



Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost

Seamus Donnelly^{1,2,3} · Patricia J. Brooks^{3,4} · Bruce D. Homer³

Published online: 27 February 2019
© The Psychonomic Society, Inc. 2019

Abstract

A contentious issue in contemporary psycholinguistics is whether bilingualism enhances executive functions. Here, we report a meta-analysis of 80 studies (253 effect sizes) comparing performance of monolinguals and bilinguals on non-verbal interference-control tasks, while examining potential moderators of effects on two dependent variables (DVs): global reaction time (RT) and interference cost. We used a multiverse approach to determine how robust conclusions were to several dataset construction and analysis decisions. In our “preferred” analysis, using a broad definition of bilinguals and standard versions of interference-control tasks, there was a very small but significant bilingual advantage for global RT ($g = .13$), which became non-significant once corrected for publication bias. For interference cost, there was a very small but significant bilingual advantage ($g = .11$). Effects were not significantly moderated by task or participant age, but were moderated by an interaction between age of second language acquisition (AoA) and the DV. Unexpectedly, larger effect sizes for interference cost were observed for studies involving bilinguals with late as opposed to early AoA. The multiverse analysis produced results largely consistent with the preferred analysis, confirming our conclusion that evidence for a bilingual advantage on interference-control tasks is weak.

Keywords Cognitive and attentional control · Language comprehension

Introduction

A hotly debated question in contemporary psycholinguistics is whether the regular use of multiple languages has consequences for cognition more generally (Adesope, Lavin, Thompson, & Ungerleider, 2010). It is widely assumed that words in each of a bilingual’s two lexicons are active during language use and compete for selection (for an overview, see Bialystok, Craik, Green, & Gollan, 2009). Theorists have proposed that competition between first language (L1) and second language (L2) lexical items is overcome using more

general cognitive processes. For example, Green (1998) proposed that words in the non-target language are inhibited during regular language use. Others have speculated that bilingual language use might involve conflict monitoring (Abutalebi et al., 2012). This raises the intriguing possibility that bilingualism may exercise and thereby strengthen these processes, yielding a so-called bilingual advantage. This hypothesis has been tested using a variety of tasks and bilingual samples; however, results of these studies have been mixed (Costa, Hernández, Costa-Faidella, & Núria Sebastián-Gallés, 2009; de Bruin, Treccani, & Della Sala, 2015b; Hilchey & Klein, 2011; Lehtonen, Soveri, Laine, Järvenpää, J., de Bruin, & Antfolk, 2018; Paap, Johnson, & Sawi, 2015).

Across many studies, interference-control tasks, such as the Simon, Flanker, and variations of the Stroop, have been used as tests of a bilingual advantage. In all interference-control tasks, participants respond to either a single bivalent stimulus or a single stimulus with a distracter stimulus present. On some trials, the distracter dimension or stimulus cues the same response as the target (congruent trials) and in other trials it cues a different response (incongruent trials). These tasks yield two dependent variables (DVs). The first DV is global RT, which is the average RT across congruent and incongruent

✉ Seamus Donnelly
seamus.donnelly@anu.edu.au

¹ Research School of Psychology, The Australian National University, Canberra, ACT, Australia

² ARC Centre of Excellence for Dynamics of Language, Canberra, ACT, Australia

³ Developmental Psychology, The Graduate Center, CUNY, New York, NY, USA

⁴ Department of Psychology, College of Staten Island, CUNY, New York, NY, USA

trials. It is generally assumed to reflect the efficiency of processing in a high-conflict environment. The second DV is interference cost, calculated as the difference in mean reaction time (RT) between incongruent and congruent trials. It is thought to reflect the amount of extra time it takes to suppress the response cued by the incongruent information.

Researchers have predicted bilingual advantages on both of these DVs. For example, Green's (1998) prominent model of bilingual lexical production assumes that lexical items for each of a bilingual's two languages compete for selection and that non-target lexical items are suppressed by a domain-general inhibition mechanism. Bialystok, Craik, Klein, and Viswanathan (2004) speculated that regular exercise of this mechanism in bilinguals would lead them to exhibit smaller interference effects (i.e., lower interference cost) than monolinguals. Costa et al. (2009) suggested that bilinguals might experience enhanced conflict monitoring skills, leading to a general advantage in processing stimuli with a high degree of conflict and a bilingual advantage on global RT. Similarly, noting inconsistent evidence for a bilingual advantage on interference effects, Hilchey and Klein (2011) suggested that bilinguals might process conflict differently than monolinguals by developing separate processing pathways for conflicting and non-conflicting stimuli, which should give bilinguals an advantage for global RT. Importantly, these predictions are not necessarily mutually exclusive. Indeed, recently Bialystok (2017) sketched out a theory where bilingualism enhances broader executive attention abilities, which presumably might lead to bilingual advantages on both interference cost and global RT across a broad range of tasks.

Many researchers have sought to synthesize this literature with an eye toward explaining inconsistencies in findings across studies. For example, Hilchey and Klein (2011) conducted a literature review and identified a consistent bilingual advantage on global RTs but not on interference costs, and developed the previously described theoretical account of how such an advantage might occur. However, in a follow-up review, Hilchey, Saint-Aubin, and Klein (2015) found inconsistent evidence of a bilingual advantage for both interference cost and global RT. Valian (2015), focusing on executive functioning (EF) more generally, argued that many cognitively challenging activities, including practicing music and exercise, involve and strengthen EF. She speculated that bilingual advantages on EF tasks are real, but as they compete with advantages conferred by other challenging activities they are difficult to detect in individual studies. If this were true, one promising approach would be to quantitatively synthesize many studies, rather than consider individual studies since, in a single study, the effect of bilingualism may be overwhelmed by other factors. In addition to three older syntheses (Adesope et al., 2010; Costa et al., 2009; Hilchey & Klein, 2011), four recent reports (de Bruin et al., 2015b;

Lehtonen et al., 2018; Paap et al., 2015; Paap et al., 2017a) have quantitatively synthesized the large database of bilingual advantage studies.

De Bruin et al. (2015b) conducted two analyses to determine whether publication bias may explain the inconsistent findings across studies exploring the bilingual advantage. First, they collected conference abstracts reporting on comparisons between bilinguals and monolinguals on any cognitive task, and conducted a logistic regression to determine whether abstracts reporting an advantage were more likely to result in publication than those that did not. Results revealed that 63% of conference abstracts reporting a bilingual advantage were published in comparison to only 36% reporting no bilingual advantage. Second, they conducted a meta-analysis on the conference submissions that had been published in a journal, which yielded a medium effect size ($d = .30$), with funnel plots indicating strong evidence of publication bias.

It is important to note that the de Bruin et al. (2015b) study was primarily an assessment of the presence of publication bias rather than an attempt to estimate an average effect size for the bilingual advantage. As noted by the authors, and by Bialystok, Kroll, Green, MacWhinney, and Craik (2015), publication bias is ubiquitous to psychology and its existence does not mean the absence of an effect. Notably, de Bruin et al. (2015) included any study that contained a bilingual group, a monolingual group, and some sort of cognitive task, and thus did not specifically focus on EF.

Paap et al. (2015) adopted a different approach in their analysis of the relationship of bilingualism to EF. They surveyed the literature and summarized the results of all studies comparing monolinguals and bilinguals on interference-control and task-switching tasks. They relied on a vote-counting procedure, in which they coded each comparison as either supporting or failing to support a bilingual advantage according to whether it yielded a significant p -value. Across all measures, proportions of comparisons yielding a significant bilingual advantage were low (the highest proportion of significant tests was $\sim .22$). Moreover, they found that small sample studies were more likely to produce bilingual advantages than were large sample studies. They argued that the few significant comparisons might reflect questionable research practices endemic to psychology, confounds and between-group differences on non-EF constructs measured by these tasks. Paap et al.'s (2015) contribution is important, especially because of its focus on questionable research practices and interpretive methods in psychology research in general, and bilingual advantage research in particular. These factors certainly account for some of the discrepant findings. However, their review does not definitively disprove the existence of a bilingual advantage. First, it is unlikely that the advantage exists for all bilingual individuals. Looking solely at aggregate results might obscure effects for specific groups. Bilingual advantages may be restricted to certain age groups or

bilinguals who began learning their L2 before a certain age. Second, as noted by Linck (2015), the vote-counting procedure is potentially misleading as it conflates large and small effect sizes.

Two meta-analyses conducted contemporaneously with the present one, however, provide stronger evidence against a bilingual advantage. First, Paap et al. (2017a) averaged global RT and interference cost scores from 101 studies including Simon, Flanker, Stroop, or Attentional Network tasks. Using raw RT difference scores, rather than effect sizes, they analyzed small sample studies ($n < 40$) and large sample studies ($n > 40$) separately. Amongst small sample studies they found bilingual advantages of 107 ms for interference costs and 114 ms for global RTs, both of which significantly differed from 0. However, amongst large sample studies, they found bilingual advantages of only 6 ms for interference costs and 9 ms for global RTs. Paap et al. (2017a) interpreted the findings as providing weak evidence for a bilingual advantage.

Second, Lehtonen et al. (2018) conducted a large-scale meta-analysis of studies comparing bilingual adults on a broad array of cognitive tasks, including scores derived from interference-control tasks, and considered several theoretically and methodologically significant moderators. They found a non-significant effect size for monitoring, which included global RT, and a very small significant effect size for inhibition, which included interference cost, that became non-significant after correcting for publication bias. Moreover, they found that task did not moderate effect sizes within either of the broader domains, suggesting that effect sizes on individual interference-control tasks were non-significant. Further, they found that age and AoA, both considered in the present analysis, did not significantly moderate effect sizes. The contribution of Lehtonen et al. (2018) is important in that it provides a broad review of different tasks, and yet it only found a small, tenuous bilingual advantage for inhibition that may be attributable to publication bias. However, a meta-analysis that is more focused on a narrower array of tasks may be more likely to uncover group differences. Furthermore, Lehtonen et al. only examined studies with adults, but it is possible that any bilingual advantage may be more pronounced in childhood, when executive functions are still developing.

The current study

While these recent reports provide clearer evidence against a bilingual advantage, the debate is still heated (Bialystok, 2017). We conducted a series of meta-analyses, using a multiverse approach (Steege, Tuerlinckx, Gelman, & Vampaemel, 2016). As is evident from various meta-analyses of research on cognitive effects of other experiential variables, such as videogame play, different meta-analyses on the same topic can reach different conclusions because of different decisions

about inclusion criteria, coding, and analytic strategies (e.g., Bediou, Adams, Mayer, Tipton, Green, & Bavelier, 2018; Powers, Brooks, Aldrich, Palladino, & Alfieri, 2013; Sala, Tatlidil, & Gobet, 2018). It is therefore important to see how robust conclusions are over variations in these factors. Our multiverse analytic approach was inspired by the recommendations of Steege, Tuerlinckx, Gelman, & Vanpaemel (2016). These authors note that in a given empirical study the dataset is often one of a large set of reasonable datasets; conducting the same analysis over the entire multiverse of datasets allows the analyst to determine whether their conclusions are specific to a particular dataset or analytic strategy. While Steege, Tuerlinckx, Gelman, & Vanpaemel (2016) promoted the use of a multiverse analysis in the context of a single empirical study, it seems particularly relevant to meta-analysis, where decisions about coding and inclusion are often imperfect judgment calls. Moreover, the multiverse approach can be extended to alternative approaches to modeling dependences between effect sizes (e.g., multiple tasks within the same study, multiple DVs from the same task), since it is likely there is no ideal model for the observed dependence.

