Are bilingual advantages dependent upon specific tasks or specific bilingual experiences?

Article *in* Journal of Cognitive Psychology · August 2014

DOI: 10.1080/20445911.2014.944914

3 authors:

Kenneth R. Paap
San Francisco State University

Hunter A. Myuz
New Mexico State University

Oliver M Sawi
University of Connecticut

Journal of Cognitive Psychology

ISSN: 2044-5911 (Print) 2044-592X (Online) Journal homepage: http://www.tandfonline.com/loi/pecp21

Are bilingual advantages dependent upon specific tasks or specific bilingual experiences?

Kenneth R. Paap, Hunter A. Johnson & Oliver Sawi

To cite this article: Kenneth R. Paap, Hunter A. Johnson & Oliver Sawi (2014) Are bilingual advantages dependent upon specific tasks or specific bilingual experiences?, Journal of Cognitive Psychology, 26:6, 615-639, DOI: 10.1080/20445911.2014.944914

To link to this article: http://dx.doi.org/10.1080/20445911.2014.944914

Published online: 08 Aug 2014.

Journal of Cognitive Psychology, 2014
Vol. 26, No. 6, 615–639, http://dx.doi.org/10.1080/20445911.2014.944914

Routledge
Taylor & Francis Group

Are bilingual advantages dependent upon specific tasks or specific bilingual experiences?

Kenneth R. Paap[1], Hunter A. Johnson[1], and Oliver Sawi[2]

[1]Department of Psychology, San Francisco State University, San Francisco, CA, USA
[2]Department of Psychology, University of Connecticut, Storrs, CN, USA

Paap and Greenberg concluded that there is no coherent evidence for bilingual advantages in executive processing. More optimistic researchers believe that the advantages may be restricted to certain types of bilinguals. Recent large-scale and lifespan investigations that tested highly fluent bilinguals from communities where the same two languages are spoken by most residents reported no bilingual advantages in any age group or in any of the tasks used to measure executive functioning. The present study takes a complementary approach by examining a sample that is quite homogeneous in terms of current life experiences, but heterogeneous in terms of its exposure to second languages. The composite database of 168 bilinguals and 216 monolinguals is used to explore for differences based on: (1) the age of acquiring a second language (L2), (2) the relative proficiency of an L2 and (3) the number of languages used. Across 12 different measures of executive function, derived from 4 different nonverbal tasks, there was no consistent evidence supporting the hypotheses that either early bilingualism, highly fluent balanced bilingualism, or trilingualism enhances inhibitory control, monitoring or switching. In fact, when statistically significant effects did occur, they more often disconfirmed than confirmed these hypotheses.

Keywords: Bilingualism; Bilingual advantage; Executive processing; Flanker task; Simon task; Switching.

Executive functions (EFs) consist of a set of general-purpose control processes believed to be central to the self-regulation of thoughts and behaviours and that are instrumental to accomplishing goals. The construct of EF is often viewed as a set of interrelated component processes all involving the prefrontal cortex with each component recruiting additional areas of cortical function. Miyake and Friedman (Friedman et al., 2008; Miyake & Friedman, 2012) reported evidence for three components of EF: updating, shifting and inhibiting. Confirmatory factor analyses (CFAs) were based on measures from three different tasks for each of the three hypothesised components. At the higher level the three latent variables correlate with one another and this is consistent with the assumption that each contributes to a common EF. When the same data are reanalyzed with a second-order CFA where the three latent variables are nested under a common EF latent variable, the nine observed measures all load on common EF with two of the components (updating and shifting) still making unique contributions. These findings support the assumption of a general EF ability with separable updating and switching components and an inhibition component that is not separable and that is weakly to moderately linked to general EF ability. The best models of the data include both common and componential levels which is why Miyake and Friedman proposed that EF has both *unity* and *diversity*.

Correspondence should be addressed to Kenneth R. Paap, Department of Psychology, San Francisco State University, 1600 Holloway Avenue, EP301, San Francisco, CA 94132, USA. E-mail: kenp@sfsu.edu

The hypothesis of bilingual advantages in EF

There is a widely held belief that bilinguals enjoy an advantage over monolinguals with regard to EF (Bialystok, Craik, & Luk, 2012; Kroll & Bialystok, 2013). In both academic journals and the mainstream media the bilingual advantage is often described in ways that naturally lead to the inference that most bilinguals outperform their monolingual counterparts and that the advantages are consistently observed in circumstances that require conflict resolution, the ability to filter out distracting or competing information, and the ability to effectively switch back and forth in multitasking situations. As reviewed later there are many studies showing no language-group differences and even occasional reports of monolingual advantages. This state of affairs has reasonably led many investigators to suggest that bilingual advantages may be contingent upon specific types of bilingual experiences or that the advantages are limited to specific components of EF. Exploring these possibilities is the main purpose of this study.

Inhibitory control. Bilingual advantages in inhibitory control have proven to be very inconsistent to the point where the existence of the phenomena has been questioned. Hilchey and Klein (2011) reviewed 31 experiments and concluded that evidence for a bilingual advantage in inhibitory control in both children and young adults is rare. More emphatically they asserted that the collective evidence "…is simply inconsistent with the proposal that bilingualism has a general positive effect on inhibitory control processes" (p. 629).

Since Hilchey and Klein's conclusion regarding the inhibitory control component of EF there have been many attempts to test for bilingual advantages using these nonverbal tasks and reports of no group differences strongly outnumber those showing advantages. Kousaie and Phillips (2012a) found no behavioural differences between groups of young adults in the Stroop, Simon or flanker tasks. Kousaie and Phillips (2012b) used both young adults and older adults and found no differences in the magnitude of Stroop interference. Similarly, Kirk, Fiala, Scott-Brown, and Kempe (2014) found no differences between older Gaelic–English bilinguals and older English monolinguals in a Simon task. A study by Humphrey and Valian (2012) using the Simon and flanker tasks followed the same pattern. Four different groups of multilinguals showed Simon and flanker effects statistically equivalent to a group of English monolinguals. Paap and Greenberg (2013) found no bilingual advantage in three Simon experiments and one flanker experiment. Testing a new set of participants, Paap and Sawi (in press) reported no bilingual advantages in either flanker or Simon effects. Mor, Yitzhaki-Amsalem, and Prior (2014) reported no language-group differences in either a numeric Stroop task or a Simon arrows task. Ryskin, Brown-Schmidt, Canseco-Gonzalez, Yiu, and Nguyen (2014) showed no language-group difference across three experiments in visuospatial perspective-taking, but most relevant to present purposes is that there were no differences in two versions of the Stroop task (Experiment 1), an antisaccade task (Experiment 2) or in the magnitude of flanker interference (Experiment 2).

The absence of bilingual advantages in inhibitory control is consistent with the emerging neuroscience work on bilinguals. For example, Branzi, Martin, Abutalebi, and Costa (2014) reported an absence of inhibitory effects in the N2 component time-window of a mixed picture-naming study despite the fact that second language (L2) naming globally slowed subsequent first language (L1) naming. Branzi et al. concluded that highly proficient bilinguals may achieve language control in a way that is qualitatively different from less proficient bilinguals and that the control does not rely on inhibitory mechanisms.

Turning to the new studies using children a study by Engel de Abreu, Cruz-Santos, Tourinho, Martin, and Bialystok (2012) did report a significant advantage in flanker interference for a group of 40 beginning Portuguese–Luxembourgish bilinguals compared to a group of 40 Portuguese monolinguals. Likewise, Yang, Yang, and Lust (2011) reported a bilingual advantage in flanker interference for Korean–English bilinguals (n = 15) in comparison to Korean monolinguals (n = 13) who were about 5-year-olds. Poarch and van Hell (2012) reported a significant advantage in Simon interference in 5- to 8-year-olds for trilinguals (n = 18) over German-speaking monolinguals (n = 20), but the advantage of a German–English bilingual group (n = 18) over the monolingual group was marginal (p = .062).

In contrast, Duñabeitia et al. (2013) compared Spanish monolinguals (n = 252) to Basque–Spanish bilinguals (n = 252) at six successive grades with respect to both a verbal Stroop task and a number-size congruency task. Bilinguals and monolinguals performed equivalently in these two tasks across all the indices of inhibitory skills explored and across all grade levels. Antón et al. (2014) compared a group of 180 Basque–Spanish bilingual children with a group of 180 carefully matched monolinguals on an attentional network task (ANT) version of the flanker task. The comparison between the language groups was consistent and null: no inhibitory advantage (incongruent–congruent), no global response time (RT) advantage, no alerting advantage and no orienting advantage.

Goldman, Negen, and Sarnecka (2014) compared English monolinguals (n = 32) to bilingual children who were either exposed to two languages in the home (n = 40) or only to a language other than English at home (n = 20). The children were 3- to 6-year-olds and all completed a numerical discrimination task that required the children to ignore area and attend to number. Half of the trials were congruent, where the numerically greater array was also larger in total area, and half were incongruent, where the numerically greater array was smaller in total area. There were no language-group differences in either overall performance or on a measure of inhibitory control computed by subtracting accuracy on the congruent trials from that on the incongruent trials. Similarly, in a study of children ranging from 6- to 15-year-olds, Kapa and Colombo (2013) found no differences in the magnitude of the flanker interference effect across three groups: early Spanish–English bilinguals (n = 21), late Spanish–English bilinguals (n = 36) and English monolinguals (n = 22).

In a lifespan study testing seven age groups (from three years of age through over 60), Gathercole et al. (2014) found no systematic language-group differences on three tasks assumed to reflect executive functioning: dimensional card sorting (N = 650), Simon (N = 557) and a grammaticality judgement task with irrelevant semantic anomalies (N = 354). Participants were classified as either English monolinguals or English–Welsh bilinguals coming from homes in which only Welsh was spoken, only English was spoken or both were spoken. The Duñabeitia et al. (2013), Antón et al. (2014) and Gathercole et al.'s (2014) studies share the strengths of using bilinguals immersed in a bilingual region, monolinguals from the same country, a very large number of participants, multiple age groups and multiple measures of EF.

Monitoring. In contrast to the results with measures of inhibitory control, Hilchey and Klein (2011) reported 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). Global RT (the overall mean across both the congruent and incongruent trials) is often used as a measure of the monitoring component of EF. Alternatively, the mean RT difference between the congruent trials from the standard mixed block of both congruent and incongruent trials and a pure block of neutral (control) trials is sometimes used.[1] Paap and Greenberg (2013) reviewed several studies conducted after the Hilchey and Klein's review and there were no significant advantages on any measure of monitoring out of 18 new tests. Two additional studies (Abutalebi et al., 2012; Yudes, Macizo, & Bajo, 2011) also report no language-group differences on congruent trials or global RTs. It appears that bilingual advantages in measures assumed to reflect monitoring are also difficult to replicate. In an update of the 2011 review, Hilchey, Saint-Aubin, and Klein (in press) observe that the influx of new data strongly repudiates their earlier conclusion that bilingualism leads to bilingual advantages in monitoring.

Switching. In contrast to their examination of inhibitory control and monitoring, Hilchey and Klein did not review the existing evidence for bilingual advantages in the switching component of EF. There have been three reports of a bilingual advantage in colour-shape switching tasks. Garbin et al. (2010) reported a bilingual advantage in a switching task (significant only in the accuracy measure) that compared 21 Spanish monolinguals to 19 Spanish–Catalan bilinguals. In the behavioural data it is surprising that the bilinguals showed no switching costs at all as switching costs are typically much larger than the interference effects observed in the Simon or flanker tasks. This oddity could be partially or completely responsible for the bilingual advantage in switching costs. Prior and MacWhinney (2010) also reported a bilingual advantage in switching costs with monolinguals incurring a 206 ms switching cost compared to only 144 ms for bilinguals. Prior and Gollan (2011) used the same colour-shape switching task to test for

[1] Because the latter is a difference score and individual differences in processing speed cancel out, it is usually credited with being a more pure measure of the monitoring construct.

differences in switching costs for English monolinguals ($n = 47$), Mandarin–English bilinguals, ($n = 43$) and Spanish–English bilinguals ($n = 41$). After controlling for differences in response speed and parent's educational level (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 had significantly smaller switching costs in a language-switching task. Prior and Gollan concluded 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).

