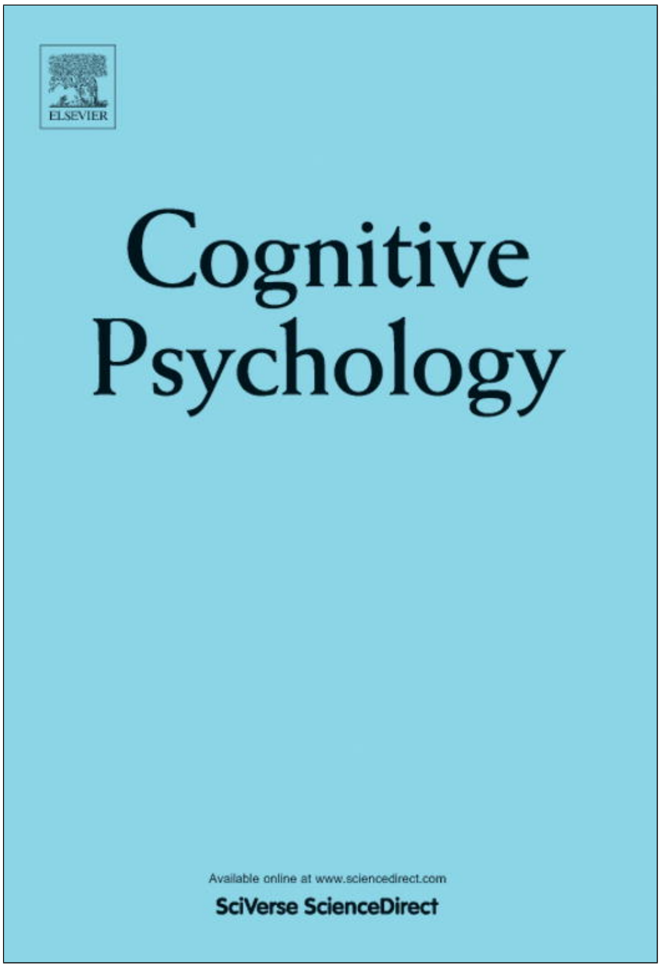
# There is no coherent evidence for a bilingual advantage in executive processing

**Article** *in* Cognitive Psychology · January 2013

DOI: 10.1016/j.cogpsych.2012.12.002 · Source: PubMed



Kenneth R. Paap
San Francisco State University


ELSEVIER
Cognitive Psychology
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Cognitive Psychology 66 (2013) 232–258

Contents lists available at SciVerse ScienceDirect

Cognitive Psychology

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

# There is no coherent evidence for a bilingual advantage in executive processing

Kenneth R. Paap *, Zachary I. Greenberg

*San Francisco State University, 1600 Holloway Avenue, EP301, Department of Psychology, San Francisco, CA 94132, USA*

A R T I C L E  I N F O

*Article history:*
Accepted 28 December 2012

*Keywords:*
Bilingual advantage
Executive processing
Cognitive control

A B S T R A C T

Three studies compared bilinguals to monolinguals on 15 indicators of executive processing (EP). Most of the indicators compare a neutral or congruent baseline to a condition that should require EP. For each of the measures there was no main effect of group and a highly significant main effect of condition. The critical marker for a bilingual advantage, the Group × Condition interaction, was significant for only one indicator, but in a pattern indicative of a bilingual disadvantage. Tasks include antisaccade (Study 1), Simon (Studies 1–3), flanker (Study 3), and color-shape switching (Studies 1–3). The two groups performed identically on the Raven's Advanced Matrices test (Study 3). Analyses on the combined data selecting subsets that are precisely matched on parent's educational level or that include only highly fluent bilinguals reveal exactly the same pattern of results. A problem reconfirmed by the present study is that effects assumed to be indicators of a specific executive process in one task (e.g., inhibitory control in the flanker task) frequently do not predict individual differences in that same indicator on a related task (e.g., inhibitory control in the Simon task). The absence of consistent cross-task correlations undermines the interpretation that these are valid indicators of domain-general abilities. In a final discussion the underlying rationale for hypothesizing bilingual advantages in executive processing based on the special linguistic demands placed on bilinguals is interrogated.



## 1. Introduction

Fluent bilinguals have extensive experience in language switching that involves monitoring the situation to select the appropriate language, activating the selected language, and inhibiting the other

* Corresponding author. Fax: +1 415 338 2398.
*E-mail address:* kenp@sfsu.edu (K.R. Paap).

0010-0285/$ - see front matter 
http://dx.doi.org/10.1016/j.cogpsych.2012.12.002

language. This extensive practice may lead to an enhanced ability in cognitive control that is general and not language specific. Indeed several investigators have reported a bilingual advantage in tasks that seem to require executive processing (EP), that is, the ability to monitor goal-setting cues, to switch attention to goal-relevant sources of information, and to inhibit those that are irrelevant or competing (Bialystok, 2006; Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik, & Luk, 2008; Costa, Hernández, & Sebastián-Gallés, 2008). However, as Hilchey and Klein (2011) enumerate in their recent comprehensive review, there have also been many failures to observe a bilingual advantage.

Although the construct of EP continues to evolve, it is often viewed as a set of interrelated component processes all involving the prefrontal cortex (PFC) with each component recruiting other constellations of cortical function. This componential framework allows for the possibility that the related components have some degree of anatomical and functional independence. Ideally an investigation of bilingual advantages in EP should be grounded in a specific conceptual framework that elucidates the nature of executive processes and guides operational definitions for manipulating and measuring them.

As discussed in Section 1.1.1 there is very little evidence that the measures and tasks typically used to test for differences between bilinguals and monolinguals in inhibitory control are tapping into the same general ability. This is somewhat surprising because the seminal work of Miyake and Friedman (Miyake et al., 2000; Friedman et al., 2008; Miyake & Friedman, 2012) shows evidence for three components of EP: updating, switching[1], and inhibition. In their large-scale studies these investigators conducted confirmatory factor analyses (CFA) using measures from three different tasks for each of the three hypothesized components. For each latent variable (viz., updating, switching, inhibition) the three observed variables (e.g., color, number, category) linked to the same latent variable (e.g., switching), are correlated with one another, and result in standardized factor loading ranging from .40 to .71. However, it is instructive to note that three of the four lowest factor loadings are for the inhibition tasks. At the higher level the three latent variables (updating, switching, inhibition) correlate with one another and this is consistent with the assumption that each contributes to a common EP. When the same data are reanalyzed with a second order CFA where the three latent variables (updating, switching and inhibition) are nested under a common EP latent variable, the nine observed measures all load on common EP with two of the nested components (updating and shifting) still making unique contributions. In summary, these studies support the theory of a general EP ability with separable updating and switching components and an inhibition component that is not separable and that is weakly linked to the general EP ability.

Most tests for a bilingual advantage in EP in adults have focused on only two of three components studied by Miyake and Friedman (viz., switching and inhibition) and frequently employ tasks not tested by Miyake and Friedman. As discussed in the next section this different mix of specific tasks has resulted in less convergent validity.

### *1.1. The Role of bilingualism in three executive processes*

#### *1.1.1. Inhibitory control*

In Bialystok et al.'s seminal article primary focus was placed on the proposition that bilinguals are better at selecting goal-relevant information and suppressing competing and distracting information. Bilinguals exercise this type of control at two levels: (1) at a high level of goal setting when one language is selected and the other is inhibited and (2) at a lower level where the lexical forms of the goal relevant language are activated and the competing translation equivalents are inhibited (e.g., Green, 1998). If this extensive practice hones a general ability, not specific to language, then bilinguals should be less vulnerable to interference in nonlinguistic tasks.

The standard marker of inhibitory control is the difference in mean response time between trials that require conflict resolution compared to those that do not. In the Stroop, Simon, and the Eriksen flanker tasks conflict occurs on a subset of trials because a potent but task-irrelevant stimulus is often

[1] Miyake and Friedman refer to this component as "shifting", but the term "switching" is used consistently in the literature on bilingualism.

paired in an incongruent manner with the task-relevant stimulus. Performance can be enhanced by boosting the influence of the goal-relevant information relative to that of the competing information. The effectiveness of this control can be inferred from differences in response time between congruent trials and incongruent trials with smaller interference effects implying superior ability. In a recent and comprehensive review of bilingual advantages in EP Hilchey and Klein (2011) conclude that evidence for a bilingual advantage in inhibitory control is rare in both children and young adults.

In the present studies inhibitory control should play a role in the Simon task (Studies 1–3), antisaccade task (Study 1), and Eriksen flanker task (Study 3). Thus, the critical test for a bilingual advantage in inhibitory control is the presence of a significant Group (bilingual versus monolingual) × Trial Type (congruent versus incongruent) interaction with the pattern of interaction showing a larger interference effect for the monolingual group. Compelling evidence for a bilingual advantage in inhibitory control would demonstrate significant advantages in two or more tasks requiring inhibitory control and further show that the interference effects correlate with each other as one would expect if each task includes a common component associated with a general ability to exercise inhibitory control.

Although researchers investigating the bilingual advantage have employed several different tasks that should require inhibitory control the same set of matched bilinguals and monolinguals typically participate in only a single task and, hence, provide only a single measure of this component. This is less than satisfactory because the three most-frequently used nonlinguistic interference tasks do not correlate with one another. Stins, Polderman, Boomsma, and de Geus (2005) tested a group of 12-year old children using the flanker, Simon, and Stroop interference tasks. The correlations between these tasks were all smaller than +0.20 and nonsignificant. Fan, Flombaum, McCandliss, Thomas, and Posner (2003) reported that the flanker, Simon, and Stroop tasks all activated the AC and the left PFC, but again the interference scores were uncorrelated. Each task also activated areas unique to that task. They conclude that "*The behavioral and fMRI results taken together seem to argue against a single unified network for processing conflict, but instead support either distinct networks for each conflict task or a single network that monitors conflict with different sites used to resolve the conflict*" p. 42. Likewise, Kousaie and Phillips (2012a) using a sample of 51 young adults report no significant correlations between the Stroop, Simon, and Eriksen flanker tasks. Keye, Wilhelm, Oberauer, and van Ravenzwaaij (2009) report a structural-equation analysis of the data obtained from 150 adults who participated in both the flanker and Simon task. There was no association between the two interference tasks. Likewise, Humphrey and Valian (2012) using a sample of 208 young adults report no significant correlation between the Simon and flanker effect. In contrast to all of the above comparisons involving the Simon, flanker, or Stroop tasks; Unsworth and Spillers (2010) in a study using college students do report a significant correlation ($r = +0.17$, $p < .05$) between a flanker and Stroop task. The weak correlation achieves statistical significance because the n of 181 is very large.

The fact that Friedman et al. (2008) found evidence for an inhibition component is at odds with the absence of significant correlations reported above for the Simon, Stroop, and flanker tasks. At first look this may appear puzzling since the nature of the inhibition required in the three tasks used by Friedman et al. (antisaccade, stop signal, and Stroop) appears to be as varied, if not more so, than the type of inhibition required across the Simon, flanker, and Stroop tasks. Although these tasks always produce robust interference effects, individual differences seem to be another matter. Friedman et al. (2008) point out that the individual difference correlations in interference tasks are usually low and seem sensitive to task variations. The present studies will provide additional tests of convergent validity between the Simon and antisaccade tasks (Study 1) and between the Simon and flanker tasks (Study 3).[2]

[2] Our use of the term inhibitory control ignores the distinction between interference control (suppression of interference due to resource or stimulus competition) and response inhibition (suppression of prepotent responses). The task-impurity problem makes it very difficult to isolate the different interference-related processes and Friedman and Miyake (2004) have shown that the latent variables for interference control (e.g., the flanker effect in their study) and response inhibition (e.g., antisaccade and Stroop effects) are highly correlated (.68).

*1.1.2. Monitoring*

Given that cognitive control dynamically changes in response to changing goals and changing affordances another important type of executive process is monitoring one's performance, internal states, and current environment.[3] This possibility was first proposed by Bialystok et al. (2004) in the context of the Simon task: "*The advantage for bilinguals, therefore, may be not in the enhanced ability to inhibit the misleading spatial cue but in the ability to manage attention to a complex set of rapidly changing task demands*" p. 292. Costa, Hernández, Costa-Faidella, and Sebastián-Gallés (2009) have also focused more on monitoring than inhibitory control: "*The bilingual advantage in overall RTs may reveal the better ability of bilinguals to handle tasks that involve mixing trials of different types: bilinguals would be more efficient at going back and forth between trials that require implementing conflict resolution and those that are free of conflict*" (p. 136). Costa et al. have also updated the underlying rationale for why managing two languages should enhance conflict monitoring: "*This better functioning of the monitoring system may come about because of the bilinguals' need to continuously monitor the appropriate language for each communicative interaction. That is, proper communication in bilingual settings involves the monitoring of the language to be used depending on the interlocutor(s) language knowledge*". (Costa et al., 2009, p. 136).