Therefore, in the present analysis, we first conducted what we call our “preferred” analysis that utilized a broad definition of bilinguals but included only standard versions of interference-control tasks. We then conducted a multiverse analysis to determine how sensitive our conclusions were to decisions about inclusion criteria, coding of variables, and plausible analytic strategies. In the multiverse analysis, we varied the definition of bilingual group, task selection, coding of the age of L2 acquisition (AoA), and the treatment of outliers.

Potential moderators of effect sizes

Age One moderator that has been considered extensively in the bilingual advantage literature is age. It is agreed upon that EF exhibits a complex developmental trajectory, developing slowly during childhood, peaking in early adulthood, and declining later in life (Zelazo & Lee, 2010). For example, Waszak, Li, and Hommel (2010) conducted a lifespan study of interference cost on the Flanker task, and observed smallest costs between the ages of 16 and 42 years. Interference cost decreased non-linearly over childhood and increased non-linearly after the age of 42 years. Bialystok, Martin, and Viswanathan (2005b) have argued that when EF is operating at peak efficiency, bilingual advantages might not be easily detected. If bilingualism accelerates the development of and/or ameliorates the decline of this system, stronger advantages for bilinguals may be seen in children and older adults, relative to young adults.

When looking at individual studies, findings are mixed amongst all age groups. For example, while several studies report significant advantages for children (Engel de Abreu,

Cruz–Santos, Tourinho, Martin, & Bialystok, 2012; Kapa & Colombo, 2013; Martin-Rhee & Bialystok, 2008; Poarch & Bialystok, 2015; Poarch & van Hell, 2012; Yang, Yang, & Lust, 2011). A few recent, large-scale studies have failed to observe a bilingual advantage (Antón et al., 2014; Duñabeita et al., 2014; Gathercole et al., 2014). Findings are even more mixed among young adults, with some studies finding a bilingual advantage on interference-control tasks (e.g., Coderre, van Heuven & Conklin, 2013; Costa, Hernández, & Sebastián-Gallés, 2008; Luk, Anderson, Craik, Grady, & Bialystok, 2010; Luk, de Sa, & Bialystok, 2011), and others failing to do so (Bialystok et al., 2004; Gathercole et al., 2014; Paap & Greenberg, 2013; Paap & Sawi, 2014). Results with older adults are also mixed. Bialystok et al. (2004) report several studies finding a bilingual advantage for older adults on the Simon task, as do Salvatierra and Rosselli (2011), but neither Kirk, Fiala, Scott-Brown, and Kempe (2014) nor Gathercole et al. (2014) found one. In their recent meta-analysis, Lehtonen et al. (2018) found no effect of age, providing the clearest available evidence against a moderating effect of age; however, this analysis considered a narrower age range than the present study (only adults) and a broader array of tasks.

In testing for a moderating effect of age, at least three patterns of results might be expected to hold. First, there could be a main effect of age, with children and perhaps older adults showing a larger effect of bilingualism on EF than young adults. Alternatively, if the bilingual advantage was specific to global RT or interference cost, there could be an interaction between age and the DV, with children and older adults differing from young adults on only one of the two measures. Or there could be larger effect sizes for children than young adults on one DV and larger effect sizes for older adults than young adults on the other DV. If this were the case, it would suggest that the two DVs reflect different EF constructs over the lifespan.

Task Several interference-control tasks have been used to compare monolinguals and bilinguals. While these tasks are often viewed as interchangeable indicators of the same constructs, there are several reasons to doubt this claim. First, while factor-analytic studies have reported that many tasks of inhibition load onto a single factor (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000), few studies have utilized multiple interference-control tasks (e.g., Simon, Flanker, and Stroop), and when they have, they have typically not used interference cost as the DV. For example, Miyake et al. (2000) found that the difference between incongruent and neutral trials (trials without distracters) on the Stroop task loaded on an inhibition factor (later re-named common EF; Miyake & Friedman, 2012), along with the stop-signal and anti-saccade tasks. In a subsequent study, Friedman and Miyake (2004) found that the difference in RT between congruent and neutral

trials on the Flanker task loaded on a shared factor with the analogous measure of cost from the Stroop task; however, it is not clear whether interference cost would follow the same pattern. Second, as reviewed by Paap and Sawi (2014), the correlations between raw scores on these tasks are often quite low, especially for interference cost. For example, Paap and Greenberg (2013) reported a non-significant correlation between interference cost in the Simon and Flanker tasks ($r = .01$); see also Salthouse (2010). Strikingly, measures of interference cost across different versions of the Stroop task appear to be uncorrelated as well (Shilling, Chetwynd, & Rabbitt, 2002). Global RT tends to exhibit larger correlations; for example, Paap and Sawi (2014) report correlations of .60 between global RT from Simon and Flanker tasks. Given this pattern of correlations, it is sensible to question whether the Simon, Flanker, and Stroop tasks are indices of the same EF constructs. If the three tasks measure separable constructs, a bilingual advantage may emerge on only a single task. It is, therefore, possible that we may observe a main effect of task on effect sizes, or an interaction of task with the DV. However, as stated above, Lehtonen et al. (2018) did not find that task moderated bilingual advantages in their meta-analysis.

Age of L2 acquisition Researchers have speculated that the effects of bilingualism on interference-control tasks might be influenced by the age at which the bilingual subject learned their L2; however, as with other potential moderators, effects have been mixed. Using the Attentional Network Test as a measure of EF, Pelham and Abrams (2014) reported a bilingual advantage of similar magnitude for Spanish-English bilinguals who acquired their L2 either in childhood or in adulthood, relative to monolinguals. In contrast, other studies suggest larger effects for bilinguals with early AoA of their L2. Luk et al. (2011) compared college-aged early bilinguals (average age of regular use of L2 = 5 years) to late bilinguals (average age of regular use of L2 = 15 years) and to monolinguals on the Flanker task. They found that early bilinguals exhibited significantly smaller interference cost than the other two groups, who did not differ significantly. Subsequent regression analyses found that AoA, treated as a continuous variable, was positively related to interference cost. Tao, Marzecová, Taft, Asanowicz, and Wodniecka (2011) compared college-aged monolinguals to early (average age of exposure to L2 = 3 years old) and late (average age of exposure to L2 = 8 years old) bilinguals on a lateralized Flanker task. Early bilinguals exhibited faster global RT and smaller interference cost than monolinguals, whereas late bilinguals only exhibited smaller interference cost relative to monolinguals. Kapa and Colombo (2013) compared 10-year-old monolinguals, early bilinguals (L2 learned before age 3 years), and late bilinguals (L2 learned after age 3 years) on the Flanker task. They found that, after controlling for age and English receptive vocabulary, early bilinguals

outperformed both the late bilinguals and the monolinguals. Note, however, that the cut-off used by Kapa and Colombo (2013) to distinguish early and late AoA is notably earlier than the other studies (presumably due to the young age of the sample).

Taken together, these studies suggest that the effect of the bilingualism may vary as a function of the AoA of the L2, with early bilinguals exhibiting a larger bilingual advantage than late bilinguals. Moreover, if the bilingual advantage is specific to a particular DV, there might be an interaction between AoA and DV, with the larger effect of AoA in modulating the effect of bilingualism on one of the two DVs. Although Lehtonen et al. (2018) did not find AoA to moderate their findings, we revisited this issue here using a multiverse approach to coding AoA.

Research questions

The current meta-analysis explored the following research questions:

- 1) What is the magnitude of the effect of bilingualism on global RT and interference cost?
- 2) Is the bilingual advantage restricted to particular interference-control tasks, such as the Simon or Flanker tasks? If so, does task interact with the DV (global RT vs. interference cost)?
- 3) Does the bilingual advantage vary as a function of the age of participants, with smaller effects for young adult samples than for children or older adults? Does the effect of age interact with the DV?
- 4) Does the bilingual advantage vary as a function of the AoA of the L2, with larger effects for early bilinguals than for late bilinguals? If so, does this effect interact with the DV?
- 5) How robust are the results given the variety of plausible data sets and coding procedures?

Method

Literature search

In this section we describe our method for finding and selecting studies to be included in the analyses.

PsycINFO was searched periodically until December 2017. Search terms included a combination of *bilingual* or *bilingualism* with *executive control*, *executive function*, *inhibition*, or *interference control*. Reference sections of the relevant studies and review articles (e.g., de Bruin et al., 2015b; Paap et al., 2015; Lehtonen et al., 2018) were examined to identify additional studies for inclusion. Of the 286

studies consulted, 80 met the following criteria for inclusion (see Fig. 1 for more details):

1. Study included a bilingual group. We defined bilingualism in two different ways, to create two separate datasets for our multiverse analysis. First, under a broad definition, we treated any group designated as bilingual by the research team as bilingual. Second, under a narrow definition of bilingual, we looked for groups who could be reasonably described as near balanced on the basis of reported proficiency.

The first dataset, which used the broad definition, included all studies that had a group designated as bilingual; see Table 1 for a description of all datasets. This included lapsed and late bilinguals, but excluded groups designated as second language learners (excluding one study and two groups from included studies) and bidialectals (excluding one study, and two groups from included studies). We favor this broad definition as we believe the researchers of the original studies were in the best position to define their groups as bilingual or monolingual, given the combination of demographic information, familiarity with the local sociolinguistic context, and familiarity with the participants. We therefore used this dataset in our preferred analysis.