From this point on the tide clearly turns. Paap and Greenberg (2013) showed 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 Paap and Greenberg's bilinguals overwhelmingly reported that they used both languages every day and switched 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 Paap and Greenberg's bilinguals (72%). From this one might infer that Paap and Greenberg's bilinguals switched 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 the Paap and Greenberg's null results.

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 1,100 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. More recently Paap and Sawi (in press) reported another large *n* study showing no language-group differences in switch costs. Finally, three large *n* experiments using Spanish–Catalan bilinguals and Spanish monolinguals (Hernández, Martin, Barcelo, & Costa, 2013) also showed no language-group differences in switch costs. In summary, the early reports of bilingual advantages in switching costs have sunk into a mire of consistently null results.

New studies reporting bilingual advantages in young adults

In contrast to the null results described earlier there have been seven new (i.e., post Hilchey and Klein's review) reports of bilingual advantages in either inhibitory control or monitoring using either the flanker task (Abutalebi et al., 2012; Kapa & Colombo, 2013; Marzecová, Asanowicz, Krivá, & Wodniecka, 2013; Pelham & Abrams, 2014; Tao, Marzecová, Taft, Asanowicz, & Wodniecka, 2011) or Simon task (Salvatierra & Rosselli, 2011; Schroeder & Marian, 2012).

Although this evidence should be given just consideration the following points should be considered in weighing the evidence. These seven studies are very typical of those reporting bilingual advantages. First, the number of participants in each group is small and this is risky for the reasons reviewed in the general discussion. The recent spate of large *n* (at least 180 participants per language group) studies described earlier (Anton et al. 2014; Duñabeitia et al., 2013; Gathercole et al., 2014) consistently reported no language-group differences across *multiple measures* and *multiple tasks*. This segues to the next contrast, namely, that all seven of the new studies reporting a bilingual advantage used a single task that yields one measure of monitoring and one of inhibitory control. With one exception (Tao et al., 2011), one measure shows a bilingual advantage and the other does not. None of the studies conform to the recommendation of Paap and Greenberg that a compelling demonstration of a bilingual advantage should show significant advantages on the same component of EF (e.g., monitoring) in two different tasks thus demonstrating convergent validity.

Some of the studies generate more specific concerns. For example, although Schroeder and Marian (2012) reported a bilingual advantage in the Simon effect, monolinguals ($M = 624$ ms) were slightly faster than bilinguals ($M = 632$ ms) on the incongruent trials. Thus, there was no bilingual advantage on the trials where conflict actually occurred. The only reason for the bilingual advantage in the magnitude of the Simon effect was because monolinguals (M = 573 ms) were substantially faster than bilinguals (M = 609) on the congruent trials. The same pattern of results was reported by Salvatierra and Rosselli (2013) for the performance of groups of older adults in a standard Simon task and again the results were interpreted as a bilingual advantage in inhibitory control despite the fact that the interaction was caused by a monolingual advantage on the congruent trials.

The interpretation of the Tao et al.'s results is compromised because the early bilinguals had significantly higher Raven's[2] scores (and significantly lower levels of parental education) compared to the monolinguals. The authors reasonably express the concern that it is important to make sure that any differences between the groups in EF were not due to preexisting differences other than language background. Unfortunately, they assume that this confounding can be statistically controlled by including the Raven's scores and PED as covariates. This violates the assumption that the covariate and groups are independent. As Miller and Chapman (2001) point out in their landmark article:

> "control" is altogether the wrong metaphor for understanding what ANCOVA accomplishes. We have found that investigators are frequently surprised when this is pointed out. Some assert that "controlling" or "removing" nontrivial group differences on the covariate is the primary use of ANCOVA… the relevant literature roundly condemns this view. (p. 42)

The independence assumption is a basic tenet of ANCOVA and when violated the regression adjustment may either obscure part of the grouping effect (e.g., language effect) or produce spurious effects; thus the ANCOVA results are uninterpretable when there are systematic differences in the covariate across groups. The bilinguals in the Marzecová, Asanowicz, et al. (2013) study also had significantly higher Raven's scores compared to the monolinguals, $t(33)$ = 2.70, p = .009, and they also used the Raven's scores as a covariate.

Kapa and Colombo (2013) reported a global RT advantage in the flanker task, but there was no advantage with respect to the flanker interference effect. As discussed earlier the global RT measure is a very impure measure of monitoring (or anything else) because it also includes differences in lower level (e.g., perceptual and motor speed) processes. A purer measure of monitoring is the difference between the mean of a block of baseline trials that are all conflict free and the mean RT of the congruent trials from a block that mixes together both congruent and incongruent trials.

In a study that concurrently acquired functional magnetic resonance imaging (fMRI) neuroimages, Abutalebi et al. (2012) reported an advantage of bilinguals (n = 17) over monolinguals (n = 14) in flanker interference, but only in the second of two blocks. This pattern was interpreted as a bilingual advantage because bilinguals "…are better able to adjust to conflict, hence, to adapt to conflicting situations" (p. 2085). However, it is difficult to replicate higher order interactions with small ns. Indeed, in this case, the Group × Block × Congruency interaction tested by Abutalebi et al. (2012) has yielded three distinctively different patterns. Costa, Hernandez, and Sebastian-Galles (2008) reported a pattern that is the opposite of Abutalebi et al., namely, a bilingual advantage appeared in the first block and disappeared in the later blocks. Costa et al. (2008) suggested that practice reduces attentional demands and the effects of bilingualism. Paap and Greenberg (2013) reported that the magnitude of the interference effect significantly increased across blocks, but with equivalent interference effects for bilinguals and monolinguals in each block. As Costa et al. (2008) had 100 participants in each language group and Paap and Greenberg had more than 50 participants in each group, it is more likely that the Abutalebi pattern was a random perturbation than a genuine "adaptation" effect on the part of bilinguals (n = 14).

One could characterise our review of the studies in this section as an effort to carefully examine the reports of bilingual advantages and finding a fault, oddity or problem with each. Nevertheless, one could also observe that this exercise counters the tendency of authors to highlight the results that "worked" and to allow the inconsistencies to either recede into the background or to be integrated into complex accounts that do not generalise.

Are bilingual advantages restricted to certain types of bilingual experiences?

The empirical landscape minimally establishes that bilingual advantages in EF are elusive. The

[2]The Raven's Advanced Matrices test is a nonverbal measure of general fluid intelligence.

empirical purpose of this study is to further explore the possibility that bilingualism can enhance EF, but that the enhancement is constrained to only certain types of bilingual experience or can be detected with only certain tasks. This is consistent with the view expressed by Marzecová, Bukowski, et al. (2013) that:

> the emergence and specific tuning of the bilingual advantage in conflict resolution may be modulated by different factors of the language experience, including the age of acquisition of the second language, proficiency, the extent of use of languages, the relative balance between languages, and the context of language use (Green, 2011; Luk et al., 2011; Tao et al., 2011). These factors could account for some of the inconsistencies (see, e.g., Tao et al., 2011), but the empirical evidence is not yet sufficient. Moreover, we lack theoretical models of how specifically these factors mediate the bilingual advantage. (p. 587)

We hope to add substantially to the empirical evidence by forming a composite database of 384 participants and investigating the following dimensions: early versus late acquisition of L2, the relative balance of L2 to L1 and the number of languages spoken.

Are bilingual advantages restricted to early bilinguals? Since the seminal study of Mechelli, Crinion, Noppeney, O'Doherty, Ashburner, Fackowiak, and Price (2004), it has become increasingly clear that bilinguals have greater density of grey matter in the left inferior parietal cortex compared to monolinguals and that the difference is more pronounced in early compared to late bilinguals. One purpose of the current study is to investigate whether significant and consistent bilingual advantages in EF are restricted to bilinguals who acquire both languages early in life. Luk, De Sa, and Bialystok (2011) reported that the magnitude of the flanker effect was significantly smaller for early bilinguals compared to both late bilinguals and monolinguals. Early and late were operationally defined in terms of the age at which each participant began using two languages actively on a daily basis. The age of onset was determined by a self-report and the cut-off age for categorising bilinguals as early or late was 10-years old. In contrast, to the difference obtained with the flanker interference effect, there were no differences between the three groups with respect to a difference measure (congruent RT–control RT) that was assumed to reflect the monitoring component of EF. In summary, Luk et al. reported the following pattern of differences across the three language groups: no language-group differences in monitoring and an advantage in inhibitory control for early bilinguals over both late bilinguals and monolinguals. The overall pattern has not been replicated in any of the following four subsequent studies. Pelham and Abrams (2014) also showed no differences in measures of monitoring, but both groups of bilinguals showed an equivalent advantage in inhibitory control (as indexed by the flanker interference effect) compared to monolinguals. Although this is evidence favouring a generic bilingual advantage it is inconsistent with the more specific hypothesis that early bilingualism is critical. Likewise, Tao et al. also reported no differences in the magnitude of the flanker effect between early and late bilinguals. Kapa and Colombo (2013) using children in the age range of 6–15 reported no language-group differences at all with respect to inhibitory control (the flanker interference effect) but did report a global RT advantage for early bilinguals compared to monolinguals. The fourth study, Humphrey and Valian (2012), had the simplest pattern of results: there were no differences between early bilinguals, late bilinguals and monolinguals in either monitoring (global RT) or inhibitory control (interference effect). Furthermore, these null results were obtained with both a Simon task and a flanker task. Not only were the interference effects non-significant, but they were also nearly identical across the language groups with flanker effects of about 87 ms and Simon effects of about 35 ms. The bilinguals were classified on the basis of their self-ratings of proficiency at two points in time in *childhood*, and currently as young adults.

To summarise, there are five studies that have conducted a total of six tests that have included separate groups of early bilinguals, late bilinguals and monolinguals. Only two of the six (Kapa & Colombo, 2013; Tao et al., 2011) showed a unique advantage for early bilinguals in measures of monitoring (global RT) and only one of the six (Luk et al, 2011) showed a unique advantage for early bilinguals in measures of inhibitory control. The hypothesis that early bilingualism enhances EF can be retested in our flanker and Simon task data and extended to the antisaccade and switching task. One likely contributor to the inconsistencies across studies comparing early bilinguals, late bilinguals and monolinguals is that different definitions of early and late were used. To sidestep the question of how to partition the bilinguals we will also analyze our data using age of first exposure to an L2 as a predictor in a regression analysis.

Do advantages occur at just the right balance between L2 and L1? Rather than age-of-acquisition, it may be the proficiency of L2 relative to L1 that determines if bilingualism enhances EF. However, a rationale argument can be constructed for predicting either that having high levels of proficiency and balance helps or that it hurts. The "helping" possibility assumes that the experience of managing two languages will enhance general EF only when the links to both translation equivalents are strong and there is a long history of managing both languages. This rationale predicts that highly proficient and balanced bilinguals should have better EF than bilinguals who have a clearly dominant language. Although this logic has intuitive appeal the argument becomes complicated if the task of coordinating two languages is viewed as one that varies along a dimension of automatic to controlled processing. Although the links between form and meaning may be stronger for highly fluent and balanced bilinguals, the mechanism for resolving the competition between translation equivalents may become automatic and more reliant on language-specific mechanisms within the bilingual lexicon (Costa & Santesteban, 2004; Costa, Santesteban, & Ivanova, 2006). If this is true, then the controlled processing regulated by EF may be exercised far more often by bilinguals with a clearly dominant language, especially when the non-dominant language is used often. This would be consistent with the evidence presented by Kroll and Rossi (2013) that although both languages are activated during speech planning, it is the dominant language that is the target of inhibition; an inhibition that can extend to a global component that goes beyond the level of individual words. It is also consistent with the results reported by Branzi et al. (2014) that picture naming in L1 is globally delayed by previous naming in the L2. Both possibilities were tested by partitioning our 384 participants into 5 groups based on their L2 to L1 proficiencies.

