Cultural effects rather than a bilingual advantage in cognition: A review and an empirical study.

Steven Samuel[a,b]

Karen Roehr-Brackin[c]

Hyensou Pak[d]

Hyunji Kim[e]

[a] Dept. of Psychology, University of Essex, U.K.

[b] Dept. of Psychology, University of Cambridge, U.K.

[c] Dept. of Language and Linguistics, University of Essex, U.K.

[d] ALLICE, Yeungnam University, South Korea

[e] Faculty of Psychology, University of Vienna, Austria

Word count: 9013 (excluding abstract, acknowledgements, references and tables, including footnotes and captions)

Key words: Bilingual advantage, Code switching, Inhibitory control, Korean culture, Simon task

Address for correspondence: Steven Samuel. University of Cambridge, Department of Psychology, Downing Street, CB2 3EB. Tel: 01223 333563. Email: ss2391@cam.ac.uk

Abstract

The bilingual advantage hypothesis contends that the management of two languages in the brain is carried out through domain-general mechanisms, and that bilinguals possess a performance advantage over monolinguals on (non-linguistic) tasks that tap these processes. Presently, there is evidence both for and against such an advantage. Interestingly, the evidence in favor has been thought strongest in children and older adults, leading some researchers to argue that young adults might be at peak performance levels, and therefore bilingualism is unable to confer an improvement. We conducted a large-scale review of the extant literature and found that the weight of research pointed to an absence of positive evidence for a bilingual advantage at any age. We next gave a large number of young adult participants a task designed to test the bilingual advantage hypothesis. Reasoning from the literature that young adults from an East Asian (Korean) culture would likely outperform those from a Western (British) culture, we also compared participants on this factor. We found no evidence for a bilingual advantage, but did find evidence for enhanced performance in the Korean group. We interpret these results as further evidence against the bilingual advantage hypotheses.

1. Introduction

In recent decades, the hypothesis that bilingualism confers performance advantages in aspects of cognitive functioning has received some theoretical and empirical support (e.g. Abutalebi et al., 2012; Bialystok, 1999, 2010; Bialystok, Craik, Klein & Viswanathan, 2004; Bialystok, Martin & Viswanathan, 2005; Costa, Hernández, Costa-Faidella & Sebastián-Gallés, 2009; Green & Abutalebi, 2013; Kroll & Bialystok, 2013; Martin-Rhee & Bialystok, 2008). The bilingual advantage hypothesis is founded on the premise that bilinguals differ from monolinguals in that they cannot produce language without first selecting which one to use[1] (Abutalebi et al., 2012), and that the way that this is managed is through domain-general mechanisms – those mechanisms that are recruited to perform a variety of different tasks – rather than mechanisms that are specific to language processing (e.g. Green, 1998; Green & Abutalebi, 2013; see Bialystok, 2017, for a review). Research has traditionally focussed on two such mechanisms: *inhibition* (the ability to ignore salient but irrelevant information in order to select a target outcome; e.g. Bialystok et al., 2005), which is engaged in order to ensure the non-target language does not intrude, and *monitoring*, engaged to check for changes in linguistic context so that a bilingual can adapt their language choices to different interlocutors (Costa et al., 2009). These two processes map broadly onto two of the components (inhibition and updating) of executive function (EF) proposed by Miyake and colleagues (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Crucially, it is the extra practice that bilinguals have with these domain-general processes that are believed to lead to

---

[1] Bimodal (verbal and signing) language users are an exception in that they can produce both simultaneously.

collateral benefits in other, non-linguistic tasks. Despite revisions in the theoretical frameworks that are thought to underpin a bilingual advantage (e.g. Costa et al., 2009; Hilchey & Klein, 2011; Grundy, Yim, Friesen, Mak, & Bialystok, 2017; Morales, Calvo, & Bialystok, 2013; Zhou & Krott, 2016b) and our ever-evolving understanding of EF itself (Paap et al., 2017), the central concept of a domain-general process (or processes) that is shared by both bilingual language management and non-linguistic tasks is a constant in the literature (e.g. Bialystok, 2017).

The bilingual advantage hypothesis has thus typically been tested by comparing groups of bilinguals and monolinguals on tasks that tap inhibition and monitoring. For example, bilingual children have been reported to outperform monolingual children on the Sun/Moon task (Bialystok, 1986), the Dimension Change Card Sort task (Bialystok, 1999), the ambiguous figures and opposite worlds tasks (Bialystok & Shapero, 2005), the Simon task (Martin-Rhee & Bialystok, 2008), and false belief task (Kovacs, 2009). In older adults, advantages have been reported on the Simon task (Bialystok et al., 2004; Salvatierra & Rosselli, 2010) and Spatial Stroop (also knowns as Simon Arrows) task (Bialystok et al., 2004). In younger adults, advantages have been reported on numerous Stroop-like tasks, which require the inhibition of salient but misleading information, such as the flanker component of the Attentional Network Task (ANT) (e.g. Costa, Hernández, & Sebastián-Gallés, 2008; Grundy, Yim, Friesen, Mak, & Bialystok, 2017; Zhou & Krott, 2016b), the numerical Stroop task (Costa, Fuente, Vivas, & Sebastián-Gallés, 2010), the colour-shape task (Prior & MacWhinney, 2010), the Simon Task (Bialystok, Craik, Klein & Viswanathan, 2004) and Spatial Stroop task (Blumenfeld & Marian, 2014).

Yet despite the evidence reported in favour of an advantage, recent investigations have cast doubt over whether bilinguals really do outperform monolinguals on these tasks, and whether it is necessarily bilingualism that is the reason when they do (e.g. Antón et al., 2014; Antón, García, Carreiras, & Duñabeitia, 2016; Duñabeitia et al., 2014; Hilchey, Saint-Aubin, & Klein, 2015; Morton & Harper, 2007; Paap, 2014; Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2014, 2015, 2016; Paap et al., 2017). For example, large-scale studies of bilingual and monolingual children failed to find any evidence of an advantage on the ANT (Antón et al., 2014). In older adults, bilinguals have failed to show an advantage over monolinguals on numerical and verbal Stroop tasks (Antón et al., 2016) and a Simon task (Bialystok, Craik, Luk & Grady, 2015). Paap and Greenberg (2013) and Paap and Sawi (2014) failed to find evidence of an advantage in young adults on the Simon task, and indeed found some evidence of a *monolingual* advantage instead. In a particularly large study with over 500 participants, including children, younger adults and older adults, Gathercole and colleagues (2014) found no convincing evidence of a bilingual advantage on the Simon task.

These discrepant findings have been difficult to square. Some have suggested that Hawthorne effects, whereby participants' knowledge of the expertise of a particular lab might bring their performance in line with expectations, explaining why some research groups that are well-known for investigating bilingual advantages are those that appear most likely to find evidence for them (Donnelly et al., 2015). Others have argued that publication bias in favour of studies supporting an advantage leads to an underestimation of the number of null results that have actually been found—a 'file drawer' effect (de Bruin et al., 2015). Coupled with the continuing theoretical debate surrounding precisely *how* bilingualism can bring about enhanced performance, and what kind or kinds of performance advantage should be expected (e.g. Green & Abutalebi, 2013; Bialystok, 2010; 2017; Blumenfeld & Marian, 2014; Costa et al., 2009; Grundy et al., 2017), the picture being painted surrounding the bilingual advantage hypothesis—an endeavour already fraught with complexity given the interplay of life experiences that could bring to bear upon such a phenomenon (Calvo & Bialystok, 2014)—appears increasingly blurred.

In the present study, we attempt to deal with two of the major sticking points in the literature. Firstly, we examine the crucial assumption that underpins whether, as some have suggested, the relative dearth of evidence for bilingual advantages in *younger* adult samples (as opposed to children and older adults) can be explained by peak performance levels at this age. Specifically, according to the 'peak performance' hypothesis, this pattern might be ascribed to the fact that young adults are "…already in control of efficient processing" (Bialystok, Martin, & Viswanathan, 2005, p.117), and hence an absence of evidence for a bilingual advantage in this age group does not represent a challenge to the bilingual advantage hypothesis (see also Bialystok, 2017; Bialystok, Craik, & Luk, 2008; Kroll, & Bialystok, 2013; Rodriguez-Fornells et al., 2012). However, to our knowledge, there has been no systematic investigation of the claim that bilingual advantages are found in children and older adults but not in younger adults. Consequently, the notion that young adults cannot show an advantage owing to their being at peak performance remains a hypothesis founded on an impression rather than, as a first step, with a review of the literature. Given that this hypothesis is crucial to how we interpret a wealth of research into bilingual advantages, it is equally crucial to establish whether such a hypothesis is in fact *necessary*. Secondly, we investigate whether factors other than bilingualism might account for results interpreted as indexing a bilingual advantage. In the present study we focus on *culture*.