The second dataset used a narrow definition of bilingual and included only studies with participants who were either of nearly balanced proficiency, or, if proficiency information was unavailable, reported nearly balanced usage between the two languages. Studies were excluded if the authors did not provide information about their bilingual participants' language proficiency or exposure. To identify groups of balanced bilinguals, we first looked at information about proficiency. If self-reported proficiency was reported for both L1 and L2, we calculated the ratio of the weaker to the stronger language. When proficiency was reported separately for speaking, listening, writing, and reading, we only considered speaking and listening, so as not to eliminate populations who had not been schooled in one of their languages. If only average proficiency across those four domains was reported, we used those scores. If proficiency was not reported for the two languages, but was reported for the weaker language, we took the ratio of the weaker language to the maximum possible score. We preferred the ratio of reported L1 to L2 proficiency to this score, as the former accounts for response bias. If self-reported proficiency was not reported, but a linguistic measure was given in L1 and L2, we took the ratio of scores on these instruments. We preferred self-reported proficiency because it is a holistic measure of language proficiency, whereas many of the standardized tasks were specific to vocabulary. Several studies did not report proficiency but did report measures of usage or

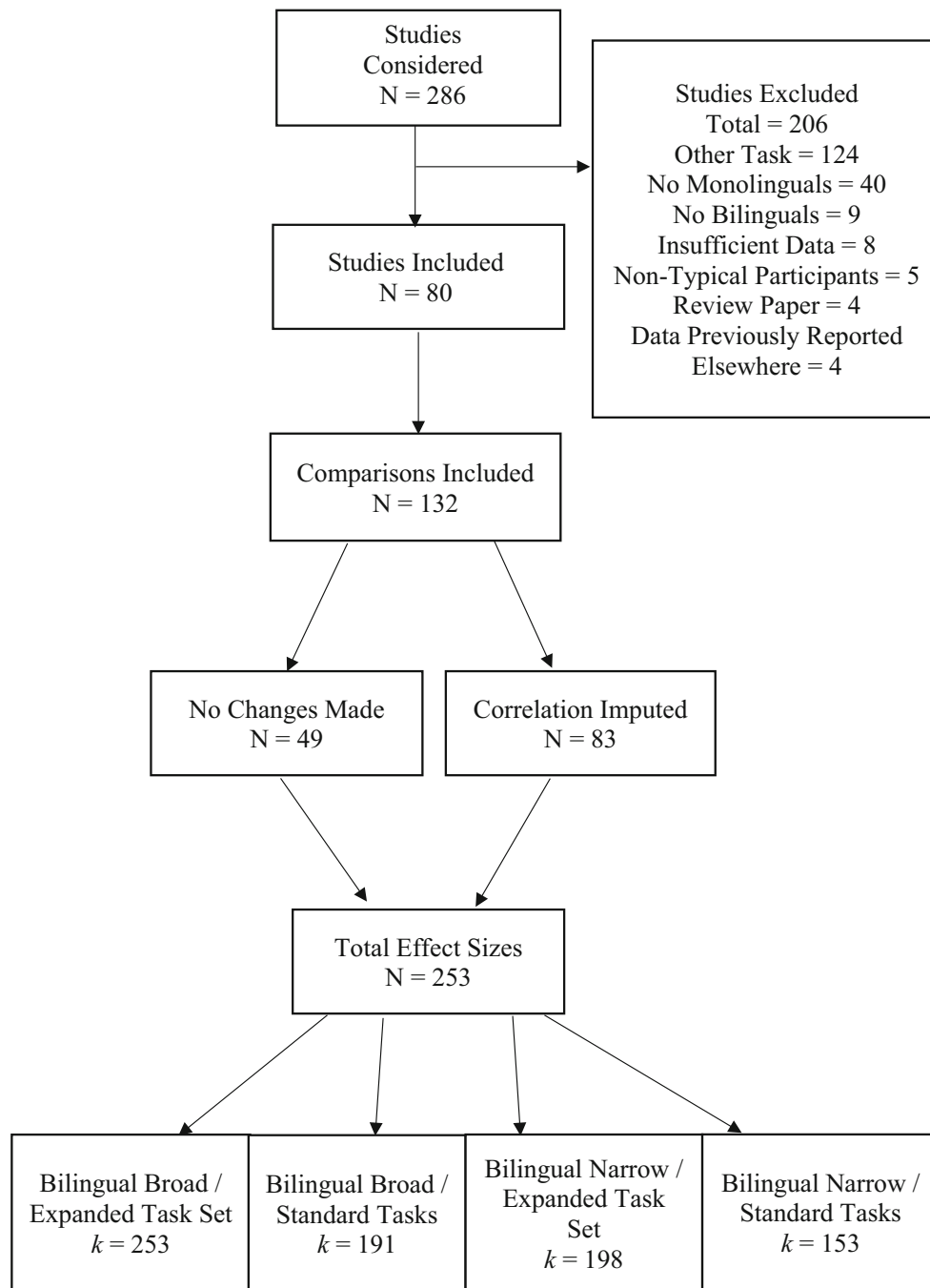


Fig. 1 PRISMA flow chart indicating the steps involved in determining the eligibility and inclusion of studies in the meta-analysis of impact of bilingualism on interference-control tasks

exposure to each language. One subset of these studies reported Likert scales ranging from all one language, to all the other language, with the middle value reflecting equal usage of the two languages. In studies with children that provided one item for the language used by the child and another item for the language spoken to the child, we averaged across those two, but ignored items about siblings and media (radio, television, and books). In studies with adults that presented Likert scales for several time

points (e.g., childhood, school age), we converted the Likert-scale scores into proportions and averaged across all time points.

As an operational definition of balanced bilingual, we selected bilingual groups who had at least a ratio of .66 for proficiency in their weaker language to their stronger language, or between .33 and .66 usage or exposure to each of the two languages. We also added bilingual groups from studies that did not report mean proficiency or mean

exposure, but said all bilinguals had above a given value of mean proficiency or exposure, if that value met our thresholds. Note that for some studies we combined a different set of bilingual groups in analyses using the broad versus narrow definitions of bilingualism; see section below on *Data preparation and effect size calculation*.

2. Study included at least one monolingual group. For both the main analysis and the multiverse analysis we treated a group as monolingual if they were designated as such by the research group.
3. Participants were at least 4.5 years old and without psychological impairment. We therefore excluded studies examining potential benefits of bilingualism as a protective factor in dementia (e.g., Bialystok, Craik, & Freedman, 2007).
4. Study contained RT data or efficiency scores (RT divided by accuracy) from at least one of five commonly used non-verbal interference-control tasks: Flanker, Simon, Simon Arrows, Numerical Stroop, and Attentional Network Test (ANT). All these tasks share the following structure: Participants are asked to make judgments about a visually presented target stimulus, where either an additional stimulus, or an irrelevant dimension of the target stimulus, can elicit a competing response. On incongruent trials the distracter (stimulus or dimension) elicited a response different from the target; on congruent trials the distracter (stimulus or dimension) elicited the same response as the target. We avoided tasks with explicitly verbal stimuli, such as the traditional Stroop and Verbal Flanker tasks, and tasks that involved a non-keypress response. In our preferred analysis, we considered standard versions of these tasks, which did not, for example, contain unbalanced numbers of congruent and incongruent trials; see section on *Moderator coding* for more details. However, in our multiverse analysis, we considered two datasets, one with standard versions and one with all versions of tasks; see Table 1 for a description of how many instances of each task type were included in each of the datasets.

Data preparation and effect size calculation

The 80 studies comprised 253 comparisons, each involving a bilingual group compared with a monolingual group on a single task; see Table 2 for a complete list of the comparisons with moderator values and effect sizes for global RT and interference cost. Hence, if a study reported on several separate bilingual and monolingual groups (e.g., varying in age), each of these groups was included as a separate comparison. Care was taken to minimize the number of statistically dependent comparisons while maximizing the number of available effect sizes, as follows:

First, if a study contained a single monolingual group and two bilingual groups and the bilingual groups did not differ on AoA of the L2 (see *Moderator coding* below), the two bilingual groups were averaged together. However, if the two bilingual groups differed in AoA and were compared to the same monolingual group, the *N* for the monolingual group was split in half across the two comparisons. For example, Luk et al. (2011) compared one monolingual group with early and late bilingual groups. Since the early and late bilingual groups differed on AoA, they were included as separate comparisons. The same monolingual mean was included in both comparisons, but the monolingual *N* for each comparison was half the *N* in the study. This is done because including the same control group in two effect sizes without correcting the *N* double counts those participants, thereby artificially inflating the number of observations (Higgins, Deeks, & Altman, 2008).

Second, if multiple tasks were given to the same sample, each task counted as a separate comparison. This amounts to assuming that scores on these tasks are independent. As described in the *Introduction*, interference costs tend to be uncorrelated, whereas global RT tends to correlate across interference-control tasks. Thus, the assumption of independence is violated for global RT from studies that included multiple interference-control tasks. Results should be interpreted with some caution because of this violation. However, if a study reported multiple blocks on the same task, and each block was reported separately, only the first block was included. This was done because averaging standard deviations across blocks

Table 1 Number of comparisons for moderator variables in each of the four datasets used in the multiverse analysis. Note that the preferred analysis used a broad definition of bilingual and standard versions of tasks

Data Set		Task				Age, y			Age of Acquisition	
Bilingual Definition	Task Set	Flanker	Simon	ANT	Stroop-like	Children	YA	OA	Early	Late
Broad	Standard	22	54	11	12	24	47	15	60	25
Broad	Expanded	24	59	29	20	29	71	17	87	29
Narrow	Standard	20	42	8	9	20	38	12	50	20
Narrow	Expanded	22	46	21	44	23	58	12	68	23

Table 2 Effect sizes and moderators for all comparisons

Study	Bilingual Broad	Bilingual Narrow	N Bilingual	N Monolingual	Task Version	AoA Strict	Age	Global RT	Interference Cost
Abutalebi et al. 2012 Block 1	Yes	No	17	14	ANT Non-Standard	Early No	Younger Adults	-0.36 (0.13)	0 (0.12)
Abutalebi et al. 2015	Yes	Yes	19	19	ANT Non-Standard	Late Yes	NA	0.82 (0.11)	0.2 (0.1)
Ansaldi et al. 2015	Yes	No	10	10	Simon Standard	Late Yes	Older Adults	0.33 (0.19)	-0.1 (0.18)
Anton et al. 2014	Yes	Yes	180	180	ANT Non-Standard	Early Yes	Children	-0.12 (0.01)	-0.09 (0.01)
Anton et al. 2016 Numerical Stroop	Yes	Yes	24	24	Stroop-like Standard	Late No	Older Adults	0.06 (0.08)	0.01 (0.08)
Antoniou et al. 2016	Yes	Yes	22	25	Simon Standard	NA NA	Children	0.53 (0.09)	0.27 (0.08)
Arredondo et al. 2017	Yes	Yes	13	14	Flanker Standard	Early Yes	Children	0.1 (0.14)	-0.02 (0.14)
Barac et al. 2016	Yes	Yes	25	37	ANT Non-Standard	Early Yes	Children	0.09 (0.07)	-0.34 (0.07)
Bialystok et al. 2004 Study 1 Young	Yes	Yes	10	10	Simon Standard	Early Yes	NA	2.3 (0.3)	2.87 (0.37)
Bialystok et al. 2004 Study 1 Old	Yes	Yes	10	10	Simon Standard	Early Yes	Older Adults	1.16 (0.21)	1.03 (0.21)
Bialystok et al. 2004 Study 2 (2 Squares) Young	Yes	Yes	32	32	Simon Standard	Early Yes	NA	4.69 (0.23)	1.73 (0.08)
Bialystok et al. 2004 Study 2 (2 Squares) Old	Yes	Yes	15	15	Simon Standard	Early Yes	Older Adults	1.07 (0.14)	2.03 (0.19)
Bialystok et al. 2004 Study 3	Yes	Yes	10	10	Simon Standard	Early No	NA	6.11 (1.04)	5.07 (0.77)
Bialystok et al. 2005a	Yes	No	19	10	Simon Standard	Early No	Younger Adults	0.79 (0.15)	0.02 (0.14)
Bialystok 2006 Players Simon	Yes	No	19	17	Simon Non-Standard	Early Yes	Younger Adults	0.18 (0.11)	0.08 (0.11)
Bialystok 2006 Non-Players Simon	Yes	No	30	31	Simon Non-Standard	Early Yes	Younger Adults	-0.16 (0.06)	-0.07 (0.06)
Bialystok 2006 Players Simon Arrows	Yes	No	19	17	Stroop-like Non-Standard	Early Yes	Younger Adults	0.29 (0.11)	-0.07 (0.11)
Bialystok 2006 Non-Players Simon Arrows	Yes	No	29	31	Stroop-like Non-Standard	Early Yes	Younger Adults	0.28 (0.07)	-0.38 (0.07)
Bialystok et al. 2008 Young	Yes	Yes	24	24	Stroop-like Standard	Early No	Younger Adults	0.47 (0.08)	0.66 (0.09)
Bialystok et al. 2008 Old	Yes	Yes	24	24	Stroop-like Standard	Late Yes	Older Adults	-0.05 (0.08)	0.46 (0.08)
Bialystok & Depape 2009	Yes	Yes	24	24	Stroop-like Standard	NA NA	Younger Adults	1.03 (0.09)	0.06 (0.08)
Bialystok et al. 2017 Non-ADHD	Yes	No	45	44	Flanker Standard	NA NA	Younger Adults	-0.31 (0.04)	-0.18 (0.04)
Billig & School 2011 Middle Aged Simon Arrows	Yes	No	20	21	Stroop-like Standard	NA No	NA	0.18 (0.09)	0.38 (0.1)
Billig & School 2011 Elderly Simon Arrows	Yes	No	21	21	Stroop-like Standard	NA No	NA	0.17 (0.09)	0.24 (0.09)

