least 60 min.[2] The first activity was obtaining written consent to participate using a form approved by the SFSU IRB. This was followed by: (1) the four nonverbal interference tasks, (2) the Raven's test, (3) the MINT task of productive vocabulary in English, and (4) the background questionnaire that was implemented as a Qualtrics survey.

[2] Appendix D presents two analyses showing that the interference effects are robust and do not decline across the four nonverbal interference tasks.

Participants

Selection criteria. The participants were 213 SFSU undergraduate students who either received extra credit or chose participation as one option for a class research assignment. For a variety of reasons ranging from an expression of "excruciating" boredom to computer failure nine participants did not complete all four parts and their incomplete data was not included in any analysis. Of the remaining 204 participants, only three were eliminated for performance reasons on the nonverbal interference tasks. The data from one participant were excluded because the overall proportion correct (0.836) was more than 6.7 standard deviations below the mean of 0.971. The data from two other participants were excluded because their overall mean RT was more than a 1000 ms and more than 7 standard deviations above the grand mean of 473 ms.

Language attributes of bilinguals and monolinguals. Extensive information was solicited from the participants about their exposure to and use of English and other languages. For each language an individual was exposed to, they were asked to rate, separately, their speaking, listening, reading, and writing proficiency using the 7-point scale developed by Paap and Greenberg (2013) and shown in Appendix A. For analyses based on partitioning the participants into bilingual and monolinguals, an individual was classified as bilingual if their average speaking and listening proficiency was 4.0 or greater on at least two languages. Value 4 was labeled *Advanced Intermediate: Can converse with little difficulty with a native speaker on most everyday topics, but with less fluency than a native speaker*. It is worth noting that this is not an arbitrary value as anyone who *"can converse with little difficulty with a native speaker…."* is, practically speaking, bilingual. This contrasts with commonly used proficiency scales which typically ask: On a scale of zero to ten, please select your level of speaking, understanding, and reading Language X. In the absence of any labels, the selection of any specific value to identify bilinguals is arbitrary.

An individual was classified as monolingual if their mean speaking and listening proficiency was 2.0 or less on all languages other than English. The value for 2 was labeled: *"Advanced Beginner – Can converse with a native speaker only on some topics and with quite a bit of difficulty."* Slightly more than 80% of the designated monolinguals claimed no exposure at all to any foreign language. By any reasonable standard, the participants designated as monolingual do not speak two languages.

Applying these selection criteria to the total set of 201 participants resulted in 104 bilinguals, 62 monolinguals, and 35 participants who were excluded from the bilingual versus monolingual comparisons. The bilinguals spoke a variety of non-English languages. The top three languages were Spanish (53%), Mandarin or Cantonese (14%), and Tagalog (8%). There were 17 multilinguals who were fluent in either 3 or 4 languages.

Most participants had a single native language and indicated greater proficiency in their native language than in any other language, but there were exceptions. The convention was adopted to use P1 to refer to the language with the highest rated proficiency regardless of whether it was English or not or whether it was a native language or not. P2 refers to the language with the next highest proficiency and so forth.[3] Table 3 shows, for each language group, the mean proficiencies for English, P1, and P2 as well as the mean number of correct responses, out of 68, on the Multilingual Naming Task (MINT, Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2011).

Inspection of Table 3 shows that the two groups are matched on P1 proficiency with means of about 6.5. Our rating scale differs from most others in that the penultimate value, 6 in this case is: *fluent: As good as a native speaker* whereas the highest value, 7 is labeled: *super fluency: Better than a typical native speaker*. The mean of 6.5 makes sense given that all participants are college students. On average the bilinguals self-rated proficiency in a P2 (M = 5.5) is far greater than for monolinguals (M = 0.3) and, on average, the bilinguals are nearly balanced with an average P2/P1 ratio of 0.84. Participants indicating that they used more than one language were also asked to indicate the percentage of time spent using each language. Given that they speak their most-used language 71% of the time, they are active bilinguals using other languages on average 29% of the time. Thus, they are not as balanced in terms of percentage of use as they are in terms of proficiency, but current use is clearly influenced by the fact that they are students at a university where English is the language of instruction. Only 2% of the bilinguals indicate that they rarely speak to others who speak both of their two languages while 40% indicate that they do so very often. A majority (55.7%) report that during the course of a typical day they switch languages *quite often* or *very often*. However, 18% do indicate that they do so *rarely*.

Language use was also assessed with respect to tendencies to use only a single language per specific context by having our bilingual participants list the languages they spoke in each of the seven contexts shown in Table 4. This may be an important metric because Hartanto and Yang (2016) reported that *dual language* bilinguals who tend to switch between languages in the same interactional context have smaller switch costs compared to those who tend to use a single language in each context. In our sample of bilinguals dual-language use occurs most often with friends and family and single-language use dominates at the university.

In summary, the participants designated as bilingual actively use two languages and self-reported as having near-native or native-like fluency in their weaker language. In contrast, the participants designated as monolinguals have little or no exposure to a second language and use only English.

To gain an understanding of the extent to which individual bilinguals engage in code switching we asked several additional questions. Although the majority (51%) will *sometimes* or *quite often* switch languages within a conversation only 9% do so *very often*. As switches within an utterance are often taken as an indication of code switching note that the same 9% report that they switch in mid utterance *very often*. When asked why they make single-word switches, these 10 bilinguals indicated that they do so both because they sometimes experience word-finding difficulty in the modal language and sometimes because the other language has a better word for the target concept. In summary, although the designated bilinguals actively use both languages and have good fluency in both languages, it would appear that few, if any, engage in the type of open control mode described in the adaptive control model (Green & Abutalebi, 2013). This may be important because, in theory, an open-control mode does not recruit a top-down mechanism that inhibits lexical representations.

Potential confounds across the language groups. Fluid intelligence was assessed using Set 1 of the Raven's Advanced Progressive Matrices (Raven, Court, & Raven, 1977). The task consisted of 12 items. Each item was composed of a pattern with a missing piece in the lower right. Participants were instructed to *"Look at the pattern, think what the missing part must be like to complete the pattern correctly, both across the rows and down the columns"*. Participants selected from a set of 8 alternatives. The task was computerized and controlled by DirectRT. Participants were given a maximum of 2 min to respond to each item. Most responses, regardless of correctness, in this self-paced computer-controlled version were made well within the deadline. The manual states that with self-pacing Set 1 can be used as a short 10-min test. A possible confound with fluid intelligence is important because Raven's scores have been shown to negatively correlate with both Simon (Paap & Greenberg, 2013) and flanker (Unsworth, Spillers, & Brewer, 2009) interference. As shown in the top row of Table 5 the mean for the monolinguals is slightly higher, but not significant at a standard alpha of 0.05.

[3] In our past work we have used L1 and L2 for this construct, but we use P (for proficiency) here to avoid confusion with the common convention that L1 refers to one's native language.

Table 3
Characteristics of and differences between bilinguals and monolinguals.

Group	N	P1 Pro	P2 Pro	P2/P1 Ratio	% Most Used	Eng Pro	MINT
Bilingual	104	6.54	5.47	0.84	70.6	6.10	59.7
Monolingual	62	6.53	0.28	0.04	99.0	6.52	64.4
B-M Diff.		+0.01	+5.18	+0.80	−28.4	−0.42	−4.7
SE		0.08	0.12	0.02	2.19	0.11	0.58
t		+0.81	+44.56**	+41.70**	−12.96**	−3.72**	−7.38**

B-M = Bilingual minus Monolingual.
MINT is an objective measure of productive vocabulary.
** Significant at alpha < 0.001; Both the P1 and P2 proficiencies were based on the mean of listening and speaking on a scale of 0–7.