1.1. Testing the claim that young adults do not show bilingual advantages.

First, we reviewed the evidence for a U-shaped curve in bilingual advantages as they relate to age. We therefore conducted a review of research that has used the three main tasks that have been identified by Zhou and Krott (2016a) as the most common in the field, namely the Simon task (Simon, 1969), the Spatial Stroop (e.g. Blumenfeld & Marian, 2014), and the Flanker Task (Eriksen & Eriksen, 1974), the latter of which often occurs as part of an ANT

(Fan, McCandliss, Sommer, Raz, & Posner 2002)[2]. In the Simon task (see Hommel, 2011) participants see squares appear one at a time on the left or right of a screen and are asked to respond not according to the square's spatial location but rather according to its colour by pressing a button aligned to either the left or right. When the position of the square on the screen and the location of the button to be pressed coincide, that trial is a congruent trial. When they do not, it is incongruent. The difference in response times between these two trial types (congruent trials minus incongruent trials) is known as the Simon Effect, and is generally thought to measure inhibitory control through the need to suppress the irrelevant spatial information contained in the stimulus (a smaller effect indicates better inhibition). The Spatial Stroop task functions in a similar manner, except that the squares are replaced by arrows that point either left or right. The participant is required to indicate the direction of the arrow while ignoring its location on the screen. Finally, in the Flanker task participants are asked to provide a left/right response according to a centrally-placed stimulus flanked by distractors, such that on incongruent trials the flankers contradict the target (e.g. < < > < <) but on congruent trials they do not (e.g. > > > > >). Crucially, each of these tasks requires inhibition of salient but task-irrelevant information, and it is a robust finding that it takes longer on average to respond on an incongruent trial than a congruent one. The tasks also require monitoring skills in order to adapt to unpredictable trial-by-trial changes, as trial presentations are either random or perceived to be so by the participants. An advantage in inhibition is typically indicated by a decreased Simon (or related conflict) Effect, and an

---

[2] For a recent review and test of bilingualism and *task-switching* experiments specifically, which better maps onto Miyake et al.'s (2000) early idea of *shifting*, see Paap et al. (2017). We focus here on the tasks that are viewed as tapping the inhibition and monitoring components of EF.

advantage in monitoring is typically seen to be evidenced by faster responding over the task as a whole, aggregated over congruent and incongruent trials.

For our review, we first selected those studies found in recent meta-analyses and reviews (Donnelly, Brooks, & Homer, 2015; Hilchey & Klein, 2011; Lehtonen et al., 2018; Paap, Johnson, & Sawi, 2015; Sanchez-Azanza, López-Penadés, Buil-Legaz, Aguilar-Mediavilla, & Adrover-Roig, 2017; Zhou & Krott, 2016). We included only studies that i) are published (excluding for example doctoral theses); ii) include one or more of the Simon Task, Flanker task (as part of the ANT or separately), or Spatial Stroop task; iii) include a healthy monolingual as well as a healthy bilingual group, or at least included a reportedly bilingual sample that incorporated participants who would be classified as monolingual (e.g. Tse & Altarriba, 2014); iv) measure or define language proficiency in a quantifiable (e.g. numeric) way, or at least descriptively in terms of competence, but not in terms of national-level exam passes (e.g. Wang, Fan, Liu, & Cai, 2016); v) are *not* re-analyses of data reported in an earlier paper (e.g. Calabria, Hernández, Martin & Costa, 2011); and vi) did not adapt the task or stimuli such that it made qualitatively different demands on participants relative to the original design (e.g. the LANT used by Marzecová and colleagues, 2013), or any instance of linguistic stimuli rather than the standard shapes or symbols (e.g. Rubio-Fernández & Glucksberg, 2012). To ensure an up-to-date assessment of the literature, we also conducted a search using Web of Science, using the two obligatory search terms “Bilingual” and “Monolingual”, and then adding one at a time each of “Simon”, “Flanker”, Spatial Stroop”, “Arrows”, and “ANT”. We then conducted the same searches once more without the term “Monolingual” to ensure we included tasks where no explicit monolingual group was present but the study included a meaningful amount of data from participants that were closer to the monolingual end of the spectrum. We set as the publication period the year 2014 to the present. These revealed a further 17 studies not included in prior reviews. We also added a total of four further studies that were a fit for our criteria but did not come up in previous reviews or in our Web of Science search, but which were known to us and published in reputable journals (Antoniou, Grohmann, Kamabanaros & Katsos, 2016 [Cognition]; Arredondo, Hu, Satterfield, & Kovelman, 2015 [Developmental Science]; De Cat, Gusnanto, & Serratrice, 2017 [Studies in Second Language Acquisition]; Woumans, Surmont, Struys, & Duyck, 2016 [Language Learning])[3]. In our review, we extracted data relating to the languages and nationalities of participants, the country of testing, participant age, sample size, and number of trials. For the ANT, we restrict our review to results relating to the flanker task specifically (ignoring effects of cueing). Next, we classified the results of each experiment according to a simple 'yes' (bilingual advantage supported), 'no' (bilingual advantage not supported), or 'mixed' based on reports of smaller conflict effects and faster global speeds (inhibition and monitoring respectively). We operationally defined the 'mixed' category as either i) only statistically marginal results, ii) conflict effect reductions that were driven not by enhanced performance on incongruent trials but by poorer performance on congruent trials, iii) results that showed an advantage on one measure but a disadvantage on another, or iv) advantages in accuracy alone. Any cases of bilingual *disadvantages* were classified as 'no'. A young adult sample was defined as having a mean age range within the 18-40 bracket. Sample means younger or older than this were classified as children and older adults respectively. The experiments making up this review are described in the supplemental material (S1). A total of 9798 participants across 124 Simon (n = 57), Flanker (n = 48), or Spatial Stroop (n = 19) tasks were included in the review. The first author coded the studies

[3] One of these studies offered support for a bilingual advantage, two offered mixed evidence, and one no evidence; thus their inclusion did not favor either the null or alternative hypotheses but did provide a fuller picture of the data to date.

for the review in the first instance, noting whether they offered support, mixed support, or no support for the bilingual advantage hypothesis, as well as the ages of participants tested (and the age category that they thus fell into), the task they performed (and the number of trials in that task), the languages of the participants and the country of testing, and the sizes of each group. Then two further coders (co-authors) did the same for 20% of the studies, finding 100% agreement with the first coder. We wish to note however that this form of review is at best *suggestive*—a first step— rather than comprehensive; the most informative approach would also incorporate a larger-scale statistical analysis of effect sizes (see the Discussion section for further comment).

Figure 1 displays the findings of the review. Specifically, panel 1A displays the gross number (count) of tasks that were classified as offering either positive support, mixed support, or rejecting the bilingual advantage hypothesis. Panel 1B displays the same data in terms of percentages of all tasks conducted with that age group. Panel 1C displays the total number (count) of participants in those studies, and panel 1D displays the same data in terms of percentages. Overall, and contrary to the *premise* for the peak performance hypothesis, the results of the review do not suggest clear support for a bilingual advantage in any age group. In fact, less than 35% of participants of any age group took part in tasks whose results supported an advantage. In sum, the review shows no clear evidence supporting a bilingual advantage regardless of age, and hence the peak performance hypothesis appears not to fit the pattern of results found in the literature.