Differences in monitoring between bilinguals and monolinguals have been inferred from two different measures. The arguably better indicator is the difference in mean RT between a pure block of easy choice RT trials that involve no conflict and the mean RT on the congruent trials in a block that mixes both congruent and incongruent trials together. This difference should reflect the cost of having to monitor for the presence of conflict, apart from the need to actually resolve the conflict on incongruent trials. Ideally, the two groups would show comparable RTs on the pure blocks of easy trials showing that they are matched in terms of speed when there is no need to monitor for potential conflict. A control condition of no-conflict trials is not always included in the experimental design. Under these circumstances a bilingual advantage in monitoring has been inferred if bilinguals are faster than monolinguals on the congruent trials.

*1.1.3. Switching*

Another special experience of bilinguals is that they get lots of experience in switching per se as they shift from one language to the other. Language switching is complicated by the common assumption that both lexicons are connected to the same conceptual system and that consequently switching leads to the need to inhibit the translation equivalents in the non-target language. Setting this complication aside, the ability to switch from one task to a completely different task is assumed to require an executive process ("*switching*") that is functionally separable from inhibitory control and monitoring and involves unique areas in addition to the anterior cingulate (AC) and PFC (Collette et al., 2005).

Tests for bilingual advantages in switching include three critical conditions: (1) pure blocks in which participants perform the same task on every trial, (2) mixed-block trials that involve a switch from one task to the other, and (3) mixed-block trials that repeat the same decision. For example, in a mixed block Prior and MacWhinney (2010) presented a precue on every trial that signaled to the participant whether to make a binary color decision with two fingers of one hand or a binary shape decision with two fingers of the other hand. The difference between repeat trials and switch trials in the mixed block was used as an indicator of "*switching costs*" whereas the difference between repeat trials and pure trials was used as an indicator of "*mixing costs*". The results showed a bilingual advantage in switching costs, but not mixing costs. These results, in isolation, are coherent if switching and monitoring are two separate EP components and bilingualism provides special and domain-independent experience only for switching.

[3] Miyake and Friedman did not include the monitoring component per se in their studies, but suggest that it could be considered a subcomponent of both switching and updating. The lack of explicit focus on monitoring is unfortunate for present purposes because the current prevailing view is that bilingual advantages in EP are more likely to occur in the monitoring component than in the inhibitory-control component.

**Table 1**
Sequence of activities for each of the three samples.

| Sample 1 | Sample 2 | Sample 3 |
|---|---|---|
| Informed consent | Informed consent | Informed consent |
| Questionnaire | Questionnaire | questionnaire |
| (Category fluency) | (Category fluency) | (Category fluency) |
| (Collectivism–individualism) | (Category-priming LDT) | Raven's advanced matrices |
| Anti-saccade task | | |
| Simon task | Simon task | Simon task |
| Color-shape switching | Color-shape switching | Color-shape switching |
| (Category-priming LDT) | Homograph interference | |

*Note*: Tasks in parentheses are not presented in this article.

### *1.2. Purpose*

The main purpose of this study is to use the framework and strategy presented above to determine if there is a coherent set of evidence for a bilingual advantage in EP. Three large samples of participants complete multiple tasks that require EP. Each task yields multiple comparisons that are typically assumed to be associated with specific components of EP. More specifically, the analyses of each sample, and the combined sample, enable multiple tests for bilingual advantages in inhibitory control, monitoring, and switching. Furthermore, by examining the correlations between performance variables one can assess the assumption that these markers are converging on domain general measures of cognitive control.

## 2. Methods

The data reported are drawn from three studies. Each study consisted of a series of seven or eight activities that required 1.5–2 h to complete. Table 1 shows the sequence of activities for each study.

### *2.1. Classification and characteristics of participants*

#### *2.1.1. Recruitment*

Participants were students taking psychology courses at San Francisco State University (SFSU). The vast majority were students in the first author's cognitive psychology or psycholinguistics courses.[4] Participation fulfilled a research project assignment that could also be satisfied by writing a brief report. Most of the participants were junior or senior psychology majors. Although SFSU students come from diverse backgrounds and cultures some homogenization through self-selection and recent life experience is likely to occur when both bilingual and monolingual samples are drawn from the same pool of students.

#### *2.1.2. Language proficiency scale and classification rule*

Participants rated their speaking and listening proficiency using the following scale:

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.

[4] One reviewer reasonably asks about possible demand characteristics given that most of our participants were drawn from students in the first author's cognitive and psycholinguistics courses. The results of any previous studies or the current study were never discussed until the last day of class. The informed-consent and recruitment protocol characterized the purpose as exploring for possible differences between bilinguals and monolinguals in several cognitive tasks. With respect to experimenter bias, the authors and research assistants all anticipated finding bilingual advantages in EP in Study 1 and were quite agnostic when the flanker task was added later. The tasks were computer controlled and this should minimize experimenter influence.

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.

If a participant rated their proficiency in two (or more) languages as a 4 or more they were classified as bilingual. If a participant rated their proficiency in English as a 4 or more and rated all other languages as 3 or less they were classified as monolingual. Participants who did not meet either classification criteria were excluded from further data analyses.[5]

### *2.1.3. Language characteristics for each sample*

For Study 1 a total of 90 participants were recruited in the spring semester of 2010. A subset of 34 was classified as bilingual and 46 were classified as native English-speaking monolinguals. Ten participants were classified as neither bilingual nor monolingual. Three were non-native English speakers who rated their proficiency in English as only a 3. As upper division students immersed in English as the language of instruction it seemed inappropriate to classify them as monolinguals. The other six participants rated their proficiency in their native language as 3 or less and their proficiency in English as 5 or higher. The families of these individuals typically moved to the United States when they were preschoolers and they never achieved fluency in their native language. One participant had a disability that prevented her from responding with the right hand, and although she completed all but the switching task, her data was not included in the analyses reported below. In summary there were 80 able participants who could be unambiguously classified as bilingual ($n = 34$) or monolingual ($n = 46$).

For Study 2 a total of 86 participants were recruited from the author's classes in the fall semester of 2010. A subset of 36 was classified as bilingual and a second subset of 50 was classified as native English-speaking monolinguals. Three participants were eliminated because they had participated in the first study, two participants did not complete the questionnaire, and one non-native English speaker rated her proficiency in English as only a 3 despite English being the current language of instruction.

For Study 3 a total of 110 participants were recruited from the author's classes in the spring semester of 2011. A subset of 55 was classified as monolingual and a second subset of 52 was classified as native English-speaking monolinguals. Three participants were eliminated from further analyses because English was not their native language and they rated their proficiency in their native language as three or less.

The mean proficiency and percentage of English usage for the bilingual and monolingual groups in each of the three samples groups are shown in Table 2. The bottom panel of the table shows the weighted means for the combined data. The table also subcategorizes the bilinguals into those who have two native languages (i.e., were exposed to both languages from birth), those whose native language is English, and those who native language is a language other than English. The latter two groups have a native language (L1) and a non-native language (L2). The median age-of-acquisition (i.e., age of initial exposure) for L2 is shown for the latter two groups. Finally, Table 2 also shows the proportion of English use compared to using other languages.

Reflecting the diversity of San Francisco the 122 bilinguals spoke 30 different languages. Our typical bilingual is a native speaker of Spanish (or Cantonese or Mandarin or Tagalog), acquired English as a second language at about age five, switch languages every day, and speak English about 70% of the time. Although English is not the native language for these modal bilinguals it is noteworthy that their rated proficiency in English (mean = 6.0) is equivalent to that in their native language (mean = 5.9) and

[5] Direct tests of language proficiency were not included in the present set of studies, but have been incorporated into ongoing research projects of the same nature. It is, however, important to note that self ratings have been shown to correlate highly with a range of objective measures of language proficiency and fluency (Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012).

**Table 2**
Language characteristics of monolinguals, bilinguals, and types of bilinguals.

| Group | Spoken proficiency | | | Median age-of-acquisition | | |
|---|---|---|---|---|---|---|
| | *N* | English | Other | English | Other | Percent English |
| *Sample 1* | | | | | | |
| Monolingual | 46 | 6.6 | 1.0 | 0.0 | 13.0 | - |
| Bilingual | 34 | 6.2 | 5.4 | 4.3 | 0.0 | 75 |
| Native both languages | 5 | 6.4 | 4.4 | 0.0 | 0.0 | 88 |
| Native other language | 25 | 6.0 | 5.8 | 5.0 | 0.0 | 70 |
| Native English language | 4 | 6.8 | 5.0 | 0.0 | 7.5 | 87 |
| *Sample 2* | | | | | | |
| Monolingual | 50 | 6.7 | 1.7 | 0.0 | 14.0 | - |
| Bilingual | 36 | 6.2 | 5.8 | 5.0 | 0.0 | 71 |
| Native both languages | 7 | 6.6 | 5.3 | 0.0 | 0.0 | 73 |
| Native other language | 26 | 6.0 | 6.0 | 6.0 | 0.0 | 71 |
| Native English language | 3 | 6.3 | 5.3 | 0.0 | 13.0 | 72 |
| *Sample 3* | | | | | | |
| Monolingual | 55 | 6.6 | 1.8 | 0.0 | 14.0 | - |
| Bilingual | 52 | 6.2 | 5.6 | 2.0 | 0.0 | 71 |
| Native both languages | 17 | 6.6 | 5.3 | 0.0 | 0.0 | 74 |
| Native other language | 29 | 6.0 | 6.0 | 5.0 | 0.0 | 70 |
| Native English language | 6 | 6.3 | 5.2 | 0.0 | 5.5 | 86 |
| *Combined* | | | | | | |
| Monolingual | 151 | 6.6 | 1.5 | 0.0 | 14.0 | - |
| Bilingual | 122 | 6.2 | 5.6 | 3.5 | 0.0 | 72 |
| Native both languages | 29 | 6.6 | 5.1 | 0.0 | 0.0 | 76 |
| Native other language | 80 | 6.0 | 5.9 | 5.3 | 0.0 | 70 |
| Native English language | 13 | 6.4 | 5.2 | 0.0 | 7.9 | 84 |

that our scale identifies a rating of 6 to mean "*As fluent as a typical native speaker*". In contrast the mean proficiency of the monolinguals in any L2 was only 1.5.

In a review article discussing the bilingual advantage in EP Bialystok (2009) noted that "all the research reported in these studies was based on individuals who were fully bilingual and used both languages regularly (often daily) to a high level of proficiency" (p. 9). As shown in Table 2 Bialystok's qualitative description applies to our bilinguals.

### *2.2. Computer control*

All of the tests of EP were programmed in and controlled by DirectRT, as was the Raven's Advanced Matrices test included in Study 3. Three identical Dell desktop computers running MS Windows XP were used as work stations separated by partitions. Viewing distance was not strictly controlled and varied between 45 and 55 cm. The visual angles reported below were based on a viewing distance of 50 cm.

### *2.3. Simon task*

#### *2.3.1. Trial definition*

Each trial began with the presentation of a center fixation (+) for 500 ms. The center fixation was immediately followed by the target stimulus which was either a "Z" or a "/". The participant's task was to press the corresponding key as quickly as possible without making errors. The left index finger rested on the "Z" key and the right index finger rested on the "/" key. In a neutral block the target was displayed either 2.3° above or below the center fixation. In a Simon block the target was displayed either 3.9° to the left or to the right of the center fixation. In a Simon block a trial was defined as congruent if the location of the target was on the same side as the correct response and as incongruent if the location of the target was on the opposite side.

#### 2.3.2. Design

The critical Simon blocks were identical across the three studies and were always the last two blocks. Each Simon block consisted of 20 congruent trials and 20 incongruent trials presented in random order. Half the trials of each type presented the target on the left with the other half presented the target on the right. Thus, the mean response time (RT) for the four conditions defined by the combination of two blocks and two levels of congruency (congruent versus incongruent) were each based on 20 trials and when collapsed across block on 40 trials.