Table 4
Mean languages spoken in each of seven interactional contexts.

Languages used in specified context	Mean (SD)
Where you are currently living	1.67 (0.65)
When speaking with family face-to-face or by telephone	1.64 (0.61)
When speaking with friends face-to-face or by telephone	1.65 (0.75)
When you are at work	1.42 (0.66)
When you are at school	1.26 (0.52)
When you are in your local community	1.55 (0.67)
When you are listening to entertainment media	1.94 (0.87)
Mean languages used in 7 interactional contexts	1.59 (0.40)

Table 5
Nonlinguistic characteristics of bilinguals and monolinguals.

	Bilinguals		Monolinguals					
Measure	n	mean	n	mean	Diff	SE	t	p
Raven's	104	8.20	62	8.89	−0.68	0.35	−1.95	.053
Age	104	23.7	62	25.7	−2.02	1.03	−1.96	.051
Mother's education	103	4.79	62	5.47	−0.68	0.30	−2.23	.002
Father's education	104	4.61	62	5.71	−1.10	0.30	−3.65	.000
Family income	103	2.89	61	3.05	−0.16	0.14	−1.17	.273
SES composite	104	4.10	62	4.75	−0.66	0.20	−3.31	.001

Diff = difference between group means, SE = standard error. See Appendix A for a full description of each measure.

Table 5 compares the two language groups on age, parents' education, and family income. Significant group differences were observed for the highest educational level obtained by the participant's mother, father, and the composite of all three measures of SES. In our earlier work (Paap & Greenberg, 2013; Paap & Sawi, 2014, 2016; Paap et al., 2017) using the same participant pool and recruiting methods, we also found differences between bilinguals and monolinguals, but the correlations between parents' educational level and measures of monitoring, inhibitory control, and switching were always non-significant and often near zero. That pattern holds in the present study as the correlations between all the measure of SES and the interference scores (based on efficiency scores) are all very close to zero: $r = +0.02$, $p = .83$, for mother's education; $r = +0.02$; $p = .79$ for father's education, $r = +0.02$, $p = .79$; for family income, $r = +0.01$, $p = .88$; and $r = +0.02$, $p = .78$ for a composite of all three measures of SES. Thus, although SES is confounded with our grouping variable it does not matter because SES has no effect on our measures of interference control.

Studies using children often report effects of SES on EF. For example, Calvo and Bialystok (2014) tested six-year olds and reported main effects for both bilingualism and SES on the flanker and Stroop effects. A possible explanation for why the relationship is consistently weak and nonsignificant in our studies is that the lower SES students in our college student population either had enriching early experience despite their parent's education and income or have otherwise managed to compensate for disadvantages in early childhood. Another possibility is that our composite measure of SES, based on the ordinal scales for mother's education, father's education, and family income have poor psychometric properties. This does not seem to be the case as the mean of 4.4 had a standard deviation of 1.3 and a Q-Q plot of an expected normal against the obtained SES scores shows a close fit to the normal (see Appendix C).

The four nonverbal interference tasks

Simon. In the Simon task (panel 1 of Fig. 1) the instruction was to press a response key located on the left side (the "z" key) if the arrow pointed up and to press a key on the right ("/") if the arrow pointed down. The target arrow was displaced either to the left or right of the fixation creating congruent trials when the correct response and ipsilateral hand matched and incongruent trials when they did not. As discussed in the introduction and as illustrated in the Venn diagrams at the bottom, the Simon task requires resolution of S-R conflict.

Spatial Stroop. In the spatial Stroop task (panel 2 of Fig. 1) the instruction was to press the left key if the arrow pointed left and the right key if it pointed to the right. When the arrow points in the opposite direction from its location the trial is incongruent. As illustrated in the Venn diagram overlap, the incongruency may involve both S-S and S-R conflict. Thus, references to this instantiation of the spatial Stroop task (either here or in prior work) as an "S-S task" should be taken as an economy of expression for the point that this task critically includes S-S conflict as well as potential S-R conflict.

Vertical Stroop. The vertical Stroop task developed by Liu, Banich, Jacobson, and Tanabe (2004) was added to the design because it is a pure S-S task in the Kornblum taxonomy. As shown in Fig. 1 (panel 3) the instruction for the vertical Stroop task is to press the left key if the vertically displaced arrow points up and the right key if it points down. The direction of the arrow (up, down) matches its location (above, below) on congruent trials and mismatches on incongruent trials. Because the up-down stimulus dimension is orthogonal to the left-right layout of the response keys there is no S-R conflict. Our vertical Stroop task is sometimes referred to as a spatial Stroop task, but for clarity we will continue to refer to this specific task as the vertical Stroop task.

Flanker. To reduce the differences between the flanker task and the other three tasks, we included only a single flanker on each side of the central target (Fig. 1, panel 4). When the flankers point in the same direction as the central arrow the trial is congruent and when they point in the opposite direction it is incongruent. The flanker task includes both S-S and S-R incompatibilities.

Trial definition for all tasks. The protocol was programmed in DirectRT. Each trial was initiated with a plus sign in the center of the display for 500 ms that served as a fixation point and warning signal. The plus sign was followed by the imperative stimulus (row of arrows for the flanker and a single arrow for the other tasks) that remained in view until a valid response was made. Any response longer than 2 s was followed by the prompt "please try to respond faster!"[4] Incorrect

[4] These prompts were rarely needed. Of the 128,640 experimental trials (640

Table 6
Means and standard errors for latency, accuracy, and efficiency for the four tasks based on the 166 designated bilinguals and monolinguals.

Measure	Simon	Spatial Stroop	Vertical Stroop	Flanker	*F*	p
Global RT	519 (7)	484 (6)	506 (6)	477 (7)	F(3, 489) = 1.59	.200
Interference RT	91 (3)	90 (2.9)	75 (3)	74 (6)	F(3, 492) = 0.92	.430
Global PC	0.934 (0.004)	0.951 (0.004)	0.955 (0.003)	0.956 (0.004)	F(3, 489) = 2.09	.101
Interference PC	0.117 (0.007)	0.085 (0.006)	0.066 (0.005)	0.073 (0.007)	F(3, 492) = 1.83	.141
Global ES	564 (7.3)	515 (5.4)	532 (5.8)	511 (11.0)	F(3, 489) = 2.23	.084
Interference ES	171 (7.8)	146 (7.0)	116 (4.2)	137 (16.0)	F(3, 489) = 1.70	.166

ES = Efficiency Score, calculated as RT/PC. The F value for the interference scores is that associated with a one-way ANOVA on the interference scores. All of the means were adjusted to take into account the Raven's scores as a covariate.

responses were followed by a "beep". The fixation point for the next trial appeared immediately after the participant responded. Thus, the response stimulus interval was 500 ms.

Display dimensions. Each arrow regardless of its location or direction was 7.5 cm (8.1°) in length and 5.4 cm (5.8°) in maximum width. The gap between the center fixation and the nearest edge of a horizontally displaced horizontal arrow (or a vertically displaced vertical arrow) was 4.5 cm (4.9°). The gap between the center fixation and a horizontally displaced vertical arrow (or a vertically displaced horizontal arrow) was 5.75 cm (6.2°). The gap between adjacent arrows in the flanker task was 2.54 cm (2.7°). The visual angles shown in parentheses assume a viewing distance of 53 cm.