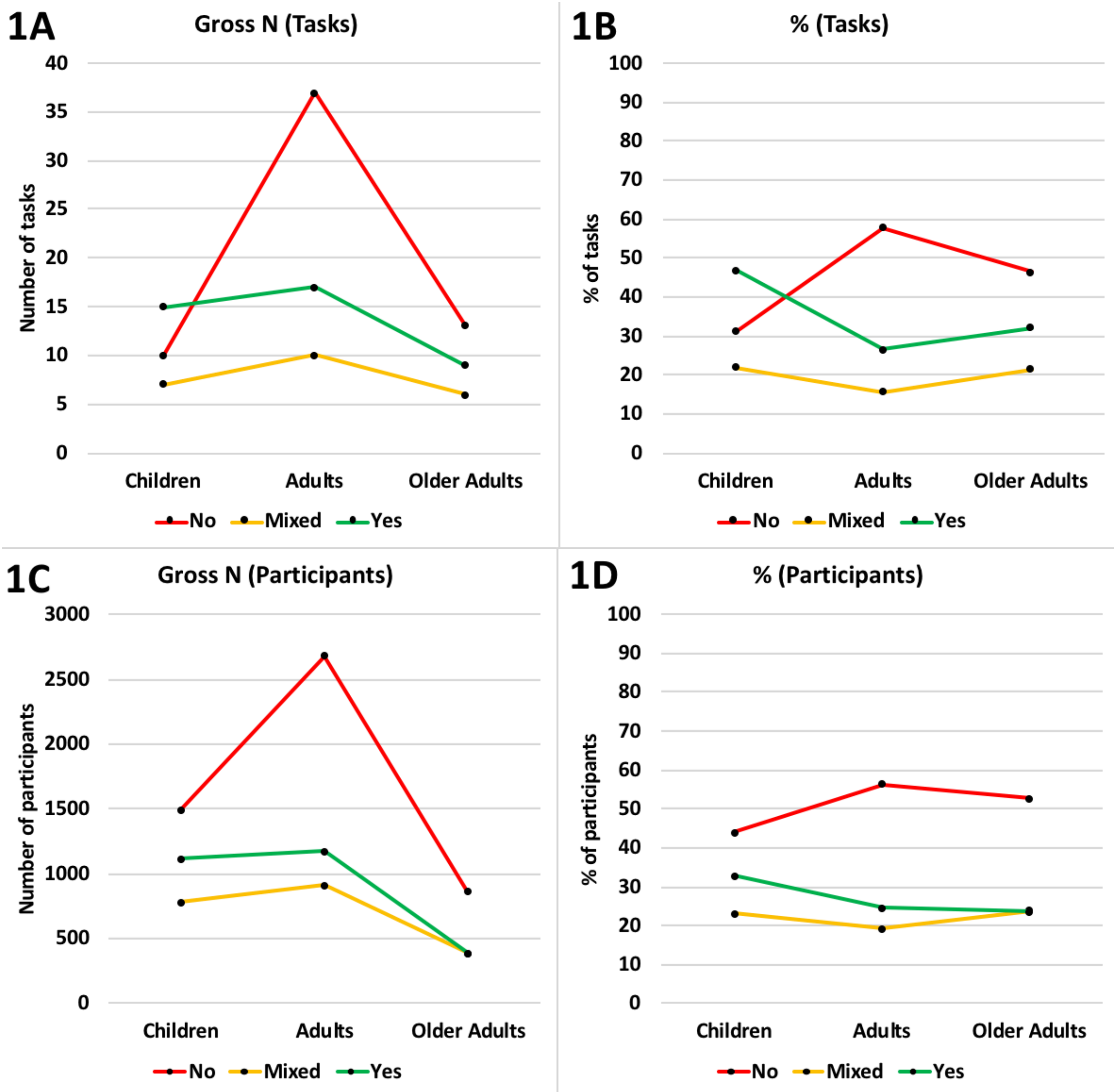

Figure 1. Is there a bilingual advantage in executive function? Results classified as supporting ('yes'), rejecting ('no'), or offering only qualified support ('mixed') for the hypothesis based on data from Simon, Spatial Stroop and Flanker tasks and classified by age group.

1.2. A potential alternative explanation for bilingual advantages: culture.

Culture has long been recognized as a potential confound in studies testing the bilingual advantage hypothesis, since many such studies have compared groups that differ in culture and language status, with bilinguals coming from one cultural background and monolinguals from another (Oh & Lewis, 2008; Hilchey & Klein, 2011). While conducting our review, it became apparent that growing up in an East Asian culture in particular may provide an alternative explanation for some reports of bilingual advantages[4]. We found that of the ten experiments where the monolinguals spoke a European language (usually English) and the bilinguals spoke either Chinese (including from Hong Kong), Japanese, or Korean, a bilingual advantage was reported in nine (Abutalebi et al., 2015; Bialystok et al., 2004; Bialystok, Craik,…et al., 2005; Ong et al., 2017; Tao et al., 2011; Yang, Yang, & Lust, 2011; Yang & Yang, 2016; Zhou & Krott (three experiments), 2016b). The tenth experiment we classified as 'mixed' owing to a reduced Simon Effect in the bilinguals but at the same time faster global speed in the monolinguals (Zhou & Krott, 2016b, experiment 3). Interestingly, the apparently enhanced EF performance of East Asian participants is also clear in a study which compared groups of *bilinguals*, where Japanese-English bilinguals produced a smaller Simon effect than Spanish-English bilinguals (Linck, Hoshino, & Kroll, 2008), and from other research using tasks not included in our review, or research done with other questions (not bilingualism) in mind, in both children (Bialystok, 1999; Sabbagh, Xu, Carlson, Moses & Lee, 2006; Tran & Yoshida, 2015; though see Barac & Bialystok, 2012) and adults (e.g. Wu & Keysar, 2007).

We chose to compare the performance of Korean participants to the performance of British participants. The Korean education system has deep roots in Confucianism, which emphasizes self-regulation and inhibition (Clarke-Stewart, Lee, Allhusen, Kim & McDowell, 2006; French & Song, 1998; Kwon, 2003), and it has been argued that such cultural practices might lead to better performance on tasks that tap inhibitory control (Oh & Lewis, 2008).

[4] Note again here that a full statistical meta-analysis would be invaluable in establishing whether culture or other variables do indeed moderate performance, and if so by how much.

Crucially, Oh and Lewis, and others, have found that Korean pre-schoolers outperformed similarly-aged British children on a cluster of tasks believed to tap executive processes, including inhibition (Lewis, Koyasu, Oh, Ogawa, Short & Huang, 2009). It follows therefore that if Korean participants outperform British participants, then this can be attributable to the experience of Korean culture, much as an advantage of bilingualism is considered attributable to the experience of bilingual language management.

1.3. The present study.

In the present study, we investigated whether young adult bilinguals would outperform young adult monolinguals on a Simon task. We reasoned that if bilingualism confers an advantage in domain-general EF, then bilinguals should outperform monolinguals. Additionally, we also compared participants' performance as a function of an additional factor that we felt also had the potential to produce an advantage in the task; namely East Asian (Korean) culture. We reasoned that if a group of Korean participants outperform a group of British participants, then findings with bilingual and monolingual groups divided along similar cultural lines and interpreted as supporting a *bilingual* advantage may instead have tapped an effect of *culture*.

Finally, to generalize our investigation beyond Korean and British participants alone, we also recruited a third, culturally-heterogeneous group of participants. This also allowed us to compare any advantage in Korean participants' performance to a culturally-heterogeneous group in addition to the British group.

2. Methods

2.1. Participants.

We collected data from 215 participants. The data from four of these were removed for response times averaging more than three times the standard deviation of the sample mean. Final groups by culture consisted of 78 allocated to the British group (M_{age} = 21, SD = 4.2, 18 males), 69 to the Mixed group (M_{age} = 23, SD = 4.5, 18 males) and 64 to the Korean group (M_{age} = 23, SD = 2.2, 28 males). Participants in the British and Mixed groups were all recruited and tested in the UK (instructions in English), while those in the Korean group were all recruited and tested in Korea (instructions in Korean). The Mixed group had spent an average 26 months in the UK or other English-speaking countries (SD = 44), and consisted of one participant each from Albania, the USA, Angola, Bosnia, Bulgaria, Egypt, Estonia, France, Italy, Japan, Korea, Latvia, Lithuania, Malta, Pakistan, Slovakia, Spain, and Taiwan, two from Brunei, Malaysia, Mexico, Portugal, and Vietnam, three from Hong Kong, India, Georgia, and Greece, four from Romania, five from Nigeria, nine from China, 11 from Norway, one Greek-German, one Italian-Spanish.

2.2 Measuring bilingualism

We obtained three measures of participants' bilingualism: second language (L2) proficiency, language dominance, and codeswitching. This was motivated by different theoretical considerations concerning what specific aspects of bilingualism may lead to an advantage. In this way, we could test for a bilingual advantage owing to higher bilingualism 'per se' (as measured by L2 proficiency), to weaker dominance of a single-language, or to more frequent codeswitching (see Blumenfeld & Marian, 2014; Costa et al., 2009; and Green & Abutalebi, 2013, for discussions of these theoretical standpoints).

L2 Proficiency. All participants rated their language proficiency on a 0-5 self-rating[5] scale, where zero equalled 'no knowledge' and five equalled 'native-like proficiency'. All participants gave '5' for their native language/s.

Language dominance. Participants also rated each of their languages as a percentage of their daily language use, such that a participant may rate their English as being used 70% of the time and their Spanish (for example) as 30% (following Pelham & Abrams, 2014, and Woumans et al., 2015). We took the highest percentage as our measure of single-language use or 'language dominance', in line with the view that dual- and single-language contexts are bipolar opposites (Hartanto & Yang, 2016). Lower language dominance scores therefore indicated greater dual-language use and hence more bilingual behaviour.

Codeswitching. Participants also rated how often they used more than one language within one sentence, on a 0-10 scale from 'never' to 'always' (following Hartanto & Yang, 2016, who used a 5-point scale but with the same end labels).

2.3 Other measures

Nonverbal IQ. Korean children have been found to outperform British children on measures of nonverbal IQ (Lynn & Song, 1994), and nonverbal IQ has previously been found to correlate with performance on the Simon task, even when bilingualism does not (Rosselli et al., 2016). It was thus important to establish that any effects related to culture or

[5] A total of 97 non-native speakers of English (34 Koreans) also took the Oxford Quick Placement Test or QPT (QPT, 2001) in order to check that self-ratings correlated with objective scores. The QPT is a short (15-30 minute) assessment of English grammar, vocabulary and collocation comprehension through multiple-choice questions. QPT. Scores on the Oxford QPT test correlated significantly and strongly with subjective L2 rating ($r_s(97) = .752$, $p < .001$), indicating that self-ratings were in line with the results of objective testing.

bilingualism were not confounded with nonverbal IQ. All participants therefore completed the short form of the Advanced Raven Progressive Matrices set (Arthur & Day, 1994), which has been used in prior research in the field (e.g. Woumans et al., 2015).