In Study 1 the two critical Simon blocks were preceded by a practice Simon block of 20 trials. In contrast to Study 1, rather than having a block of practice in the standard Simon task, Study 2 started with two blocks of trials where the target was displaced either above or below the center fixation. This change introduces a "neutral" condition because the location of the target is neither compatible nor incompatible with pressing the "Z" key on the left or the "/" key on the right. Block 1 provided 20 trials of practice in the neutral condition and was followed by a 40-trial Block 2. Displacements above and below the fixation were randomly ordered with the constraint that within the block there were 20 displacements above and 20 below.

### 2.4. Color-shape switching task

The color-shape switching task was identical across Studies 1–3. The task was patterned on that used by Prior and MacWhinney (2010).

#### 2.4.1. Trial definition

Each trial began with the presentation of a center fixation (+) for 350 ms and then a blank screen for 150 ms. The left middle and index fingers rested on the "Z" and "X" key, respectively. The right index and middle fingers rested on the "." and "/" keys, respectively. In a pure color block the participant's task was to press the "Z" key if the target was blue and the "X" key if it was red. In a pure shape block the task was to press the "." key if the target was a circle and the "/" key if it was a triangle. The target set consisted of a blue circle, a blue triangle, a red circle, and a red triangle.

In a mixed block the target was preceded by a precue for 250 ms that remained in view until the participant responded to the target. If the precue was a rainbow then the participant had to make a color decision when the target appeared. If the precue was a black circle embedded within a black triangle then the participant had to make a shape decision when the target appeared. Participants were instructed to respond as quickly as they could on the basis of the precued dimension (viz., color or shape). Each trial was designated as a "repeat" trial if the cued decision was the same as on the previous trial and a "switch" trial if it was different. Each target and precue subtended about 1.83° of visual angle with the center of the precue appearing 2.3° above the center of the fixation stimulus and the upcoming target.

#### 2.4.2. Design

The task consisted of six blocks. The first block of 16 trials was "pure" color. Each of the four targets appeared four times in random order. The second block of 16 trials was "pure" shape with each of the targets appeared in random order. Following Block 2 the "mixed" task was introduced with detailed instructions regarding the use of the precue to signal whether a color or shape would be required on each specific trial. Each of the four "mixed" blocks started with two buffer trials that were not analyzed. Block 3 was a practice block and consisted of 18 trials (including the two buffers). Blocks 4, 5, and 6 each consisted of 50 trials (including the two buffers). A single random order was used for every participant. Each of the four targets appeared 36 times across Blocks 4 to 6 and there were 72 repeat trials and 72 switch trials.

### 2.5. Antisaccade task

The design, materials, and procedure for the antisaccade tasks were closely modeled from those used by Kane, Bleckly, Conway, and Engle (2001) who showed that individual differences in working-memory capacity predicted performance on an antisaccade blocks, but not prosaccade blocks.

The task on each trial was to identify the target stimulus (i.e., "B", "P", or "R") by pressing the key with the corresponding label using three fingers of the right hand. The briefly presented target is followed by a visually similar mask ("8"). The target and mask subtended about 0.9° of visual angle. In the anti-saccade condition a distracter stimulus is always blinked (presented and represented for 100 ms with a blank ISI of 50 ms) before and on the opposite side from the target stimulus. The distractor appeared about 2.0° to one side of fixation and the target 2.0° to the opposite side. Because the eventual target is always presented on the opposite side the best strategy is to inhibit the natural predisposition to attend to (or saccade toward) any peripheral stimulus with an abrupt onset. If bilinguals have superior inhibitory control, then they should respond faster in the antisaccade task. The antisaccade trials are preceded by a block of control trials that used a centered target and no distracting stimulus. The control trials provide a baseline response time (RT) that should require little or no EP.

Experimental trials consisted of the following sequence of events: (1) a center fixation (∗∗∗) was presented for a variable duration (i.e., 600, 1000, 1400, 1820, 2200 ms) in order to introduce temporal uncertainty; (2) a blank field for 100 ms; (3) a "#" sign for 100 ms displaced 2° to the opposite side from the eventual target; (4) a blank field for 50 ms; (5) the "#" sign in the same location for 100 ms; (6) a target letter ("B", "P", or "R) for 150 ms displaced a comparable extent on the opposite side; (7) a mask ("8") presented until the response.

The baseline trials presented no opposite field distracter and consisted of these events: (1) a center fixation (∗∗∗) was presented for a variable duration (i.e., 600, 1000, 1400, 1820, 220 ms; (2) a blank field for 100 ms; (3) a centered target-letter ("B", "P", or "R) for 150 ms; and (4) a mask ("8") presented until the response.[6]

The trials were organized and presented in the following order. A practice block consisted of 15 baseline trials, one at each combination of 5 fixation durations and 3 target letters and presented in random order. Block 2 was identical to the first block and provided the baseline RTs. Block 3 was 30 anti-saccade trials formed by the random combination of: 5 fixation durations by 3 target letters by 2 sides (left and right).

### *2.6. Ravens advanced progressive matrices task*

Similar to Costa et al. (2009) nonverbal general intelligence was assessed using Set 1 of the Ravens 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.

### *2.7. Eriksen flanker task*

Costa et al. (2008) used an elaboration of the Eriksen flanker task developed by Fan, McCandliss, Sommer, Raz, and Posner (2002) that is referred to as the attentional network task (ANT). Costa et al. reported that bilinguals were faster than monolinguals on both congruent and incongruent trials (a global RT advantage), showed smaller interference effects, and were aided more by the presence of an alerting cue. The advantage in inhibitory control was present in Blocks 1 and 2, but not in Block 3. Given that tests for bilingual advantages in EP do not often use the flanker task and usually do not result in advantages in inhibitory control, the flanker task used in this study is patterned very closely on the one used by Costa et al. (2008).

[6] We used this centered condition rather than a prosaccade condition, and always tested it first, because Kane et al. showed that low-span participants performed poorly when switching from antisaccade to prosaccade blocks. Our goal was to establish a neutral baseline that could show that our groups were matched in speed and accuracy when identifying masked targets in the absence of exogenous distractors.

*2.7.1. Trial definition*

The cues, arrows, and flankers were implemented in DirectRT using Costa et al.'s (2009) Figure 1 as the model. The congruent display consisted of a central arrow pointing either left or right and two flankers on each side pointing in the same direction. A single arrow subtended about .9° of visual angle and the entire horizontal extent of the five-arrow stimulus was about 6.3°. In the incongruent displays the flankers pointed in the opposite direction from the central target arrow. The sequence of events was as follows: (a) a fixation point (a plus sign) appeared at the center of the screen and remained throughout the trial, (b) a cue (described below) was presented for 100 ms, (c) followed by the fixation field for an additional 400 ms, and then (d) the target display until the participant's response or for up to 1700 ms. The target was vertically displaced either 1.2° above or below the fixation point. Participants were instructed to press the "z" key with their left index finger if the target arrow pointed left and to press the "/" key with their right index finger if the target arrow pointed right.

Consistent with the ANT methodology four types of cues were used. On "no cue" trials the 100 ms cue display is simply a continuation of the centered fixation point (+). Obviously it affords no information about the temporal onset or spatial location of the upcoming target. The "double cue" display consists of a two ◇ symbols above and below the fixation point. This provides no information about the location of the upcoming target, but does reduce the temporal uncertainty. Subtracting the means of the double cue trials from the no cue trials yields the alerting effect. The third type of cue is the "central cue" that simply replaces the + fixation point with the ◇ symbol. It does reduce temporal uncertainty, but provides no cue to spatial location. In contrast, the "spatial cue" display adds a valid diamond cue above or below the fixation point. As both the "central cue" and "spatial cue" displays provide the same advantages in alerting, the mean of the "spatial cue" trials can be subtracted from the mean of "central cue" trials to derive the orienting effect.

*2.7.2. Design*

Block 1 consisted of 20 neutral trials where all the targets consisted of a centered arrow and the flankers were dashes. Each target was randomly preceded by one of the four cue types. Block 1 is similar to the block of neutral trials that initiated the Simon task and, likewise, enables the computation of mixing costs by subtracting the mean of these neutral trials from the mean of the congruent trials in the experimental blocks that randomly mix conflict and no-conflict trials.

Blocks 2 through 5 were standard ANT blocks with 50% congruent and incongruent trials. Block 2 consisted of 16 trials and was considered practice. Blocks 3, 4, and 5 each consisted of 64 trials with 8 repetitions of the combinations formed by 2 target types (congruent versus incongruent) × 4 cue displays. Thus, given standard practice for analyzing each attentional network (executive attention, alerting, and orientating) in the ANT each block provided 32 trials of each condition (e.g., 32 congruent and 32 incongruent trials) and overall means were based on 96 trials. The trials within each block were randomized.

*2.8. Computation of mean RT per condition for all tasks*

Other than the Raven's test, the primary dependent variable is correct response time (RT). For each RT task and each individual participant the mean and standard deviation (SD) for correct responses was computed across the experimental trials. RTs exceeding 2.5 SDs above the mean were eliminated. The trimming was done separately for the neutral blocks (e.g., Simon blocks with vertical target displacement) and the blocks assumed to require EP (e.g., Simon blocks with horizontal target displacement). The number of participants included in the analysis of a specific study and task is sometimes less than the totals shown in Table 2 for two general reasons. One class is missing data caused by the participant failing to return for session 2 or because the task was interrupted by an external event such as a power outage. The data from some participants was removed because their performance in terms of RT, accuracy, or both were clear outliers. The number of bilingual and monolingual participants eliminated for performance reasons is stated for each task.

## 3. Results and discussion by task

### 3.1. Simon task

The data for one bilingual was eliminated in Study 3 because her mean RT was more than 300 ms longer than the next slowest participant. The trimmed correct RTs were analyzed using a mixed design ANOVA with group (bilingual versus monolingual) and congruency (congruent versus incongruent) as factors. In testing for bilingual advantages the main effect of congruency (i.e., the magnitude of the Simon effect) plays a less important role than the main effect of group and the Group × Congruency interaction, but it is important to show that it is significant and falls within the range of typical findings. This is also true for the main effects defining the flanker effect, antisaccade effect, switching costs, and mixing costs. Table 3 aggregates the main effects in each task showing the means, standard errors, F statistic, exact probabilities, and effect sizes. As shown in Table 3 there is always a highly significant Simon effect of about +32 ms with an associated effect size of about .7. The magnitude of the interference effect is very typical for the Simon task (Lu & Proctor, 1995).

The Group × Congruency interaction and the main effect of group provide tests for bilingual advantages in inhibitory control and monitoring, respectively. A bilingual advantage in inhibitory control can be inferred if the interference effect for monolinguals is greater than that for bilinguals. Thus, the key test statistic is the F ratio for the Group × Congruency interaction. Table 4 shows separately for bilinguals and monolinguals the means and standard errors for the congruent and incongruent trials, the F statistic for the interaction, and the magnitude of the bilingual advantage in inhibitory control as indexed by differences in the size of the Simon effect for each group. A plus sign indicates that the difference is consistent with a bilingual advantage in inhibitory control (i.e., bilinguals show a smaller Simon effect). There is no evidence supporting a bilingual advantage in inhibitory control. In fact, Table 4 shows a small, but significant bilingual disadvantage in both Study 3 (−10 ms) and in the combined analysis (−5 ms). This was an unanticipated outcome, but note that the effect sizes associated with these disadvantages were extremely small. Parallel analyses of proportion correct are shown in Table 5. Overall accuracy in the Simon task is very high, averaging 98% correct. All of the F statistics associated with the Group × Congruency interaction are nonsignificant.