Design. The number and proportion of trials of each type were the same across tasks and the same as that used by Blumenfeld and Marian. Each task started with a practice block of 20 trials where the imperative arrow was centered at fixation. Practice was followed by an experimental block of 160 trials. Half the trials required pressing the left key and half the right key. However, 75% (120 trials) of the trials were congruent compared to only 25% (40 trials) that were incongruent. Making incongruent trials less frequent usually increases the interference scores. The order of the four tasks was counterbalanced across participants using a Latin square whereby each task appears an equal number of times in each position and is preceded by and followed by each of the other three tasks an equal number of times.

Results and discussion

Trimming of response latencies

Consistent with Blumenfeld and Marian's statistical analysis, RTs less than 200 ms or more than 2.5 standard deviations above the participant's mean were removed. However, given Zhou and Krott (2015) hypothesis that bilingual advantages are primarily driven by the right tail of the RT distribution all analyses were repeated trimming only RTs that were greater than three seconds. In our tasks, a trial did not terminate until there was a valid response and, consequently, a distracting event might lead to an RT of 20 or even 30 s and these distant outliers would have a large impact on the corresponding condition mean if we conducted analyses on completely untrimmed latencies. Given that the grand mean across all four tasks was 473 ms and that the average standard deviation across individual participants was 70 ms, it seems safe to conclude that individual trials taking longer than 3 s were caused by an extraneous event or mind wandering and were not the product of normal individual and error variation in the sequence of cognitive processes under study.

The main analyses are based on the 640 experimental trials for each of the 201 participants yielding a grand total of 128,640 trials. Only 80 of those trials (< 0.07%) had RTs longer than 3 s. Fifty-five of the 201 participants had at least one response longer than 3 s and the maximum number for a participant was five. There were four anticipatory responses less than 200 ms. With respect to the 2.5 SD trimmed response latencies, 2.5% of the correct RTs were removed for being too long. All of the descriptive and inferential statistics reported in the text and tables that follow use the 2.5 SD trim. However, all the statistical tests were repeated using the much longer 3 s trim and this never changed the pattern of significance.

Three-way ANCOVAS based on 104 bilinguals and 62 monolinguals

A mixed ANOVA was conducted separately for RT, proportion correct (PC), and efficiency scores (ES) calculated as RT/PC. Language Group (bilingual versus monolingual) was a between-subject factor and the four Tasks and Congruency (congruent versus incongruent) were repeated measures. As reported in the method section, Raven's scores correlated with a composite measure of inhibitory control but by conventional standards did not differ across the language groups. Miller and Chapman (2001) cogently argue that ANCOVA is invalid for pre-existing disparate groups that significantly differ on the variable to be covaried. Field (2018), while discussing Miller and Chapman, reinforced the strategy of checking whether the groups differ on the potential covariate (using a standard alpha of 0.05) before fitting the model and, if they do not significantly differ, deeming it reasonable to use the covariate. On this basis, all the analyses reported in this section use Raven's as a covariate. In a subsequent section these analyses are complemented by a reanalysis on groups matched on their Raven's scores.

Response Times. Table 6 shows the mean global RTs and mean interference scores for each of the four tasks and for each of the three dependent variables. These are the adjusted means based on the Raven's covariate. The Simon task is numerically the slowest, but the main effect of task was not significant, $F(3, 489) = 1.59$, $p = .191$, partial $\eta^2 = 0.01$. The main effect of language group was also not significant, $F(1, 163) = 0.45$, $p = .506$, partial $\eta^2 = 0.003$. This shows that the groups were matched on global speed. As expected the main effect of congruency is large and significant, $F(1, 163) = 159.60$, $p < .001$, partial $\eta^2 = 0.495$; but the Congruency × Task interaction was not significant, $F(3, 489) = 0.69$, $p = .556$, partial $\eta^2 = 0.004$. The non-significant interaction signals that the magnitude of the interference effect was the same across the tasks.

Equivalent bilingual advantages across all tasks would result in a Group × Congruency interaction, but this interaction was not significant, $F(1, 163) = 0.69$, $p = .793$, partial $\eta^2 < 0.001$. Central to the hypothesis that bilingualism affects S-S tasks more than S-R tasks is the Group × Task × Congruency interaction. This too was not significant, $F(3, 489) = 1.27$, $p = .284$, partial $\eta^2 = 0.008$. This indicates that group differences in interference scores do not significantly differ across the four tasks. In summary, in the RT data there is no evidence that bilinguals perform differently from monolinguals on any of these nonverbal interference tasks.

Accuracy. Accuracy was measured as the proportion of correct (PC) responses. Although the Simon task was numerically the least accurate,

(*footnote continued*)
trials × 201 participants) only 0.1% yielded RTs longer than 2 s. The median and modal number per participant was zero. Following the prompt participants pressed the space bar when they were ready to resume.

the main effect of task was not significant, $F(3, 489) = 2.09$, p = .101, partial $\eta^2 = 0.01$. The main effect of language group was also non-significant, $F(1, 163) = 0.58$, p = .448, partial $\eta^2 = 0.004$. This suggests that the groups adopted similar and fairly high accuracy criteria, namely, about 95% correct. As expected the main effect of congruency was significant, $F(1, 163) = 33.30$, $p < .001$, partial $\eta^2 = 0.170$; but the Congruency × Task interaction was not, $F(3, 489) = 2.08$, p = .102, partial $\eta^2 = 0.013$. This shows that the magnitudes of the PC interference effects were the same across the tasks.

The Group × Task × Congruency interaction was not significant, $F(3,489) = 1.24$, $p = .294$, partial $\eta^2 = 0.01$. Furthermore, there was no main effect of group, nor did group enter into a significant two-way interaction with congruency or task, all F's < 0.8. There was no evidence for a bilingual advantage in proportion correct, nor was accuracy across tasks more differentiated for the bilinguals.

Efficiency Scores. Efficiency scores, computed as RT/PC, were also analyzed. This composite measure of speed and accuracy showed the same pattern of null outcomes as already reported for speed and accuracy. Neither the main effect of task, $F(3, 489) = 2.23$, p = .084, partial $\eta^2 = 0.013$, nor language group, $F(1, 163) = .098$, p = 0.755, partial $\eta^2 = .001$ was significant. The critical interactions involving language group were not significant: $F(1, 163) = 0.04$, p = .849, partial $\eta^2 < 0.001$ for the Group × Congruency interaction and $F(3, 489) = 1.14$, p = .332, partial $\eta^2 = 0.007$ for the Group × Task × Congruency interaction. The absence of a three-way interaction shows that the group differences in the magnitude of the ES interference scores do not differ across the four tasks. In contrast to the results of Blumenfeld and Marian's first experiment, there were no bilingual advantages in efficiency scores, nor were the task differences greater for the bilinguals.

Fig. 2 shows the nonsignificant Group × Congruency interaction for each of the four tasks. The top-right figure shows the outcome for the spatial Stroop task that provided evidence for a bilingual advantage in inhibitory control in Blumenfeld and Marian's Experiment 1. The lines for bilinguals and monolinguals are nearly on top of one another. The bottom row, shows a slight trend in slope favoring monolinguals in the pure S-S vertical Stroop task (left) and a slight tend in slope trend favoring bilinguals in the S-S flanker task. Perhaps the more important point is that the Group × Congruency interactions in each task do not even approach significance.