Socio-economic status (SES). SES is usually defined as a composite of financial wealth, education, and social prestige (e.g. Hilchey & Klein, 2011; Morton & Harper, 2007). SES has been found to correlate positively with performance on tasks that tap EF in children (Hackman & Farah, 2009; Mezzacappa, 2004; Morton & Harper, 2007) and crucially also in children and adults in the absence of a similar effect of bilingualism (Morton & Harper, 2007; von Bastian, Souza, & Gade, 2015). It was therefore important that we measured SES and either controlled for it in our analyses or established empirically that SES could not account for results linked to bilingualism. We asked participants to describe their level of education and that of one of their parents (the highest-scoring), each on a four-point scale. A *1* indicated High School education, *2* undergraduate-level education, *3* Masters-level education, and *4* PhD-level education. The two scores were tallied to make a Composite Education Score. We included the education of one parent (the highest educated) in order to capture variance owing to the participant's upbringing. Most participants were undergraduates, and hence would have spent most of their lives within the socio-economic context of their parents. Nevertheless, given that all our participants were recruited through universities, we felt it was unlikely that SES would show much variance.

2.4 Materials and procedure.

The Simon task. Participants sat at their own comfortable distance from the screen (approx 60cm). Red and green squares appeared randomly one by one on either the left or right side (counterbalanced) of the screen. Each square commanded approximately 6° 0' 0.63" of view, and the centre of each square was 10° 0' 0.08" from fixation. The instructions were to press 3 on the top row of the keyboard (located so that the number '6' aligned to the centre of the screen) when they saw a red square and 9 on the same row when they saw a green square, and to be as quick and accurate as possible. There was a 50/50 ratio of congruent to incongruent trials. Separate control blocks of centrally-placed squares, where participants still responded according to color, were included to generate a measure of general response speed. In total there were 11 blocks of 52 trials, in the order of two Simon blocks followed by a control block. There were therefore eight Simon blocks and three control blocks. Each block was separated by a 7-second interval.

Each trial proceeded as follows. First, a fixation cross appeared in the centre of the screen for 150ms, followed by a 350ms blank screen interval. The stimulus square then appeared and remained on screen for 400ms, during which time the participant could respond. The background was always black, and the fixation cross was white. A further 900ms of blank screen followed, during which a response could still be entered, before the presentation cycle began again. Each trial therefore lasted 1800ms. This procedure was closely modelled on Bialystok et al.'s (2005) study (52 trials per block, 8 experimental blocks, red and green squares). We removed the 'warning cue' for the next trial, and removed the last control block (cut from 4 to 3) to reduce participation time[6]. The trials were presented in the same fixed pre-randomised order for each participant. No feedback was provided

---

[6] There are of course many variations, but the 'signature' of a Simon Task is that there must on occasion be a conflict between stimulus location and response location (Hommel, 2011). Subtle differences between Simon task procedures include small changes in inter-trial intervals, the colour of the squares, the distance of the square to the left or right of fixation, whether or not there are control (central-square-) trials, and of course the number of trials, among other things.

during the task, and responding neither extinguished the stimulus nor terminated the trial early so that inter-trial intervals were uniform both within and across participants. The Simon task was run using E-prime 2.0 software and performed on two 21.5" Apple desktop computers, one in the U.K. and one in Korea[7]. The Simon task was always performed first, followed by the short form of the Advanced Raven Progressive Matrices set.

2.5. Analyses.

With the exception of responses faster than 100ms, incorrect trials, and the first four trials of each experimental block, we retained all responses and response times and analysed them using linear mixed-effects regression (LMER) modelling, which relaxes the requirement for normality of the residuals of each model (Gellman & Hill, 2007). As a result, we neither transformed nor trimmed response time data, giving bilingual advantages the best environment in which to emerge (Zhou & Krott, 2016a). Following von Bastian et al. (2015), all continuous predictor variables entered into the models (Ravens scores, each of the three bilingualism variables) were grand-mean centred, as were response times. Unless otherwise stated, all models with RT as the dependent variable included Congruency (congruent vs. incongruent), Ravens scores, and Group (British vs. Mixed vs. Korean) as standard fixed factors. Given the correlation between all three measures of bilingualism (all *p*s < .001, see results section) and the potential therefore for multicollinearity to affect any results if we

[7] The precise specifications of each machine are as follows: Processor: 3.1GHz Intel (R) Core(TM) i7-3770S (UK), i7-4770S (Korea), 16GB RAM; graphics: NVIDIA GeForce GT 650M (UK), 750M (Korea), 60 Hz refresh rate and 32-bit colour depth, 64-bit operating system.

included more than one measure of bilingualism in a single model, we created separate models for each measure of bilingualism. Thus, there were a total of three LMER models; one LMER with only L2 Proficiency as the independent bilingualism variable, one with only Codeswitching, and one with only Language Dominance. In every case, all fixed factors[8] within a single model were allowed to interact with each other.

In addition to the fixed-effect structure, the random-effect structures always included Participant and Trial as random intercepts, and in accordance with the convention that random effect structures should maximally reflect the fixed factor structure (Barr, 2013), the factor Congruency was allowed to vary by Participant in the form of random slopes. More complex random effects structures were discarded either i) for increasing the Akaike criterion or failing to reduce this figure by at least a value of 2, or ii) for issues with model construction by the software ('convergence'). Since the sequence of trials was fixed across every participant, the inclusion of the Trial factor served not only to help account for any effects of trial sequences (such as when the same stimulus was repeated more than twice), but also for performance changes as the task progressed, as (for example) the final trial of the task was allowed to have a different intercept in the model from the first trial (e.g. Costa et al., 2008, 2009).

All of these model characteristics were retained when accuracy was the dependent variable, with the exception that we used generalised linear modelling (GLM) using the logit function and analysed the data with fixed factor structure only, as random effect structures

[8] Ravens scores were converted into a 1-10 scale (from 1-12) and Language Dominance scores divided by ten in order to avoid issues with model convergence due to widely varying scales. Language Dominance scores are always reported after conversion back to percentages.

caused models to fail to converge. For control trials, we repeated the analyses used for experimental trials but with the redundant factor Congruency omitted.

In all cases, higher-order interactions (for example, between Group, Ravens scores and a bilingualism variable), and other interactions that were not of direct relevance to the hypotheses in question (for example, an interaction between nonverbal IQ scores and a bilingualism variable) were omitted from reporting for reasons of brevity and, relative to the research questions, irrelevance[9]. All models were created using the R statistical software environment (version 3.4.0, The R Foundation for Statistical Computing, 2017). We obtained the results for each group by resetting the reference levels (intercepts) without changing the model.

3. Results

3.1 Initial analyses

To begin, we investigated whether SES and nonverbal IQ were likely to be predictors of performance in our main analyses using LMERs. This was important in order to avoid the potential for unnecessary over-parameterization by incorporating variables that were unlikely to add much value to the model. A Kruskal-Wallis test found that the three groups differed on our SES measure, $H(2) = 26.621$, $p < .001$, with the Mixed culture group ($Mdn = 5$) scoring higher than both the British group ($Mdn = 3$), $Z = 5.005$, $p < .001$, and the Korean group ($Mdn = 4$), $Z = 3.610$, $p < .001$, but no difference between the British and Korean groups, $Z = 1.190$, $p = .702$. However, SES itself failed to correlate with any accuracy or response time variable on the Simon task, including the crucial Simon Effect scores for both response times, $r_s = .043$, $p = .538$, and accuracy $r_s = .092$, $p = .297$, as well as global response speed, $r_s =$

[9] Full details of all results from all models can be found in Supplemental Materials (S2).

.073, $p = .289$, and global accuracy, $r_s = .074$, $p = .287$. This was not unexpected, given that our sample was fairly homogenous with respect to SES (participants were recruited through advertising in universities), that the median education score for participants discounting parental education was 2 for each group (indicating undergraduate-level education). We excluded SES from our main analyses as a result.

A Kruskal-Wallis test for Ravens scores was also significant, $H(2) = 32.637$, $p < .001$, with follow-up pairwise comparisons showing that the Korean group scored more highly than the British group, $Z = 5.631$, $p < .001$, and Mixed group, $Z = 3.823$, $p < .001$, but no difference between the Mixed culture group and British group, $Z = 1.732$, $p = .250$. This result is consistent with the finding that Korean children outperform British children on the standard (rather than advanced) version of the Ravens test (Lynn & Song, 1994). Crucially, higher Ravens scores correlated with smaller Simon Effect scores both in response times, $r = .338$, $p = < .001$, and accuracy $r = .248$, $p = < .001$, as well as faster overall speed, $r_s = -.373$, $p = < .001$, and higher overall accuracy, $r = .274$, $p = < .001$. It was thus clearly important to include our measure of nonverbal IQ in all models so that any results relating to bilingualism and/or culture could be interpreted while controlling for this factor. Thus the final models included a total of four fixed factors: Congruency (congruent vs. incongruent), one bilingualism variable, Nonverbal IQ, and Culture (British vs. Korean vs. Mixed).

The overall distributions of responses to each of the three bilingualism variables across the sample as a whole are displayed in Fig 2. Means were an L2 Self-Rating of 2.5 (SD = 1.6), which is precisely at the mid-point in the measure, a Codeswitching score of 4.8 (SD = 3.4), and a Language Dominance score of 83% (SD = 13%). These were therefore the values at which these variables were mean-centred for entry into the LMERs. As expected, all three variables were highly correlated with each other (L2 Self Ratings and Codeswitching, $r_s = .510$, $p < .001$; L2 Self Ratings and Language Dominance, $r_s = -.683$, $p <$

.001; Codeswitching and Language Dominance, $r = -.532$, $p < .001$), justifying their use in separate models.