**Table 3**
Main effects for the difference-score indicators of EP.

| Data set | *N* | Control Trial RT | Control Trial SE | EP Trial RT | EP Trial SE | Difference | *F* | Exact *p* | Effect Size |
|---|---|---|---|---|---|---|---|---|---|
| | | Congruent | | Incongruent | | Interference effects | | | |
| Simon 1 | 79 | 449 | 7 | 480 | 7 | +31 | 183.00 | .000 | .704 |
| Simon 2 | 86 | 448 | 6 | 481 | 6 | +33 | 221.88 | .000 | .725 |
| Simon 3 | 106 | 466 | 5 | 497 | 5 | +31 | 203.70 | .000 | .662 |
| Simon combined | 271 | 455 | 3 | 487 | 3 | +32 | 605.54 | .000 | .692 |
| Flanker 3 | 104 | 539 | 7 | 635 | 9 | +96 | 495.45 | .000 | .829 |
| | | Pure trials | | Repeat trials | | Mixing costs | | | |
| Mixing cost 1 | 74 | 537 | 17 | 848 | 40 | +311 | 81.83 | .000 | .532 |
| Mixing cost 2 | 80 | 534 | 16 | 793 | 24 | +259 | 115.73 | .000 | .597 |
| Mixing cost 3 | 99 | 590 | 19 | 834 | 29 | +244 | 85.64 | .000 | .469 |
| Mixing cost combined | 253 | 558 | 10 | 826 | 18 | +268 | 273.57 | .000 | .522 |
| | | Repeat trials | | Switch trials | | Switching costs | | | |
| Switching cost 1 | 74 | 848 | 40 | 1061 | 51 | +213 | 91.89 | .000 | .561 |
| Switching cost 2 | 80 | 793 | 24 | 995 | 30 | +202 | 248.31 | .000 | .761 |
| Switching cost 3 | 99 | 834 | 29 | 1052 | 36 | +218 | 255.59 | .000 | .725 |
| Switching cost combined | 253 | 826 | 18 | 1060 | 34 | +234 | 519.34 | .000 | .674 |

*Note*: Numbers following task refer to Studies 1, 2, or 3. Effect size is partial $\eta^2$.

**Table 4**
Language-group RT differences in the Simon (S) and Flanker (F) tasks.

| Data set | LG | *N* | Con | | Incon | | Interference effect (inhibitory control) | | | | | Global RT (monitoring) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | RT | SE | RT | SE | Diff | Adv | *F* | *p* | $\eta^2$ | GRT | Adv | *F* | *p* | $\eta^2$ |
| Simon 1 | B | 34 | 459 | 10 | 489 | 10 | +30 | +1 | 0.08 | .780 | .001 | 474 | −19 | 2.00 | .161 | .025 |
| | M | 45 | 440 | 9 | 471 | 9 | +31 | | | | | 455 | | | | |
| Simon 2 | B | 36 | 445 | 9 | 481 | 6 | +36 | −5 | 1.19 | .278 | .014 | 463 | +3 | 0.08 | .772 | .001 |
| | M | 50 | 451 | 7 | 482 | 7 | +31 | | | | | 466 | | | | |
| Simon 3 | B | 51 | 464 | 8 | 500 | 7 | +36 | −10 | 5.12 | .026 | .047 | 482 | −1 | 0.01 | .918 | .000 |
| | M | 55 | 468 | 8 | 494 | 7 | +26 | | | | | 481 | | | | |
| S All | B | 121 | 457 | 5 | 491 | 5 | +34 | −5 | 3.89 | .050 | .014 | 474 | −5 | 0.74 | .390 | .003 |
| | M | 150 | 454 | 5 | 483 | 4 | +29 | | | | | 469 | | | | |
| S High F | B | 61 | 466 | 9 | 504 | 9 | +38 | −7 | 4.32 | .039 | .029 | 485 | −12 | 1.19 | .276 | .008 |
| | M | 84 | 457 | 8 | 488 | 8 | +31 | | | | | 473 | | | | |
| S PED M | B | 90 | 458 | 6 | 492 | 6 | +34 | −4 | 1.44 | .232 | .008 | 475 | −10 | 1.47 | .227 | .008 |
| | M | 90 | 450 | 6 | 480 | 6 | +30 | | | | | 465 | | | | |
| Flanker 1 | B | 49 | 560 | 13 | 643 | 15 | +83 | −1 | 0.01 | .909 | .000 | 602 | −6 | 0.09 | .762 | .001 |
| | M | 55 | 555 | 12 | 637 | 14 | +82 | | | | | 596 | | | | |
| Flanker 2 | B | 49 | 534 | 11 | 638 | 14 | +104 | −6 | 0.34 | .561 | .003 | 586 | −6 | 0.11 | .737 | .001 |
| | M | 55 | 531 | 11 | 629 | 13 | +98 | | | | | 580 | | | | |
| Flanker 3 | B | 49 | 529 | 10 | 635 | 15 | +106 | −4 | 0.12 | .731 | .001 | 582 | −3 | 0.04 | .835 | .000 |
| | M | 55 | 528 | 10 | 630 | 14 | +102 | | | | | 579 | | | | |
| F All | B | 49 | 541 | 11 | 639 | 14 | +98 | −4 | 0.18 | .676 | .002 | 590 | −5 | 0.00 | .767 | .001 |
| | M | 55 | 538 | 10 | 632 | 13 | +94 | | | | | 585 | | | | |
| F High F | B | 28 | 571 | 18 | 697 | 32 | +126 | −31 | 1.36 | .248 | .023 | 634 | −39 | 1.32 | .255 | .023 |
| | M | 31 | 548 | 17 | 643 | 31 | +95 | | | | | 595 | | | | |
| F PED M | B | 40 | 548 | 12 | 649 | 16 | +101 | −9 | 0.78 | .381 | .010 | 598 | −7 | 0.14 | .711 | .002 |
| | M | 40 | 545 | 12 | 637 | 16 | +92 | | | | | 591 | | | | |

*Note*: Numbers for Simon refer to study number and for flanker refer to block number. High F = high fluency subset; PED M = PED matched subset; B = bilingual, M = monolingual; Con = congruent; Incon = incongruent; LG = language group; Adv = bilingual advantage; GRT = global RT.

The main effect of group (within a block that includes both congruent and incongruent trials) was used by Hilchey and Klein (2011) as the primary measure for a bilingual advantage in monitoring.[7] Table 4 also shows the means and standard errors for the main effect of group, the critical F statistic, and the effect size. There were no significant main effects (Global RT or accuracy differences) across the three studies and in two of the three cases the trend is toward a bilingual disadvantage.

### *3.1.1. Highly-fluent bilinguals*

Although our criteria for classifying participants as bilingual or monolingual are similar to most studies investigating the bilingual advantage, it is fair to ask if highly-fluent and balanced bilinguals may produce advantages in inhibitory control or monitoring compared to monolinguals with very little exposure to a second language. To answer that question a high-fluency group of 61 bilinguals was formed from participants who rate their fluency as 6 or 7 on two or more languages. Recall that a 6 on our rating scale means as fluent as a typical native speaker and that 7 represents a level of super fluency that is better than a typical native speaker. In mirror-image fashion a subset of 84 monolinguals was formed that rated their proficiency in languages other than English as a 0 or 1.

As shown in Table 4 these subsets show exactly the same pattern of results, a significant, main effect of congruency; a significant Group × Congruency interaction; and a nonsignificant main affect

[7] Hilchey and Klein refer to this difference as a Bilingual Executive Processing Advantage (BEPA) which involves monitoring and managing trial to trial variation with respect to the presence or absence of conflict and which they clearly distinguished from the hypothesized Bilingual Inhibitory Control Advantage (BICA). In both their meta analysis and in the analyses of our data the main effect collapsed across both congruent and incongruent trials and congruent trials alone yield very similar results.

**Table 5**
Language-group differences in proportion correct in the Simon (S) and Flanker (F) tasks.

| Data set | LG | *N* | Con | | Incon | | Interference effect (inhibitory control) | | | | | Global RT (monitoring) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | PC | SE | PC | SE | Diff | Adv | *F* | *p* | $\eta^2$ | GPC | Adv | *F* | *p* | $\eta^2$ |
| Simon 1 | B | 34 | .989 | .004 | .974 | .004 | +.015 | −.001 | 0.00 | .972 | .000 | .982 | +.001 | 0.02 | .896 | .000 |
| | M | 45 | .988 | .004 | .974 | .003 | +.014 | | | | | .981 | | | | |
| Simon 2 | B | 36 | .988 | .004 | .961 | .007 | +.027 | −.009 | 1.18 | .280 | .014 | .975 | +.004 | 0.40 | .528 | .005 |
| | M | 50 | .980 | .003 | .962 | .006 | +.018 | | | | | .971 | | | | |
| Simon 3 | B | 51 | .987 | .003 | .972 | .005 | +.015 | +.007 | 0.72 | .396 | .007 | .980 | +.004 | 0.58 | .447 | .006 |
| | M | 55 | .987 | .003 | .965 | .005 | +.022 | | | | | .976 | | | | |
| S all | B | 121 | .988 | .002 | .969 | .003 | +.019 | −.001 | 0.06 | .811 | .000 | .978 | +.002 | 0.63 | .439 | .002 |
| | M | 150 | .985 | .002 | .967 | .003 | +.018 | | | | | .976 | | | | |
| S high F | B | 61 | .991 | .003 | .966 | .005 | +.025 | −.009 | 1.44 | .232 | .010 | .978 | +.003 | 0.85 | .357 | .006 |
| | M | 84 | .983 | .002 | .967 | .004 | +.016 | | | | | .975 | | | | |
| S PED M | B | 90 | .987 | .003 | .971 | .004 | +.016 | +.001 | 0.02 | .884 | .000 | .979 | +.005 | 2.07 | .152 | .012 |
| | M | 90 | .982 | .003 | .965 | .004 | +.017 | | | | | .974 | | | | |
| Flanker 1 | B | 49 | .996 | .002 | .977 | .006 | +.019 | −.001 | 0.01 | .913 | .000 | .986 | −.001 | 0.05 | .818 | .001 |
| | M | 55 | .996 | .002 | .978 | .005 | +.018 | | | | | .987 | | | | |
| Flanker 2 | B | 49 | .996 | .002 | .978 | .005 | +.018 | −.001 | 0.01 | .913 | .000 | .987 | −.001 | 0.05 | .818 | .002 |
| | M | 55 | .996 | .002 | .979 | .005 | +.017 | | | | | .988 | | | | |
| Flanker 3 | B | 49 | .992 | .002 | .974 | .005 | +.018 | −.004 | 0.55 | .458 | .005 | .983 | −.005 | 1.79 | .184 | .017 |
| | M | 55 | .995 | .003 | .981 | .004 | +.014 | | | | | .988 | | | | |
| F all | B | 49 | .994 | .001 | .976 | .004 | +.018 | −.002 | 0.14 | .712 | .001 | .985 | −.003 | 0.48 | .489 | .005 |
| | M | 55 | .996 | .001 | .980 | .004 | +.016 | | | | | .988 | | | | |
| F high F | B | 28 | .993 | .002 | .974 | .005 | +.019 | −.002 | 0.10 | .751 | .002 | .983 | −.003 | 0.60 | .440 | .010 |
| | M | 31 | .995 | .002 | .978 | .005 | +.017 | | | | | .986 | | | | |
| F PED M | B | 40 | .996 | .001 | .978 | .004 | +.018 | −.008 | 1.65 | .202 | .021 | .987 | −.004 | 1.56 | .216 | .020 |
| | M | 40 | .996 | .001 | .986 | .004 | +.010 | | | | | .991 | | | | |

*Note*: Numbers for Simon refer to study number and for flanker refer to block number. High F = high fluency subset; PED M = PED matched subset; B = bilingual, M = monolingual; Con = congruent; Incon = incongruent; LG = language group; Adv = bilingual advantage; GPC = global proportion correct.

of group. There are no hidden bilingual advantages in either global RT or inhibitory control when only highly-fluent balanced bilinguals are compared to monolinguals with very little exposure to an L2.

*3.1.2. Groups matched on parent's education level*

Fair tests for differences between bilinguals and monolinguals rely on the groups being matched with respect to other factors that may influence EP. Hilchey and Klein (2011) provide an extensive discussion of possible hidden factors, the need for appropriate demographic controls, and meticulously chart the controversies surrounding the performance of bilingual and monolingual children using the Simon task. Bialystok, Martin, and Viswanathan (2005) reported global RT advantages for English-French bilingual 5-year olds compared to English speaking monolinguals who lived in "*similar middle-class neighborhoods*" in the same city and who were also matched on digit span. Morton and Harper (2007) raise the concern that this recruitment protocol did not ensure that the groups were matched on socioeconomic status (SES), ethnicity, and immigrant status.[8] When Morton and Harper replicated Bialystok et al.'s Simon task, but recruited all non-immigrant participants and showed that the groups were equal in SES with respect to a composite measure of parent's education and income there were no significant differences in global RT and the numerical advantage favored the monolingual by about 70 ms.