Are advantages most likely to occur with multilinguals? Yet another plausible hypothesis is that the experience of managing three or more languages may provide the greatest enhancement of EF. Poarch and van Hell (2012) reported advantages in Simon interference of trilinguals over monolinguals and advantages in flanker interference of trilinguals over second-language learners. However, the interference effect obtained with trilinguals was not significantly smaller compared to bilinguals in either task. Also inconsistent with the hypothesis that trilingualism generates the greatest enhancement of EF are the results of Humphrey and Valian, using young adults, showing no significant differences between trilinguals, bilinguals and monolinguals.

On the other hand, in an intriguing report Bak, Nissan, Allerhand, and Deary (2014) compared a measure of general fluid intelligence (gF) at age 73 to a baseline measure of child intelligence taken when the cohort was 11-year-olds. The cohort of 85 multilinguals showed elevated gF scores compared to the 162 bilinguals or the monolinguals. The implications of this for possible gains in EF are ambiguous. As discussed by Paap and Sawi (in press), gF and EF frequently show high degrees of association but are conceptually and empirically separable to some degree (Salthouse, 2005; Salthouse, Pink, & Tucker-Drob, 2008). Bak et al. (2014) did not include any of the standard non-verbal measures of EF that are the focus of this study, but the relationship they report between multilingualism and gF reinforce the importance of looking at this dimension. In addition, there have been recent reports of structural differences in the brain between multilinguals and bilinguals. For example, Grogan et al. (2012) reported that multilinguals had higher grey matter density in the right posterior supramarginal gyrus (a brain region associated with vocabulary size) compared to bilinguals.

THE COMPOSITE DATABASE

Participants

A pool of 384 participants was formed by combining the participants from Paap and Greenberg (2013) and Paap and Sawi (in press). Each participant completed two to four of the following tasks: antisaccade (Studies 1 and 2), flanker (Studies 3 and 4), Simon (Studies 1–4) and colour-shape switching (Studies 1–4). Most participants were upper-division psychology majors. Language characteristics for the entire set of participants are shown in Table 1. These are representative of each individual study as each study sampled from the same participant pool with the same recruitment methods. Proficiency in understanding and speaking was self-rated using

TABLE 1
Language characteristics for early versus late analysis: Mean (*M*) and standard error (*SE*)

		Age		*L1 Pro.*		*L2 Pro.*		*L2 AoA*		*% English use*	
Group	n	M	SE	M	SE	M	SE	M	SE	M	SE
Native 2L	34	23.9	0.8	6.6	0.1	5.7	0.2	0.0	0.0	71.0	2.4
B L2 before 6	57	22.9	0.6	6.5	0.1	5.4	0.1	3.9	0.2	69.8	2.1
B L2 at/after 6	55	25.6	1.0	6.6	0.1	5.2	0.1	10.8	0.6	63.9	2.8
M L2 before 6	33	22.3	1.6	6.5	0.1	2.7	0.2	1.8	0.3	90.7	2.5
M L2 at/after 6	60	23.1	0.8	6.7	0.1	2.4	0.1	13.4	0.4	93.8	1.8
M L2 no L2	119	24.8	0.4	6.6	0.0	–	–	–	–	99.9	0.1

B, bilingual; m, monolingual; L1, first language; L2, second language; Pro., rated proficiency; AoA, age-of-acquisition.

the 7-point scale described in Paap and Greenberg (2013). For the purpose of presenting the characteristics shown in Table 1 participants were classified as bilingual if they rated their proficiency as 4 or higher on two languages.

For the bilinguals the mean proficiency was 6.25 for English and 5.70 for their other language. A rating of 6 represents *Fluent: As good as a typical native speaker* and a rating of 7 represents *Super Fluency: Better than a typical native speaker*. As a group those classified as bilinguals were highly fluent in at least two languages and 22% were fluent in three or more languages. The bilinguals currently use both languages and speak English 72% of the time. This reflects greater balance between languages than, for example, the 75% use of Catalan reported by Hernández, Martin, Barceló, and Costa (2013) for Spanish–Catalan bilinguals living in Barcelona. The bilinguals spoke 38 distinct non-English languages, but 40% were Spanish–English bilinguals, 14% spoke a variant of Chinese and 6% Tagalog. In summary, the participants classified in Table 1 as "bilinguals" were fluent in at least two languages and actively used at least two languages.

Some researchers are sceptical about the accuracy of self-ratings of language proficiency, but self-ratings are highly correlated with a range of objective and standardised measures of language proficiency. For example, a study by Marian, Blumenfeld, and Kaushanskaya (2007) correlated self-report measures of reading, speaking and listening proficiency with eight different standardised measures of language skill involving reading, writing, speaking and listening and covering both comprehension and production. These correlations were obtained for both L1 and L2 where L1 was defined as the language a bilingual acquired first. For L2 (the proficiency of greatest concern in classifying an individual as bilingual), all 24 correlations between the three subjective measures and the eight objective measures were significant with Pearson r values ranging from .29 to .74 with a mean of .59. Taking all of their results into account Marian et al. concluded that self-ratings are "an effective, efficient, valid, and reliable tool for assessing bilingual language status" (p. 960). In a similar study, Francis and Strobach (2013) reported that self-ratings in both English and Spanish are highly predictive of standardised objective measures.

Although our bilinguals are fluent in two languages, their rated proficiency in English (M = 6.3) is less than that of the monolinguals (M = 6.5), $t(382) = 3.83$, $p < .001$. This occurred because more than half of the monolinguals consider themselves to be "super fluent", that is, better than a typical native speaker. Studies 1–3 from Paap and Greenberg did include an objective measure of category fluency (Gollan, Montoya, & Werner, 2002) as participants were asked to name as many instances in 1 min as they could for the following categories: musical instruments, vegetables and animals. An independent samples t-test confirmed that the mean number of correct responses (across all three categories) for monolinguals ($M = 50.02$) was greater than the mean for bilinguals ($M = 42.7$), $t(269) = 5.06$, $p < .001$. Thus, the group difference in self-reported English proficiency correctly predicted the group differences in the category fluency task. The self-ratings of English proficiency also significantly correlated with the objective measure of category fluency, $r(282) = +0.33$, $p < .001$. Thus, a correlation of medium size, following Cohen's (1992) guidelines, was observed even though the range of the English proficiency variable was very restricted (namely from 4 to 7).

TABLE 2
Performance on 12 measures of EF based on 3 aspects of multilingualism

Task	*Six groups based on age-of-acquisition of L2*				*Five groups based on L1 proficiency relative to L2 proficiency*				*Three groups based on number of languages spoken fluently*			
Measure	F (df)	MSE	p	*ES*	F (df)	MSE	p	*ES*	F (df)	MSE	p	*ES*
Antisaccade (*N* = *191*)												
RT	1.89 (5, 172)	33,479	.098	.052	2.76 (4, 186)	32,577	**.029**	.056	3.16 (2, 188)	33,032	**.045**	.033
RT cost	1.78 (5, 172)	9,403	.119	.049	1.17 (4, 186)	9,489	.325	.025	0.03 (2, 188)	9,622	.030	.000
PC	0.22 (5, 172)	.011	.952	.006	0.81 (4, 186)	.011	.521	.017	0.05 (2, 188)	1.994	.948	.001
PC cost	1.90 (5, 172)	.015	.096	.052	1.98 (4, 186)	.007	.099	.041	3.12 (2, 188)	.007	**.046**	.032
Flanker (*N* = *212*)												
Effect	0.78 (5, 194)	1,294	.563	.020	1.00 (4, 207)	1,363	.406	.019	0.10 (2, 209)	1,375	.909	.001
Mixing costs	0.54 (5, 194)	4,376	.743	.014	0.97 (4, 207)	4,356	.427	.018	0.25 (2, 209)	4,384	.777	.002
Global RT	0.86 (5, 194)	5,461	.510	.022	1.75 (4, 207)	6,264	.140	.033	1.23 (2, 209)	6,340	.294	.012
Simon (*N* = *379*)												
Effect	2.45 (5, 347)	845	**.033**	.034	3.29 (4, 374)	815	**.011**	.034	3.85 (2, 376)	822	**.022**	.020
Mixing costs	1.40 (5, 263)	1.467	.224	.026	1.22 (4, 283)	1,473	.303	.017	1.98 (2, 285)	1,468	.140	.014
Global RT	0.90 (5, 347)	4,123	.478	.013	1.77 (4, 374)	4,426	.133	.019	0.32 (2, 376)	4,478	.725	.002
Switching (*N* = *351*)												
Switch cost	1.02 (5, 322)	19,506	.405	.016	0.39 (4, 346)	19,451	.819	.004	0.32 (2, 348)	19,390	.724	.002
Mixing cost	2.67 (5, 322)	60,732	**.022**	.040	0.39 (4, 346)	63,722	.816	.004	0.65 (2, 348)	63,404	.522	.004

N, total number of participants; PC, proportion correct; RT, response time; ES, partial η squared.

Method, measures and procedures for the four tasks

Details of the procedures and methods used for each of the four tasks are provided in Paap and Greenberg (2013).

Antisaccade task. The antisaccade task is often used as a measure of inhibitory control because participants must suppress the reflexive urge to shift attention and gaze to a visual stimulus that appears suddenly in the periphery. The antisaccade task used by Paap and Greenberg was modelled on those used by Kane, Bleckley, Conway, and Engle (2001). On antisaccade trials a distractor (#) blinked on the opposite side of fixation before the appearance of a target letter (B, P or R) that was followed by a visually similar mask (8). The task was to identify the target letter by pressing the corresponding key. Participants were instructed to respond as quickly as possible without making errors.

The block of antisaccade trials was preceded by a block of control trials that used a centred target and no distracting stimulus. These control trials provide a baseline that should require little or no inhibitory control. Unlike the other three tasks, the overall accuracy (proportion correct = .93) on the antisaccade trials was not pinned tightly against the ceiling of 1.00. Thus, the antisaccade task yields four potentially important measures: the speed and accuracy on the block of antisaccade trials and also the "costs" reflected in the difference between the antisaccade block and the baseline block (see Table 2) for both speed and accuracy. All four of these indices have been used as measures of inhibitory control in previous studies.

Flanker task. The flanker task used by Paap and Greenberg was closely modelled after the flanker task used by Costa et al. (2008) which, in turn, was a version of the ANT developed by Fan, McCandliss, Sommer, Raz, and Posner (2002). The central target and flankers were arrows pointing either left or right. Participants were instructed to press the left key if the centre arrow pointed to the left and the right key if it pointed to the right. The flanking arrows were either congruent or incongruent with respect to the centre target. The flanker interference effect (mean incongruent RT − mean congruent RT) is a frequently used measure of inhibitory control. There were a total 96 congruent and 96 incongruent trials. The difference between the mean RT on congruent trials from a mixed block and the mean on a block of neutral trials that have no flankers was used as the primary measure of monitoring. However, global RT (the mean across both congruent and incongruent trials) was also reported because it is a common measure of monitoring in the flanker task.

Simon task. The Simon task was described by Paap and Greenberg in Study 2. In the two Simon blocks the target "Z" or "/" appeared to the left or right of fixation and the location was either congruent or incongruent with the hand required for a correct response. Participants were instructed to press the Z key with their left index finger if the target was Z and the "/" key with their right index finger if the target was a "/". Each block consisted of a randomised mix of 20 congruent and 20 incongruent trials. Although the location of the target is task irrelevant, location is automatically processed and presumably must be suppressed on the incongruent trials. Thus, the difference between congruent and incongruent trials is often used as a measure of inhibitory control.