Table 2 (continued)

Study	Bilingual Broad	Bilingual Narrow	N Bilingual	N Monolingual	Task	Standard Task	AoA Strict	Age	Global RT	Interference Cost
Blom et al. 2017 Polish-Frisian	no	Yes	88	44	Flanker	Standard	Early	Children	NA	-0.35 (0.03)
Blom et al. 2017 All 3	Yes	No	132	44	Flanker	Standard	Early	Children	NA	-0.22 (0.03)
Blumenfeld & Marian 2014 Study 1 Simon Arrows	Yes	Yes	30	30	Stroop-like	Non-Standard	Early	Younger Adults	0.36 (0.07)	0.59 (0.07)
Blumenfeld & Marian 2014 Study 1: Simon	Yes	Yes	30	30	Simon	Non-Standard	Early	Younger Adults	-0.2 (0.07)	-0.31 (0.04)
Blumenfeld & Marian 2014 Study 2 Stroop	Yes	Yes	60	60	Stroop-like	Non-Standard	Early	Younger Adults	0.06 (0.03)	-0.04 (0.03)
Blumenfeld & Marian 2014 Study 2: Simon Arrows	Yes	Yes	60	60	Simon	Non-Standard	Early	Younger Adults	-0.36 (0.03)	-0.21 (0.03)
Blumenfeld et al. 2016	Yes	Yes	30	30	Stroop-like	Non-Standard	NA	NA	NA	-0.23 (0.07)
Bogulski et al. 2015 Full & Lapsed	Yes	No	52	27	Flanker	Standard	NA	Younger Adults	0.22 (0.06)	0.07 (0.06)
Bogulski et al. 2015 Full only	no	Yes	30	27	Flanker	Standard	NA	Younger Adults	0.23 (0.07)	0.02 (0.07)
Calvo & Bialystok 2014 Working Class	Yes	Yes	44	20	Flanker	Standard	Early	Children	-0.01 (0.07)	-0.2 (0.07)
Calvo & Bialystok 2014 Middle Class	Yes	Yes	65	46	Flanker	Standard	Early	Children	0.02 (0.04)	0.01 (0.04)
Chabal et al. 2015	Yes	Yes	21	17	Simon	Standard	NA	Younger Adults	NA	0.16 (0.1)
Clare et al. 2016 Simon	Yes	Yes	50	49	Simon	Standard	Early	NA	NA	0.05 (0.04)
Coderre & van Heuven 2014 A Simon	Yes	Yes	58	18	Simon	Standard	Late	Younger Adults	0	0.69 (0.07)
Coderre et al. 2016 Non-linguistic Flanker	Yes	Yes	14	15	Flanker	Non-Standard	Early	Younger Adults	-0.54 (0.14)	-0.16 (0.13)
Costa et al. 2008	Yes	Yes	100	100	ANT	Standard	Early	Younger Adults	0.4 (0.02)	0.25 (0.02)
Costa et al. 2009 Study 1 8% Condition Block 1	Yes	Yes	30	30	ANT	Non-Standard	Early	Younger Adults	0.06 (0.06)	-0.21 (0.07)
Costa et al. 2009 Study 1 92% Condition Block 1	Yes	No	30	30	ANT	Non-Standard	Early	Younger Adults	0.16 (0.06)	0 (0.07)
Costa et al. 2009 Study 2 75% condition block 1	Yes	Yes	31	31	ANT	Non-Standard	Early	Younger Adults	0.78 (0.07)	0.61 (0.07)
Costa et al. 2009 Study 2 50%	Yes	Yes	31	31	ANT	Non-Standard	Early	Younger Adults	0.88 (0.07)	-0.03 (0.06)
Cox et al. 2016	Yes	No	26	64	Simon	Non-Standard	NA	Older Adults	NA	0.56 (0.05)
de Bruin et al. 2015a Full & Lapsed	Yes	No	52	24	Stroop-like	Non-Standard	Early	Older Adults	0.08 (0.06)	0.09 (0.06)
De Leeuw & Bogulski 2016	Yes	Yes	28	14	Flanker	Standard	Early	Younger Adults	0.1 (0.1)	-0.3 (0.1)
Desideri & Bonmalfacci 2018	Yes	No	25	25	ANT	Standard	Early	Younger Adults	0.66 (0.08)	0.59 (0.08)

Table 2 (continued)

Study	Bilingual Broad	Bilingual Narrow	N Bilingual	N Monolingual	Task	Standard Task	AoA	AoA Strict	Age	Global RT	Interference Cost
Duñabeitia et al. 2014 Numerical Stroop	Yes	Yes	252	252	Stroop-like	Standard	Early	Yes	Children	-0.06 (0.01)	0.02 (0.01)
Duncan et al. 2016	Yes	Yes	19	22	Simon	Standard	Late	No	Younger Adults	0.7 (0.1)	0.1 (0.09)
Emmorey et al. 2008 Conflict blocks	Yes	Yes	15	15	Flanker	Standard	Early	Yes	NA	0.99 (0.14)	-0.06 (0.13)
Engel de Abreu et al. 2012	Yes	Yes	40	40	Flanker	Standard	Early	Yes	Children	0.79 (0.05)	0.38 (0.05)
Freeman et al. 2017	Yes	Yes	21	23	Stroop-like	Non-Standard	Early	Yes	Younger Adults	NA	-0.41 (0.09)
Gathercole et al. 2014 5 year olds All Bilingual	Yes	No	51	20	Simon	Standard	Early	Yes	Children	-0.71 (0.07)	-0.11 (0.07)
Gathercole et al. 2014 Primary Schoolers All Bilingual	Yes	No	51	11	Simon	Standard	Early	No	Children	-0.25 (0.11)	-0.03 (0.11)
Gathercole et al. 2014 Teenagers All Bilingual	Yes	No	94	20	Simon	Standard	Early	No	NA	0.29 (0.06)	-0.03 (0.06)
Gathercole et al. 2014 Young Adults All Bilingual	Yes	No	65	20	Simon	Standard	Early	No	Younger Adults	-0.41 (0.07)	0.01 (0.06)
Gathercole et al. 2014 Elderly All Bilingual	Yes	No	64	20	Simon	Standard	Early	No	Older Adults	-0.07 (0.06)	0.03 (0.06)
Gathercole et al. 2014 5-year-olds WEH only	no	Yes	19	20	Simon	Standard	Early	Yes	Children	-0.19 (0.1)	1.58 (0.13)
Gathercole et al. 2014 Primary Schoolers WEH only	no	Yes	17	11	Simon	Standard	Early	No	Children	-0.22 (0.14)	-0.13 (0.14)
Gathercole et al. 2014 Teenagers WEH only	no	Yes	31	20	Simon	Standard	Early	No	NA	0.31 (0.08)	-0.1 (0.08)
Gathercole et al. 2014 Young Adults WEH only	no	Yes	23	20	Simon	Standard	Early	No	Younger Adults	-0.36 (0.09)	-0.04 (0.09)
Gathercole et al. 2014 Elderly WEH only	no	Yes	17	20	Simon	Standard	Early	No	Older Adults	-0.19 (0.1)	-0.05 (0.1)
Grady et al. 2015	Yes	No	14	14	Simon	Standard	Late	No	Older Adults	0.15 (0.14)	-0.05 (0.13)
Grant & Dennis 2017	Yes	No	26	25	Flanker	Standard	NA	NA	Younger Adults	0.72 (0.08)	-0.4 (0.08)
Grundy et al. 2017 Study 1 Flanker	Yes	Yes	31	28	Flanker	Standard	Early	Yes	Younger Adults	0.14 (0.07)	0.24 (0.07)
Grundy et al. 2017 Study 2 Flanker	Yes	Yes	58	53	Flanker	Standard	Early	Yes	Younger Adults	-0.12 (0.04)	0.12 (0.04)
Grundy et al. 2017 Study 3 Flanker	Yes	Yes	56	55	Flanker	Standard	Early	Yes	Younger Adults	-0.14 (0.04)	0.07 (0.04)
Hernandez et al. 2010	Yes	No	41	41	Stroop-like	Standard	Early	No	Younger Adults	0.49 (0.05)	0.03 (0.05)
Jiao et al. 2017 Flanker Standard	Yes	Yes	31	27	Flanker	Standard	NA	NA	Younger Adults	0.58 (0.07)	0.06 (0.07)
Kapa & Colombo 2013	Yes	Yes	57	22	ANT	Non-Standard	Early	Yes	Children	0.1 (0.06)	0.05 (0.06)
Kirk et al. 2014 Asian-English	Yes	Yes	16	16	Simon	Standard	Late	Yes	Older Adults	-0.67 (0.13)	-0.05 (0.12)

Table 2 (continued)