Analyses based on a composite measure of ES interference scores

Given the absence of task main effects or task interactions, the analyses in this section are restricted to a composite measure that is the mean of the standardized efficiency scores for each of the four tasks.

Groups matched on Raven's Scores. In this analysis, the two language groups were precisely matched on Raven's scores and this reduced the sample size from 104 bilinguals and 62 monolinguals to 54 in each group. In the full sample, the difference between the composite means for the two groups (favoring monolinguals) was −0.039 z compared to −0.045 z in the matched sample. These very small differences between the groups were not significant, $t(163) = -0.36$, $p = .709$ for either the full set or the matched set, $t(106) = -0.37$, $p = .719$. The means and t values were entered into Rouder's Bayes Factor (BF) calculator (pcl.missouri.edu; Rouder, Speckman, Sun, Morey, & Iverson, 2009). The BF values favoring the null result were 5.43 for the complete set and 4.63 for the matched set. Thus, the ratio of the probability of the null result given the data compared to the probability of the alternative given the data favors the null by more than a 4 to 1 ratio and this falls well within the range that Jeffreys (1961) considered to be substantial evidence for the null.

Tests of Spanish-English bilinguals. Our bilinguals spoke a variety of languages whereas Blumenfeld and Marian tested Spanish-English bilinguals. To investigate the possible role of language pairing the designated bilinguals were partitioned into Spanish-English bilinguals (n = 53) and other-English bilinguals (n = 35) and compared to each other and the designated monolinguals (n = 65). Separate ANOVAs (with Raven's scores as a covariate) were run on the composite interference scores derived from RT, PC, and ES. In each analysis there were no differences between these three language groups. For the case most relevant to Blumenfeld and Marian's findings (ES scores from the spatial Stroop task) the main effect of language group was not significant, $F(2, 149) = 0.34$, $p = .715$, partial $\eta^2 = 0.004$. The null result signals that the interference scores for monolinguals are the same as those for bilinguals regardless of whether the bilinguals speak Spanish or some other non-English language.

Aspects of Bilingual Experience. Contrasts treating bilingualism as a dichotomy can be supplemented with regression analyses that treat bilingualism as a set of continuous dimensions. The potential predictors based on bilingual experience were: P2 proficiency, P2/P1 ratio, number of native languages, the mean number of languages used per context, the frequency of switching per day, frequency of switching within a conversation, frequency of switching within an utterance, and the percentage use of the most used language. The analysis was based on the 104 participants who reported a proficiency of 4 or more on a P2 for speaking or listening and, consequently, were asked the detailed questions concerning how often, when, and why they switched languages. Raven's scores and the composite measure of SES were also included as predictors. The descriptive statistics associated with each predictor are shown in Table 7a and the correlations and standardized regression coefficients (beta) are shown in Table 7b. Raven's was the only predictor to enter the stepwise regression model and accounted for 11.1% ($R = 0.333$, $F(1, 99) = 12.34$, $p = .001$) of the variance in the composite interference effect. For the nine predictors that failed to enter the model, Table 7b shows the hypothetical "Beta In" coefficient that would be observed if that predictor was forced into a model with the Raven's scores. Note that the bilingualism predictor with the strongest Beta In, *Daily frequency of switching*, is in the direction opposite of the expectation that increased language switching should be associated with better interference control. Thus, for those participants who use two languages, specific aspects of their bilingual experience do not correlate with the composite interferences scores derived from the four nonverbal interference tasks.

A second stepwise regression analysis included all 201 participants and used Raven's, P2, and the P2/P1 dominance ratio as predictors of the composite ES interference scores. Monolinguals who reported no exposure to a foreign language were assigned a P2 of zero and likewise for their P2/P1 ratio. The final stepwise model consisted of only the Raven's score, Beta = −0.326, $t = -4.87$, p < .001. Neither P2, nor the P2/P1 ratio, exceeded the 0.05 entry criteria and their "Beta In" values were Beta In = 0.022, t = 0.32, p = .747 and Beta In = 0.020, t = 0.30, p = .767. When language proficiency is treated as a continuous variable and there are no issues regarding partitioning, there is no relationship between degree of bilingualism and the magnitude of the interference scores.

Correlations between tasks

The Kornblum taxonomy predicts that interference scores in S-S tasks, *ceteris paribus,* should correlate with one another and show smaller correlations with interference scores in S-R task. Table 8 shows the bivariate correlations between the interferences scores based on ES for the four tasks. Three of the six correlations are statistically significant and of moderate strength. This is consistent with the view that interference control was not task specific. However, inspection of the rightmost column shows that it is the flanker task, not the S-R Simon task, that fails to correlate with the other tasks. This pattern is more consistent with the hypothesis that what distinguishes one form of inhibition from another is whether the source of the conflict is another dimension of the target stimulus or the same dimension from a distractor.