If mismatches in SES, as Morton and Harper suggest, can produce spurious bilingual advantages, then it follows that they could also conceal genuine differences.

[8] Immigrant status might influence EP in either direction because Canadian national averages show that immigrants tend to have more education, but lower incomes.

To guard against this possibility SES was evaluated using parent's educational level (PED). The PED score was coded (1 to 6) based on the participant's response to the following item on our questionnaire. *Circle the number that best describes the highest educational level obtained by your most highly educated parent: (1) did not graduate from high school, (2) graduated from high school, (3) attended college, but did not earn a degree, (4) earned an associate of arts or other 2-year degree, (5) earned a bachelor's degree, and (6) earned a graduate or professional degree that required additional education beyond a bachelor's degree.* The mean PED for monolinguals (4.2) is significantly greater that that for bilinguals (3.4), $t(267) = 4.30$, $p < .001$, SE = .19. However, in the combined Simon data there is no correlation between the size of the Simon effect and PED, $r = +.042$, $p = .698$.

The near zero correlation makes it highly unlikely that our failure to observe bilingual advantages in the Simon task were due to these differences in PED, but just to make sure, we precisely matched 90 monolinguals to 90 bilinguals on PED scores. For each level (e.g. graduated with a 4-year degree) we took all the participants from the smaller group and randomly selected the same n from the larger group. For these matched groups the only significant effect was congruency. As show in Table 4 neither the main effect of group, nor the Group $\times$ Congruency interaction was significant. Thus, there is no support at all, for the otherwise reasonable conjecture, that a bilingual advantage in our Simon data may have been concealed by differences in SES.

#### 3.1.3. Summary of Simon results

There is always a highly significant Simon effect (+32 ms). Study 3 and the combined results show a significant Group $\times$ Congruency interaction. However, the pattern is consistent with a bilingual disadvantage in inhibitory control. Although statistically significant the effect size is miniscule. The significant interaction is also observed for high-fluency bilinguals, but is not significant when the groups are matched on PED. There is never a main effect of groups. Thus, there is no evidence for a bilingual advantage in either inhibitory control or monitoring.

### 3.2. Flanker task

The same participant eliminated from the Simon analyses in Study 3 was eliminated in the Flanker analyses because her mean RT was more then 400 ms longer than the next slowest participant. The trimmed correct RTs were analyzed using a mixed designed ANOVA with group (bilingual versus monolingual), congruency (congruent versus incongruent), and block as factors. The inclusion of blocks is essential given that Costa et al. (2008) report bilingual advantages in inhibitory control (the magnitude of the flanker effect) only in Blocks 1 and 2 and Costa et al. (2009) report an advantage only in Block 1.

As shown in Table 3 the main effect of congruency was highly significant. The magnitude of the flanker effect (+96 ms) is very similar to those reported in the Costa studies using similar proportions of incongruent trials. In contrast to the results of the Costa studies the three-way interaction was not significant, $F(2,204) = 0.32$, $p = .73$. Table 4 shows the analysis of the flanker effect for each separate block and the combined data. There are no trends for an early bilingual advantage and, in fact, each block shows a very small and non-significant bilingual disadvantage.

The Costa studies also reported bilingual advantages in global RT. In contrast, there was no main effect of groups in our flanker study in the analyses of individual blocks or in the combined data. As shown in Table 4 the small global RT differences actually favored the monolinguals. The overall proportion correct in the flanker task was .987. Thus, it is not surprising that the results of the accuracy analyses shown in Table 5 are all non-significant.

#### 3.2.1. Flanker effects in highly-fluent bilinguals

The flanker task was only used in Study 3 and consequently there is no combined data to analyze. However, there were 55 monolinguals and 49 bilinguals in the flanker analysis reported above and, for completeness, one can also select subsets of the Study 3 that compare the best bilinguals to the most extreme monolinguals. Using the same stringent criteria yields subsets of 28 bilinguals and 31 monolinguals in the flanker task. As shown in Table 4 these subsets have the same pattern as the full analysis. There is a significant effect of congruency, but neither a significant effect of group nor a

**Table 6**
Language-group differences in RT and proportion correct in antisaccade task.

| LG | *N* | Neutral trials | | | | | Antisaccade trials | | | | | Antisaccade effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | RT/PC | SE | Adv | *t* | *p* | RT/PC | SE | Adv | *t* | *p* | Diff | Adv | *F* | *p* | $\eta^2$ |
| *Reaction time* | | | | | | | | | | | | | | | | |
| B | 35 | 530 | 23 | −42 | −1.48 | .144 | 567 | 20 | −34 | −1.29 | .199 | +37 | +8 | 0.15 | .701 | .002 |
| M | 45 | 488 | 17 | | | | 533 | 18 | | | | +45 | | | | |
| *Proportion correct* | | | | | | | | | | | | | | | | |
| B | 35 | .941 | .009 | −0.023 | +1.83 | .071 | .948 | .006 | −.008 | −0.89 | .376 | −.007 | +.015 | 1.40 | .240 | .018 |
| M | 45 | .964 | .008 | | | | .956 | .006 | | | | +.008 | | | | |

*Note*: PC = proportion correct; LG = language group; Adv = bilingual advantage; Diff = antisaccade trials − neutral trials.

Group × Congruency interaction. There is no support for the conjecture that the most highly fluent and balanced bilinguals in our sample enjoy an advantage in either inhibitory control or monitoring.

### *3.2.2. Flanker effects in groups matched on PED*

The same matching procedure described above for the combined data was applied to the Study 3 participants in order to precisely match the groups on PED. This produced subsets of 40 bilinguals and 40 monolinguals. There is a significant effect of congruency, but no significant effect of group, nor a Group × Congruency interaction (Table 4). Thus, there is no support for the possibility that the absence of bilingual advantages in flanker interference is due to differences in PED.

### *3.2.3. Group differences in alerting or orienting*

With respect to the other attentional networks separate analyses were conducted to test for possible difference in alerting and orienting effects. Alerting effects are computed by subtracting the double-cue trials from the no-cue trials. An ANOVA on the RT data showed a main effect of cue type, $F(1,102) = 118.75$, $p = .000$, MSE = 352.11, and partial $\eta^2 = .54$. The double-cue reduces temporal uncertainty and produces RTs that are 29 ms faster than the no-cue trials. However, the alerting effect is nearly the same for both groups and the Group × Cue Type is not significant, $F(1,102) = 0.36$, $p = .550$, MSE = 352.11, and partial $\eta^2 = .00$. Orienting effects are computed by subtracting the mean for spatial-cue trials from the central-cue trials. There was a significant main effect of orienting (22 ms), $F(1,102) = 78.26$, $p = .000$, MSE = 336.42, and partial $\eta^2 = .43$; but neither a main effect of group, ($F(1,102) = 0.21$, $p = .644$, MSE = 14,491.86, partial $\eta^2 = .00$; nor a Group × Cue Type interaction, $F(1,102) = 0.14$, $p = .706$, MSE = 336.42, and partial $\eta^2 = .00$. Thus both groups take advantage of cues that reduce temporal or spatial uncertainty, but neither group does so more effectively. Costa et al. did report a bilingual advantage in alerting in 2008, but not in 2009. There has never been a bilingual advantage reported for orienting.

### *3.2.4. Sequential dependencies (shifting costs) in the flanker task*

Sequential dependencies can occur when congruent (C) and incongruent (I) trials are mixed together. The congruency of the current trial is either the same as the previous trial (represented as cC or iI) or different (cI or iC). Costa et al. (2008) compared the different trials that require shifting[9] (cI and iC) to no-shift trials (cC and iI) and reported a Trial Type × Group interaction that showed a larger cost of shifting for monolinguals compared to bilinguals.

Consistent with the analysis used by Costa et al. a mixed ANOVA was conducted on our flanker RT data with groups and shifting (shift trials versus no-shift trials) as factors. The main effect of shifting was significant, $F(1,102) = 46.26$, $p = .000$, MSE = 238.96, and partial $\eta^2 = .31$. The overall shifting cost was +14.6 ms. Neither group, $F(1,102) = 0.36$, $p = .548$, MSE = 13,916.74, and partial $\eta^2 = .00$; nor the Group × Shifting interaction were significant, $F(1,102) = 1.48$, $p = .227$, MSE = 238.96, and partial

[9] Costa et al. refer to these as "switch" trials and it could be the case that switching contexts in a flanker task shares processing with switching tasks in the color-shape switching task. However, this relationship has not been empirically demonstrated and for clarity we will using "shifting" for changing context and "switching" for changing tasks.

**Table 7**
Language-group RT differences in the switching task.

| Data set | LG | *N* | Pure | | Repeat | | Switch | | Mixing costs (monitoring) | | | | | Switching costs | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | RT | SE | RT | SE | RT | SE | Diff | Adv | *F* | *p* | η² | Diff | Adv | *F* | *p* | η² |
| Study 1 | B | 30 | 553 | 21 | 877 | 61 | 1077 | 79 | +324 | −14 | 0.05 | .819 | .001 | +200 | +2 | 0.00 | .963 | .000 |
| | M | 44 | 504 | 17 | 814 | 50 | 1016 | 66 | +310 | | | | | +202 | | | | |
| Study 2 | B | 31 | 531 | 25 | 796 | 38 | 1013 | 46 | +265 | −13 | 0.08 | .781 | .001 | +217 | −31 | 1.37 | .245 | .017 |
| | M | 49 | 538 | 20 | 790 | 30 | 976 | 37 | +252 | | | | | +186 | | | | |
| Study 3 | B | 48 | 610 | 28 | 838 | 38 | 1072 | 52 | +228 | +2 | 0.00 | .955 | .000 | +234 | −7 | 0.21 | .649 | .002 |
| | M | 51 | 573 | 27 | 803 | 37 | 1030 | 51 | +230 | | | | | +227 | | | | |
| All | B | 109 | 572 | 15 | 837 | 26 | 1060 | 34 | +265 | −3 | 0.01 | .919 | .000 | +223 | −12 | 0.06 | .806 | .000 |
| | M | 144 | 540 | 13 | 802 | 23 | 1013 | 30 | +262 | | | | | +211 | | | | |
| High F | B | 55 | 559 | 20 | 859 | 39 | 1094 | 47 | +300 | −32 | 0.56 | .454 | .004 | +235 | −41 | 3.34 | .070 | .025 |
| | M | 77 | 547 | 17 | 815 | 33 | 1009 | 40 | +268 | | | | | +194 | | | | |
| PED M | B | 83 | 581 | 20 | 869 | 32 | 1090 | 37 | +288 | −25 | 0.37 | .545 | .002 | +221 | −27 | 2.03 | .157 | .012 |
| | M | 83 | 557 | 20 | 820 | 32 | 1014 | 37 | +263 | | | | | +194 | | | | |

*Note*: B = bilingual, M = monolingual; High F = high fluency subset; PED M = PED matched subset. Diff for mixing costs is RT repeat − RT Pure. Diff for switching costs is RT switch − RT switch. Adv = bilingual advantage.

$\eta^2 = .01$. In summary, the present results do not replicate Costa et al.'s report of a larger shifting costs for monolinguals compared to bilinguals.

### 3.3. Antisaccade task

The most straightforward test for a bilingual advantage in inhibitory control is to simply compare speed and accuracy in the antisaccade condition. As shown in Table 6 the differences in mean RT favor monolinguals, but the difference is non-significant. The proportion of correct responses was almost identical for the two groups.

Another and arguably better measure of inhibitory control is the difference in performance between the centered block and antisaccade block as the subtraction compensates for individual differences in processes not requiring cognitive control. A mixed design ANOVA with group (bilingual versus monolingual) and distraction (centered versus antisaccade) as factors was used to test for a bilingual advantage with this indicator. The means, standard errors, and critical test statistics for the parallel analyses of speed and accuracy are also shown in Table 6. For the analysis of trimmed RT scores the antisaccade effect is 8 ms less for bilinguals compared to monolinguals, but the Group × Trial Type interaction is not significant. The accuracy measure also shows a small bilingual advantage that is non-significant.[10]

### 3.4. Switching task

About six percent of the participants found the switching task very challenging. The overall proportion correct for 9 bilinguals and 5 monolinguals in the mixed block (where there are four response alternatives) fell between .35 and .73 and their data were not included in the analyses. An additional bilingual was unusually slow in the pure block with a mean of 1571 ms compared to the next slowest mean of 884 ms. The data for these participants were not included in the analyses reported in detail below.[11]

[10] Three related concerns can be raised about our instantiation of the antisaccade task. First, the main effect of distraction (41 ms) in the RT analysis, while significant, had a much smaller effect size (partial $\eta^2 = .192$) compared to either the Simon (partial $\eta^2 = .692$) or flanker (partial $\eta^2 = .829$) task. Furthermore, there was simply no difference at all in proportion correct between the centered ($m = .953$) and antisaccade ($m = .952$) blocks. The opposite field distractor in the antisaccade condition usually generates a substantial error rate. For example, the mean proportion correct in Kane et al.'s antisaccade condition was .636. Given these concerns in Studies 2 and 3 we replaced the antisaccade task with other tasks.