Study	Bilingual Broad	Bilingual Narrow	N Bilingual	N Monolingual	Task	Standard Task	AoA Strict	Age	Global RT	Interference Cost
Kirk et al. 2014 Gaelic-English	Yes	Yes	16	16	Simon	Standard	Early	Older Adults	0.1 (0.12)	-0.05 (0.12)
Kousate & Phillips 2012 Simon	Yes	Yes	26	25	Simon	Standard	NA	Younger Adults	-0.58 (0.08)	-0.08 (0.08)
Kousate & Phillips 2012 Flanker	Yes	Yes	26	25	Flanker	Standard	NA	Younger Adults	-0.26 (0.08)	0.09 (0.08)
Kousate et al. 2014 Older Adults	Yes	Yes	36	61	Simon	Standard	Late	Older Adults	-0.04 (0.04)	0.01 (0.04)
Kousate et al. 2014 Younger Adults	Yes	Yes	51	70	Simon	Standard	Late	Younger Adults	-0.04 (0.03)	0 (0.03)
Kousate & Phillips 2017 Flanker	Yes	Yes	20	20	Flanker	Standard	Early	Older Adults	0.28 (0.1)	-0.24 (0.1)
Kousate & Phillips 2017 Simon	Yes	Yes	21	21	Simon	Standard	Early	Older Adults	0.52 (0.09)	-0.18 (0.09)
Kramer & Meota 2015 Early Young	Yes	No	10	10	Simon	Standard	Early	Younger Adults	0.04 (0.18)	0.08 (0.18)
Kramer & Meota 2015 Early Middle	Yes	No	14	14	Simon	Standard	Early	NA	0.02 (0.13)	-0.08 (0.13)
Kramer & Meota 2015 Early Older	Yes	No	14	14	Simon	Standard	Early	Older Adults	-0.4 (0.14)	0.35 (0.14)
Kramer & Meota 2015 Later Older	Yes	No	14	14	Simon	Standard	Late	Younger Adults	0.56 (0.14)	0.64 (0.14)
Ladas et al. 2015 Study 1	Yes	Yes	26	24	ANT	Standard	Early	Children	-0.36 (0.08)	-0.09 (0.08)
Ladas et al. 2015 Study 2	Yes	Yes	28	32	ANT	Standard	Early	Children	0.11 (0.07)	0.03 (0.07)
Luk et al. 2011 early	Yes	Yes	43	19	Flanker	Standard	Early	Younger Adults	-0.02 (0.07)	0.28 (0.07)
Luk et al. 2011 late	Yes	Yes	42	19	Flanker	Standard	Late	Younger Adults	-0.07 (0.07)	0.03 (0.07)
Luk et al. 2010	Yes	Yes	10	10	Flanker	Non-Standard	Early	Younger Adults	0.4 (0.19)	0.17 (0.18)
Marian et al. 2014	Yes	Yes	17	18	Simon	Standard	Early	Younger Adults	NA (0.11)	0.18 (0.11)
MartinRhee & Bialystok 2008 Study 1	Yes	Yes	17	17	Simon	Standard	Early	Children	1.1 (0.13)	-0.05 (0.11)
MartinRhee & Bialystok 2008 Study 2	Yes	No	20	21	Simon	Standard	Early	Children	0.44 (0.1)	0.44 (0.1)
MartinRhee & Bialystok 2008 Study 3	Yes	No	13	19	Simon	Standard	Early	Children	0.29 (0.12)	0.44 (0.13)
Marzecova et al. 2013	Yes	Yes	18	17	ANT	Non-Standard	Early	Younger Adults	0.03 (0.11)	0.41 (0.11)
Mohades et al. 2014 Simon	Yes	No	19	14	Simon	Standard	Early	Children	0.32 (0.12)	-0.24 (0.12)
Mohades et al. 2014 Numerical Stroop	Yes	No	19	14	Stroop-like	Standard	Early	Children	-0.02 (0.12)	-0.54 (0.12)
Mor et al. 2014 Simon Arrows	Yes	Yes	20	20	Stroop-like	Standard	Late	Younger Adults	-0.8 (0.1)	0.25 (0.1)

Table 2 (continued)

Study	Bilingual Broad	Bilingual Narrow	N Bilingual	N Monolingual	Task	Standard Task Version	AoA Strict	Age	Global RT	Interference Cost
Mor et al. 2015 Numerical Stroop	Yes	Yes	20	20	Stroop-like	Standard	Late No	Younger Adults	0.28 (0.1)	0.9 (0.11)
Morales et al. 2013	Yes	Yes	27	29	Simon	Standard	Early Yes	Children	0.47 (0.07)	0.13 (0.07)
Morton & Harper 2007	Yes	Yes	17	17	Simon	Standard	Late Yes	Children	-0.23 (0.11)	-0.23 (0.11)
Nair et al. 2017 Simon	Yes	Yes	18	18	Simon	Standard	Early Yes	NA	1.37 (0.13)	0.13 (0.11)
Nair et al. 2017 ANT	Yes	Yes	18	18	ANT	Standard	Early Yes	NA	0.41 (0.11)	0.42 (0.11)
Niamazi & Thordarottir 2010	Yes	Yes	15	30	Simon	Standard	Early Yes	Children	-0.04 (0.1)	-0.04 (0.1)
Paap & Greenberg 2013 Study 1 Simon	Yes	Yes	34	45	Simon	Standard	Early Yes	Younger Adults	-0.33 (0.05)	0.02 (0.05)
Paap & Greenberg 2013 Study 2 Simon	Yes	Yes	36	50	Simon	Standard	Early Yes	Younger Adults	0.08 (0.05)	-0.13 (0.05)
Paap & Greenberg 2013 Study 3 Simon	Yes	Yes	51	55	Simon	Standard	Early Yes	Younger Adults	-0.02 (0.04)	-0.23 (0.04)
Paap & Greenberg 2013 ANT	Yes	Yes	49	55	ANT	Non-Standard	Early Yes	Younger Adults	-0.06 (0.04)	-0.06 (0.04)
Paap & Sawi 2014 ANT	Yes	Yes	58	62	ANT	Non-Standard	Early Yes	Younger Adults	-0.28 (0.03)	0.28 (0.03)
Paap & Sawi 2014 Simon	Yes	Yes	58	62	Simon	Standard	Early Yes	Younger Adults	-0.5 (0.03)	-0.46 (0.03)
Pelham & Abrams 2014 Early	Yes	Yes	30	15	ANT	Standard	Late Yes	Younger Adults	0.33 (0.1)	0.38 (0.1)
Pelham & Abrams 2014 Late	Yes	Yes	30	15	ANT	Standard	Late Yes	Younger Adults	0.2 (0.1)	0.34 (0.1)
Poarch & Van Hell 2013	Yes	Yes	36	20	Simon	Standard	Early Yes	Children	0.04 (0.08)	0.33 (0.08)
Poarch & Bialystok 2015	Yes	Yes	99	60	Flanker	Standard	Early No	Children	-0.01 (0.03)	0.56 (0.03)
Rainey et al. 2016	Yes	Yes	66	26	Flanker	Standard	Early No	Children	-0.11 (0.05)	-0.11 (0.05)
Rodrigues & Zimmer 2016 Simon	Yes	No	20	20	Simon	Standard	NA NA	NA	0.56 (0.1)	-0.35 (0.1)
Rodrigues & Zimmer 2016 ANT	Yes	No	20	20	ANT	Standard	NA NA	NA	0.42 (0.1)	0.23 (0.1)
Ross & Melinger 2017 Simon	Yes	Yes	54	45	Simon	Standard	Early Yes	Children	0.04 (0.04)	-0.01 (0.04)
Ross & Melinger 2017 Flanker	Yes	Yes	54	45	Flanker	Standard	Early Yes	Children	-0.08 (0.04)	0.25 (0.04)
Roselli et al. 2016 High Proficiency vs. High Proficiency Mono	Yes	Yes	20	20	Simon	Standard	Early Yes	Younger Adults	NA (0.1)	-0.2 (0.1)
Roselli et al. 2016 Low Proficiency & Unbalanced vs. Low Proficiency Mono	Yes	Yes	54	20	Simon	Standard	Late Yes	Younger Adults	NA (0.07)	-0.39 (0.07)
Ryskin et al. 2014 Exp 2	Yes	Yes	20	21	ANT	Non-Standard	Early Yes	Younger Adults	NA (0.1)	0.42 (0.1)

Table 2 (continued)

Study	Bilingual Broad	Bilingual Narrow	N Bilingual	N Monolingual	Task	Standard Task Version	AoA Strict	Age	Global RT	Interference Cost
Salvatierra & Rosselli 2011 Young (2 squares)	Yes	Yes	67	66	Simon	Standard	Late	Younger Adults	-0.21 (0.03)	0.22 (0.03)
Salvatierra & Rosselli 2011 Old (2 Squares)	Yes	Yes	58	42	Simon	Standard	Late	Older Adults	-0.1 (0.04)	0.71 (0.04)
Schroeder & Marian 2012	Yes	Yes	18	18	Simon	Standard	Late	Older Adults	0.12 (0.11)	0.69 (0.11)
Schroeder et al. 2016 Non-Musicians	Yes	Yes	43	64	Simon	Standard	Early	Younger Adults	-0.12 (0.04)	0.09 (0.04)
Schroeder et al. 2016 Musicians	Yes	Yes	69	42	Simon	Standard	Early	Younger Adults	0 (0.04)	0.19 (0.04)
Shulley & Shake 2016	Yes	Yes	52	52	Stroop-like	Non-Standard	Late	Younger Adults	-0.49 (0.04)	-0.03 (0.04)
Tao et al. 2011 Early	Yes	Yes	36	17	ANT	Standard	Early	Younger Adults	1.02 (0.09)	0.12 (0.08)
Tao et al. 2011 Late	Yes	Yes	30	17	ANT	Standard	Late	Younger Adults	0.32 (0.09)	0.41 (0.09)
Vivas et al. 2017	Yes	No	45	45	ANT	Standard	Late	Younger Adults	-0.52 (0.05)	0.01 (0.04)
Woumans et al. 2015 Simon - Balanced Group	Yes	Yes	31	30	Simon	Standard	Early	Younger Adults	-0.46 (0.07)	0.66 (0.07)
Woumans et al. 2015 Simon - Unbalanced Group	Yes	No	34	30	Simon	Standard	Late	Younger Adults	-0.08 (0.06)	0.59 (0.06)
Woumans et al. 2015 ANT - Balanced Group	Yes	Yes	31	15	ANT	Non-Standard	Early	Younger Adults	0.57 (0.1)	0.41 (0.1)
Woumans et al. 2015 ANT - Unbalanced Group	Yes	No	34	15	ANT	Non-Standard	Late	Younger Adults	0.79 (0.1)	0.15 (0.09)
Xie & Dong 2015 Flanker	Yes	Yes	33	32	Flanker	Standard	Late	Younger Adults	0.06 (0.06)	-0.04 (0.06)
Xie & Dong 2015 Number Stroop	Yes	Yes	32	31	Stroop-like	Standard	Late	Younger Adults	-0.17 (0.06)	-0.3 (0.06)
Yang Yang & Lust 2011	Yes	No	15	41	ANT	Non-Standard	Early	Children	0.79 (0.09)	0.27 (0.09)
Yang & Yang 2016 Children	Yes	No	32	31	ANT	Non-Standard	Early	Children	0.71 (0.07)	0.22 (0.06)
Yang & Yang 2016 College Adults	Yes	Yes	20	19	ANT	Non-Standard	Late	Younger Adults	0.88 (0.11)	1.23 (0.12)
Yudes et al. 2011	Yes	Yes	16	16	Simon	Standard	NA	Younger Adults	0.27 (0.12)	-0.09 (0.12)

Note. In some cases, a comparison had to be defined differently under broad vs. narrow definitions of bilingual. As a result, some comparisons in the above table are repeated, with one row corresponding to the broadly defined bilingual group and another corresponding to the narrowly defined bilingual group. For rows with “no” in the ilingual Broad column, the relevant row was included only in analyses using a narrow definition of bilingual where the comparison had been defined differently from the analysis using the broad definition

would require assuming a correlation between consecutive blocks on the same task. Moreover, if a study included more than one condition of the same task (e.g., with varying working memory demands), only trials from the standard condition were chosen. The reason we did not include multiple blocks or conditions from the same task is that these are far more likely to be correlated than different tasks. Averaging them together would have required knowing the correlations between separate blocks or conditions of the same task.