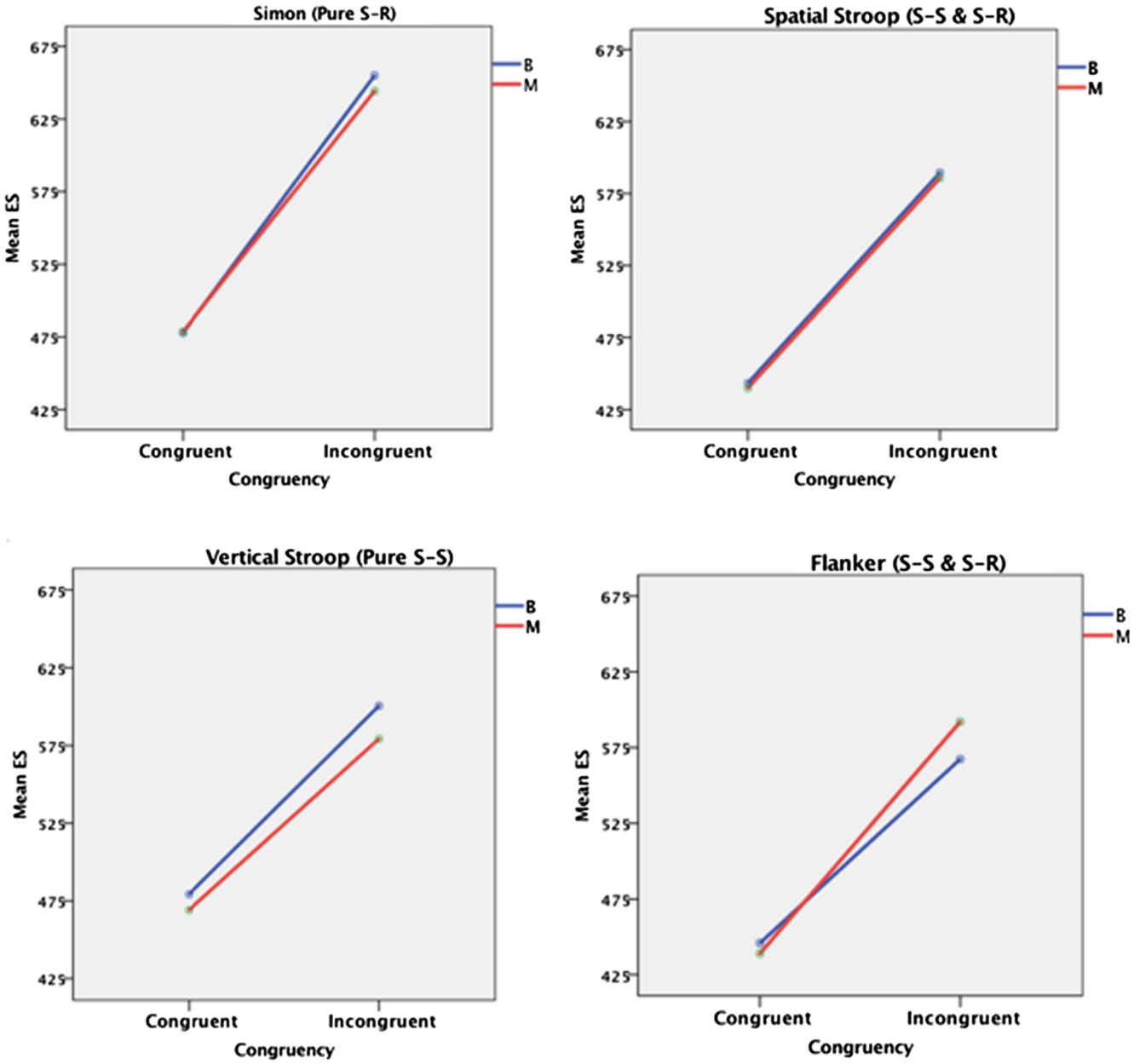


Fig. 2. The Group (bilingual versus monolingual) × Congruency (congruent versus incongruent) interaction for efficiency scores for each of the four nonverbal interference tasks. The Group × Task × Congruency interaction was not significant. The plots show adjusted means after taking Raven's scores as a covariate.

Table 7a
Descriptive statistics of the Raven's scores, the composite SES scores, and eight measures of bilingual experience.

Statistic	Min	Max	Mean (SD)	Skew (SE)	Kurtosis (SE)
*Raven's scores	3	12	8.2 (2.3)	−0.54 (0.23)	−0.43 (0.47)
*SES composite scores	1.3	7.3	4.1 (1.3)	+0.20 (0.24)	−0.65 (0.47)
Daily frequency of switching	1	5	3.6 (1.1)	−0.33 (0.24)	−0.84 (0.47)
Frequency of switching within a conversation	1	5	3.2 (0.98)	0.05 (0.24)	−0.50 (0.47)
Frequency of switching within an utterance	1	5	3.2 (0.93)	0.07 (0.24)	0.02 (0.48)
Mean languages used per context	1.0	2.7	1.6 (0.40)	0.57 (0.24)	0.17 (0.47)
P2 proficiency	4	7	5.5 (0.86)	−0.19 (0.24)	−0.83 (0.47)
P2/P1 ratio	0.57	1.00	0.84 (0.13)	−0.42 (0.24)	−0.86 (0.47)
Percentage use of most used language	40	100	70.6 (17)	0.03 (0.24)	−1.40 (.47)
Number of native languages	1	3	1.4 (0.60)	1.2 (0.24)	0.46 (0.47)

* The Raven's and SES scores are based on all 201 participants, not just the bilinguals.

Table 7b
Correlations and beta values treating Raven's scores, SES and eight aspects of bilingual experience as predictors of a composite interference score in a stepwise linear regression.

Variable	Zero-Order	Partial	Beta/Beta In	t	p
Raven's scores	−0.333	−0.333	−0.333	−3.51	.001
SES composite	+0.093	+0.107	+0.101	+1.07	.288
Daily frequency of switching	+0.199	+0.192	+0.181	+1.94	.056
Frequency of switching within a conversation	+0.253	+0.167	+0.167	+1.68	.096
Frequency of switching within an utterance	+0.078	+0.085	+0.080	+0.84	.401
Mean languages used per context	−0.130	−0.088	−0.084	−0.88	.384
P2 proficiency	+0.024	+0.008	+0.008	+0.08	.937
P2/P1 dominance ratio	+0.062	−0.006	−0.006	−0.06	.952
Number of native languages	+0.004	−0.113	−0.106	−1.12	.264
Percentage use of most used language	−0.017	+0.038	+0.036	+0.37	.710

Note. Because Raven's was the only predictor to enter the model the values for the other predictors in the Beta column are the hypothetical "Beta In" values computed in IBM SPSS Version 24.

Table 8
Intertask correlations between ES interference scores.

	Spatial Stroop	Vertical Stroop	Flanker
Simon	+0.399**	+0.354**	+0.033
Spatial Stroop		+0.425**	+0.046
Vertical Stroop			+0.120

** p < .01.

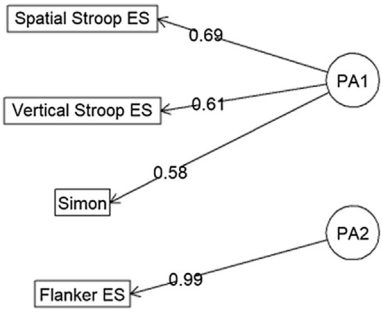


Fig. 3. An exploratory factor analysis of the interference scores (based on efficiency scores) from each of the four tasks.

The interference scores based on ES were also used as input to an exploratory factor analysis. A KMO test showed that the interference scores were adequate for exploratory factor analysis, KMO overall = 0.65. Eigenvalues were calculated without normalization for all four dimensions. Using Cattell's criterion two factors were extracted for further analysis. The two factor solution shown in Fig. 3 was run using the principal axis method with a varimax solution and accounted for 54% of the cumulative variance. The first factor includes both S-S and S-R tasks, but shares the attribute that the conflict on incongruent trials is based on two dimensions of the same target stimulus rather than between target and distractors. The second factor is a pseudofactor in the sense that a single data variable cannot establish a latent variable, but the EFA is consistent with the interpretations drawn from the intertask correlations shown in Table 8.

General discussion

If all four tasks had produced robust bilingual advantages this would imply that there is a general inhibitory control ability that also plays a prominent role in bilingual language control. Alternatively, if one of the tasks or a subset of the tasks had produced robust bilingual advantages this would imply that there are at least two conflict resolution mechanisms and that the tasks producing advantages employ the same mechanism that is prevalent during bilingual language control. Unfortunately, the finding that no task produced an advantage disallows such intriguing conclusions. Rather, the present results fall in line with the meta-analyses showing no evidence for the bilingual advantage hypothesis.

We offered two explanations for the absence of bilingual advantages in inhibitory control: (1) bilingual language control is encapsulated within the language-processing system and does not involve top-down inhibitory control and (2) inhibitory control is generally task specific. Our intertask correlations require a more nuanced interpretation. As reviewed in the introduction intertask correlations tend to be weak and nonsignificant, but previous tests were usually restricted to only two nonverbal interference tasks per study and the displays, event timing, and other procedures were not always closely matched. In contrast, this study produced moderate size correlations between the interference effects for the Simon, spatial Stroop, and vertical Stroop tasks. This favors the interpretation that these three tasks share, at least to some extent, the same inhibitory-control mechanism; but one that is different from that used in either the flanker task or during bilingual language control.

Perhaps with the benefit of hindsight, it makes sense that the flanker task was the odd man out showing weak correlations with the other three tasks. The nature of conflict resolution may depend on whether the conflict arises from two dimensions of the same stimulus or between adjacent but separate stimuli. The flanker task fits the latter because participants must select the relevant central arrow among the irrelevant flankers using visuospatial attention. Many theorists have suggested that conflict in the flanker task is resolved by spatially attending to the target stimulus (e.g., Magen and Cohen's Dimension Action model, 2007). If spatial attention is construed as a filter or the upregulation of task relevant information then it clearly contrasts with inhibition. This interpretation of the flanker task is timely with respect to the bilingual advantage controversy as Bialystok (2017) has recently reframed her hypothesis by jettisoning inhibition in favor of selective attention (but see Paap, Anders-Jefferson, Mason, Alvarado, & Zimiga, 2018, for a challenge to this new formulation).