For Studies 2, 3 and 4 the two blocks of Simon trials were preceded by two practice blocks of "neutral trials" where the target was displaced above or below the centre fixation and, accordingly, was neither congruent nor incongruent with the left- or right-motor responses. The difference in mean RT on the congruent trials in the block of standard Simon trials compared to the neutral block where the targets were centred was used as the primary measure of the monitoring component of EF. Language-group differences in global RT (mean RT across both congruent and incongruent trials in the standard Simon task) are also reported in Table 2 because global RT is a very common measure of monitoring when bilingual advantages are investigated with the Simon task.

Switching task. The colour-shape switching task used by Paap and Greenberg was based on the task used by Prior and MacWhinney (2010). During a pure block of colour decisions participants responded with two fingers of the left hand to indicate if the target was blue or red. During a pure block of shape decisions participants responded with two fingers of the right-hand to indicate if the target was a circle or a triangle. In the mixed blocks, a precue was presented on each trial to signal whether colour or shape was the relevant dimension. Thus, each trial in a mixed block was either a repeat (same dimension as the previous trial) or a switch (different from the previous trial). *Switching costs* (mean RT on switch trials – mean RT on repeat trials) were assumed to measure the switching component of EF; whereas the differences between the pure blocks and the repeat trials are referred to as *mixing costs* and are assumed to provide a measure of the monitoring component of EF.

ANALYSES OF COMPOSITE DATABASE

NHST and Bayesian analysis plan

Consistent with the vast majority of research testing for bilingual advantages in EF both ANOVA and regression analyses were used to examine the possible roles of age-of-acquisition, L2/L1 proficiency and multilingualism in enhancing EF. Additionally, a Bayes factor (BF) analysis was interwoven in three situations: (1) when the present study yielded a null result in a targeted replication, (2) when the present study yielded a significant but anomalous language-group difference and (3) for a set of comparisons between the most highly proficient bilinguals and monolinguals with minimal exposure to an L2.

The BF is a likelihood ratio of the probability of the data given the null hypothesis data over the probability of the data given the alternative[3]:

$$\mathrm{BF_{NA}} = \frac{p(\mathrm{Data/null})}{p(\mathrm{Data/alternative})}$$

Unlike null hypothesis statistical testing (NHST) the BF directly compares the evidence favouring the null to that favouring the alternative. Thus, the BF is directly interpretable, that is, if $\mathrm{BF_{NA}} = 10$, then the empirical data are 10 times more probable if the null were true then if the alternative were true. Thus, it is especially informative in a situation where a sceptic might want to convince a believer that the data provide more evidence for the null than for the alternative. As Rouder, Morey, Speckman, and Province (2012) put it "If the null is true, the best outcome of a significance test is a statement about a lack of evidence for an effect. It

[3]The reciprocal Bayes Factor, $\mathrm{BF_{AN}}$ is often reported and this reverses the interpretation. A second variation is to transform raw BF to the natural logarithm of the BM in order to create identical scales above and below a raw value of 1. We chose the convention matching the output of the BF calculator on Rouder's web site: pcl.missouri.edu.

would be desirable to state positive evidence for a lack of an effect" (p. 357). When using NHST the sceptic is on the wrong side of the null because the null may be rejected but not supported.

Baysian analyses can counter the biased impression that significant p-values mean that the null hypothesis is far less likely than the alternative. "This bias is highly problematic because researchers may reject the null without substantial evidence against it … a p-value of 0.05 may correspond to as much evidence for the alternative as the null" (Rouder et al., 2012, pp. 357–360). To be more specific, Wetzels et al. (2012) compared p-values, BFs and effect sizes for 855 t-statistics published in two major journals. For the subset of tests with p-values between .01 and .05 the BF values for 70% were only in the range offering "anecdotal" support for the alternative hypothesis. Nine of these actually yielded BFs greater than 1 indicating that the data were more probable given the null than given the alternative.

The Bayesian approach sets up two competing models. One model posits that only the null value is possible. The alternative model posits that a broad range of other values is also possible. The value of BF depends on the selection of the prior distribution on effect size for the alternative hypothesis. One school of thought is that the priors should be noninformative such that they can be used as a default in all tests. Following Rouder, Speckman, Sun, Morey, and Iverson (2009), the default BF values we report use the Jeffreys-Zellner-Siow (JZS) as an objective prior. Wetzels et al. observe that this default test "… avoids informed specification of prior distributions that other researchers may contest" (p. 294).

A final observation is that the odds ratio delivered in a BF does not impose a conclusion on the reader in the way that the author's choice of an alpha level does in NHST. The reader, based on his/her own prior probabilities (e.g., ranging from sceptic to agnostic to believer) can adjust them on the basis of the objective BF. For those who would rather rely on an established convention most contemporary experts and proponents of Bayesian analysis still endorse Jeffrey's guidelines that odds ranging from 1 to 3 are "barely worth mentioning" (translated as "anecdotal" by Wetzels et al., 2012), from 3 to 10 are "substantial", from 10 to 30 are "strong" and greater than 30 as "very strong or decisive".

Early versus late bilinguals

Using our standard criteria participants were classified as bilinguals if they had a proficiency of at least four in two or more languages. In order to approximate the additional restrictions used by Luk et al. (2011) in seeking "active" bilinguals, we removed the 23 bilinguals who reported currently using English more than 94% of the time. Explicit reports of frequency of switching were solicited only in our two most recent studies but for that subset (n = 90 bilinguals) the median and modal scale value for switching was 3, *I usually switch from one language to the other a couple of times a day*.

Table 1 shows the characteristics of the six groups formed to investigate the role that early bilingualism plays in EF. The first group of bilinguals had two native languages (Native 2L) as they were exposed to two languages since birth. The Native 2L group could be viewed as the quintessential representatives of "early bilinguals". The second group of bilinguals had one native language and started to acquire an L2 before the age of 6. This group seems most comparable to the "early bilinguals" tested by Luk et al. who were actively using two languages before the age of 10. The third group of "late bilinguals" started to acquire their L2 at or after the age of 6. For present purposes the most important dimension shown in Table 1 is that the mean age-of-acquisition of L2 increased from 0.0 to 3.9 to 10.8 years for the three groups of bilinguals.

Three subgroups of monolinguals were also created. One group consisted of participants who indicated that either they had no exposure to an L2 or rated their proficiency as a 1: *Beginner—know some words*. The other two groups of monolinguals rated their proficiency as either a 2 or 3 and were first exposed to an L2 either before the age of 6 or at/after the age of 6. Thus, this first set of analyses partition the total set of participants into six groups. A separate one-way ANOVA comparing the six groups was conducted on each of the 12 measures of EF the results of which are shown in Table 2 with p values less than .05 highlighted in bold.

Flanker and Simon task. The flanker and Simon effects are both defined as the difference in mean RT between the congruent and incongruent trials. Each measure was analyzed with a separate one-way ANOVA with the six language groups defined earlier. As shown in Table 2 the magnitude of the flanker effect does not differ across groups. To enable a more exact comparison to the results of Luk et al., Figure 1 shows the mean flanker effect for each of the six groups with 95% confidence intervals. The pattern of nonsignificant differences in Figure 1 is actually opposite of the pattern reported by Luk et al. (2011) in that the late bilinguals (L2 at/after 6) have the smallest flanker effects. Together with the results of Humphrey and Valian, it appears that the advantages of early bilinguals over both late bilinguals and monolinguals in the magnitude of the flanker effect do not replicate.

For the Simon effect (Table 2 and Figure 2), there was a significant main effect of groups, but the group with Native 2L (+42 ms) had the largest Simon effect and the purest monolingual group had the smallest interference effect (+22 ms). This pattern is incompatible with the hypothesis that early bilingualism enhances the inhibitory control component of EF. Although the NHST implies a monolingual advantage rather than a null result for this comparison, the BF value of .625 shows that the empirical data are only 1.6 times more probable if the alternative were true then if the null were true. Thus, from the BF perspective this surprising difference is "barely worth mentioning" or "anecdotal".

One-way ANOVAs were also conducted on the monitoring measures for each task. As shown in Table 2 there were no differences across groups with respect to the monitoring component in either the flanker or Simon task.

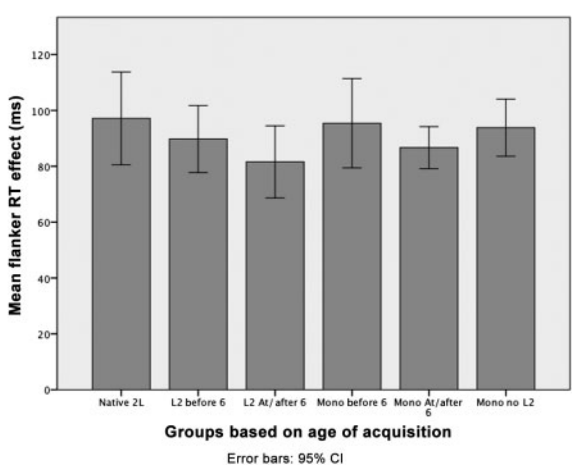

Figure 1. Mean flanker effect (incongruent–congruent) for groups based on age-of-acquisition of L2. Earlier bilinguals on the left.

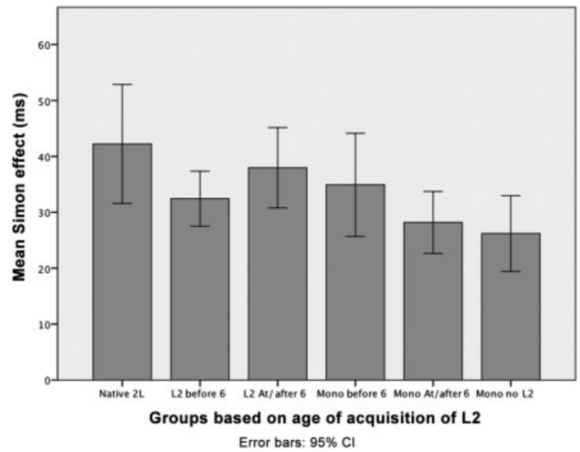

Figure 2. Mean Simon effect (incongruent–congruent) for groups based on age-of-acquisition of L2. Earlier bilinguals on the left.

Colour-shape switching task. Switching costs are computed as the difference in mean RT between the repeat trials and switch trials from a block that randomly mixes the two types of trials. Switching costs are assumed to reflect the cost of actually having to shift from one task to another, but are usually assumed to also require inhibitory control in order to suppress the tendency to respond based on the rule exercised on the preceding trial. As shown in Table 2 there were no significant differences in switching costs across groups.

Mixing costs are computed as the difference in mean RT between blocks of single-task trials and the mean on repeat trials from a mixed block. This measure is assumed to reflect the cost of having to monitor and prepare for a switch without actually having to implement a switch. There was a significant main effect of group; as shown in Figure 3, the group with Native 2L (+150 ms) had smaller mixing costs than the late bilinguals (+338 ms). In a simple t-test this advantage of 188 ms is significant with $p = .005$. The corresponding BF is .142 indicating that the results are seven times more likely under the alternative hypothesis than the null and the evidence in favour of the alternative does fall into the conventional category of "substantial". Using ANOVA and the Games–Howell post-hoc test this is the only pairwise comparison among the 15 comparisons that is significant with $p < .05$. Among the most relevant other comparisons are those between early and late sequential bilinguals (where a significant early advantage was reported by Luk et al.) and between early sequential bilinguals and monolinguals with minimal exposure to an L2 (where a

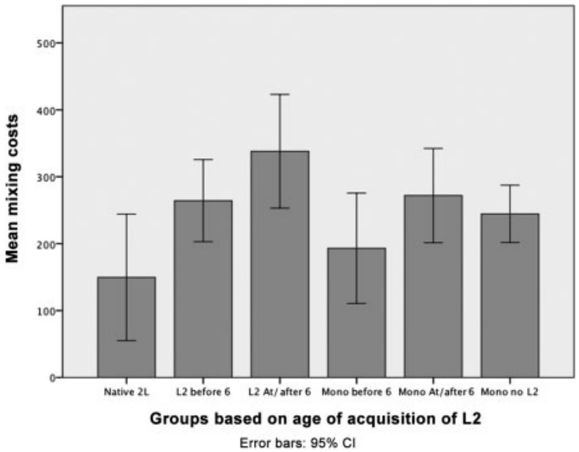

Figure 3. Mean mixing costs (repeat trials–single-task trials) for groups based on age-of-acquisition of L2. Earlier bilinguals on the left.

significant early advantage was reported by Luk et al.). The BFs for these two comparisons are 2.54 and 6.81, respectively. Thus, the overall pattern of BF analyses suggest that there is strong evidence for a bilingual advantage in mixing costs for the *bilinguals with two native languages*, but strong evidence for no differences between *early sequential bilinguals* and monolinguals.