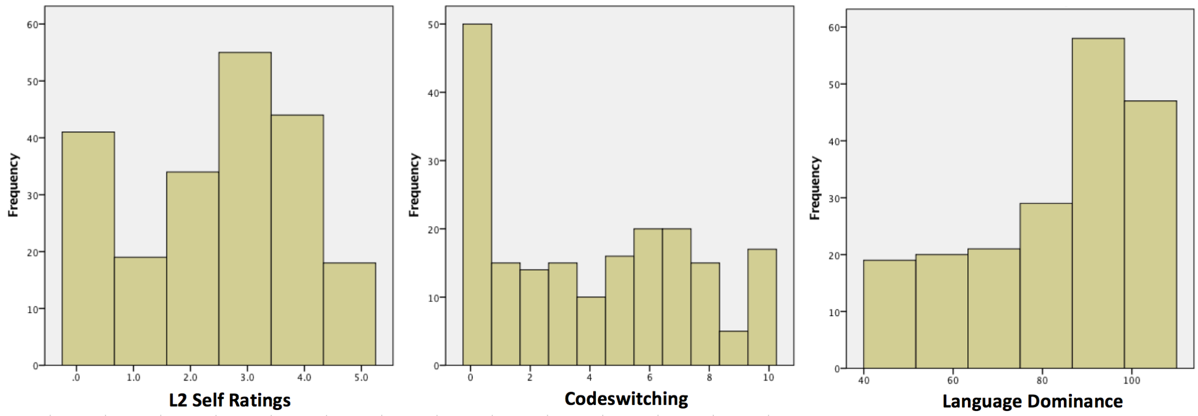

Figure 2: Distribution of responses for the three bilingualism variables.

Since L2 Self-ratings, Codeswitching and Language Dominance scores were not normally distributed within each culture group, we used Kruskal-Wallis tests to investigate between-group differences, the results of which are illustrated in Fig 3. The test for L2 Self ratings was significant, $H(2) = 90.626$, $p < .001$, as were all pairwise contrasts, with the Mixed culture group rating higher than both the British group, $Z = 9.405$, $p < .001$, and Korean group, $Z = 6.022$, $p = .008$, and the Korean group in turn rating higher than the British, $Z = 6.022$, $p < .001$. This was expected, since the Mixed culture group were all living abroad, whereas the British and Korean participants were not. The test for Codeswitching was also significant, $H(2) = 68.406$, $p < .001$, as were all pairwise comparisons (Koreans > Mixed, $Z = 2.585$, $p = .029$; Koreans > British, $Z = 7.964$, $p < .001$; Mixed > British, Z = 5.663, $p < .001$). We discuss the unusually high result in the Korean group below. Finally, the test for Language Dominance was also significant, $H(2) = 92.142$, $p < .001$, with all contrasts significant (British > Korean, $Z = 4.043$, $p < .001$; British > Mixed, $Z = 9.586$, $p < .001$; Korean > Mixed, $Z = 5.028$, $p < .001$).

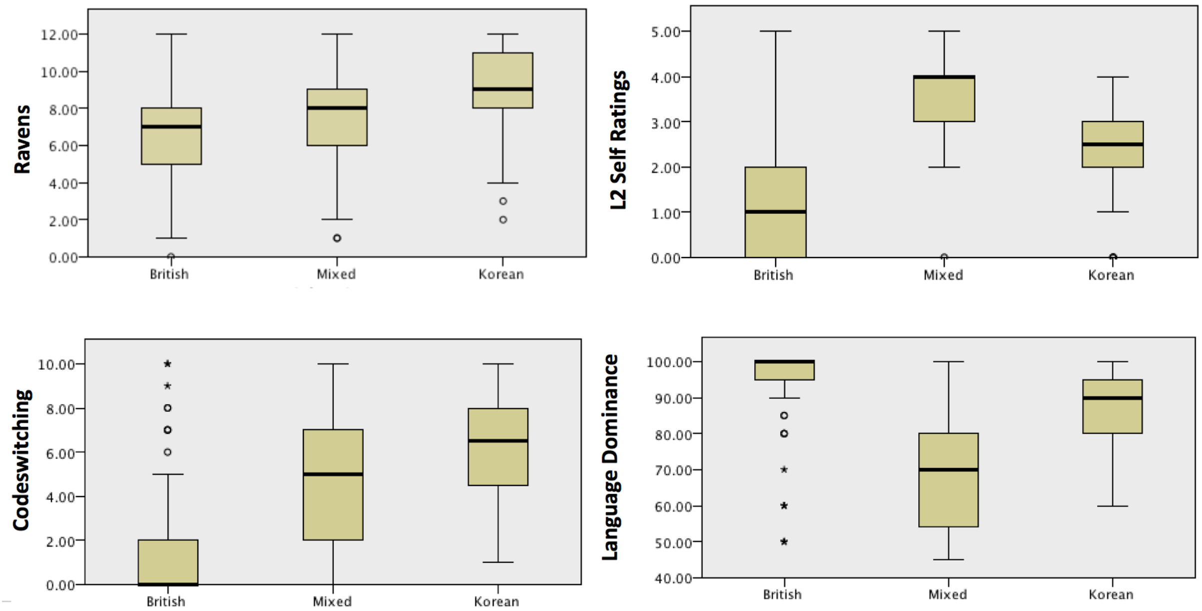

Figure 3: Results for Nonverbal IQ and bilingualism measures.

These analyses suggested that although the Koreans were less bilingual in both proficiency and in terms of dual-language use than the Mixed culture group, they reported codeswitching more frequently than any other group. This pattern suggested that the Koreans considered themselves to be speaking Korean most of the time that they mixed words from other languages into their sentences. This may be because the Koreans in our sample, being resident in their home country at testing, use English words not in the context of second-language speaking but as loan words instead (e.g. Shim, 1994). Participants in the Mixed group were using a foreign language (English) in a foreign country (the UK), and hence codeswitching scores in this group are highly likely to involve genuine language switches. As a result, we suggest interpretations of any effect of codeswitching in the Korean group specifically be approached with caution.

3.2 Main analyses

Our main analyses consisted of three models for response times and three for accuracy; each with either L2 Proficiency, Codeswitching scores, or Dual-Language Use as the bilingualism variable. This latter variable was calculated by inverting Language Dominance scores, since doing so meant that higher scores in all three bilingualism variables represented higher bilingualism, which facilitates the reading of the results.

All results relating to nonverbal IQ, bilingualism and culture come from the same models, meaning the effects of one can be interpreted as controlling for the effects of the others. Before we report the effects of these variables, we report first the simple estimated outputs for congruent trials, incongruent trials, and control trials for each model. Recall that the Simon Effect is calculated as the RT on congruent trials minus the RT on incongruent trials, where the expected slower performance on incongruent trials is represented by a negative score. The presence of a Simon Effect is thus verified by a significant main effect of the factor Congruency in each model, and for each group within that model. This initial analysis is important in order to establish that participants on this task produce the expected Simon Effect. The estimates for each trial type, and the test of the Simon Effect through these initial analyses are displayed in Table 1. As these outputs are model estimates, and since the independent bilingualism variable in them changed, results vary subtly between models. The expected pattern of fastest performance on control trials, followed by congruent trials, and then slowest performance on incongruent trials, was clear in all cases. Crucially, statistically significant Simon Effects were found for every group in all three models, varying in size from -16 ms to -36 ms.

Table 1. Accuracy (%) and RTs (ms) estimates across trial types and analyses in the Simon task, in all three LMER analyses. Cong = Congruent; Inc = Incongruent; Simon = Simon Effect.

Model estimates of Simon Task performance

Group	Accuracy (%) Cong	Inc.	Simon	Control	Response times (ms) Cong.	Inc.	Simon	Control
Estimates from model which included *L2 Proficiency* as bilingualism variable								
British	91	82	-9***	91	383	419	-36***	376
Mixed	88	87	-1 ns	92	399	419	-20***	386
Korean	93	91	-2***	96	381	397	-16***	370
Estimates from model which included *Codeswitching* as bilingualism variable								
British	91	83	-8***	91	385	419	-34***	379
Mixed	90	88	-3***	93	402	422	-20***	387
Korean	93	90	-2***	96	374	390	-16***	363
Estimates from model which included *Dual-Language Use* as bilingualism variable								
British	92	83	-8***	91	385	420	-35***	381
Mixed	90	88	-3***	93	407	428	-21***	390
Korean	92	90	-2***	95	379	396	-17***	367

Effects of nonverbal IQ on performance. As can be seen in Table 2, higher nonverbal IQ predicted better performance covering both higher global accuracy, faster global speed, and reduced Simon Effects, the latter for both accuracy and response times. The most frequent relationship was with global response speeds and global accuracy, which occurred in each of the three groups. The Mixed group showed the most prevalent influence of this variable, as it predicted smaller Simon Effect scores regardless of which bilingualism variable was in the analysis. The Korean group showed the most limited influence of this variable. Overall, the results suggested that higher nonverbal IQ predicted better performance on the Simon task. The inclusion of this factor in all models thus controlled for this variable when looking at the results of bilingualism and culture (all estimates are based on models including Nonverbal IQ, Culture, and Congruency, which are all allowed to interact).