Each comparison typically yielded two DVs, for global RT and interference cost, respectively. Recall that interference cost is calculated as the difference in RTs for congruent versus incongruent trials, and global RT is calculated as the average RT across these two trial types. In some studies, these scores and their standard deviations were reported. However, in many cases, means and standard deviations (or standard errors) were reported for the congruent and incongruent trials, but not for global RT or interference cost. Because RTs for congruent and incongruent trials are correlated, calculating the standard deviation for global RT and interference cost from these scores requires the correlation between trial types. Two strategies were employed to estimate this correlation: When the mean and standard deviations of RTs for congruent and incongruent trials and for interference cost were reported, it was possible to solve algebraically for the correlation coefficients. However, in most cases correlation coefficients had to be stipulated. In order to do so, a dataset of correlation coefficients was created. First, authors who were e-mailed to provide information for an earlier version of this project (Donnelly, Brooks, & Homer, 2015), were asked to provide correlations between RTs for congruent and incongruent trials within the same task. We then conducted a meta-analysis on these correlations, and estimated an average correlation coefficient of .72, which was used for the remaining studies. In three studies, no standard deviations were provided. Standard deviations for these studies were estimated in the following manner: A linear regression predicting the standard deviation from the mean RT was fit to the available means and standard deviations, and then used to predict the standard deviations for the three studies based on the published means.

Moderator coding

In this section we describe how the moderator variables were coded. The numbers presented below represent the dataset utilizing the broad definition of bilingual; see Table 1 for numbers of comparisons included in each of the four datasets used in the multiverse analysis.

Task Task was coded as a factor variable with four levels: Flanker, Simon, ANT, and Stroop-like. Task was further coded as standard or non-standard (see below). Flanker tasks included the traditional and alternating position Flanker task. Standard

versions of the Flanker task presented the distracters on the left and right sides of the target arrow and did not include additional EF manipulations within the same block ($k = 22$). However, in the expanded dataset, we included all versions of the Flanker task, including those with go/no-go trials within block and unconventional placements of distracter arrows ($k = 24$). The Simon task included the traditional Simon task, and a version of the Simon task used by Blumenfeld and Marian (2014), in which arrows rather than colored squares were used as the stimuli, but, unlike in the Simon Arrows task, participants had to remember an arbitrary mapping between upward and downward pointing arrows and left and right key presses. Standard versions of the task included equal numbers of congruent and incongruent trials and did not explicitly manipulate the number of switch trials ($k = 54$; if the number of congruent and incongruent trials were not explicitly mentioned, they were assumed to be equal). In the expanded set, we included versions of the task with switch manipulations and unbalanced numbers of congruent and incongruent trials ($k = 59$). The ANT and lateralized ANT were coded as the ANT. The lateralized ANT was coded as a non-standard version of the task. In general, ANT data were presented in one of two formats, with data for each cueing condition presented either separately or aggregated across the cueing conditions. When each cueing condition was presented separately, we used the no-cue condition to calculate effect sizes. When data were aggregated across cueing conditions, we used the aggregated data. We coded data from the no-cue condition as the standard version of the task ($k = 11$), but included both sources of data in our expanded dataset ($k = 29$) because the aggregated scores were also influenced by the dynamics of attentional cueing. We coded both the Numerical Stroop and Simon Arrows (or Spatial Stroop) as Stroop-like because there were relatively low numbers, and since, unlike the Simon task, the Simon Arrows task does not have arbitrary stimulus-response mappings and, therefore, may be more Stroop-like. We treated versions with equal numbers of congruent and incongruent trials as standard versions ($k = 12$). In our expanded data set, we additionally included trials with unbalanced numbers of congruent and incongruent trials, and one study that randomly interspersed mind-wandering probes between trials ($k = 20$).

Age Age was coded categorically with the prediction that larger effect sizes would be observed in children and older adults than in adults. Hence, we coded participants younger than 13 years as children, participants aged 18–40 years as younger adults, and participants over 60 years as older adults. Thirteen comparisons were excluded from the analysis of age: in 12 cases the participants' ages were outside these three categories and in one case RTs were aggregated across multiple age groups in the original study (Blumenfeld, Schroeder, Bobb, Freeman, & Marian, 2016). Note that these 13 comparisons were included in tests of other moderators.

AoA of the L2 Sabourin, Brien, and Burkholder (2014) found empirical evidence of differences in lexical organization between participants who began learning their L2 before and after age 7 years; thus, for the purpose of defining early versus late bilinguals, we aimed to use this age as a cut-off. However, because the studies varied widely in what information was reported, there was no single rule that could be applied to every single study. For our multiverse analysis, we coded AoA of the L2 according to two schemes: lax and strict.

In applying our lax scheme, we considered any relevant information that was reported and applied a set of rules in a fixed order to determine AoA. First, if mean participants' self-reported age of L2 acquisition, immersion, or immigration or the mean number of years of L2 exposure was reported, this number was used ($k = 36$). Next, if the maximum age of acquisition was reported, this was used ($k = 16$). Next, if the authors stated participants used languages in school, participants were treated as early AoA ($k = 9$). Next, if participants were described as early or late bilinguals, they were coded as such ($k = 12$). Finally, if participants were 6 years of age or younger, they were coded as early AoA ($k = 12$; i.e., all studies with participants 6 years of age or younger). The lax coding scheme resulted in 60 early and 25 late AoA comparisons. We used this definition of AoA in our preferred analysis.

According to our strict scheme, we calculated AoA if one of four conditions were met. First, if the mean participants' self-reported age of L2 acquisition, immersion, or immigration or the mean number of years of L2 exposure was reported, this number was used ($k = 33$). Second, if the paper reported the latest AoA of their bilingual participants, and this age was below 7 years, this number was used ($k = 8$). Third, if the paper reported that participants learned both languages in school, and reported the age at which both languages were used in the curriculum, we used this age ($k = 6$); note this could overestimate AoA, but did not lead to any late AoA categorizations. Fourth, if participants were 6 years of age or younger, they were coded as early AoA ($k = 12$). The strict coding scheme resulted in 44 early and 15 late AoA comparisons.

Analytic strategy

In both our preferred and multiverse analyses, we first fit models to estimate average effect sizes for global RTs and interference costs. We then tested each moderator separately. To do so, we included both its main effect and its interaction with DV, and conducted an omnibus test to determine whether they jointly improved fit.

As global RT and interference cost are DVs taken from the same studies, they are likely to be correlated at both the level of the true effect and the sampling error. The ideal way to model these effect sizes would be to conduct a multivariate meta-analysis. However, these models require that the analyst

knows or assumes a value of the correlation between the two variables within a study. As these are generally not reported, we tried four separate multivariate meta-analyses. First, based on available unpublished data, we calculated the correlation between the global RTs and interference costs in Flanker, Simon, and Stroop tasks and got an average correlation coefficient of .14. In our preferred analysis, we stipulated this as the correlation between the two DVs. Second, in our multiverse analysis we considered the unlikely possibility that the DVs were uncorrelated, and conducted a multivariate meta-analysis with a within-comparison correlation of 0. Third, we considered the possibility that the true within-comparison correlation was double the estimated value and assumed a within-comparison correlation of .30. Fourth, a helpful reviewer pointed out that the cluster robust variance estimation could be used to estimate the correlation. In all of these models, in addition to estimating random effects by comparison, we also estimated random effects by publication. Another possible way of analyzing these data is to conduct a multilevel univariate meta-analysis with cluster-robust variance estimation within comparisons (Hedges, Tipton, & Johnson, 2010). We considered this fifth model in our multiverse analysis as well. All analyses were conducted using the *metafor* package in R (Viechtbauer, 2010); the code is available in the Online Supplemental Materials.

Results

We report the results in two parts. First, we report in detail the results of our preferred analysis, which utilized a broad definition of bilingual but included only standard versions of interference-control tasks. Second, we present a multiverse analysis, examining how inferences about the overall effect sizes and moderators are impacted by different plausible decisions for the construction and analysis of the dataset. A preliminary model was a multivariate meta-analysis, allowing effect sizes to vary by comparison and by paper. However, this yielded a highly skewed distribution of internally standardized residuals with several large positive values (seven standardized residuals between 2.63 and 7.98). Ideally we would have calculated externally standardized residuals, but at this point *metafor* only calculates internally standardized residuals for multivariate models. As a result, these residuals are likely underestimates. All five of these large standardized residuals came from a landmark study that reported extremely large effect sizes (Bialystok et al., 2004). The question of what to do with outlying effect sizes is difficult. On the one hand, these effect sizes likely violate the assumption of normally distributed random effects and exert an outsized influence on the overall effect size estimates. On the other hand, this is a landmark study in the field and does not contain obvious methodological shortcomings that would justify excluding

them entirely. We therefore decided to exclude this study as an “outlier” in our preferred analysis, but include a dimension in our multiverse analysis to examine how the results are influenced by the inclusion of this study.