Proficiency of L2 relative to L1

Two different hypotheses were introduced earlier. One assumes that enhancements in EF will simply accumulate as L2 proficiency grows. The other hypothesis assumes that high levels of L2 proficiency are accompanied by more automated and language-specific control mechanisms such that enhanced EF is most likely to occur with speakers who often use a less proficient L2 and engage in ubiquitous controlled suppression of their dominant L1. Both possibilities were tested by partitioning our composite database into five groups (see Table 3) based on the proficiency of participant's L2 relative to L1. The *high proficiency* group was either as fluent as a native speaker (rating = 6) or super fluent (rating = 7) in both languages. The remaining groups rated their L2 proficiency as: a 5 (*Mid Pro*), a 4 (*Low Pro*) a 2 or 3 (*Very Low*) or as either a 0 or 1 (*No Pro*). The mean L2/L1 ratio for each group is shown in bold in Table 3.

As shown in Table 2 there were no differences in the magnitude of the flanker effect across groups. In contrast, the main effect of group on the Simon effect was significant. As shown in Figure 4, the highest proficiency groups (High Pro and Mid Pro) had the largest interference effects (about + 38 ms) and the monolingual group with minimal exposure to an L2 had the smallest (+26 ms). The BF for the comparison of the most proficient bilinguals compared to the monolingual group was 0.342 showing that the probability of observing this difference given the alternative was about 2.9 times greater than given the null. This value is very close to the conventional threshold of "substantial" evidence.

There were no significant differences across groups on the monitoring measures for the flanker task or the Simon task. Likewise, there were no differences in either the magnitude of switching costs or mixing costs.

Although the ANOVAs reported in this section were intended to investigate the potential role of L2 proficiency in enhancing EF they also speak to concerns and questions about the results reported by Paap and Greenberg (2013) and Paap and Sawi (in press) that consistently showed no advantages for bilinguals over monolinguals. Specifically, might our bilinguals be insufficiently bilingual? Or, might our monolinguals not be sufficiently pure? By using the composite database we can maintain large n comparisons but test our most proficient bilinguals (the High Pro group) against the monolinguals with little or no exposure to an

TABLE 3
Language characteristics for L2 proficiency/balance analysis: Mean (*M*) and standard error (*SE*)

		Age		*L1 pro.*		*L2 pro.*		*Balance*		*L1 AoA*		*L2 AoA*		*% English use*	
Group	n	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
B high pro.	120	25	.1	6.64	.05	6.2	.04	**.94**	.00	.04	.04	4.9	0.5	67.6	1.8
B mid pro.	91	24	1.2	6.49	.07	5.0	.00	**.77**	.01	.00	.00	6.4	0.8	74.2	2.8
B low pro.	40	24	.7	6.57	.05	4.0	.00	**.61**	.01	.00	.00	6.2	1.0	77.9	3.1
M very low	49	23	1.1	6.65	.05	2.4	.05	**.09**	.01	.08	.06	9.2	0.6	92.6	1.7
M no pro.	84	25	.8	6.59	.05	0.3	.04	**.07**	.01	.00	.00	–	–	99.8	0.1

B, bilingual; m, monolingual; L1, first language; L2, second language; Pro., rated proficiency; AoA, age-of-acquisition.

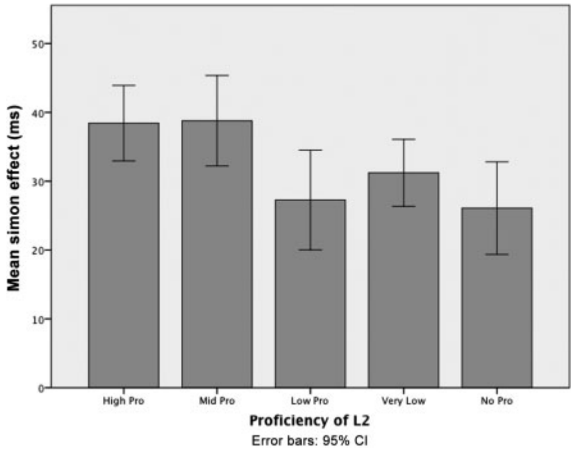

Figure 4. Mean Simon effect (incongruent–congruent) for groups based on proficiency of L2. Most balanced bilinguals on the left.

L2 (the No Pro group). Any significant differences between these two groups should have been detected with the one-way ANOVAs including all five groups, but for completeness the High Pro and No Pro groups were directly compared in a series of *t*-tests, one for each of the 12 measures of EF.

The group means, group differences, *t*-statistics, *p*-values and BF values for each measure are shown in Table 4. Eleven of the 12 differences were non-significant with all 11 *p*s > .20. The only exception was the significant bilingual disadvantage reported earlier for the Simon effect. Another informative aspect of Table 4 is the dance of the *p*-values. As one would expect if the null hypothesis of no language-group differences was true, the *p*-values appear evenly distributed across the range of 0 to 1. Setting aside the monolingual advantage for the Simon effect the BFs range from 2.84 to 8.92 with a mean BF of 6.22. Thus, for 11 measures of EF the comparisons between highly proficient bilinguals and monolinguals with minimal exposure to an L2 the BFs shows substantial support for the null hypothesis over the alternative. The alternative favoured by the data from the Simon effect surprisingly favours monolinguals over bilinguals.

TABLE 4
Differences between highly proficient bilinguals (B) and monolinguals (M) with minimum exposure to an L2 on 12 measures of EF

Task measure of EF	n_B	n_M	*Diff.*	t	p	*BF*
Antisaccade						
RT	38	54	−47.6	−1.05	.237	2.84
RT cost	38	54	+4.8	+0.22	.831	6.03
PC	38	54	−0.01	−0.56	.577	5.32
PC cost	38	54	−0.23	−1.22	.227	3.10
Flanker						
Effect	49	66	−2.6	−0.33	.741	6.52
Mixing costs	49	66	+5.9	+0.46	.649	6.22
Global RT	49	66	−12.6	−0.76	.450	5.24
Simon						
Effect	83	119	−12.3	−2.64	.009	0.32
Mixing costs	83	119	+7.0	+0.99	.326	7.98
Global RT	83	119	−0.7	−0.07	.946	8.92
Switching						
Switch cost	78	115	−6.6	−0.35	.726	8.22
Mixing cost	78	115	−15.1	−0.42	.675	8.01

Difference is calculate such that + represents a bilingual advantage; n_B, number of bilinguals; n_M,, number of monolinguals; *t*, the value of the *t*-statistics; *p*, exact probability; BF, Bayes factor.

Trilinguals

The hypothesis that trilinguals may have enhanced EF compared to bilinguals was tested using the same 12 measures of EF. Thirty-seven participants who self-rated their proficiency as a four or more on at least three languages were classified as trilingual. Most of the trilinguals spoke three quite different languages (e.g., English/Spanish/Tagalog or English/German/French). However, two of the trilinguals spoke both Hindi and Urdu and one of the trilinguals spoke Cantonese and Mandarin. As a group their mean English proficiency was 6.1 (*SD* = 1.1) where a value 6 is as fluent as a typical native speaker. When L1, L2 and L3 are defined in terms of self-rated proficiency, the mean proficiencies for the trilinguals were 6.7 (*SD* = 0.5), 5.9 (*SD* = 0.9) and 4.7 (*SD* = 0.9), respectively. The trilinguals reported that in their current language environment they speak English 64.8% of the time. Their median and modal scale value for switching was 3, *I usually switch from one language to the other a couple of times a day*.

A separate independent groups ANOVA was performed on each of the 12 measures of EF used in the previous analyses. As shown in Table 2 two of the significant omnibus *F*s involved measures from the antisaccade task, namely antisaccade RT and antisaccade costs in proportion correct (PC). In each case the performance of the trilinguals falls in-between that for the monolinguals and bilinguals and consequently provides no evidence that trilingualism enhances inhibitory control compared to bilingualism. Follow-up comparisons showed that, with each measure, there is an advantage for monolinguals over bilinguals, *p* = .035 for PC costs and *p* = .067 for antisaccade RT. However,

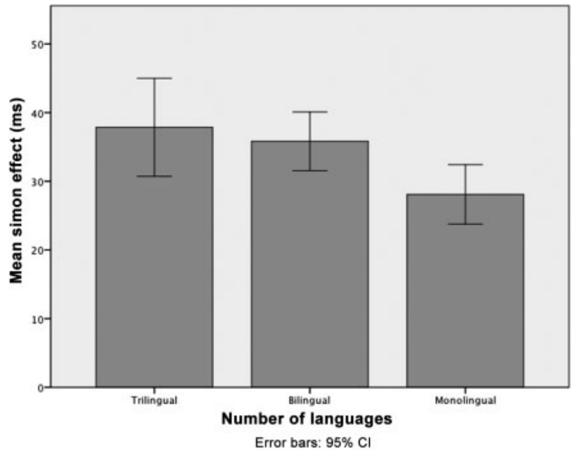

Figure 5. Mean Simon effect (incongruent–congruent) for groups based on number of languages.

corresponding BF values are .411 and .748 suggesting that the evidence for the alternative is only in the "barely worth mentioning" range. The magnitude of the Simon effect also significantly differed across groups, but as shown in Figure 5 there is a monolingual advantage (smaller Simon effect) over both bilinguals (p = .034) and trilinguals (p = .005).

Regression analyses on bilinguals

Rather than partitioning variables like age-of-acquisition of L2 into categories and analyzing for group differences using ANOVA, the same variable can be treated as a continuous predictor in a regression analysis. Stepwise regression analyses were conducted on each of the 12 possible measures of EF shown in Table 2 and using the following four predictors: age-of-acquisition of L2, L2 to L1 balance, percentage of English use and PED. Only participants classified as bilinguals were included in these analyses.

Age-of-acquisition of L2 is intended to capture the early-to-late dimension that was the focus of the Luk et al. investigation. The L2 to L1 balance ratio was defined as the minimum proficiency divided by the maximum proficiency.[4] The balance ratio is highly related to the partitioning of the bilinguals shown in Table 1 that defined the high-, medium- and low-proficiency groups in the ANOVA analyses. The percentage of English spoken was included as a predictor intended to reflect the degree to which both languages are currently active.

The final predictor, PED, is often used as a proxy measure of socio-economic status. In our earlier work, using San Francisco State University (SFSU) students, we have shown repeatedly that PED is not a significant predictor of any of our performance measures and, in fact, the correlation is often near zero (Paap & Greenberg, 2013; Paap & Sawi, in press). Nonetheless, we included PED for its potential role as a covariate because there is no doubt that socio-economic status predicts executive functioning in more diverse populations (e.g., Carlson & Meltzoff, 2008). Our conjecture is that most of the SFSU participants who came from families with low PED benefited from countervailing opportunities and experiences that enabled the development of EF and supported the pathway to higher education.

Only 3 of the 12 models were significant and each of those consisted of a single predictor. Given the importance of Luk et al.'s results showing advantages for early bilinguals in the flanker task it is noteworthy that age-of-acquisition of L2 was the only significant predictor of the flanker effect, standardised $\beta = -.25$, $t(93) = -2.49$, $p = .017$. Critically, the β is negative indicating that as age of acquiring L2 increased the magnitude of the flanker effect decreased. This is opposite from the pattern reported by Luk et al., but consistent with the direction of the non-significant group trends shown in Figure 1.