A study by de Bruin and Della Sala (2018) provides further evidence that inhibitory control in the flanker task is functionally different from that in the spatial Stroop task. Overall RTs were unsurprisingly slower in older adults across three tasks, but effects of age on interference scores were only found in a spatial Stroop task and not on the two flanker tasks. The same pattern of results was reported by Kawai, Kubo-Kawai, Kubo, Terazawa, and Masataka (2012) who also had younger and older adults complete both the spatial Stroop and an arrows flanker task. This also fits the pattern of results reported by Rey-Mermet et al. In summary, age effects are observed relatively often in the spatial Stroop and Simon tasks, but not in flanker tasks. Thus, the effects of aging on interference control are consistent with the clustering results reported in the present study. The effects of bilingualism do not show the same pattern (as no task yielded a group difference) and this favors the interpretation that bilingual language control is encapsulated and

different from the control mechanisms used in any of the nonverbal interference tasks.

The effects of specific types of bilingualism

If bilingual language control is specific to language comprehension and production, then there should be no systematic relationships between specific dimensions of bilingualism and nonverbal measures of inhibitory control. Indeed, as shown in Table 7b, the present study found that the composite interference effect was unrelated to eight different aspects of bilingual experience.

Blumenfeld and Marian speculated that the inconsistencies between their two experiments might be due to differences in the amount of code switching in their two samples. In that context, it is relevant to note that Verreyt, Woumans, Vandelanotte, Szmalec, and Duyck (2016) concluded that the frequency with which bilinguals switch language is the key determinant in potentiating a bilingual advantage and reported that high-switch bilinguals had smaller Simon and flanker interference effects compared to low-switch individuals. However, using much larger samples sizes Paap, Johnson, and Sawi (2014) and Paap et al. (2017) found no significant relationships between the frequency of language switching on a typical day and the magnitude of flanker or Simon effects.

Be that as it may, the Blumenfeld and Marian hypothesis is quite different in focusing on code switching rather than the frequency of switching on a typical day. Although they do not explicitly refer to Green and Abutalebi's (2013) Adaptive Control hypothesis they appeal to the possibility that when bilinguals code switch within utterances they may be operating in an open-control mode that does not establish a single-language task schema and therefore does not routinely inhibit entries in a non-target language. From this perspective, once the frequency-of-language switching crosses over from many controlled and intentional switches throughout the day to continuous code switching the consequences switch from intensive practice in controlled switching to little or no practice. Blumenfeld and Marian speculated that controlled processing characterizes their Experiment 1 sample, but code-switching often occurred in their Experiment 2 sample. This would account for why clear bilingual advantages in S-S conflict resolution were solely nested in Experiment 1. The evidence for this is quite indirect as Blumenfeld and Marian did not ask their bilinguals about the frequency and reasons for switching languages. They do know that the percentage of time spent in non-English speaking environments was only 21% in Experiment 1 compared to 35% in Experiment 2. Several caveats loom. First, the percentage of time spent in a specific language environment may correlate with the amount of code switching, but it is an indirect measure at best. Our language background questionnaire included two items that may provide more direct operational measures of the degree of code switching: (1) the frequency of switching within a conversation and (2) the frequency of making single-word, cross-language substitutions within an utterance. Neither was a significant predictor of the composite RT interference scores. All things considered, it does not appear that the inconsistencies can be resolved by appealing to differences in how often bilinguals code switch and are operating in an open control mode.

Limitations

Our investigation of inhibitory control in nonverbal interference tasks was limited to one pure S-R task (Simon), one pure S-S (vertical Stroop), and two tasks (spatial Stroop and flanker) where both types of conflict can occur. Although significant intertask correlations were obtained between the three non-flanker tasks, it would be instructive to expand the set of pure S-R and pure S-S tasks. Even if this entailed different versions of the same task, it may be worthwhile as different versions often fail to correlate with one another and may not load on the same factor. Our tasks were deliberately homogeneous in terms of visual form, event timing, proportion of congruency, number of trials, and response keys. Although these are obvious advantages in comparing simple correlations between the interference scores obtained in individual tasks, as Rey-Mermet et al. point out homogeneity can be disadvantageous in latent variable analyses and produce spurious loadings by creating non-target sources of shared variance.

Conclusion

The first major purpose was to determine the psychometric structure among a set of four tasks that are frequently used to measure inhibition. The three tasks where conflict is between two dimensions of the same stimulus formed a coherent latent variable that excluded the flanker effect where the conflict is between different stimuli. This pattern was counter to predictions based on Kornblum's taxonomy, but consistent with the assumption that selective attention to objects or spatial locations entails control mechanisms different from those that resolve conflict between two dimensions of the same stimulus. The second purpose was to determine if bilingual advantages in interference scores would occur in three tasks involving S-S conflict, but not in a task involving only S-R conflict. There were no group differences in any of the four tasks. Our results and the new wave of meta-analyses are consistent with the assumptions of the BIA + /Multilink model that bilingual language control is encapsulated within the language processing system.

Appendix A

Language proficiency rating scale

0 no exposure to a language other than English

1. Beginner - Know some words and basic grammar.
2. Advanced Beginner – Can converse with a native speaker only on some topics and with quite a bit of difficulty.
3. Intermediate – Can converse with a native speaker on most everyday topics, but with some difficulty
4. Advanced Intermediate – Can converse with little difficulty with a native speaker on most everyday topics, but with less fluency than a native speaker.
5. Near Fluency – Almost as good as a typical native speaker on both everyday topics and specialized topics I know about.
6. Fluent – As good as a typical native speaker.
7. Super Fluency – Better than a typical native speaker.