Table 2. Effect of Nonverbal IQ (as measured by Ravens scores) on Simon Task performance.

	Effects of Nonverbal IQ on Simon Task performance																	
	Simon Effect in RT (ms) (negative scores indicate smaller Simon Effect with higher Ravens scores)									**Overall Speed (ms)** (negative scores indicate globally faster performance with higher Ravens scores)								
	Model w. L2 Proficiency			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*			*Model w. L2 Proficiency*			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*		
Group	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*
British	-0.5	0.601	549	-0.6	0.786	.433	-0.6	0.715	.475	-4.6	1.946	.053	-5.1*	2.186	.030	-5.2*	2.271	.024
Mixed	-2.4*	2.280	.024	-1.9**	2.632	.009	-2.0	2.453*	.015	-9.7**	3.257	.001	-10.3***	4.912	<.001	-9.6***	4.133	<.001
Korean	-0.8	0.916	.361	+0.2	0.195	.846	-0.9	0.653	.514	-5.5*	2.089	.038	-6.5	1.889	.060	-4.5	1.212	.227
	Simon Effect in Accuracy (%) (negative scores indicate smaller Simon Effect with higher Ravens scores)									**Overall Accuracy (%)** (positive scores indicate globally higher accuracy with higher Ravens scores)								
	Model w. L2 Proficiency			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*			*Model w. L2 Proficiency*			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*		
Group	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*
British	-0.2	1.369	.171	-0.2	1.773	.076	-0.1	1.203	.229	+0.6***	4.853	<.001	+0.9***	7.262	<.001	+0.5***	6.062	<.001
Mixed	-0.5*	2.515	.012	+0.0	0.030	.976	-0.1	0.833	.405	+0.6***	4.150	<.001	+0.3*	2.486	.013	+0.5***	4.520	<.001
Korean	0.2	1.321	.186	+0.2	1.039	.299	-0.1	0.420	.675	+0.1	0.976	.329	-0.1	0.739	.460	+0.6**	2.967	.003

Effects of culture on performance. As can be seen in Table 3, there were also effects related to culture. Consistent with our hypothesis, the Korean group outperformed the British group on every measure and on every model, displaying higher global accuracy, faster global RT, and smaller Simon Effects both in RT and accuracy. The Korean group also performed faster overall than the Mixed group in two out of three models. The Mixed group also outperformed the British group on every measure except global speed, across every model.

Table 3. Group-based comparisons on performance on the Simon Task.

	Effects of Culture on Simon Task performance																	
	Simon Effect in RT (ms) (negative scores indicate smaller Simon Effect in *first* named group)									**Overall Speed (ms)** (negative scores indicate globally faster performance in *first* named group)								
	Model w. L2 Proficiency			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*			*Model w. L2 Proficiency*			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*		
Comparison	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*
Korean vs. British	-19.4***	5.079	<.001	-18.2***	4.206	<.001	-18.3***	4.349	<.001	-21.8*	2.006	.046	-29.3***	2.357	.019	-24.3*	2.052	.042
Korean vs. Mixed	-3.5	0.682	.496	-4.2	1.063	.289	-4.3	0.958	.339	-22.3	1.537	.126	-32.2**	2.794	.006	-32.3*	2.561	.011
Mixed vs. British	-16.0**	3.070	.002	-13.9***	4.078	<.001	-14.0**	3.280	.001	+0.4	0.029	.977	+2.8	0.290	.772	+8.0	0.665	.507
	Simon Effect in Accuracy (%) (negative scores indicate smaller Simon Effect in *first* named group)									**Overall Accuracy (%)** (positive scores indicate globally higher accuracy in *first* named group)								
	Model w. L2 Proficiency			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*			*Model w. L2 Proficiency*			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*		
Comparison	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*
Korean vs. British	-4.3***	5.477	<.001	-4.1***	4.555	<.001	-5.1***	6.405	<.001	+8.6***	15.656	<.001	+7.9***	12.060	<.001	+6.8***	11.481	<.001
Korean vs. Mixed	+2.4*	2.468	.014	+0.4	0.471	.638	-0.0	0.039	.969	+3.6***	5.492	<.001	+3.3***	5.964	<.001	+2.0***	3.504	<.001
Mixed vs. British	-7.5***	7.042	<.001	-4.6***	7.118	<.001	-5.1***	6.426	<.001	+5***	6.188	<.001	+4.6***	9.105	<.001	+4.8***	7.600	<.001

Effects of bilingualism on performance. The results related to bilingualism are displayed in Table 4. Overall, the results offered no support for a bilingual advantage; indeed, where there was evidence of an effect of this variable it was typically of a bilingual *disadvantage*. Both the British and Korean groups showed poorer overall accuracy with higher L2 proficiency. The Korean group also showed lower accuracy with greater dual-language use, and the British group a larger Simon Accuracy effect with greater dual-language use. The Mixed group showed poorer overall accuracy with greater levels of codeswitching. Moreover, higher L2 proficiency and more frequent codeswitching also predicted a larger Simon Effect in accuracy in the Mixed group.

Additionally, unlike the effects of culture, all the statistically significant effects of bilingualism were restricted to accuracy analyses alone and not response times. When taking into account marginally significant results, the negative effect of bilingualism spread to response times in the form of a larger Simon Effect in the British group as a function of higher L2 proficiency. The sole evidence for an advantage was a smaller Simon Effect in accuracy in the Korean group that was related to higher L2 proficiency.

Table 4: Performance on the Simon task as it related to bilingualism.

Effects of Bilingualism on Simon Task performance

	Simon Effect in RT (ms) (negative scores indicate smaller Simon Effect with greater bilingualism)									**Overall Speed (ms)** (negative scores indicate globally faster performance with greater bilingualism)								
	Model w. L2 Proficiency			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*			*Model w. L2 Proficiency*			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*		
Group	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*	Est	*t*	*p*
British	+2.6	1.767	.079	+1.0	1.329	.186	+0.3	1.570	.118	+3.2	0.765	.445	+2.0	0.990	.324	+0.5	1.008	.315
Mixed	+0.6	0.228	.820	+0.4	0.638	.524	-0.0	0.011	.991	+2.6	0.322	.747	+0.6	0.309	.758	-0.3	0.656	.513
Korean	-1.4	0.486	.628	+0.2	0.210	.834	+0.1	0.165	.869	-9.2	1.108	.269	+3.1	1.050	.295	-0.0	0.047	.963

	Simon Effect in Accuracy (%) (negative scores indicate smaller Simon Effect with greater bilingualism)									**Overall Accuracy (%)** (positive scores indicate globally higher accuracy with greater bilingualism)								
	Model w. L2 Proficiency			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*			*Model w. L2 Proficiency*			*Model w. Codeswitching*			*Model w. Dual-Lang. Use*		
Comparison	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*	Est	*Z*	*p*
British	+0.4	1.752	.080	+0.0	0.041	.967	+0.0*	2.022	.043	-0.9***	3.535	<.001	-0.0	0.23	.818	-0.0	0.657	.511
Mixed	+1.7**	3.201	.001	+0.5***	3.930	<.001	+0.0	1.135	.257	+0.0	0.118	.906	-0.5***	4.72	<.001	-0.0	1.895	.058
Korean	-1.9***	3.643	<.001	-0.1	0.922	.357	-0.0	0.274	.784	-1.2**	3.140	.002	-0.1	0.861	.389	-0.0***	4.296	<.001

Additional analyses[10]**.** Finally, an analysis of control trial performance (where squares appeared centrally) found that the Korean group were faster than the Mixed group in the LMER with Codeswitching (SE = -11ms, $t(197) = 2.311$, $p = .022$) and the LMER with Language Dominance (SE = -12ms, $t(194) = 1.988$, $p = .048$). No other between-culture differences were found on control trials, and bilingualism did not show any influence on response speed (*p*s > .1). Thus, the Korean and British groups, as well as bilinguals and monolinguals, showed no evidence of performing at different baseline levels for simple motor response speed.

4. Discussion

We gave young adult participants a Simon task in order to investigate whether bilingualism confers an advantage in aspects of executive function. We found no support for a bilingual advantage, whether measured in terms of higher proficiency in a second language,

---

[10] Our design also allowed us to test a recently-suggested reformulation of the bilingual advantage hypothesis, whereby bilinguals may be better at disengaging from a previous trial type, whether that trial be congruent or incongruent (Grundy et al., 2017), suggesting bilingual advantages not specifically related to inhibition, but rather to the ability to constantly monitor performance and disengage attention (see Costa et al. 2008, 2009, for an early formulation of this hypothesis). We tested this additional possibility by splitting the Congruency factor in our analysis into two factors, Prime (congruent vs. incongruent) and Target (Congruent vs. Incongruent). The results of this analyses were in line with the results described above, indicating no evidence for a bilingual advantage (see supplemental material S3 for full details).

greater within-utterance codeswitching, or greater usage of two languages rather than one in daily life. In fact, we found greater evidence to the contrary; bilingualism more frequently displayed a *disadvantage* on Simon task performance. In contrast, our comparisons based on culture were clear. Korean participants outperformed their British counterparts on every performance measure, whether RT- or accuracy-based. This pattern suggests that culture may drive some reported cases of an advantage previously attributed to bilingualism (c.f. Oh & Lewis, 2008). Indeed, the only measure in which the Korean group did not outperform the British group was on control trials, which suggested that variance on experimental trials could not be due to differences in simple motor response speeds. The apparent cultural advantage in performance also extended to a comparison with a culturally-heterogeneous group of participants, albeit in terms of a global response time advantage alone. In sum, these results corroborate our literature review, providing no clear evidence of bilingualism conferring an advantage in domain-general executive function.