The second regression model predicting the magnitude of the Simon effect was also significant, but in this case L2 to L1 balance was the only significant predictor; standardised $\beta = +.19$, $t(161) = 2.45$, $p = .015$. The positive β coefficient means that as the degree of balance increases the magnitude of the Simon effect increases. Thus, managing two languages with high proficiency is actually associated with larger interference effects. The regression analysis of the Simon effect is consistent with the group analysis (see Figure 2): bilinguals with high- and mid-level L2 proficiencies had larger Simon effects compared to those with low L2 proficiency and the monolingual groups.

The third significant regression model used "mixing costs" as the outcome variable. The only significant predictor was age-of-acquisition of L2, standardised $\beta = +.21$, $t(149) = +2.62$, $p = .010$. The positive β coefficient shows that as the age of acquiring L2 increases, the magnitude of mixing costs increases. This is consistent with the ANOVA showing that bilinguals with Native 2L had smaller mixing costs compared to late bilinguals.

[4]For participants who spoke more than two languages, the max was the language with the highest rated proficiency and the "min" was the second highest proficiency.

DISCUSSION OF EARLY VERSUS LATE BILINGUALISM

Inhibitory control

The primary impetus for examining age-of-acquisition was Luk et al.'s conclusion that early bilingualism enhances EF, specifically the inhibitory control component. That conclusion was probably premature as the cumulative evidence is far from compelling. The most obvious problem is that the critical finding does not replicate. Humphrey and Valian (2012), Kapa and Colombo (2013) and the present study tested for differences in the magnitude of the flanker effect and found no advantages for early bilinguals compared to either late bilinguals or monolinguals. Tao et al. (2011) and Pelham and Abrams (2014) reported no significant differences in the flanker effects between early and late bilinguals. Finally, our regression analysis treating the magnitude of the flanker effect as the outcome variable revealed a significant negative relationship between age of acquiring L2 and the magnitude of the flanker effect. Thus, for this sample early bilingualism was associated with larger- not smaller-interference effects.

A second potential problem with the Luk et al. study is that it relied on a single task. Evidence that a specific type of linguistic experience hones general-purpose (rather than task-specific) processes is far more compelling if the performance advantages can be produced in two different tasks. The Humphrey and Valian study employed two tasks and our composite database includes data from four tasks. Thus, not only do Humphrey and Valian and the present study fail to replicate the advantage for early bilinguals reported by Luk et al. in the flanker task, but they also failed to show an advantage for early bilinguals in measures of inhibitory control obtained from three other tasks (namely Simon, antisaccade and switching). In the present study bilinguals with Native 2L actually showed a larger Simon interference effect than the pure monolinguals. The other five measures of inhibitory control derived from the antisaccade and switching tasks simply showed no effects of age-of-acquisition.

Monitoring or mental flexibility

Another problem with the conclusion offered by Luk et al. is that it focuses only on the inhibitory control component of EF. Recall that Luk et al. reported significant advantages of early bilinguals over both late bilinguals and monolinguals with respect to the magnitude of the flanker effect, but not for the difference between the congruent trials and the neutral control trials. The latter is usually assumed to reflect differences in monitoring and preparing for conflict. The fact that Luk et al. found advantages of early bilingualism on conflict resolution, but not conflict monitoring complicates their account. Granted, the role of bilingualism in the development of EF may not be parsimonious; however, ignoring the null results for the measure of conflict monitoring appears to be at odds with the current thinking of Bialystok and her collaborators. For example, Kroll and Bialystok (2013) suggested that the key differences between bilinguals and monolinguals are in *coordination* or *mental flexibility* rather than *inhibitory control*. Thus, the special-experiences hypothesis becomes increasing complex, namely that early bilingualism is the critical experience for producing a bilingual advantage in inhibitory control, but other special bilingual experiences must produce the advantage in monitoring.

With respect to the effects of early bilingualism on monitoring the present analyses showed no effects of age-of-acquisition in any of the four measures obtained with the flanker and Simon tasks. The results obtained by Kapa and Colombo (2013) and Tao et al. (2011) are ambiguous. Although both studies showed a significant advantage of early bilinguals over monolinguals in global RT, the differences between early and late bilinguals were not significant.

The only results that were consistent with the hypothesis that early bilinguals are advantaged compared to late bilinguals were the mixing costs obtained in our switching task as the bilinguals with Native 2L had smaller mixings costs compared to the late bilinguals. This trend for an advantage of early bilingualism was also observed in the regression analysis. Although this finding merits attention, there are multiple reasons to interpret the outcome with caution. First, mixing costs are usually assumed to reflect the monitoring component of EF and our set of measures included four other presumed measures of monitoring (namely flanker global RT, flanker congruent – neutral, Simon global RT, Simon congruent – neutral). As discussed earlier there were no differences in age-of-acquisition in any of the ANOVA or regression analyses using these measures of monitoring. Second, although both Prior and MacWhinney (2010) and Prior and Gollan (2011) reported bilingual advantages in *switching* costs, neither found differences in *mixing* costs. Thus, mixing costs are simply an "odd" place for advantages of early bilingualism on the monitoring component of EF to surface.

If the result is not spurious then the overall pattern would suggest a specific resonance between simultaneous bilingualism (but not early sequential bilingualism) and the type of monitoring required in switching tasks (but not the type of monitoring required by interference tasks). Perhaps monitoring for cues to switch tasks is more akin to the bilingual experience of monitoring for cues to switch languages than is the monitoring for potential conflict required by the Simon or flanker task, but it does not seem plausible to argue that positive transfer from language switching to general task switching requires simultaneous bilingualism and that even early acquisition of L2 is too late.

Discussion of role of L2 to L1 proficiency

The way in which the relative balance between L2 and L1 proficiency might affect EF depends on one's preferred hypothesis. As discussed earlier, one view is that maximal benefit should be derived by bilinguals who achieve very high levels of proficiency in both languages and actively use both languages. This hypothesis would be supported if highly proficient and balanced bilinguals showed better performance than both less proficient bilinguals and monolinguals.

An alternative view is that highly fluent and balanced bilinguals are likely to have developed automated and language-specific control mechanisms that do not boost or sustain general EF. This view predicts that bilinguals with a dominant L1 and who actively use both languages are the most likely to show bilingual advantages. This hypothesis would be supported if bilinguals with a dominant L1 showed better performance compared to both balanced bilinguals and monolinguals.

Ten of the 12 measures of EF showed no language-group differences across the five groups. These included all five measures of monitoring and five measures of inhibitory control. There was a main effect of language group in the analysis of the Simon interference effect. However, follow-up comparisons showed that the purest monolingual group with a Simon effect of +26 ms was significantly better than either the high (+38 ms) or low-proficiency (+39) bilinguals. Likewise, follow-up comparisons on the other significant ANOVA—the mean of the antisaccade trials—showed that one of the monolingual subgroups (very low L2 proficiency) was marginally faster compared to two of the bilingual subgroups (with no significant differences between the bilingual subgroups). Obviously bilingual disadvantages cannot support either hypothesis under investigation.

The only language-group difference that shows up consistently across our four studies and in the analysis of our composite database is the monolingual advantage in the Simon task. Note that this advantage does not occur when the same sample of participants are tested with other measures of inhibitory control: the flanker task, antisaccade task or switching costs. This dissociation between the Simon task and the other tasks may be less surprising when viewed in the context of the low levels of convergent validity shown between these tasks (Paap & Sawi, in press). The small cross-task correlations involving the Simon effect and other measures of interference suggest that conflict resolution in the Simon task involves different mechanisms (and perhaps task-specific mechanisms) from those used in other tasks. Even if this is true, we have no explanation for the observed monolingual advantages in the Simon task.

GENERAL DISCUSSION

As Paap and Greenberg (2013) observed, there are two mutually exclusive viewpoints regarding the hypothesis that bilingualism enhances EF. One perspective, "the bilingual advantages are artifacts perspective", assumes that when bilinguals do outperform monolinguals, they are due to causes unrelated to bilingualism enhancing EP. This perspective on reconciling the empirical inconsistencies is to attribute the performance advantages, when they do occur, to factors other than bilingualism enhancing EP.

Before further exploring these two perspectives it would be useful to examine the relative frequency of significant and nonsignificant results and their association with various sample sizes. There is ample evidence that psychological science in general, and research on the bilingual advantage more specifically, is underpowered due to the use of risky small *n*s. As described by Paap (2014), small *n*s dominated the early-published reports of bilingual advantages. A more comprehensive examination of the contribution of small *n*s to the appearance of a steady stream of bilingual advantages is provided in the Appendix and illustrated

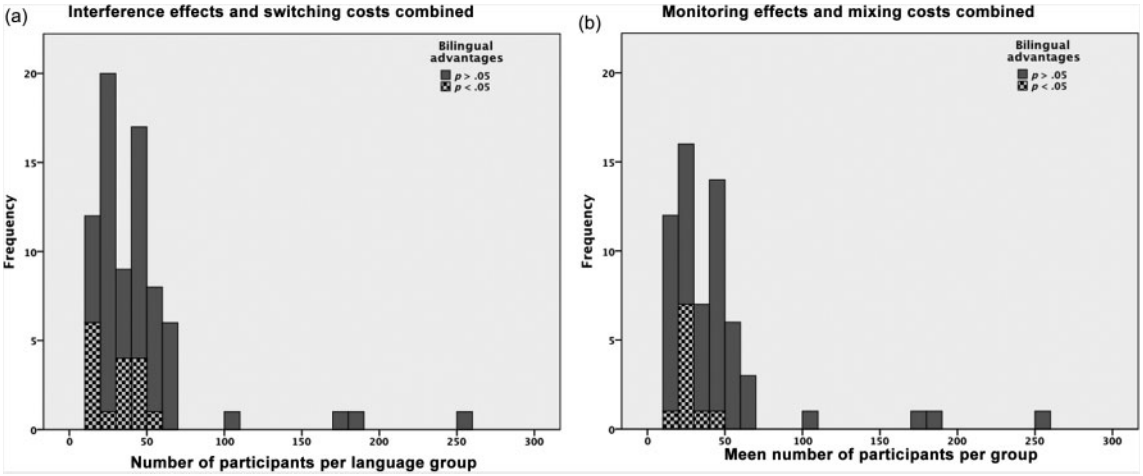

Figure 6. The number of significant ($p < .05$) and nonsignificant ($p > .05$) tests for bilingual advantages in nonverbal interference tasks reported after Hilchey and Klein's (2011) review and for all nonverbal switching tasks. Panel A, on the left, presents the results for interference effects and switching costs and Panel B on the right presents the results for monitoring effects and mixing costs.

in Figure 6(a) and (b). The Appendix shows the outcome of all of the new tests for bilingual advantages using nonverbal interference tasks reported since Hilchey and Klein's review. It also includes all the tests involving nonverbal switching tasks, a set of tasks not systematically reviewed by Hilchey and Klein. The focus of the Appendix is on whether each test resulted in a statistically significant bilingual advantage ($p < .05$) for two reasons. First, Hilchey et al. (in press) present a similar update, but they focus on the effect sizes in individual studies rather than the pattern of statistically significant and nonsignificant results. Each perspective is valuable and our focus on statistical significance is what is needed to evaluate our concern that the prevailing view (i.e., that bilingual advantages in EF are ubiquitous and that the phenomena is well established) has been distorted by the chronic use of small *n*s and a confirmation bias on the part of researchers, reviewers and editors to favour positive results over null results.

Each figure is a histogram constructed from the information in the Appendix showing the frequency of significant and nonsignificant bilingual advantages as a function of the mean number of participants per language group. Figure 6(a) displays the histogram for tests of interference effects and switching costs combined while Figure 6(b) displays the histogram for tests of monitoring and mixing costs combined. It is obvious from visual inspection that bilingual advantages cluster at the low *n* end and are overwhelmed by those not showing bilingual advantages.