[11] All of the RT analyses evaluating mixing costs and switching costs in each of the three studies and in the combined data were also run without eliminating any participants for performance reasons. The analyses produced exactly the same pattern of results (i.e., a main effect of trial type, no main effect of group, and no Group × Trial Type interaction.

**Table 8**
Language-group differences in proportion correct the switching task.

| Data set | LG | *N* | Pure | | Repeat | | Switch | | Mixing costs (monitoring) | | | | | Switching costs | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | PC | SE | PC | SE | PC | SE | Diff | Adv | *F* | *p* | $\eta^2$ | Diff | Adv | *F* | *p* | $\eta^2$ |
| Study 1 | B | 30 | .956 | .010 | .954 | .009 | .914 | .040 | +.002 | −.002 | 0.04 | .836 | .001 | .040 | +.001 | 0.05 | .820 | .001 |
| | M | 44 | .969 | .008 | .969 | .007 | .928 | .041 | .000 | | | | | .041 | | | | |
| Study 2 | B | 31 | .938 | .016 | .927 | .012 | .898 | .029 | +.011 | −.020 | 0.06 | .440 | .008 | .029 | +.011 | 1.11 | .295 | .014 |
| | M | 49 | .945 | .013 | .954 | .010 | .914 | .040 | −.009 | | | | | .040 | | | | |
| Study 3 | B | 48 | .947 | .009 | .970 | .004 | .948 | .022 | −.023 | +.013 | 1.17 | .282 | .012 | .022 | .000 | 0.02 | .896 | .000 |
| | M | 51 | .966 | .009 | .976 | .004 | .954 | .022 | −.010 | | | | | .022 | | | | |
| All | B | 109 | .947 | .007 | .953 | .005 | .924 | .009 | −.006 | .000 | 0.00 | .993 | .000 | .029 | +.005 | 0.99 | .320 | .004 |
| | M | 144 | .960 | .006 | .966 | .004 | .932 | .005 | −.006 | | | | | .034 | | | | |
| High F | B | 55 | .952 | .011 | .951 | .008 | .928 | .008 | +.001 | −.011 | 0.57 | .450 | .004 | .023 | +.011 | 3.62 | .059 | .027 |
| | M | 77 | .954 | .009 | .965 | .007 | .931 | .013 | −.012 | | | | | .034 | | | | |
| PED M | B | 83 | .948 | .007 | .953 | .006 | .929 | .007 | −.005 | +.002 | 0.02 | .904 | .000 | .024 | −.002 | 0.05 | .818 | .000 |
| | M | 83 | .960 | .007 | .963 | .006 | .941 | .007 | −.003 | | | | | .022 | | | | |

*Note*: B = bilingual, M = monolingual; high F = high fluency subset; PED M = PED matched subset. Diff for mixing costs is PC repeat – PC pure. Diff for switching costs is PC repeat – PC Switch. Adv = bilingual advantage.

Switching tasks are typically used to derive measures of both mixing costs (an indicator of the monitoring component) and switching costs (an indicator of the switching component). Mixing costs were analyzed with a mixed design ANOVA including group (bilingual versus monolingual) and trial type (pure-block trials versus the repeat trials in the mixed-block). The means, standard errors, critical test statistics, and effect sizes for both speed and accuracy are shown in Tables 7 and 8, respectively. In each of the three studies and in the combined data the group differences in mixing costs were very small and nonsignificant.

Switching costs were analyzed with a mixed design ANOVA including group and trial type (repeat trials versus switch trials in the mixed block). None of the group differences approached significance and the largest individual difference (−31 ms in Study 2) trends toward a bilingual disadvantage. As shown in Table 3 the absence of bilingual advantages cannot be attributed to atypically small costs of either mixing (overall mean = +268 ms) or switching (+234 ms).

### *3.4.1. Mixing and switching costs for highly-fluent bilinguals*

Using the same criteria described for the Simon task, subsets of highly fluent bilinguals and monolinguals with limited exposure to any L2 were selected from the combined analyses of the switching task. This resulted in a set of 55 bilinguals and 77 monolinguals. For the analysis of mixing costs there was a significant main effect of mixing, but as shown in Table 7 no significant differences for group or for the Group × Mixing interaction. The same was true for the analysis of switching costs. These subsets yield the same ANOVA pattern for RT as the combined data and offer no support to the possibility that more stringent classification criteria might yield bilingual advantages in either mixing costs or switching costs.

### *3.4.2. Mixing and switching costs for PED matched bilinguals*

In the combined data the correlation between PED and mixing costs and between PED and switching costs is very near zero, $r(266) = -.005$ and $r(266) = +.014$, respectively. Nonetheless, analyses of mixing costs and switching costs were also conducted on the subsets of 83 bilinguals and 83 monolinguals that were precisely matched in terms of PED. As shown in Table 7 the pattern of results remained exactly the same as in the combined data. Thus, there is no evidence for bilingual advantages in either mixing or switching costs when groups are matched in terms of PED.

## *3.5. Post-hoc power analyses*

The G∗Power 3 tool (Faul, Erdfelder, Lang, & Buchner, 2007) was used to estimated the power to detect significant Group × Trial Type interactions in each of our tasks based on an alpha of .05, the obtained correlation between the repeated measures, the actual sample sizes, and the desire to detect a

**Table 9**
Step-wise regression predicting six indicators of EP from Study 3.

| Indicator | EP component | $R$ | $R^2$ | Predictor | Beta | $p$ |
|---|---|---|---|---|---|---|
| Simon effect | Inhibitory control | .371 | .137 | Raven's | −.23 | .013 |
| | | | | Music and homework | −.21 | .023 |
| | | | | Multilingualism | +.20 | .031 |
| Flanker effect | Inhibitory control | .288 | .083 | Multilingualism | +.22 | .022 |
| | | | | Team sports | −.21 | .030 |
| Switching costs | Switching | .327 | .107 | Team sports | −.33 | .001 |
| Mixing costs (switching) | Monitoring | .166 | .028 | Raven's | −.17 | .085 |
| Simon global RT | Monitoring | .311 | .096 | Team sports | −.22 | .001 |
| Flanker global RT | Monitoring | .390 | .152 | Raven's | −.31 | .001 |
| | | | | Team sports | −.22 | .019 |

*Note*: The step-wise criteria for probability of F-to-enter was .05 and to remove was .10, except for Mixing Costs where the criteria needed to be relaxed to .10 and .15 to enter any predictors. Beta = standardized beta coefficient.

very small effect size (viz., one based on a partial $\eta^2$ = .10). Because our *N*'s are quite large it is not surprising that the post hoc estimates of power are very good, greater than .999 for our combined Simon or switching data. At the low end the antisaccade task has estimated power equal to .892. To further illustrate the sensitivity we have in rejecting the null hypothesis note that the Group × Congruency interaction or the Simon effect was significant both in Study 3 for a 10 ms bilingual disadvantage with an associated partial $\eta^2$ of .047 and in the combined data for a 5 ms bilingual disadvantage with a partial $\eta^2$ of .014.

### *3.6. Raven's advanced progressive matrices test*

The mean number of correct responses was 8.9 for the bilinguals and 8.5 for the monolinguals. The difference was not significant, $t(100) = 0.95$, $p = .345$, SE = .40. Thus, the two groups tested in Study 3 do not differ with respect to nonverbal intelligence and the nonsignificant difference actually favors the bilinguals.

### *3.7. Predicting indicators of EP from demographic factors*

Hilchey and Klein observe that the field would profit greatly from the use of a comprehensive survey that associated a host of life experiences that might be associated with executive control. In Study 3 we expanded our questionnaire in the direction advocated by Hilchey and Klein and explored 11 predictors of executive processing. The first three were objective measures: score on the Raven's Advanced Matrices test, PED, and chronological age. A fourth predictor was multilingualism.[12] Several factors focused on the frequency of playing videogames or the frequency of engaging in different types of multitasking. Another factor focused on the ability to excel in team sports and a final factor on attitudes toward multitasking. All 11 factors were used as predictors in a stepwise multiple regression for each of the indicators of EP available in Study 3. The models, together with the standardized beta coefficients, are shown in Table 9.

For the six indicators used as outcome variables, four of the models result in significant negative beta coefficients for excelling at team sports. This is consistent with the expectation that either playing team sports enhances components of EP or that those with high genetic ability in EP are recruited and nurtured to play team sports. In three models high Raven's scores are significantly associated with better measures of EP. Note that PED never enters any of the models and this is consistent with the

[12] Multilingualism is a multivalued measure of what is usually treated as a bilingual versus monolingual dichotomy. It was operationally defined as the summed rated proficiency across all languages (e.g., a bilingual who rates her proficiency in English as 7 and in Cantonese as 6 has a multilingualism score of 13; a monolingual who rates her proficiency in English as 6 and in German as 1 has a score of 7).

interpretation that individual differences in PED are not related to performance differences in our sample. The surprising outcomes were the +.20 and +.22 standardized coefficient for multilingualism in predicting the magnitude of the Simon and flanker effects, respectively. This shows that greater summed proficiencies across multiple languages are associated with lower levels of inhibitory control in these two tasks. Thus, when other variables are taken into account through regression analyses there are no language-group differences on indicators of monitoring or switching and a bilingual disadvantage on two measures of inhibitory control.

## 4. Implications for bilingual advantages in three components of EP

### 4.1. Inhibitory control

In their 2011 review of 31 experiments Hilchey & Klein concluded that "The absence of a bilingual advantage in..." children and young adults... "is simply inconsistent with the proposal that bilingualism has a general positive effect on inhibitory control processes" p. 629. Two reports by Kousaie and Phillips (2012a, b) reinforce this conclusion. The 2012a study uses a multiple-task approach similar to ours and found no behavioral differences between groups of young adults in the Stroop, Simon, or flanker tasks (i.e., 0 group differences out of three tests). The 2012b study used both young adults and older adults and found no differences in the magnitude of Stroop interference (i.e. 0 group differences out of two tests). A similar study by Humphrey and Valian (2012) using the Simon and flanker tasks follows the same pattern. Four different groups of multilinguals (lifelong balanced bilinguals, late balanced bilinguals whose native language is English, late balanced bilinguals whose native language is not English, and trilinguals) show Simon and flanker effects statistically equivalent to a group of English monolinguals (i.e., 0 bilingual advantages out of eight tests). When all of these new findings are added to our three Simon experiments, our flanker experiment, and our antisaccade experiment these results sum to 17 new tests yielding no advantages and one that shows a bilingual disadvantage.

### 4.2. Monitoring for conflict or for the possible need to change goals or tasks

In contrast to their conclusions regarding inhibitory control, Hilchey and Klein conclude that: "In young adults, the global RT advantage is detected ubiquitously on spatial Stroop and flanker interference tasks, though seemingly not in the Simon task" p. 645. There were no global RT advantages in any of our three Simon experiments and no advantage in the flanker experiment. In fact, the nonsignificant differences usually favor the monolingual group. Finally, there was no bilingual advantage in "mixing costs" in any of our three color-shape switching experiments and "mixing costs" are also assumed to reflect the need for monitoring. These null results are echoed by Kousaie and Phillips (2012a) who report no significant differences in global RT for the Stroop, Simon, or flanker tasks. Likewise, Humphrey and Valian compared the global RT in both the Simon and flanker tasks for each of their four multilingual groups in comparison to the group of English monolinguals. Seven of these comparisons yielded no language-group differences and the trilinguals actually showed a disadvantage compared to the monolinguals. These new studies surely soften Hilchey and Klein's conclusion regarding the ubiquity of global RT advantages. This is especially true if one focuses on the findings with young adults as Hilchey and Klein's review found only seven cases testing young adults: four showed significant bilingual advantages in global RT and three simply showed a nonsignificant numerical advantage. In contrast, the new results reported in the present, Kousaie and Phillips, and Humphrey and Valian sum to 0 global RT bilingual advantages out of the 18 new tests, with 14 of the 18 showing a numerical advantage for the monolinguals. Finally, as discussed more fully in Section 4.4 Prior and MacWhinney (2010) also found no global RT advantages in either a Simon or flanker task.