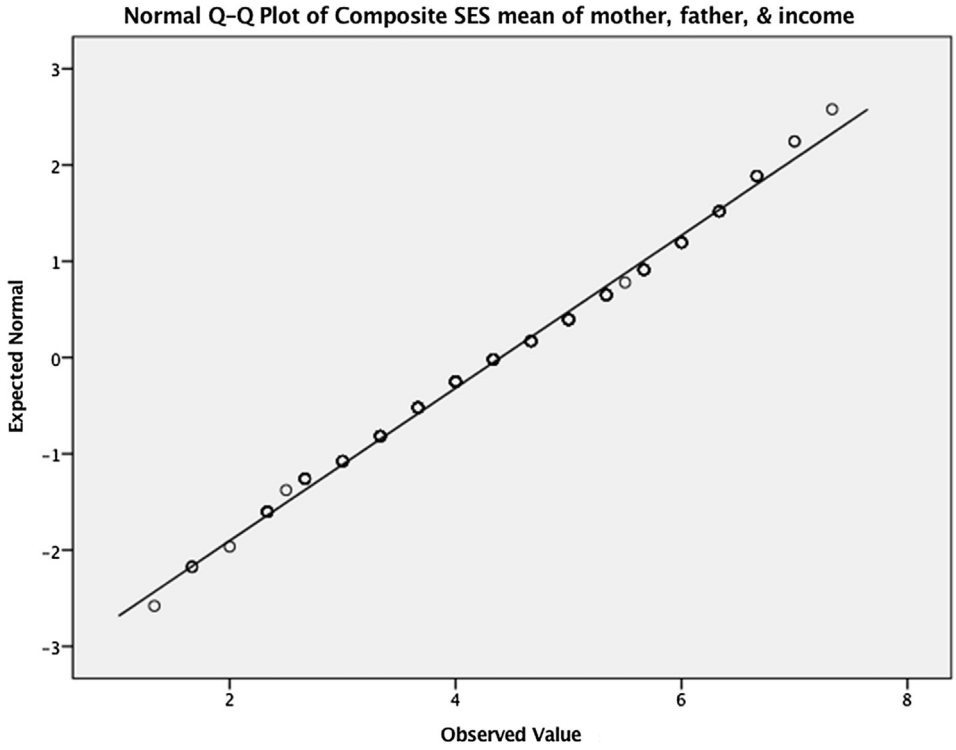
Appendix B

Background questions

What gender were you assigned at birth?
What is your current gender?
What is your age?
What country were you born in
How many years have you resided in the United States?
Which of the following best describes the highest educational level obtained by your mother?
(1) no formal education,
(2) less than 8th grade education,
(3) did not graduate from high school,
(4) graduated from high school,
(5) attended college, but did not earn a degree,
(6) earned an associate of arts degree,
(7)earned a bachelor's degree,
(8) earned a graduate or professional degree that required additional education beyond a bachelor's degree
Which of the following best describes the highest educational level obtained by your father?
Relative to other families in the country where I grew up, my family's income would be considered:
(1) very low
(2) low
(3) medium
(4) medium high
(5) high
How often do you play video games that require you to attend to many things at the same time and make fast appropriate responses?
(1) never, (2) rarely, (3) sometimes, (4) quite often (5) very often
How many years of musical training have you had?
How often do you play a musical instrument?
(1) never, (2) rarely, (3) sometimes, (4) quite often (5) very often
How often in a typical day do you engage in two or more tasks at the same time (multitask)?
(1) never, (2) rarely, (3) sometimes, (4) quite often (5) very often
How often in a typical week do you exercise, work out, or participate in a sport?
(1) never, (2) rarely, (3) sometimes, (4) quite often (5) very often
How often in a typical week do you meditate or practice mindfulness?
(1) never, (2) rarely, (3) sometimes, (4) quite often (5) very often
Team sports often involve dividing your attention between a ball, a goal, your opponents, and your teammates. Do you excel at these sports?
(1) not at all
(2) I am below average
(3) I am average
(4) I am better than average
(5) I am significantly better than average
How do you feel when you need to focus on an important task, but there are lots of things going on that could be distracting?
(1) very frustrating and my performance is usually not as good as it could be
(2) somewhat frustrating and my performance is sometimes not as good as it could be
(3) neither frustrating nor stimulating
(4) somewhat stimulating and it sometimes improves my performance
(5) very stimulating and it usually helps me to perform better

Appendix C

A Q-Q plot of the composite SES measure showing close approximation to normality



Appendix D

Supplementary analyses testing for a dilution effect across the four tasks

Although our participants completed the four nonverbal interference tasks early in the session there were two more S-S tasks in our design compared to that used by Blumenfeld and Marian. This could deplete the resources needed to resolve the competition. Alternatively, practice effects might attenuate the interference effects to the point where language group differences would be difficult to detect. This concern was addressed in two ways. First, by showing that the magnitude of the interference effects were very similar when both studies were analyzed with the same ANOVA: 2 groups (bilingual, monolingual), 2 tasks (Simon, spatial Stroop), and 2 trial types (congruent, incongruent). The main effect of congruency in our study was $F(1, 164) = 618$, $p < .001$, partial $\eta^2 = 0.79$. The corresponding analysis from Blumenfeld and Marian is, $F(1, 58) = 200$, $p < .001$, partial $\eta^2 = 0.80$. Thus, our congruency effects (for the same two tasks) is exactly the same size as those reported by Blumenfeld and Marian. Either a fatigue or practice explanation for the absence of language-group differences in our study appears unlikely given that both studies show very similar and robust interference effects.

To address the possibility that the S-S interference effect is declining over the administration of the four tasks several analyses of the spatial Stroop task were conducted. This is the task that showed a bilingual advantage in Experiment 1 of Blumenfeld and Marian. For each dependent variable (RT, PC, & ES) we analyzed the means for the subgroups who received the spatial Stroop task first (n = 49), second (n = 50), third (n = 52), and fourth (n = 50). There are no differences across these means. For example, for ES interference scores the means across positions are: 158 ms, 137 ms, 122 ms, and 148 ms; $F(3, 197) = 1.575$, $p = .197$. The magnitude of the interference effect neither increases nor decreases across the four task positions.

Appendix E. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jml.2018.12.001.

References

Antón, E., Carreiras, M., & Duñabeitia, J. A. (2018). On the effects of bilingualism on executive functions and working memory in young adults (submitted for publication).

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60*, 68–79. https://doi.org/10.1037/cjep2006008.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin, 143*, 233–262. https://doi.org/10.1037/bul0000099.

Blackburn, A. M. (2013). A study of the relationship between code switching and the bilingual advantage: Evidence that language use modulates neural indices of language processing and cognitive control (Doctoral dissertation, University of Texas at San Antonio). Retrieved from https://pqdtopen.proquest.com/pubnum/3607533.html?FMT=AI.

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in stimulus-stimulus inhibition. *Bilingualism: Language and Cognition, 17*, 610–629.