Part 1: Preferred analysis

To examine the overall effect sizes, a multivariate random effects meta-analysis with unstructured random effects by comparison and paper was fit to the subset of the data containing the studies with broadly defined bilinguals, standard versions of tasks, and without the study mentioned above.¹ For analyses involving age of L2 acquisition, the preferred analysis used the lax coding of AoA as this included more comparisons, especially of bilinguals with late AoA. The preferred model revealed a very small, but significant, effect of bilingualism on global RT ($g = .13$, $Z = 2.51$, $p = .012$, $CI = .03; .23$), and a small, but significant, effect of bilingualism on interference cost ($g = .11$, $Z = 2.71$, $p = .007$, $CI = .03; .18$). The model also revealed a significant amount of residual heterogeneity, $Q(179) = 382.00$, $p < .001$. To determine the extent of publication bias on these effect sizes, Egger’s regression tests, predicting effect sizes from standard errors, were conducted on global RT and interference cost separately. Note that this approach does not account for the dependence between sampling errors and true effect sizes in these studies and, therefore, it should only be taken as a heuristic. There was significant funnel plot asymmetry for global RT ($Z = 3.23$, $p = .001$), but not interference cost ($Z = .51$, $p = .607$). In order to assess the influence of publication bias, we followed Lehtonen et al. (2018) and used the PEESE method to correct the effect sizes. This method entails predicting effect sizes from their variances in a weighted least squares regression (weighted by standard errors). The intercept in these models can be conceived of as the zero variance effect size. Because the test above does not account for the dependence between effect sizes, we applied the PEESE correction to both global RT and interference cost, though we were more concerned about the funnel plot asymmetry for global RT. The PEESE method revealed a corrected global RT effect size that was negative and did not significantly differ from 0 ($g = -.17$, $t(86) = -1.86$, $p = .067$, $CI = -.34; .01$). Consistent with the results of Egger’s regression test, there was also a significant

positive relationship between each effect size and its variance ($b = 3.59$, $t(86) = 3.26$, $p = .002$, $CI = 1.40; 5.78$). For interference cost, the effect size was of similar magnitude to the previous analysis ($g = .08$, $t(91) = 1.17$, $p = .25$, $CI = -.05; .21$) but no longer significant. The relationship between each effect size and its variance was also not significant ($b = .32$, $t(91) = .39$, $p = .699$, $CI = -1.33; 1.64$), which is consistent with the results of Egger’s regression test above. Importantly, neither Egger’s regression test nor the PEESE correction accounts for the dependence between observations and should therefore be considered very rough approximations of the unbiased effect sizes. Because of the evidence of publication bias for global RT, we report PEESE corrected effect sizes for all subsequent analyses, as well as slopes of the relationship between variance and effect size in order to determine whether publication bias is likely in a given analysis.

To examine the effects of the three moderator variables (age group, task, and AoA), three models were fit testing for the main effects of each moderator as well as their interaction with DV (global RT or interference cost). The effects of these coefficients were then examined using omnibus tests. Including age group and its interaction with the DV did not significantly improve model fit ($Q(4) = 1.55$, $p = .818$), nor did including task and its interaction with the DV ($Q(6) = 5.49$, $p = .483$). However, including lax-coded AoA as a dummy variable with 0 for early AoA and 1 for late AoA and its interaction with the DV did significantly improve fit ($Q(2) = 6.52$, $p = .038$). This effect appears to be due to a significant interaction between AoA and DV ($b = .31$, $Z = 2.52$, $p = .012$, $CI = .07; .55$). In order to understand this interaction, we conducted a separate multilevel univariate meta-analysis on global RT and interference cost with AoA as a moderator. For global RT, there was a very small, significant effect size for early AoA comparisons ($g = .15$, $Z = 2.48$, $p = .013$, $CI = .03; .26$) and AoA did not significantly moderate effect sizes ($b = -.05$, $Z = -.06$, $p = .952$, $CI = -.16; .16$). For interference cost, there was a very small, non-significant effect size for early AoA comparisons ($g = .09$, $Z = 1.85$, $p = .064$, $CI = -.01; .18$). Effect sizes were significantly larger in AoA comparisons with late as opposed to early bilinguals ($b = .18$, $Z = 2.64$, $p = .008$, $CI = .05; .32$). In order to determine whether effect sizes from the early late AoA comparisons were significantly different than 0, we conducted a multivariate meta-analysis for early and late AoA comparisons separately. Amongst, early AoA comparisons, there was a very small but significant effect size for global RT ($g = .14$, $Z = 2.89$, $p = .022$, $CI = .02; .26$) and a non-significant effect size for interference cost ($g = .07$, $Z = 1.47$, $p = .142$, $CI = -.02; .15$). When the PEESE-method was applied to global RTs, the resulting effect size was non-significant ($g = -.15$, $t(49) = -1.3$, $p = .132$, $CI =$

¹ In general, these models in our “preferred analysis” yielded unique variance estimates for global RT and interference cost by both comparison and paper. However, in some cases the correlation between random effects by comparison was 1, while the correlation between random effects by paper was a plausible value. We suspect this implausible value suggests problems with identification. According to a likelihood ratio test, models with random effects by paper and comparison fit significantly better than models with random effects by comparison alone ($p = .02$). Moreover, results from these analyses are not markedly different from results from the multilevel univariate model with cluster robust variance estimation, which has a considerably different random effects structure. We, therefore, feel that our inferences are robust.

-.36; .07) and the relationship between effect size variances and effect sizes magnitudes was significant ($b = 3.68$, $t(49) = 2.69$, $p = .010$, $CI = .94; 6.42$). Amongst late AoA comparisons there was a non-significant effect for global RT ($g = -.02$, $Z = -.27$, $p = .789$, $CI = -.16; .13$), and a small but significant effect size for interference costs ($g = .25$, $Z = 3.29$, $p = .001$, $CI = .10; .40$). When the PEESE-method was applied to the effect size for interference cost for late AoA comparisons, the resulting effect size was no longer significant ($g = .14$, $t(23) = .81$, $p = .43$, $CI = -.22; .50$). However, there was no evidence of publication bias ($b = .81$, $t(23) = .40$, $p = .694$, $CI = -3.39; 5.01$), so this non-significant corrected effect size may reflect the relatively small number of effect sizes included in this analysis ($k = 25$).

Part 2: Multiverse analysis

To examine the sensitivity of conclusions to plausible analysis and design decisions, we conducted sets of analyses influenced by Steegen, Tuerlinckx, Gelman, & Vanpaemel (2016) multiverse analysis approach. More specifically, we addressed each research question between 40 and 80 times, making different decisions about the construction of the dataset and analysis. We considered four datasets with varying inclusion criteria, defined by broad versus narrow definitions of bilingual and exclusion/inclusion of non-standard versions of interference-control tasks (see Table 1). We considered five statistical models: four multivariate meta-analyses stipulating different values of the correlations and one univariate multilevel meta-analysis with cluster robust variance estimation. Model 1 used an estimate of .14 for the correlation between the sampling errors for global RT and interference cost; Model 2 used an estimate of 0 for the correlation; Model 3 used an estimate of .30 for the correlation; Model 4 used cluster robust variance estimation; and Model 5 was the multilevel meta-analysis with robust variance estimation. We compared these methods with and without the inclusion of the Bialystok et al. (2004) study that was identified as an outlier in our preferred analysis. Moreover, for analysis of the moderator AoA of the L2, we considered both strict and lax coding of AoA.

Tables 3 and 4 show the effect size estimates for global RT and interference cost, respectively, for each of the analyses in the multiverse. As shown in Table 3, global RT effect sizes ranged between .07 and .17 when the outlier effect sizes were excluded and between .13 and .20 when the outliers were included. In general, the effect sizes were also a bit smaller in the univariate multilevel meta-analyses than in the multivariate meta-analyses. As shown in Table 4, the interference cost effect sizes ranged between .10 and .12 when outliers were excluded, and

between .16 and .20 when outliers were included.² Tables 5, 6, and 7 show the p -values for analyses examining the effects of the three moderators (age group, task, AoA). All models include the main effects for the moderators and their interactions with the DV. As shown in Table 5, age did not significantly moderate effect sizes in any cells of the multiverse. As shown in Table 6, task also did not significantly moderate effect sizes in any cells of in the multiverse. However, as shown in Table 7, when using lax coding of AoA, AoA tended to significantly moderate effect sizes when outliers were excluded; when using strict coding of AoA, AoA tended to significantly moderate effect sizes only under a broad definition of bilingual with the expanded task set.

Two results from the multiverse analysis warrant follow-up. First, when outliers were included, effect sizes increased. In order to determine whether these larger effect sizes were affected by publication bias, we first applied the PEESE correction to the global RT and interference cost with outliers included. Because these outliers were large effect sizes with large standard errors, they strongly influenced the regression line, and resulted in publication bias corrected effect sizes that seemed implausible because the coefficients were negative (indicative of better performance for monolinguals). The corrected effect sizes were significant for both global RT ($g = -.35$, $t(91) = -4.52$, $p < .001$, $CI = -.50; -.19$; $b = 6.49$, $t(91) = 8.73$, $p < .001$, $CI = 5.01; 7.97$) and interference cost ($g = -.24$, $t(96) = -3.53$, $p = .001$, $CI = -.36; -.09$; $b = 5.13$, $t(96) = 7.45$, $p < .001$, $CI = 3.76; 6.49$).

Second, as AoA interacted with the DV in the preferred analysis (with AoA coded using a lax scheme), it was necessary to see if the effect was similar when AoA was coded using a strict scheme. We therefore re-ran the models with strict coding of AoA. The models used a broad definition of “bilingual,” and both standard and non-nonstandard tasks without outliers, and assumed a correlation of .14 between the sampling errors for global RT and interference cost. As in the previous analysis with lax coding, including AoA and its interaction with DV significantly improved model fit ($Q(2) = 6.42$, $p = .040$) with a significant interaction between AoA and DV ($b = .23$, $Z = 2.09$, $p = .037$, $CI = .01; .45$). Again, we ran separate analyses on global RT and interference cost. For global RTs, there was a non-significant effect size for early AoA comparisons ($g = .11$, $Z = 1.72$, $p = .086$, $CI = -.02; .23$) and a non-significant effect of AoA ($b = .05$, $Z = .50$, $p = .61$, $CI =$

² Note that in Table 4, there are no effect sizes for Model 5. This is because the univariate model is parameterized differently than the multivariate models. The multivariate models produce effect sizes for global RT and interference cost. However, the univariate model produces an effect size for global RT and a difference score between interference cost and global RT. There is, therefore, no comparable coefficient in the univariate model to the interference cost in the multivariate models. Across all cells in the multiverse analysis, the confidence interval for this difference score overlapped with 0.