As was made clear from our review, the finding that young adult bilinguals do not show an advantage over young adult monolinguals on this task is not new (e.g. Bialystok, 2006; Kousaie & Phillips, 2012; Paap & Greenberg, 2013), and our results are thus consistent with these. Crucially, given that the review also found no clear support for the peak performance argument, our findings (and others') are highly unlikely to be the result of any ceiling performance.

The results of our study stand up to a number of potential counterarguments. We investigated whether bilingualism could enhance performance in any of three ways: through higher L2 proficiency, more frequent code-switching, and greater balance of language use. It is telling that despite performing separate analyses for each variable, which increases the chances of discovering a false positive in favor of a bilingual advantage, we still found no evidence to support the hypothesis. Additionally, we were also able to establish through the analysis of control trials that the Korean advantage was not related to general processing speed (cf. Paap, 2017), and we showed that the absence of a bilingual advantage was not related to the trimming of response time data (Zhou & Krott, 2016a). By including a measure of nonverbal IQ as a factor in each model, our findings are also free of this potential confound[11]. Finally, our Simon Task produced a clear Simon Effect (faster performance on congruent than incongruent trials), indicating that the participants did indeed find incongruent trials more difficult.

The primary result of the present study was the consistently better performance of the group of Korean participants compared to the group of British participants. Given that each of these groups was recruited and tested in their home countries and in an entirely L1 context, we can rule out second-language contexts as a potential explanation for this difference. What then can explain this cultural advantage? We speculate that the source lies in the socio-educational background of participants that we described in the introduction, by which the relatively intensive education experienced by Koreans from a young age serves to improve their cognitive performance more generally. This view is also supported by the higher nonverbal IQ scores we found in the Korean group than in either the British or Mixed groups. It is important to note, however, that the Korean group did not outperform the British group

---

[11] Nonverbal IQ might also have captured aspects of the socio-economic status of our sample that our direct measure of SES did not, since the two variables are frequently related (e.g. Brooks-Gunn, Klebanov, & Duncan, 1996). We are grateful to a reviewer for pointing this out. In the present dataset, a correlation analysis found no significant relationship between Raven's scores and our measure of SES (r_s = .056, p = 418). Given that Nonverbal IQ was included as a covariate, any influence of SES captured within this variable would in any case be controlled for in the results relating to bilingualism and culture.

simply because they displayed a higher nonverbal IQ score. A benefit of the regression modelling is that our analysis found dissociable effects of culture (Korean vs. British) and nonverbal IQ (high vs. low).

Critically, our findings suggest important ramifications for previous studies that have compared bilinguals and monolinguals and reported an advantage when including a significant proportion of bilingual participants from East Asian cultures. For example, the bilingual advantage on the Simon task reported in Bialystok et al. (2004) was based on a sample which included a large number of bilinguals recruited in Hong Kong, and in Prior and MacWhinney's (2010) study, just over half the bilingual group spoke Chinese or Korean, whereas the monolingual group was English-speaking. The apparent support for a bilingual advantage from the other studies we described in our introduction as potentially tapping a cultural rather than language-related effect may also need to be reconsidered. However, we do not claim that culture is responsible for *all* those cases where an advantage is attributed to bilingualism and groups differ simultaneously in culture. This is because our data is limited to a direct comparison of participants from Korean and British cultures. These two cultures clearly form only a tiny fraction of known cultures, and it would be unwise to extrapolate further than the data allow. Additionally, there is evidence that culture alone cannot explain findings of bilingual advantages in children. For example, Bialystok and Viswanathan (2009) report enhanced EF in bilingual children in both India and Canada relative to monolingual children in Canada. Barac and Bialystok (2012) report bilingual advantages in Canadian schoolchildren who were bilingual in French, Spanish or Chinese with English relative to monolingual English children recruited from the same schools, but found no differences in EF performance attributable to culture. Tse and Altarriba (2014) reported that higher levels of L2 English proficiency in L1 Cantonese speakers in Hong Kong was related to a smaller Simon Effect, and Yang and colleagues (2011) found both an advantage of Korean culture and bilingualism in children, suggesting dissociable benefits of both. However, it is true that there is also a great deal of evidence suggesting an absence of a bilingual advantage in within-culture studies. For example, Wu, Zhang, and Guo (2016) found no evidence of increased bilingualism influencing EF in an L1 Chinese sample tested in China, nor did Yow and Li (2015), who looked at English-Mandarin bilinguals in Singapore. Ye, Mo, and Wu (2016) found only limited evidence for a bilingual advantage in a similar study with Chinese-English bilinguals in China.

In contrast to the effects of Korean culture, the effects of bilingualism—where they were found—were almost always negative. Although it was unexpected that bilingualism should tend towards an effect of a disadvantage, it is not without precedent in the literature. Paap and Sawi (2014) and Paap and Greenberg (2013) reported evidence of bilingual disadvantages, and there was also some evidence of poorer performance in bilinguals than monolinguals in studies by Zhou and Krott (2016b) and Kousaie and Phillips (2012). However, far more common than these 'reverse' outcomes are null results, and we interpret our findings as failing to support the bilingual advantage hypothesis, rather than supporting a bilingual *disadvantage* hypothesis.

Overall, given the results of our study and our review, it is hard to see how the bilingual advantage hypothesis as it is commonly formulated is clearly supported by the combined evidence. At the theoretical level, we note that in a recent response to a critique of a study that *had* found bilingual advantages, Bialystok and Grundy pointed out that "If we are to understand cognition, then it is imperative to understand how it is impacted by experience. Yet, of all activities in which humans engage, *nothing is as intense or sustained as using language* [italics added]." (Bialystok & Grundy, 2018, p.330). If language is the most intense cognitive experience of all, as Bialystok and Grundy suggest, then finding an effect of culture, which clearly incorporates a strong experiential factor, but not bilingualism, and moreover within the same sample, raises in our view serious questions about whether bilingualism really does confer a performance advantage. Overall, therefore, we feel that there is now sufficient evidence to suggest that a rethinking of research into bilingual advantages, both past and future, may be required; previous studies with young adults that have failed to support a bilingual advantage should not be dismissed as non-evidence, and indeed require explanation in any future investigations.

Finally, we wish to put forward two further possibilities for future research. Firstly, our review—though extensive—lacks a truly meta-analytical approach which incorporates not only sample size but also effect sizes and unpublished data. A such, future work that extends the remit of our review to include such parameters would help clarify our necessarily suggestive rather definitive findings.

Secondly, given the clear difference in performance between the Korean and British groups, another potentially fruitful area of research here is to investigate in more detail precisely which experiences relating to Korean culture lead to enhanced performance. We have already described how a reasonable candidate for this difference might be educational practices, whether in terms of the drive for achievement, hours of education, or the emphasis on self-discipline and control. Nevertheless, the possibility that it is not the Korean participants that show better-than-expected performance but the British participants who show *poorer*-than-expected performance should also not be discounted. Indeed, the differences we found between the British, Korean *and* Mixed groups suggest that broader constructs such as ‘Western’ or ‘Eastern’ culture are likely far too simplistic a distinction to draw. It may also be the case that culture itself is confounded with some other experience or experiences that give rise to enhanced EF. A recent example concerns the finding that EF might be crucial to the processing of logographic writing systems such as Chinese (Chung, Lam, & Cheung, 2018). Written Korean is a combination of both a phonemic system

(Hangul) and a logographic system (Hanja) (Cho et al., 2014). One avenue for future research may thus be to investigate whether logographic systems may account for an enhancement in EF relative to purely alphabetized systems such as English[12]. Any such research might also be in a position to inform investigations into bilingualism, since as already discussed it may be that aspects of experiences relating to culture account for some results previously interpreted as evidence of a bilingual advantage. Through examining culture-related effects and effects related to the written form that certain languages take, it should be easier to find the right empirical approach to test the bilingual advantage hypothesis in the absence of potential confounds. This could go some way towards resolving the debates around methodology and clear a path to a more informed research program in future.

Acknowledgments

We are grateful to Monika Schmid for comments on an earlier version of this manuscript and for her assistance with aspects of the analyses.

Disclosure of interest

The authors report no conflicts of interest.

References

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., Cappa, S. F., & Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex, 22*, 2076-2086.

[12] We are grateful to an anonymous reviewer for this suggestion.

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*, *5*, 398.

Antón, E., García, Y. F., Carreiras, M., & Duñabeitia, J. A. (2016). Does bilingualism shape inhibitory control in the elderly?. *Journal of Memory and Language*, *90*, 147-160.