### 4.3. Switching-costs

There have been several recent reports of a bilingual advantage in color-shape switching tasks. For example, Garbin et al. (2010) report a bilingual advantage in a switching task that compared 21 Spanish monolinguals to 19 Spanish–Catalan bilinguals. Interestingly, the bilinguals activated left hemisphere networks thought to underlie language control. An empirical peculiarity of the Garbin et al. RT data is that the bilinguals show no switching costs at all. This oddity could be partially or completely responsible for the bilingual advantage in switching costs.

Prior and MacWhinney (2010) also report a bilingual advantage in switching costs. Mean switching costs for a group of 45 monolinguals was 206 ms compared to only 144 ms for a group of 45 bilinguals. Besides English, the bilinguals spoke a variety of different other languages, but for 24 of them the other language was either Mandarin or Korean. Robust mixing effects (314 ms) were also reported, but there were no group differences.

Prior and Gollan (2011) used the same color-shape switching task to test for advantages in mixing costs or switching for English monolinguals ($n = 47$), Mandarin–English bilinguals, ($n = 43$) and Spanish–English bilinguals ($n = 41$). After controlling for differences in response speed and PED there was a bilingual advantage in switching costs for the Spanish–English bilinguals, but not for the Mandarin–English bilinguals. The Spanish–English group reported that they switched languages more frequently than the Mandarin–English bilinguals and, not surprisingly, had significantly smaller switching costs in language-switching task. Prior and Gollan conclude that there is "... *a tight link between language-switching and general switching ability, and that certain aspects of bilingual language use, which are not universal to all bilinguals, introduce the advantage*" p. 6. Like Prior and MacWhinney, they found no language-group differences in mixing costs.

To this pattern we show no differences in three separate studies testing a total of 109 bilinguals. Given Prior and Gollan's evidence implicating the importance of frequency-of-switching languages it is informative that our bilinguals overwhelmingly report that they use both languages every day and switch every day. Furthermore, the mean percentage of current daily use of English reported by Prior and Gollan for their Spanish–English bilinguals (84.6%) is higher than the percentage of English use reported by our bilinguals (72%). From this one might infer that our bilinguals switch as often, if not more often, than Prior and Gollan's Spanish–English bilinguals. Thus, Prior and Gollan's reasonable hypothesis that the presence or magnitude of a switching advantage is determined by the frequency of language switching appears to be inconsistent with our data.

These inconsistencies are vexing and not easily set aside or understood. Tare and Linck (2011) perhaps shed some additional light with their report that no bilingual switching-costs were observed for a set of 35 bilinguals who were individually matched from a pool of more than 1100 monolinguals. The technique of propensity score matching was used to statistically equate the two groups to the greatest extent possible on demographic factors such as age, education, and pay grade, as well as measures of general intelligence and verbal ability. With these controls in place there were no group differences in switching costs and there was a bilingual disadvantage on a measure of inhibitory control. Tare and Linck conclude "*These results suggest that factors other than bilingualism* per se *may be driving any purported bilingual cognitive advantages*" p. 132.

### 4.4. Patterns of language-group differences across the components of EP

In Hilchey and Klein's (2011) review of the bilingual advantage in EP the most perplexing aspect of the empirical pattern is that the bilingual advantage in nonlinguistic interference tasks is usually equivalent for congruent and incongruent trials. This pattern could only occur if there was an advantage in monitoring that applies equally to both types of trials and, at the same time, there was no advantage in inhibitory control. This provokes Hilchey and Klein to observe that it is somewhat ironic that the global RT advantages occur "...*so long as the* [overall] *task entails some level of conflict*" p. 634.

Against the backdrop of Green's inhibitory control model it is surprising that the ubiquitous inhibitory control exercised by bilinguals in making lexical selections might turn out to have no influence on EP, but that the monitoring demands of managing two languages does. This is particularly true if the special experience in monitoring for bilinguals (as suggested by Costa et al. in the earlier quotation) simply involves monitoring for a new interlocutor who doesn't speak or prefer the language of the ongoing conversation. This type of monitoring does not seem to have the same tempo and intensity in comparison to the need to inhibit the context-inappropriate translation equivalent at each step in the production of a sentence.

**Table 10**
Reliability of indicators of EP components.

| EP component | Indicator | Study | *N* | SBP | *p* |
|---|---|---|---|---|---|
| Inhibitory control | Antisaccade trials | 1 | 85 | .942 | .000 |
| | Antisaccade effect | 1 | 85 | .900 | .000 |
| | Simon effect | 1–3 | 284 | .448 | .000 |
| | Flanker effect | 3 | 108 | .942 | .000 |
| Monitoring | Simon global RT | 1–3 | 284 | .932 | .000 |
| | Simon mixing | 2–3 | 201 | .952 | .000 |
| | Flanker global RT | 3 | 108 | .944 | .000 |
| | Flanker mixing | 3 | 108 | .944 | .000 |
| | Mixing costs (switching) | 1–3 | 267 | .912 | .000 |
| Switching | Switching costs | 1–3 | 267 | .734 | .000 |

*Note*: SBP = Spearman–Brown Prophecy formula.

The participants in the Prior and MacWhinney switching study also participated in a Simon task and flanker task that were not reported in the article. However, there was no bilingual advantage with respect to the magnitude of the interference effect in either task, nor were there any group differences in global RT (A. Prior, personal communication). Thus, the full set of findings reveals the surprising outcome that bilingual advantages in switching do not co-occur with similar advantages in indicators of either inhibitory control or monitoring. The dissociation between switching advantages and monitoring advantages is also observed within the color-shape switching task. That is, when switching advantages did occur (Prior & MacWhinney and the Spanish–English bilinguals in Prior and Gollan) they are never accompanied by advantages in mixing costs.

These failures to find bilingual advantages across components of EP seem inconsistent with expectations derived from emphasizing the unity of EP (Miyake & Friedman, 2012) or advocating for a holistic view of EP (Bialystok, 2011). Friedman and Miyake (2004) suggest that one reason for the lack of separability of the inhibition component is that all interference-related processes share the requirement of actively maintaining task goals (usually in the face of interference from external stimuli) and that this basic ability is necessary for all components of EP. If this line of reasoning is on track, then one would expect bilingual advantages across components of EP to frequently co-occur. In fact, most reports of a bilingual advantage involve a single component in a single task.

## 5. Correlations between indicators of the same executive process

The argument advanced in the introduction was that a coherent demonstration of a bilingual advantage in inhibitory control, or any other EP component, would show the advantage in two different tasks and that the markers for the two tasks would correlate with each other. If the two indicators did not correlate with one another, then the bilingual advantages were likely to be task specific rather than providing evidence of a shared and domain-general ability. Although the present set of three studies provide multiple markers of both inhibitory control and monitoring there were no bilingual advantages. Nonetheless, it is informative to explore the degree to which individual differences in one indicator predict other indicators of the same component.

All of the indicators, save global RT, are based on subtractive logic. For example, markers of inhibitory control are differences between congruent and incongruent conditions. The advantage of using subtractions is that they take out individual differences in the speed of performing easy choice tasks that presumably do not require or recruit EP. The disadvantage of using differences is that owing to the variance-sum law they are notoriously noisy measures. Consequently it is important to assess the reliability of each indicator. In each task the experimental conditions were balanced and randomized within blocks. Table 10 shows the reliability of each of our indicators based on block-to-block correlations adjusted by the Spearman–Brown prophecy formula.[13] As shown in Table 9 all the correlations

[13] This is consistent with the reliability measures for the antisaccade and color-shape switching tasks reported by Friedman et al. (2008) in their seminal article on individual differences in EP. Because our Simon and color-shape switching tasks used three blocks the reliability correlations shown in Table 6 for these two tasks are the mean of the three pairs of blocks.

**Table 11**
Cross-task correlations for components of EP.

| EP component | Indicator 1 | Indicator 2 | Study | *N* | *r* | *p* |
|---|---|---|---|---|---|---|
| Inhibitory control | Antisaccade trials | Simon effect | 1 | 86 | −0.12 | .911 |
| | Simon effect | Flanker effect | 3 | 107 | −0.01 | .959 |
| Monitoring | Simon mixing | Mixing costs (switching) | 2 | 87 | 0.00 | .998 |
| | Simon mixing | Mixing costs (switching) | 3 | 108 | −0.04 | .677 |
| | Flanker mixing | Mixing costs (switching) | 3 | 106 | −0.04 | .758 |
| | Simon mixing | Flanker mixing | 3 | 107 | +0.10 | .310 |
| | Simon global RT | Mixing costs (switching) | 1 | 82 | +0.26 | .018 |
| | Simon global RT | Mixing costs (switching) | 2 | 87 | +0.26 | .015 |
| | Simon global RT | Mixing costs (switching) | 3 | 109 | +0.22 | .020 |
| | Flanker global RT | Mixing costs (switching) | 3 | 107 | +0.23 | .017 |
| | Simon global RT | Flanker global RT | 3 | 107 | +0.73 | .000 |

are greater than .90 [with the exception of the Simon effect (+0.448) and switching effect (+0.734)] and therefore appear to be highly reliable.

### *5.1. No convergent validity in inhibitory control*

Flanker effects and Simon effects are often used as more-or-less equivalent markers of inhibitory control (e.g., Mullane, Corkum, Klein, & McLaughlin, 2009). However, as shown in Table 11 the correlation between these two effects in Study 3 is $r$ = -0.01. There is no hint in this tiny negative correlation that both effects are influenced by a domain-general contribution to inhibitory control. The complete absence of an association between these two interference effects has important implications for the standard practice of using these tasks as possible indicators of a bilingual advantage in inhibitory control. If there is no association, then it logically follows that, at best, only one of the two tasks requires executive control. The other alternative is that the inhibitory control exercised in both tasks is completely task dependent. This lack of convergence undermines the confidence that a bilingual advantage in the magnitude of the flanker or Simon effect signals differences in domain-independent inhibitory control. Similarly, the correlation between the magnitude of the Simon effect and antisaccade effect in Study 1 was near zero ($r = -.012$).

The cross-task correlations reported by Friedman and Miyake (2004) for inhibition always involved pairs of indicators different from ours. They do report significant correlations between antisaccade accuracy and Stroop (+.23) and between flanker and Stroop (+.18), but no correlation between antisaccade accuracy and flanker effects (+0.04). In contrast, Unsworth and colleagues have consistently found a small but significant correlation between antisaccade accuracy and flanker interference (Unsworth, McMillan, Brewer, & Spillers, 2012; Unsworth & Spillers, 2010; Unsworth, Spillers, & Brewer, 2009). The present study does not provide an opportunity for replication as the antisaccade and flanker tasks were given to separate samples of participants. Given that any pair of tasks formed from the set of antisaccade Stroop, and flanker sometimes yields a significant correlation, it appears that the Simon task may be the primary cause of near-zero correlations between measures of inhibitory control. Lest the Simon task be cast in the role of a scapegoat it is worth emphasizing that correlations among measures found with the Stroop, flanker, and antisaccade task are always weak and that conflict resolution mechanisms appear to be far more task specific then shared.

### *5.2. Weak convergent validity in the executive process of monitoring*

Across the flanker, Simon, and color-shape switching task there are three difference scores that are used as indicators of the ability to engage in effective monitoring. For the color-shape switching task (Studies 1–3) the difference between RTs in the pure blocks and the repeat-trials (of the mixed bocks) defines *mixing costs*. This difference is assumed to reflect the cost of having to prepare to switch tasks when, in fact, no switch is required. A similar indicator is available for both the Simon (Studies 2 and

3) and flanker (Study 3) task by subtracting mean RTs from a pure block of no-conflict trials from the congruent trials in the conflict block. These are referred to as *mixing Simon* and *mixing flanker*, respectively. As shown in Table 11 the four cross-task correlations involving these difference scores are all less than .10 and nonsignificant. Because these difference scores are highly reliable (Table 10) the low cross-task correlations cannot be dismissed on grounds of lack of sensitivity.