This is not the expected pattern if bilingualism truly does enhance EF. If the null is false, then as the sample size becomes arbitrarily large—the *t*-value grows without bound, and the *p*-value converges to zero. This is a good property because it implies that the null will always be rejected in the large-sample limit if there is a real difference to detect. As Rouder et al. (2009) put it: "Researchers, therefore, can rest assured that increasing sample size will, on average, result in a gain of evidence against the null when the null is, indeed, false". Thus, all others things being equal, one would expect significant effects (especially for a small effect size) to cluster at the higher end of sample sizes. Another weakness of NHST and its reliance on *p*-values is that this good property does not hold if the null is true because the *t*-values do not converge to any limit with increasing sample size and all *p*-values are equally likely—that is, they are distributed uniformly between 0 and 1. In fact, this distribution holds regardless of sample size and thus, at all sample sizes a true null is rejected with the probability alpha. The consequence of this fact for researchers using NHST is that increasing the sample size does not lead to a gain in evidence for the null, "because increasing the sample size does not affect the distribution of *p* values" (Rouder et al., 2009, p. 226).

Are bilingual advantages artefacts?

It would seem that the best explanation for the patterns observed in Figure 6 is the combined assumption that the null is true and that there is a bias on the part of researchers and reviewers to prefer positive results over null results as argued by Pashler and Harris (2012) and Paap (2014). One rationale strategy for generating more significant results is to run many small *n* experiments. Figure 6 shows that tests yielding significant bilingual advantages tend to have lower *n*, but is it correct to characterise them as being underpowered? If the effect of bilingualism on EF was generously estimated to be of medium size (Cohen's $d = .5$), if the effect was tested with a two-tailed alpha of .05, and if a researcher was willing to accept a power of only .67, then one would need 48 participants in each of 2 language groups. For the results tallied in Figure 6(a) the mean number of participants per language group for those studies reporting a bilingual advantage is 29 ($SD = 12$).[5] For those not showing a bilingual advantage the mean *n* per group is 45 ($SD = 41$).

Bakker, van Dijk, and Wicherts (2012) explain through simulations that the use of several small underpowered samples, rather than the use of one larger and more powerful sample, is a more efficient strategy for generating significant ($p <$ .05) findings. The funnel plots of the simulations shown in the top row of Bakker et al.'s Figure 4 show the striking differences between multiple small studies and one large study even if the true effect size is zero.

Small *n*s are especially risky when random assignments cannot be used because the comparison of interest is between two naturally occurring populations. In addition, some of the performance advantages may be due to systematic differences between the language groups with respect to culture or immigrant status. These possibilities were raised early on by Morton and Harper (2007) and Carlson and Meltzoff (2008) and more recently by Hilchey and Klein (2011), Kousaie and Phillips (2012a), Paap and Greenberg (2013) and Paap and Liu (2014).

Is there evidence that published research is a biased sample of all research? For all of psychological science Bakker et al. (2012) determined that 96% of research articles report significant results but typical studies are insufficiently powerful for such a track record. More narrowly, De Bruin, Treccani, and Della Sala (2014) reported that only 29% of a set of national or international conference abstracts fully challenging the replicability of bilingual advantages were published compared to 68% fully supporting bilingual advantages. This is probably the tip of the iceberg as most null results were probably placed in the "file drawer" rather than presented at major conferences. Another way in which a bias favouring statistically significant results can enhance the apparent robustness of a phenomena is via the selective reporting of dependent variables and analyses that "work" (Cumming, 2014) in the sense of yielding *p*-values less than .05. Simmons, Nelson, and Simonsohn (2011) refer to these as questionable research practices (QRPs) that are, nevertheless, practiced by a substantial number, if not the majority, of research psychologists (John, Lowenstein, & Prelec, 2012).

A confirmation bias can be very subtle and unintentional. Take for example, a set of three studies that concluded that the results supported *bilingual advantages* in *inhibitory control* (Bialystok, Craik, & Luk, 2008; Salvatierra & Rosselli, 2011; Schroeder & Marian, 2012). In each case the basis for the conclusion was a statistically significant and smaller interference effect for the bilingual group compared to the monolingual group. However, looking at the RTs for each of the four conditions, rather than the difference scores, it is apparent the group differences are due to a monolingual advantage on the non-conflict trials. If bilinguals have superior inhibitory control then the most straightforward prediction is that there should be a Group × Conflict interaction and that the interaction should be driven by monolinguals having longer RTs in comparison to the bilinguals on the conflict trials.[6] The actual forces at work leading to the common conclusion that this pattern of results supports a bilingual advantage in inhibitory control is indeterminate, but they do illustrate outcomes consistent with a confirmation bias favouring bilingual advantages.

The bilingual advantages are artefacts perspective does not rely solely on appeals to Type 1 errors, confirmation biases and inadequately matched groups (see the discussion in Paap, 2014; Paap & Greenberg, 2013; Paap & Liu, 2014). This outlook also raises concerns that the obtained empirical advantages may be task specific and not indicative of a general advantage in EF. For example, since the interference effects in the Simon task do not correlate with those in the

[5]This is not unusual as Wetzels et al. reported that the median cell size in psychology is only 24.

[6]This set of examples was independently described by Hilchey et al. (in press) who raised the same concern based on the same argument.

flanker task or the Stroop task (see Paap & Greenberg, 2013; Paap & Sawi, in press, for a review of the convergent validity problem) any bilingual advantage obtained in these nonverbal interference tasks may be measuring task-specific inhibitory mechanisms, not differences in a domain-general ability in inhibitory control.

Even if bilingual advantages turn out to be artefacts rather than a genuine phenomena, the exercise could have interesting implications for the language processing system. Paap and Greenberg (2013) offered multiple reasons why controlling multiple languages may not enhance general EF. In brief reprise, one possibility is that the additional demands on EF involved in coordinating two languages do not appreciably exceed those involved in speaking a single language. That is, within a single language conversational partners must monitor for signals regarding turn-taking, misunderstandings, possible use of sarcasm and changes of topic or register. Similarly monolinguals must incessantly make choices among activated lexical candidates during production and suppress the irrelevant meaning of homophones during comprehension. If general cognitive control mechanisms are required for the monitoring, switching and conflict resolution required within a single language, then it may be the total amount of language use, not the number of languages that determines the degree to which language enhances EF. This argument can be further extended to suggest that language control is simply one among many challenging tasks (e.g., music performance, computer gaming) that interact in complex ways to optimise individual levels of EF. To try to isolate the contribution of any one task while controlling for all of the others becomes exceedingly difficult and may explain why the transferability of "brain training" remains a controversial topic (Owen et al., 2010).

A second and dramatically different possibility is that controlling a single language or managing multiple languages relies on specialised mechanisms within a specialised module that, therefore, would not enhance domain-independent components of general EF. This line of argument could rely on the strong sense of modularity proposed by Fodor (1983) for specific input-systems like speech perception, but might alternatively embrace novel strategies and new mini-modules, constructed on the fly so to speak, for resolving conflict in new tasks. Despite the gathering neuroscience evidence that the dorsal anterior cingulate cortex is generally involved in conflict monitoring and the allocation of control (Shenhav, Botvinick, & Cohen, 2013) it may be that it is the efficiency of task-specific mechanisms of conflict resolution (regulation in Shenhav et al.'s terminology) that drives individual differences. This type of account would explain the lack of convergent validity between reputed measures of the same component of EF that were discussed earlier and the dismaying report by Salthouse (2010) that the flanker interference effects obtained from 265 participants in 2 very similar versions of the flanker task do not at all correlate with each other (r = +.03).

Could bilingual advantages be real?

The alternative, "the bilingual advantages are real perspective", assumes that there are genuine bilingual advantages that happen to be somewhat elusive given our current understanding of how and why they develop but that the march of cumulative research will eventually clarify the conditions and mechanisms of the bilingual experience that lead to enhancements in EF. Although this may turn out to be the case, the following appears to fairly capture the current state of affairs. *Ideal* bilinguals who acquire both languages early and live in language communities that speak the same two languages are not likely candidates for enhanced EF. As discussed earlier, three recent and large lifespan studies reported no language-group differences (Anton et al., 2014; Duñabeitia et al., 2013; Gathercole et al., 2014).

In contrast to samples of *ideal* bilinguals, the advantage of our composite database is that the participants are homogeneous in terms of their current life experience (undergraduate psychology majors at the same university) and heterogeneous in terms of the other languages they speak, the age of acquiring L2 and their degree of language dominance. Thus, it provided an excellent test bed for investigating whether a critical bilingual experience depends on one of these generic dimensions. It appears that early or late, balanced or not, bilinguals do not systematically differ from monolinguals in measures of EF. One cannot, of course, rule out the possibility that there exists some other special aspect (or combination of aspects) of bilingualism that reliably leads to enhanced EF, but current trends make it likely that this would involve only a small proportion of those who consider themselves bilingual and that those benefits might be detectable only in a quite restricted set of tasks or tests.

Original manuscript received May 2014
Revised manuscript received July 2014
Revised manuscript accepted July 2014
First published online August 2014

REFERENCES

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., … Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, *22*, 2076–2086. doi:10.1093/cercor/bhr287

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., … Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology: Language Sciences*, *5*, 1–12. doi:10.3389/fpsyg.2014.00398

Bak, T. H., Nissan, J. J., Allerhand, M. M., & Deary, I. J. (2014). Does bilingualism influence cognitive aging? *Annals of Neurology*, *75*, 959–963.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives in Psychological Science*, *7*, 543–554. doi:10.1177/1745691612459060

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Development*, *83*, 413–422.

Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development*, *11*, 485–508. doi:10.1080/15248372.2010.516420

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

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*, 240–250. doi:10.1016/j.tics.2012.03.001

Bilig, J. D., & Scholl, A. P. (2011). The impact of bilingualism and aging on inhibitory control and working memory. *Organon*, *51*, 39–52.

Branzi, F. M., Martin, C. D., Abutalebi, J., & Costa, A. (2014). The after-effects of bilingual language production. *Neuropsychologia*, *52*, 102–116. doi:10.1016/j.neuropsychologia.2013.09.022

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

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*(1), 59–86. doi:10.1016/j.cognition.2006.12.013

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*, 491–511. doi:10.1016/j.jml.2004.02.002

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly-proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1057–1074. doi:10.1037/0278-7393.32.5.1057

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. doi:10.1177/0956797613504966

DeBruin, A., Treccani, B., & Della Sala, S. (2014). *Cognitive advantage in bilingualism: An example of publication bias*. Manuscript submitted for publication.

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2013). The inhibitory advantage in bilingual children revisited. *Experimental Psychology*, *61*, 234–251.

Engel de Abreu, P. M. J., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor: Enhanced cognitive control in low-income minority children. *Psychological Science*, *23*, 1364–1371. doi:10.1177/0956797612443836

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. doi:10.1073/pnas.080070897

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

Francis, W. S., & Strobach, E. N. (2013, November). *Prediction of objective bilingual language proficiency: Can self-report measures suffice?* Paper presented at the 54th Annual Meeting of the Psychonomic Society. Toronto, Canada.

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. doi:10.1037/0096-3445.137.2.201

Garbin, G., Sanjuan A., Forn, C., Bustamante J. C., Rodriguez-Pujada, A., Belloch, V., … Avila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *Neuroimage*, *53*, 1272–1278. doi:10.1016/j.neuroimage.2010.05.078

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N. Vinas-Guasch, R., … Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*, 11.