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition*, *149*, 18-30.

Arredondo, M. M., Hu, X. S., Satterfield, T., & Kovelman, I. (2017). Bilingualism alters children's frontal lobe functioning for attentional control. *Developmental science*, *20*(3), e12377.

Arthur, W. Jr., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement, 54*(2), 394-403.

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child development*, *83*(2), 413-422.

Bialystok, E. (1986). Children's concept of word. *Journal of Psychololinguistic Research, 15*(1), 13-32.

Bialystok, E. (1999). Cognitive Complexity and Attentional Control in the Bilingual Mind. *Child Development, 70*(3), 636-644.

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

Bialystok, E. (2010). Global-local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology, 46*(1), 93-105.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological bulletin*, *143*(3), 233.

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effects of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage, 24,* 40–49.

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

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, memory, and cognition*, *34*(4), 859.

Bialystok, E., & Grundy, J. G. (2018). Science does not disengage. *Cognition*, *170*, 330-333.

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism, 9*(1), 103-119.

Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: The effect of bilingualism on reversing ambiguous figures. *Developmental Science*, *8*(6), 595-604.

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, *112*(3), 494-500.

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Bilingualism: Language and Cognition*, *17*(03), 610-629.

Brooks-Gunn, J., Klebanov, P. K., & Duncan, G. J. (1996). Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child development*, *67*(2), 396-408.

Calabria, M., Hernández, M., Martin, C. D., & Costa, A. (2011). When the tail counts: the advantage of bilingualism through the ex-Gaussian distribution analysis. *Frontiers in psychology*, *2*, 250.

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278-288.

Cho, Z.-H., Kim, N., Bae, S., Chi, J.-G., Park, C.-W., Ogawa, S., & Kim, Y.-B. (2014). Neural Substrates of Hanja (Logogram) and Hangul (Phonogram) Character Readings by Functional Magnetic Resonance Imaging. *Journal of Korean Medical Science*, *29*(10), 1416–1424. http://doi.org/10.3346/jkms.2014.29.10.1416

Chung, K. K. H., Lam, C. B., & Cheung, K. C. (2018). Visuomotor integration and executive functioning are uniquely linked to Chinese word reading and writing in kindergarten children. *Reading and Writing*, *31*(1), 155-171.

Clarke-Stewart, K. A., Lee, Y., Allhusen, V. D., Kim, M. S., & McDowell, D. J. (2006). Observed differences between early childhood programs in the US and Korea: Reflections of "developmentally appropriate practices" in two cultural contexts. *Journal of applied developmental psychology*, *27*(5), 427-443.

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

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

Cox, S. R., Bak, T. H., Allerhand, M., Redmond, P., Starr, J. M., Deary, I. J., & MacPherson, S. E. (2016). Bilingualism, social cognition and executive functions: A tale of chickens and eggs. *Neuropsychologia*, *91*, 299-306.

De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological science*, *26*(1), 99-107.

De Cat, C., Gusnanto, A., & Serratrice, L. (2017). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, 1-33.

Donnelly, S., Brooks. P.J., & Homer, B.D. (2015). Examining the bilingual advantage on conflict resolution tasks: A meta-analysis. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.) (2015). *Proceedings of the 37th Annual Meeting of the Cognitive Science Society,* 596-601. Austin, TX: Cognitive Science Society.

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited. *Experimental psychology*.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention, Perception, & Psychophysics*, *16*(1), 143-149.

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*(3), 340-347.

French, L., & Song, M. (1998). Developmentally appropriate teacher-directed approaches: Images from Korean kindergartens. Journal of Curriculum Studies, 30, 409 – 430.

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., Roberts, E. J., Hughes, E. K., 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*, art. 11.

Gellman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical* *models*. Cambridge: Cambridge University Press.

Giezen, M. R., Blumenfeld, H. K., Shook, A., Marian, V., & Emmorey, K. (2015). Parallel language activation and inhibitory control in bimodal bilinguals. *Cognition*, *141*, 9-25.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515-530.

Grundy, J. G., Chung-Fat-Yim, A., Friesen, D. C., Mak, L., & Bialystok, E. (2017). Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults. *Cognition*, *163*, 42-55.

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Science, 13*(2), 65-73.

Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition*, *150*, 10-19.

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

Hilchey, M. D., Saint-Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance cognitive fitness in non-linguistic executive processing tasks. *Cambridge handbook of bilingual processing*.

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, *136*(2), 189-202.

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

Kovacs, A. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning.

Developmental Science, 12(10), 48-54.

Kroll, J.F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology, 25*(5), 497-514.

Kwon, Y.-I. (2003) A Comparative Analysis of Preschool Education in Korea and England, *Comparative Education, 39*:4, 479-491.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological bulletin*.

Lewis, C., Koyasu, M., Oh, S., Ogawa, A., Short, B., & Huang, Z. (2009). Culture, executive function, and social understanding. In C. Lewis & J. I. M. Carpendale (Eds.), Social interaction and the development of executive function. *New Directions in Child and Adolescent Development*, *123*, 69–85.

Linck, J. A., Hoshino, N., & Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *Mental Lexicon, 3*(3), 349-374.

Lynn, R., & Song, M. J. (1994). General intelligence, visuospatial and verbal abilities in Korean children. *Personality and individual differences*, *16*(2), 363-364.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*(1), 81-93.

Marzecová, A., Asanowicz, D., Kriva, L. U., & Wodniecka, Z. (2013). The effects of bilingualism on efficiency and lateralization of attentional networks. *Bilingualism: Language and Cognition*, *16*(3), 608-623.

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child development*, *75*(5), 1373-1386.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of experimental child psychology*, *114*(2), 187-202.

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental science*, *10*(6), 719-726.

Oh, S. and Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. *Child Development*, *79*(1), 80-99

Ong, G., Sewell, D. K., Weekes, B., McKague, M., & Abutalebi, J. (2017). A diffusion model approach to analysing the bilingual advantage for the Flanker task: The role of attentional control processes. *Journal of Neurolinguistics, 43*, 28-38.

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*(3), 242-255.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*, 232-258.

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, *26*(6), 615-639.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265-278.

Paap, K. R., Johnson, H. A., & Sawi, O. (2016). Should the search for bilingual advantages in executive functioning continue? *Cortex*, *74*(4), 305-314.

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in psychology*, *5*, 962.

Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 313.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition, 13*(2), 253-262.

Quick Placement Test. (2001). Oxford: Oxford University Press.

Rodriguez-Fornells, A., Krämer, U. M., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Bilingualism and cognitive control*, 123.

Rubio-Fernández, P., & Glucksberg, S. (2012). Reasoning about other people's beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 211.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind a comparison of Chinese and US preschoolers. *Psychological science*, *17*(1), 74-81.

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

Sanchez-Azanza, V. A., López-Penadés, R., Buil-Legaz, L., Aguilar-Mediavilla, E., & Adrover-Roig, D. (2017). Is bilingualism losing its advantage? A bibliometric approach. *PloS one*, *12*(4), e0176151.

Shim, R. J. (1994). Englishized Korean: structure, status, and attitudes. *World Englishes*, *13*(2), 225-244.

Simon, J. R. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology, 81*, 174-176.

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. *Bilingualism and Cognitive Control*, *2*, 83-99.

Tran, C. D., & Yoshida, H. (2015). Differential Effects of Bilingualism and Culture on Early Attention: A Longitudinal Study in the US, Argentina, and Vietnam. *Name: Frontiers in Psychology*, *6*, 795.

Tse, C. S., & Altarriba, J. (2014). The relationship between language proficiency and attentional control in Cantonese-English bilingual children: evidence from Simon, Simon switching, and working memory tasks. *Frontiers in psychology*, *5*, 954.

von Bastian, C. C., Souza, A. S., Gade, M (2015). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology (General)*, 10.1037/xge0000120

Wang, R., Fan, X., Liu, C., & Cai, Z. G. (2016). Cognitive control and word recognition speed influence the Stroop effect in bilinguals. *International Journal of Psychology*, *51*(2), 93-101.

Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1579.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological science*, *18*(7), 600-606.

Wu, Y. J., Zhang, H., & Guo, T. (2016). Does Speaking Two Dialects in Daily Life Affect Executive Functions? An Event-Related Potential Study. *PloS one*, *11*(3), e0150492.

Yang, S., & Yang, H. (2016). Bilingual effects on deployment of the attention system in linguistically and culturally homogeneous children and adults. *Journal of experimental child psychology*, *146*, 121-136.

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*(3), 412-422.

Ye, Y., Mo, L., & Wu, Q. (2016). Mixed cultural context brings out bilingual advantage on executive function. *Bilingualism: Language and Cognition*, 1-9.

Yow, W. Q., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: why variations in bilingual experiences matter. *Frontiers in psychology*, *6*, 164.

Zhou, B., & Krott, A. (2016a). Data trimming procedure can eliminate bilingual cognitive advantage. *Psychonomic bulletin & review*, *23*(4), 1221-1230.

Zhou, B., & Krott, A. (2016b). Bilingualism enhances attentional control in non-verbal conflict tasks–evidence from ex-Gaussian analyses. *Bilingualism: Language and Cognition*, 1-19.