To this point it would appear that the mixing costs computed from different tasks are not measuring a general and task-independent ability to monitor conflict or to prepare for a change in goals. The four correlations that include global RT as one of the two correlated measures and a difference score as the second measure produce a different pattern. As shown in Table 11 this set of cross-task correlations clusters between 0.22 and 0.26 and, given the large Ns, are all highly significant. On the one hand this is consistent with the assumption that global RT provides a window into the monitoring component of EP. Alternatively, the small but significant correlations involving global RT may occur only because the global RT measure fails to subtract out individual differences in processing speed that are unrelated to EP. The question of convergent validity for alternative measures of the monitoring component deserves additional investigation given the ambiguity of the present results and the fact that monitoring was not included as a separate latent variable in the Miyake and Friedman studies.

## 6. Evaluation of empirical evidence

The empirical focus of this article is on bilingual advantages in inhibitory control, monitoring, and switching obtained with young adults[14] engaged in nonlinguistic interference tasks. There are two perspectives on reconciling the reports of significant bilingual advantages with the many failures to replicate.

### *6.1. The bilingual advantages are real perspective*

One perspective attributes each failure to replicate to a critical methodological difference with respect to the task (e.g., proportion of incongruent trials) or the type of bilingual (e.g., frequency of language switching). Although further work may clarify the conditions that reliably lead to bilingual advantages the prospects are not encouraging. Reports of bilingual advantages are scattered across different tasks, measuring different components of EP, and testing different types of bilinguals. Although the amount of incremental progress in understanding the conditions that produce a bilingual advantage may be "*in the eyes of the beholder*" it is worthwhile to review a recent and initially promising attempt to identify important constraints, namely, Costa et al.'s (2009) "*Now you see it, now you don't*" article. Using a series of systematic, sophisticated, and well-designed experiments they found that having the right proportion of congruent trials was critical to obtaining the bilingual advantage in the flanker task. Furthermore, they found that the advantages are transient and appear in early blocks, but not later blocks. But in close methodological replications (the present Study 3 and Kousaie & Phillips, 2012a) no trace of a bilingual advantage was detected in any block for any of the indicators (flanker effect, global RT, shifting costs, orienting) showing bilingual advantages in Costa et al.

### *6.2. The bilingual advantages are artifacts perspective*

The alternative perspective on reconciling the empirical inconsistencies is to attribute the performance advantages, when they do occur, to factors other than bilingualism enhancing EP. Some significant differences are likely Type I errors. Several may reflect task-specific performance differences on measures that lack convergent validity. Others may be due to hidden demographic factors that were not matched as proposed by Morton and Harper (2007, 2009), Morton (2010), Morton and Carlson (in press), and Tare and Linck (2011).

Hilchey and Klein provide an extensive and insightful examination of the influential reports of bilingual advantages in nonlinguistic interference tasks. They highlight many aspects of these reports

[14] Given that the vast majority of our participants are young adults we remain agnostic with respect to the possibility that more coherent evidence may be forthcoming from older adults.

that make the results vulnerable to alternative interpretations. These include experiments where the interference effects (particularly for the monolingual group) were "*extraordinarily large*", language groups that were drawn from different countries and cultures, samples of only 10–15 participants per language group, and means based on as few as 14 trials per experimental condition.[15]

As Morton and Carlson (in press) point out the role of bilingualism in the development of EP is particularly difficult to isolate because: (a) early/native bilinguals are usually bicultural and (b) many factors known to guide and influence the development of EP will vary across cultures (e.g., parenting/caregiving styles, methods of formal schooling, discipline, emphasis on self-control, SES, educational level, and values). Carlson and Choi (2009) have reported a dramatic demonstration of this entanglement between culture and bilingualism. Using six different measures of EP (including ANT) they show significant bilingual advantages comparing a group of Korean–English bilinguals living in the United States to a "matched" sample of American monolinguals. However, the performance of the Korean–American bilinguals was undistinguishable from a third group of matched Korean monolinguals. This clearly questions the interpretation that the obtained group differences were due to bilingualism and strongly supports the view that cultural differences play an influential role in the development of EP.

The *bilingual advantages are artifacts* perspective does not rely solely on appeals to Type 1 errors, inadequately matched groups, and cultural differences. It also raises important and legitimate concerns that the obtained empirical advantages may be task specific and not indicative of superior EP. For example, since the interference effects in the Simon task do not correlate with those in the flanker task (or the Stroop task) the bilingual advantages in the flanker effect reported by Costa may be measuring task-specific inhibitory mechanisms, not differences in a domain-general ability in inhibitory control.

### *6.3. A plan for possible reconciliation*

To reiterate, the two opposing views are that either there are genuine bilingual advantages that happen to be quite elusive given our current understanding of how and why they develop; or that the performance advantages, when they occur, are due to causes unrelated to bilingualism enhancing EP. The first view questions why null results occur while the second asks what may have caused performance differences favoring bilinguals. In our view the evidence points in the direction of no genuine bilingual advantage in EP. But, we are open to new and compelling evidence that follows the protocol for the following hypothetical study: (1) identify the specific component(s) of executive processing that should be enhanced by managing two languages, (2) show a bilingual advantage in an indicator of that component across two different tasks, (3) show that the indicators correlate with one another and have some degree of convergent validity, (4) show no differences between the two groups on a pure block of easy choice-RT trials, (5) match the groups on SES and (6) minimize cultural differences between the groups.

## 7. Why might there be no bilingual advantages in executive processing?

The idea that a bilingual's language experience generalizes and enhances EP is very attractive. However, in the hindsight of the lack of coherent evidence favoring a bilingual advantage in EP, it is easy to see that there is a sequence of three assumptions that have to be sustained in order for this received story to be true.

### *7.1. Do bilinguals recruit sufficiently more executive control during language use?*

A first assumption is that the amount of EP recruited by bilinguals during language comprehension and production is greater than that employed by monolinguals because of the need to monitor the communication environment for changes that trigger a language switch, because of the switching itself, and because of the need to inhibit the translation equivalents in the other language. On the one hand, all three of these aspects of bilingualism have no direct counterpart for the monolingual. On the

[15] Readers concerned that the seminal findings are being brushed aside cavalierly are encouraged to consider pages 635 to 641 of the Hilchey and Klein review.

other hand, speaking any language appears to require substantial amounts of monitoring, switching, and inhibitory control. To provide just a few examples, conversational participants must monitor the environment for signals regarding turn-taking, misunderstandings, possible use of sarcasm, changes of topic, or changes in register contingent upon who enters or leaves the conversation. These lead to switches from speaker to listener, switches from one knowledge domain to another, and so forth. Although monolinguals do not need to suppress translation equivalents during production, they incessantly make word choices among semantically and syntactically activated candidates that include synonyms, hypernyms, and hyponyms. In addition monolinguals must use context to suppress the irrelevant meaning of homographs during comprehension. In summary, fluent bilinguals have additional needs for monitoring, switching, and inhibitory control, but these unique requirements may not be substantial enough to generate group differences in cognitive control.

### *7.2. Do processing differences inside the language module generalize?*

In Section 7.1, a plausible argument is advanced that the unique requirements of bilinguals may not be sufficient to generate bilingual advantages in EP. If managing two languages was known to be sufficient, then would this guarantee an enhancement of EP? No, not unless the monitoring, switching, and inhibitory control used during language production and comprehension uses the domain-independent executive controller. An alternative, of course, is that these functions are specialized within the language module (Fodor, 1983; Frazier, 1987) and are, therefore, specific to tasks that depend on linguistic representations.

### *7.3. Have most young adults already optimized their capacity for cognitive control?*

If it was certain that bilinguals engaged in significantly more monitoring, switching, and inhibitory control during language use AND if it was clearly the case that these were generalized executive processes that were employed (rather than being task specific); would these conditions necessarily generate bilingual advantages in EP? The answer to this is no if each individual's capacity for cognitive control is restricted to some asymptotic limit that is likely to be reached by most individuals. Independent of the number of languages spoken most of us fill our days pursuing goals and subgoals, attending to relevant sources, ignoring distraction, inhibiting inappropriate responses, switching goals, planning and sequencing behaviors, and monitoring our performance. It may be the case that a wide range of normal life activities leads a vast majority of individuals to reach their innately determined asymptote of skilled EP by the time they become adults. Both Bialystok et al. (2005) and Costa et al. (2009) have offered similar interpretations as to why it is difficult to reliably demonstrate bilingual advantages in young adults.

## 8. Conclusions

The research findings testing for bilingual advantages in EP do not provide coherent and compelling support for the hypothesis that the bilingual experience causes improved EP. Because individual studies tend to use only one task and use only one indicator for each EP component there is usually no test of convergent validity. Those studies that have used multiple tasks show no bilingual advantages and little or no convergent validity. Matching language-groups on factors that influence the development of EP is a serious challenge, and particularly difficult when the bilingual group is either from a different culture or bicultural. These prescriptions for a more ideal and comprehensive approach to confirming bilingual advantages in EP are admittedly daunting. They will also be unachievable if the special experiences of bilinguals truly do not cause enhancements in executive processes that are general and domain-independent.

## Acknowledgments

We thank the following members of the LACE (Language, Attention, & Cognitive Engineering) laboratory for their individual contributions to this project: Edgar Alcaine, Nick Alvarez, Tavi Alvarez, Kirstin Anderson, Olimpia Andrade, Jack Darrow, Frank Du, Eddie Ferrero, Lynne Freeman, Jenesis Imai, James Keenan, Jessica Koernke, Yunyun Liu, Brenda Mejia, Sati Morgan, Lindsay Pellichia, Oliver Sawi, Monique Toledo, and Carlos Urtecho.

## References


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

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3–11.

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology, 65*(4), 229–235.

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

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism, 9*, 103–119.

Carlson, S. M., & Choi, H. P. (2009, April). Bilingual and bicultural: Executive function in Korean and American children. In *Paper presented at the 2009 biennial meeting of the society for research in child development*, Denver, Colorado.

Collette, F., Van Der, L. M., Laureys, S., Delfiore, G., Degueldre, C., Luxen, A., et al (2005). Exploring the unity and diversity of the neural substrates of executive functioning. *Human Brain Mapping, 25*, 409–423.

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

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

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.

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.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: The MIT Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Erlbaum.

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.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*, 201–225.

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., et al (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *Neuroimage, 53*(4), 1272–1278.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). 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*(3), 594–615.

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

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for plasticity of executive control processes. *Psychonomic Bulletin & Review, 18*, 625–658.

Humphrey, A. D., & Valian, V. V. (2012). Multilingualism and cognitive control: Simon and flanker task performance in monolingual and multilingual young adults. In *53rd Annual meeting of the psychonomic society*, Minneapolis, MN.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169–183.

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.

Kousaie, S., & Phillips, N. A. (2012a). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research, 1446*, 71–90.

Kousaie, S., & Phillips, N. A. (2012b). Aging and bilingualism: Absence of a "bilingual advantage" in Stroop interference in a nonimmigrant sample. *The Quarterly Journal of Experimental Psychology, 65*(2), 356–369.

Lu, C.-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review, 2*(2), 174–207.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychology, 21*(1), 8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wagner, T. D. (2000). The unity and diversity of executive functions and their contribution to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Morton, J. B. (2010). Language, bilingualism, and executive functioning in early development. *Psychological Reports, 107*(3), 1–3.

Morton, J. B., & Carlson, S. (in press). In M. Hoskyn, G. Iarocci, & A. Young (Eds.), *The bilingual advantage: Evidence and alternative views*. Oxford University Press.

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science, 12*, 719–726.

Morton, J. B., & Harper, S. N. (2009). Bilinguals show an advantage in cognitive control: The question is why. *Developmental Science, 12*, 502–503.

Mullane, J. C., Corkum, P. V., Klein, R. M., & McLaughlin, E. (2009). Interference control in children with and without ADHD: A systematic review of flanker and Simon task performance. *Child Neuropsychology, 15*, 321–342.

Prior, A., & Gollan, T. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of International Neuropsychological Society, 17*, 1–10.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition, 13*, 253–262.

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..

Tare, M., & Linck, J. (2011). Exploring bilingual cognitive advantages when controlling for background variables. In *52nd Annual meeting of the psychonomic society*, Seattle, WA.

Unsworth, N., McMillan, B. D., Brewer, G. A., & Spillers, G. J. (2012). Everyday attention failures: An individual differences investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(6), 1765–1772.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*, 392–406.

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. *Psychology Science Quarterly, 51*(4), 388–402.