https://doi.org/10.1017/S1366728913000564.
Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition, 130*, 278–288. https://doi.org/10.1016/j.cognition.2013.11.015.
Curtis, C. E., & D'Esposito, M. (2013). The inhibition of unwanted actions. In J. Bargh, P. Gollwitzer, & E. Morsella (Vol. Eds.), *The psychology of action: Vol. 2*, (pp. 1–16). Oxford: Oxford University Press.
de Bruin, A., & Della Sala, S. (2018). Effects of age on inhibitory control are affected by task-specific features. *Quarterly Journal of Experimental Psychology, 71*, 1219–1233. https://doi.org/10.1080/17470218.2017.1311352.
Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*, 175–197. https://doi.org/10.1017/S1366728902003012.
Dijkstra, T., Wahl, A., Buytenhuijs, F., van Halem, N., Al-Jibouri, Z., de Korte, M., & Rekké, S. (2018). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition,* 1–23. https://doi.org/10.1017/S1366728918000287.
Donnelly, S., Brooks, P., & Homer, B. (2018). Is there a bilingual advantage on interference-control tasks? A meta-analysis. Psychological Bulletin & Review [in press].
Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in Cognitive Science, 12*, 374–380. https://doi.org/10.1016/j.tics.2008.07.001.
Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. *NeuroImage, 18*, 42–57. https://doi.org/10.1006/nimg.2002.1319.
Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*, 340–347. https://doi.org/10.1162/089892902317361886.
Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). London: Sage Publications.
Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*, 101–135. https://doi.org/10.1037/0096-3445.133.1.101.
Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2011). Self-ratings of spoken language dominance: A multilingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition, 15*, 594–615. https://doi.org/10.1017/S1366728911000332.
Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67–81. https://doi.org/10.1017/S1366728998000133.
Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*, 515–530. https://doi.org/10.1080/20445911.2013.796377.
Guido Mendes, C. (2016). The impact of bilingualism on conflict control (Doctoral dissertation). University of Otago, Dunedin, New Zealand. Retrieved from WorldCat Theses Database.
Hampshire, A., & Sharp, D. J. (2015). Contrasting network and modular perspectives on inhibitory control. *Trends in Cognitive Science, 19*, 445–452. https://doi.org/10.1016/j.tics.2015.06.006.
Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion model analysis of the effects of interactional context on switch costs. *Cognition, 150*, 10–19. https://doi.org/10.1016/j.cognition.2016.01.016.
Hull, R., Martin, R. C., Beier, M. E., Lane, D., & Hamilton, A. C. (2008). Executive function in older adults: A structural equation modeling approach. *Neuropsychology, 22*, 508–522. https://doi.org/10.1037/0894-4105.22.4.508.
Humphrey, A. D., & Valian, V. V. (2012, November). Multilingualism and cognitive control: Simon and flanker task performance in monolingual and multilingual young adults. Paper presented at the meeting of the Psychonomic Society, Minneapolis, MN.
Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press, Clarendon Press.
Kawai, N., Kubo-Kawai, N., Kubo, K., Terazawa, T., & Masataka, N. (2012). Distinct aging effects for two types of inhibition in older adults: A near-infrared spectroscopy study on the Simon task and the flanker task. *NeuroReport, 23*, 819–824. https://doi.org/10.1097/WNR.0b013e3283578032.
Keye, D., Wilhelm, O., Oberauer, K., & van Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: testing means and covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychological Research, 73*, 762–776. https://doi.org/10.1007/s00426- 008-0188-9.
Klauer, K. C., Schmitz, F., Tiege-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *The Quarterly Journal of Experimental Psychology, 63*, 595–619. https://doi.org/10.1080/17470210903076826.
Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop- and Simon-like stimuli. *Psychological Research, 56*, 130–135. https://doi.org/10.1007/BF00419699.
Kousaie, S., & Phillips, N. A. (2012). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research, 1446*, 71–90. https://doi.org/10.1016/j.brainres.2012.01.052.
Kousaie, S., & Phillips, N. A. (2017). A behavioral and electrophysiological investigation of the effect of bilingualism on aging and cognitive control. *Neuropsychologia, 94*, 23–35. https://doi.org/10.1016/j.neuropsychologia.2016.11.013.
Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence, 37*, 347–364. https://doi.org/10.1016/j.intell.2009.02.003.
Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? *Psychological Bulletin, 144*, 394–425. https://doi.org/10.1037/bul0000142.
Li, Q., Nan, W., Wang, K., & Liu, X. (2014). Independent processing of stimulus-stimulus and stimulus-response conflicts. *PLOS ONE, 9*, e89249. https://doi.org/10.1371/journal.pone.0134466.
Liu, X., Banich, M. T., Jacobson, B. L., & Tanabe, J. L. (2004). Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI. *NeuroImage, 22*, 1097–1106. https://doi.org/10.1016/j.neuroimage.2004.02.033.
MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Vol. Ed.), *The psychology of learning and motivation: Advances in research and theory: Vol. 43*, (pp. 163–214). New York, NY: Elsevier Science.
Magen, H., & Cohen, A. (2007). Modularity beyond perception: Evidence from single task interference paradigms. *Cognitive Psychology, 55*, 1–36. https://doi.org/10.1037/a0017174.
Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*, 40–48. https://doi.org/10.1037/0021-843X.110.1.40.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*, 8–14. https://doi.org/10.1177/0963721411429458.
Mohades, S. G., Struys, E., Van Schuerbeek, P., Baeken, C., Van De Craen, P., & Luypaert, R. (2014). Age of second language acquisition affects nonverbal conflict processing in children: An fMRI study. *Brain and Behavior, 4*, 626–642. https://doi.org/10.1002/brb3.246.
Paap, K. R. (2018). Bilingualism in cognitive science. In A. De Houwer, & L. Ortega (Eds.). *Handbook of bilingualism* (pp. 435–465). Cambridge: Cambridge University Press.
Paap, K. R. (2019). The bilingual advantage debate: Quantity and quality of the evidence. In J. W. Schwieter (Ed.). *The handbook of the neuroscience of multilingualism* (pp. 701–735). West Sussex, UK: Wiley-Blackwell.
Paap, K. R., Anders-Jefferson, R., Mason, L., Alvarado, K., & Zimiga, B. (2018). Bilingual advantages inhibition or attentional control: More challenges. *Frontiers in Psychology, 9*, 1409. https://doi.org/10.3389/fpsyg.2018.01409.
Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*, 232–258.
Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology, 26*, 615–639. https://doi.org/10.1016/j.cogpsych.2012.12.002.
Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex, 69*, 265–278. https://doi.org/10.1016/j.cortex.2015.04.014.
Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology, 29*, 89–112. https://doi.org/10.1080/20445911.2016.1248436.
Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology, 5*, 962. https://doi.org/10.3389/fpsyg.2014.00962.
Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods, 274*, 81–93. https://doi.org/10.1016/j.neumeth.2016.10.002.
Pettigrew, C., & Martin, R. C. (2014). Cognitive declines in healthy aging: Evidence from multiple aspects of interference resolution. *Psychology and Aging, 29*, 187–204. https://doi.org/10.1037/a0036085.
Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition, 13*, 253–262. https://doi.org/10.1017/S1366728909990526 [augmented with personal communication from A. Prior].
Raven, J. C., Court, J. H., & Raven, J. (1977). *Manual for Raven's advanced progressive matrices: Sets I and II.* London, England: H.K. Lewis & Co., Ltd.
Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 501–526. https://doi.org/10.1037//xlm0000450.
Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review, 16*, 225–237. https://doi.org/10.3758/PBR.16.2.225.
Simon, J. R., & Small, A. M., Jr. (1969). Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology, 53*, 433–435. https://doi.org/10.1037/h0028034.
Stins, J. F., Polderman, J. C., Boomsma, D. I., & de Geus, E. J. C. (2007). Response interference and working memory in 12-year old children. *Child Neuropsychology, 11*, 191–201. https://doi.org/10.1080/092970490911351.
Struys, E., Duyck, W., & Woumans, E. (2018). The role of cognitive development and strategic task tendencies in the bilingual advantage controversy. *Frontiers in Psychology, 9*, 1790. https://doi.org/10.3389/fpsyg.2018.01790.
Tabori, A. A. T., Mech, E. N., & Atagi, N. (2018). Exploiting language variation to better understand the cognitive consequences of bilingualism. *Frontiers in Psychology, 9*, 1686. https://doi.org/10.3389/fpsyg.2018.01686.
Unsworth, N., Spillers, G. J., & Brewer, G. A. (2009). Examining the relations among working memory capacity, attention control, and fluid intelligence from a dual-component framework. *Psychological Science, 51*, 388–402. https://doi.org/10.1037/t38798-000.
van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*,

427–449. https://doi.org/10.1016/j.intell.2006.09.001.
Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognitions, 19*, 181–190. https://doi.org/10.1017/S1366728914000352.
Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*, 779–804. https://doi.org/10.3758/BF03194105.
Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*, 433. https://doi.org/10.3389/fpsyg.2013.00433.
Wöstmann, N. M., Aichert, D. S., Costa, A., Rubia, K., Möller, H.-J., & Ettinger, U. (2013). Reliability and plasticity of response inhibition and interference control. *Brain and Cognition, 81*, 82–94. https://doi.org/10.1016/j.bandc.2012.09.010.
Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1579–1586. https://doi.org/10.1037/xlm0000107.
Zhou, B., & Krott, A. (2015). Data trimming procedure can eliminate the bilingual cognitive advantage. *Psychonomic Bulletin and Review, 23*, 1221–1230. https://doi.org/10.3758/s13423-015- 0981-6.