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *Journal of Neuroscience*, *33*, 387–396. doi:10.1523/JNEUROSCI.3837-12.2013

Goldman, M. C., Negen, J. & Sarnecka, B. W. (2014). Are bilingual children better at ignoring perceptually misleading information? A novel test. *Developmental Science*. Advance online publication. doi:10.1111/desc.12175

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish–English bilinguals. *Neuropsychology*, *16*, 562–576. doi:10.1037/0894-4105.16.4.562

Green, D. W. (2011). Language control in different contexts: The behavioural ecology of bilingual speakers. *Frontiers in Psychology: Language Sciences*, *2*(103), 1–4. doi:10.3389/fpsyg.2011.00103

Grogan, A., Jones, O. P., Ali, N. Crinion, J., Orabona, S., Mechias, M. L., … Price, C. J. (2012). Structural correlates for lexical efficiency and number of languages in non-native speakers of English. *Neuropsychologia*, *50*, 1347–1352. doi:10.1016/j.neuropsychologia.2012.02.019

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

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. doi:10.3758/s13423-011-0116-7

Hilchey, M. D., Saint-Aubin, J., Klein, R. M. (in press). Does bilingual exercise enhance cognitive fitness in non-linguistic executive processing tasks. In J. W. Schwieter (Ed.), *Cambridge handbook of bilingual processing*. Cambridge University Press.

Humphrey, A. D., & Valian, V. V. (2012, November). *Multilingualism and cognitive control: Simon and flanker task performance in monolingual and multilingual young adults*. Paper presented at the 53rd Annual Meeting of the Psychonomic Society. Minneapolis, MN.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*, *23*(5), 524–532.

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*, 169–183. doi:10.1037/0096-3445.130.2.169

Kapa, L. L., & Colombo, J. (2013). Attentional control in early and later bilingual children. *Cognitive Development*, *28*, 233–246. doi:10.1016/j.cogdev.2013.01.011

Kirk, N. W., Fiala, L., Scott-Brown, K., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology*. Advance online publication. doi:10.1080/20445911.2014.929580

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. doi:10.1016/j.brainres.2012.01.052

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*, 356–369. doi:10.1080/17470218.2011.604788

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*, 497–514.

Kroll J. F., & Rossi, E. (2013, November). *How the mind and brain negotiate competition for selection in bilingual speech*. Paper presented at the 54th Annual Meeting of the Psychonomic Society. Toronto, Canada.

Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, *14*, 588–595. doi:10.1017/S1366728911000010

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940–967. doi:10.1044/1092-4388(2007/067)

Marzecová, A., Asanowicz, D., Krivá, L., & Wodniecka, Z. (2013). The effects of bilingualism on efficiency and lateralization of attentional networks. *Bilingualism: Language and Cognition*, *16*, 608–623. doi:10.1017/S1366728912000569

Marzecová, A., Bukowski, M., Correa, Á., Boros, M., Lupiáñez, J., & Wodniecka, Z. (2013). Tracing the bilingual advantage in cognitive control: The role of flexibility in temporal preparation and categorical switching. *Journal of Cognitive Psychology*, *25*, 586–604. doi:10.1080/20445911.2013.809348

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Fackowiak, R. S., & Price, C. J. (2004). Structural plasticity in the bilingual brain: Proficiency in a second language and age at acquisition affect grey-matter density. *Nature*, *431*, 757. doi:10.1038/431757a

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*(1), 40–48. doi:10.1037/0021-843X.110.1.40

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. doi:10.1177/0963721411429458

Mor, B., Yitzhaki-Amsalem, S., & Prior, A. (2014). The joint effect of bilingualism and ADHD on executive functions. *Journal of Attention Disorders*. Advance online publication. doi:10.1177/1087054714527790

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114*, 187–202. doi:10.1016/j.jecp.2012.09.002

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*, 719–726. doi:10.1111/j.1467-7687.2007.00623.x

Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism*, *13*, 597–616. doi:10.1080/13670050.2010.488288

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., … Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*, 775–778. doi:10.1038/nature09042

Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology*, *26*, 242–255. doi:10.1080/20445911.2014.891597

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

Paap, K. R., & Liu, Y. (2014). Conflict resolution in sentence processing is the same for bilinguals and monolinguals: The role of confirmation bias in testing for bilingual advantages. *Journal of Neurolinguistics*, *27*(1), 50–74. doi:10.1016/j.jneuroling.2013.09.002

Paap, K. R., & Sawi, O. (in press). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology: Language Sciences*.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*, 531–536. doi:10.1177/1745691612463401

Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 313–325. doi:10.1037/a0035224

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*, 535–551. doi:10.1016/j.jecp.2012.06.013

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society*, *17*, 682–691. doi:10.1017/S1355617711000580

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*, 253–262. doi:10.1017/S1366728909990526

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374. doi:10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237. doi:10.3758/PBR.16.2.225

Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language*, *74*, 46–76. doi:10.1016/j.jml.2014.04.003

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*, 532–545. doi:10.1037/0894-4105.19.4.532

Salthouse, T. A. (2010). Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. *Brain and Cognition*, *73*(1), 51–61. doi:10.1016/j.bandc.2010.02.003

Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, *36*, 464–486. doi:10.1016/j.intell.2007.10.003

Salvatierra, J. L., & Rosselli, M. (2011). The effect of bilingualism and age on inhibitory control. *International Journal of Bilingualism*, *15*(1), 26–37. doi:10.1177/1367006910371021

Schroeder, S. R., & Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology*, *24*, 591–601. doi:10.1080/20445911.2012.669367

Shenhav, A., Botvinick, M. W., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*, 217–240. doi:10.1016/j.neuron.2013.07.007

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632

Tao, L., Marzecová, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: The role of age of acquisition. *Frontiers in Psychology*, *2*, 123. doi:10.3389/fpsyg.2011.00123

Tare, M., & Linck, J. (2011, November). *Exploring bilingual cognitive advantages when controlling for background variables*. Poster presented at the 52nd Annual Meeting of the Psychonomic Society, Seattle, WA.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science*, *6*, 291–298. doi:10.1177/1745691611406923

Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition*, *14*, 412–422. doi:10.1017/S1366728910000611

Yudes, C. Macizo, P., & Bajo, T. (2011). The influences of expertise in simultaneous interpreting on non-verbal executive processes. *Frontiers in Psychology*, *2*, 1–9. doi:10.3389/fpsyg.2011.00309

APPENDIX

Study	*Task*	*B*, n	*M*, n	*Interference*	*Monitoring*	*Switch costs*	*Mixing costs*
Abutalebi et al. (2012)							
Block 1	Flanker	17	14	ns	ns		
Block 2	Flanker	17	14	B+	ns		
Anton et al. (2014)	Flanker	180	180	ns	ns		
Barac and Bialystok (2012)							
M vs. Chinese–English B	Switching	30	26			ns	B+
M vs. French–English B	Switching	28	26			ns	B+
M vs. Spanish–English B	Switching	20	26			ns	B+
Bialystok, Barac, Blaye, and Poulin-Dubois (2010)	Flanker	29	46	ns	ns		
Bilig and Scholl (2011)							
Older adults	Simon arrow	21	21	ns			
Young adults	Simon arrow	21	20	ns			
Dunabetia et al. (2013)	Numerical congruency	252	252	ns	ns		
Engel de Abreu et al. (2012)	Flanker	40	40	B+	B+		
Garbin et al. (2010)	Switching	21	19			B+	ns
Gathercole et al. (2014)							
Preschool	Simon	148	60	ns	M+		
School age and above	Simon	274	73	ns	ns		
Gold, Kim, Johnson, Kryscio, and Smith (2013)							
Exp. 1, older adults	Switching	15	15			ns	B+
Exp. 2, younger adults	Switching	20	20			ns	ns
Exp. 2, older adults	Switching	20	20			ns	ns
Goldman et al. (2014)	Number discrimination	46	46	ns			
Hernández et al. (2013)							
Exp. 1, Implicit cue	Switching	50	50			ns	ns
Exp. 1, Explicit cue	Switching	37	37			ns	ns
Exp. 2	Switching	20	21			ns	B+
Exp. 3	Switching	38	39			ns	ns
Humphrey and Valian (2012)							
Flanker	Flanker	32	49	ns	ns		
Simon	Simon	32	49	ns	ns		
Kapa and Colombo (2013)							
M vs. early B	Flanker	21	22	ns	B+		
M vs. late B	Flanker	36	22	ns	ns		
Kirk et al. (2014)	Simon	16	16	ns	ns		
Kousaie and Phillips (2012a)							
Simon	Simon	25	26	ns	ns		
Flanker	Flanker	25	26	ns	ns		
Luk et al. (2011)							
M vs. early B	Flanker	43	38	B+	ns		
M vs. late B	Flanker	42	38	ns	ns		
Marzecová et al. (2013)	Flanker	18	17	B+	ns		
Mor et al. (2014)							
Number Stroop	Number Stroop	20	20	ns	ns		
Simon arrows	Simon arrows	20	20	ns	ns		
Switching	Switching	20	20			ns	ns
Morales, Calvo, and Bialystok (2013)							
Study 1: Simon 5-year-olds	Simon RT	27	29	ns	B+		
Study 1: Simon+ 5-year-olds	Simon+	27	29	ns	B+		
Namazi and Thordardottir (2010)							
French–English B vs. English M	Simon	15	15	ns	ns		
French–English B vs. French M	Simon	15	15	ns	ns		

(*Continued*)

Study	*Task*	*B*, n	*M*, n	*Interference*	*Monitoring*	*Switch costs*	*Mixing costs*
Paap and Greenberg (2013)							
Study 1: antisaccade RT	Antisaccade	34	46	ns			
Study 1: antisaccade accuracy	Antisaccade	34	46	ns			
Study 1: Simon	Simon	34	46	ns	ns		
Study 1: switching	Switching	34	46			ns	ns
Study 2: Simon	Simon	36	50	ns	ns		
Study 2: switching	Switching	36	50			ns	ns
Study 3: flanker block 1	Flanker	52	55	ns	ns		
Study 3: flanker block 2	Flanker	52	55	ns	ns		
Study 3: flanker block 3	Flanker	52	55	ns	ns		
Study 3: switching	Switching	52	55			ns	ns
Paap and Sawi (in press)							
Antisaccade RT	Antisaccade	58	62	M+			
Flanker	Flanker	58	62	ns	ns		
Simon	Simon	58	62	M+	M+		
Switching	Switching	58	62			ns	ns
Pelham and Abrams (2014)							
M vs. early B	Flanker	30	30	B+	ns		
M vs. late B	Flanker	30	30	B+	ns		
Poarch and Van Hell (2012)							
B vs. M: Exp. 1	Simon	18	20	ns	ns		
Trilingual vs. M: Exp. 1	Simon	18	20	B+			
Prior and Gollan (2011)							
Spanish–English B vs. M	Switching					B+	ns
Chinese–English B vs. M	Switching					ns	ns
Prior and MacWhinney (2010)							
Switching	Switching	44	44			B+	ns
Flanker	Flanker	44	44	ns	ns		
Simon	Simon	44	44	ns	ns		
Ryskin et al. (2014)							
Exp. 2: flanker	Flanker	20	21	ns			
Exp. 2: antisaccade	Antisaccade	20	21	ns			
Salvatierra and Rosselli (2011)							
Younger Simon	Simon	67	66	ns			
Older Simon	Simon	58	42	B+			
Younger Simon+	Simon+	67	66	ns			
Older Simon+	Simon+	58	42	ns			
Schroeder and Marian (2012)	Simon	18	18	B+	ns		
Tare and Linck (2011)							
Switching	Switching	35	35	ns			
Antisaccade	Antisaccade	35	35	ns			
Tao et al. (2011)							
M vs. Early B	Flanker	36	34	B+	B+		
M vs. Late B	Flanker	30	34	B+	ns		
Yang et al. (2011)	Flanker	15	13	B+	B+		
Yudes et al. (2011)	Simon	16	16	ns	ns		

Null hypothesis tests (at $p < .05$) for bilingual advantages using nonverbal interference tasks since Hilchey and Klein's (2011) review and for all nonverbal switching tasks.

B, bilinguals; M, monolinguals; B+, bilingual advantage; M+, monolingual advantage; ns, nonsignificant; *n*, number of participants per language group.