Table 3 Average global reaction times (RTs) for different possible analyses

Outliers	Standard Task Set		Expanded Task Set	
	Excluded	Included	Excluded	Included
Bilingual Broad Definition				
Model 1	.13 (.03; .23)	.20 (.02; .38)	.16 (.07; .26)	.19 (.05; .34)
Model 2	.13 (.03; .24)	.20 (.02; .38)	.16 (.07; .25)	.19 (.05; .34)
Model 3	.13 (.02; .23)	.20 (.01; .38)	.17 (.07; .26)	.19 (.05; .34)
Model 4	.13 (.04; .23)	.20 (.07; .33)	.16 (.08; .24)	.19 (.09; .30)
Model 5	.09 (.01; .17)	.13 (.04; .22)	.10 (.03; .17)	.14 (.06; .22)
Bilingual Narrow Definition				
Model 1	.13 (.00; .25)	.18 (-.05; .40)	.16 (.05; .27)	.19 (.01; .38)
Model 2	.13 (.01; .25)	.18 (-.05; .40)	.16 (.05; .27)	.19 (.01; .38)
Model 3	.12 (-.00; .25)	.18 (-.05; .41)	.16 (.04; .27)	.19 (.01; .38)
Model 4	.13 (.02; .24)	.18 (.02; .34)	.16 (.06; .26)	.19 (.06; .32)
Model 5	.07 (-.02; .16)	.13 (.02; .23)	.07 (-.00; .15)	.13 (.03; .22)

Note. In all tables from multiverse analysis, Model 1 is the multivariate random effects meta-analysis with an assumed correlation of .14. Model 2 is the multivariate random effects meta-analysis with an assumed correlation of 0. Model 3 is the multivariate random effects meta-analysis with an assumed correlation of .30. Model 4 is the multivariate random effects meta-analysis with cluster-robust variance estimation. Model 5 is multi-level univariate meta-analysis with robust variance estimation

Table 4 Average interference cost for different possible analyses

Outliers	Standard Task Set		Expanded Task Set	
	Excluded	Included	Excluded	Included
Bilingual Narrow Definition				
Model 1	.11 (.03; .18)	.16 (.03; .28)	.10 (.04; .16)	.16 (.05; .28)
Model 2	.11 (.03; .18)	.16 (.04; .28)	.10 (.04; .16)	.16 (.05; .28)
Model 3	.11 (.03; .19)	.15 (.03; .28)	.10 (.04; .16)	.16 (.05; .28)
Model 4	.11 (.04; .17)	.16 (.07; .24)	.10 (.05; .16)	.16 (.09; .24)
Bilingual Strict Definition				
Model 1	.12 (.04; .20)	.20 (.03; .37)	.12 (.05; .19)	.19 (.05; .33)
Model 2	.12 (.04; .20)	.20 (.03; .37)	.12 (.06; .19)	.19 (.05; .33)
Model 3	.12 (.03; .21)	.20 (.03; .38)	.12 (.05; .20)	.19 (.05; .34)
Model 4	.12 (.04; .19)	.20 (.10; .31)	.12 (.06; .18)	.19 (.09; .28)

-.14; .23). For interference costs, there was a non-significant effect size for early AoA comparisons ($g = .04, Z = 1.14, p = .252, CI = -.03; .12$) and a significant effect of AoA ($b = .22, Z = 3.35, p < .001, CI = .09; .35$). In order to see if late AoA effect sizes differed from 0, a multivariate meta-analysis on the late AoA comparisons yielded a non-significant effect for global RT ($g = .08, Z = .80, p = .43, CI = -.12; .28$) and a small-to-medium effect for interference cost ($g = .23, Z = 2.72, p = .007, CI = .06; .40$). When the PEESE-method was applied to effect sizes for interference cost for the late AoA comparisons, the corrected effect size was smaller and non-significant ($g = .12, t(20) = .56, p = .583, CI = -.32; .55$). However, since there was no evidence of publication bias ($b = 1.57, t(20) = .65, p = .229, CI = -3.50; 6.6$), and this analysis was based on very few effect sizes ($k = 22$), we are cautious about interpreting the corrected effect sizes.

Discussion

The purpose of this meta-analytic study was to determine under which circumstances bilinguals outperform monolinguals on interference-control tasks. In particular, we sought to estimate the average effect sizes for global RT and interference cost, two DVs associated with interference-control tasks, and determine whether effect sizes were moderated by age, task, and age of L2 acquisition. After conducting our preferred analysis, which utilized a broad definition of bilingual and only standard versions of interference-control tasks, we conducted a multiverse analysis to determine how sensitive our conclusions were to plausible constructions of the dataset and analyses. We examined how the results were affected by: (1) strictness of the criteria used to define bilinguals, (2) whether both standard and non-standard versions of interference-control task were included, (3) how strictly age of L2 acquisition was coded, (4) how the dependence between effect sizes was modeled, and (5) inclusion of effect sizes from two studies that were identified as outliers (Bialystok et al., 2004; Martin-Rhee & Bialystok, 2008). Results were largely consistent across all of the dataset construction and design possibilities.

Preferred analysis

Our preferred meta-analysis revealed very small but statistically significant effect sizes for both global RT and interference cost ($g = .13$ and $g = .11$, respectively). However, once corrected for publication bias, only the effect size for interference cost remained significant. The overall effect sizes observed were of similar magnitude to the monitoring and inhibition effect sizes reported by Lehtonen et al. (2018) in a recent meta-analysis of evidence for the bilingual advantage. Lehtonen et al.'s (2018) meta-analysis differed from the

Table 5 P-values for age across different possible analyses

Outliers	Standard Task Set		Expanded Task Set	
	Excluded	Included	Excluded	Included
Bilingual Broad Definition				
Model 1	.82	.65	.68	.40
Model 2	.83	.66	.69	.41
Model 3	.78	.62	.67	.39
Model 4	.87	.76	.51	.39
Model 5	.95	.79	.78	.54
Bilingual Narrow Definition				
Model 1	.94	.85	.97	.67
Model 2	.94	.78	.94	.69
Model 3	.94	.91	.98	.65
Model 4	.91	.88	.94	.89
Model 5	.93	.84	.95	.86

present study in several important ways. First, their meta-analysis only considered studies involving adult bilinguals and encompassed a wider range of tasks than the current analysis. Second, they used a different approach to model the dependence between observations. They conducted univariate multilevel meta-analyses of each domain separately, and included four levels of random effects. Third, when descriptive statistics for interference cost were not available, they used either the difference between incongruent and neutral trials or average RT on incongruent trials, whereas we estimated the relevant standard deviations by assuming a correlation of .72 between congruent and incongruent trials.

Our estimates of the overall effect size for the bilingual advantage confirm the small magnitude of effects reported in a recent meta-analysis conducted by Paap et al. (2017a).

Table 6 P-values for task across different possible analyses

Outliers	Standard Task Set		Expanded Task Set	
	Excluded	Included	Excluded	Included
Bilingual Broad Definition				
Model 1	.48	.89	.26	.92
Model 2	.47	.85	.30	.90
Model 3	.47	.92	.23	.93
Model 4	.20	.40	.41	.34
Model 5	.86	.97	.45	.72
Bilingual Narrow Definition				
Model 1	.40	.92	.37	.86
Model 2	.38	.91	.40	.84
Model 3	.40	.94	.34	.87
Model 4	.28	.47	.54	.36
Model 5	.56	.74	.41	.71

When Paap et al. considered only large-sample studies, they found an average bilingual advantage of 9 ms for global RT and 6 ms for interference cost. Paap et al. (2017a, b) analysis is substantially different from our own: Their study analyzed raw RTs, rather than effect sizes, and analyzed the data by calculating means weighted by sample sizes rather than using a formal meta-analysis, which weights effect sizes by their variances.

Lehtonen et al. (2018) identified publication bias in their dataset and, once this was corrected for, effect sizes for monitoring and inhibition were indistinguishable from 0. Moreover, Paap et al. (2017a, b) found that effects of bilingualism were much larger for small-sample than large-sample studies for both global RTs and interference costs. In contrast, our analysis only revealed publication bias for global RT, but not interference cost. However, our methodology for testing publication bias did not account for the dependence between global RT and interference cost. When PEESE corrections were applied to each DV separately, the effects of bilingualism were not significant for either DV. Given this complex pattern of results, one might conclude either that there is not an overall bilingual advantage on interference-control tasks, or that there is a very small bilingual advantage on interference cost, but not global RT.

Effect sizes were not moderated by age in our preferred analysis or in any cells of the multiverse analysis. This result is also consistent with the results of Lehtonen et al. (2018), who did not observe differences between younger or older adults when age was coded categorically or continuously, nor did they observe effects of age when each EF construct was considered separately. This result is also consistent with the qualitative literature reviews conducted by Paap et al. (2015) and Hilchey, Saint-Aubin, and Klein (2015), who did not find evidence that age moderated effects of bilingualism. Such findings may be viewed as surprising, as it has been suggested that the influence of bilingualism on interference control should be more apparent during childhood and late adulthood than during young adulthood when EF processes are expected to be at their peak (Bialystok, Martin, & Viswanathan, 2005b; Bialystok, 2017). Given the variability in performance on the Flanker task across ages (Waszak et al., 2010), it was conceivable that the effect of bilingualism would vary across age groups. However, de Bruin and Della Sala (2017) have shown that the pattern of age-related decline is not identical across interference costs from different tasks. It is therefore possible that age effects would be evident on some tasks but masked by null effects on others. However, testing such a hypothesis would require a model with a 2 (DV) × 3 (Age Group) × 4 (Task) interaction for which our data set is too small.

We did not observe an effect of task in our preferred analysis or in any cells of our multiverse analysis. This result is consistent with the analysis of Lehtonen et al. (2018), who

Table 7 P-values for AoA across different possible analyses

	AoA Lax Coding				AoA Strict Coding			
	Standard Task Set		Expanded Task Set		Standard Task Set		Expanded Task Set	
Outliers	Excluded	Included	Excluded	Included	Excluded	Included	Excluded	Included
	Bilingual Broad Definition				Bilingual Broad Definition			
Model 1	.04	.04	.02	.03	.09	.16	.04	.08
Model 2	.04	.04	.02	.03	.11	.17	.05	.11
Model 3	.04	.04	.02	.03	.06	.15	.03	.05
Model 4	.05	.01	.02	.02	.11	.10	.06	.08
Model 5	.04	.03	.02	.02	.02	.01	.02	.02
	Bilingual Narrow Definition				Bilingual Narrow Definition			
Model 1	.03	.18	.07	.12	.15	.22	.39	.30
Model 2	.03	.17	.07	.12	.17	.21	.38	.30
Model 3	.04	.19	.07	.13	.13	.23	.40	.31
Model 4	.12	.06	.07	.05	.11	.06	.38	.11
Model 5	.14	.05	.08	.05	.04	.02	.06	.04

found that task did not significantly moderate effect sizes within either their inhibition or monitoring domains, which closely map onto interference cost and global RT, respectively. Bialystok's (2017) framework, which locates the bilingual advantage in general executive attention abilities rather than in specific executive functions, seems consistent with the lack of a task effect, but is difficult to reconcile with the very small and non-significant overall effect sizes.

AoA significantly moderated effect sizes in our preferred analysis, which appeared due to a significant interaction between AoA and the DV. We had expected early bilinguals to show larger advantages than late bilinguals on EF, and suggested that this AoA effect might interact with the DV. However, our results revealed a different pattern. While global RT was unaffected by AoA, the effect sizes for interference cost were larger for late than for early AoA comparisons. When late AoA comparisons were considered alone, there were non-significant effect sizes for global RT, but a small-to-medium effect size for interference cost. While the PEESE correction suggested this effect size was not different than 0, there was no evidence of publication bias, so the lack of significance may have reflected the relatively small number of observations in that analysis. Note that Lehton et al. (2018) also considered AoA as a moderator in their meta-analysis, but found that it did *not* moderate effect sizes.

While the direction of the effect of AoA on interference cost was unanticipated, it appears to be consistent with some existing theorizing: It has been argued that late learned second languages might incur more interference than early learned ones, thereby leading to larger bilingual advantages on measures of inhibition (Bak, Vega-Mendoza, & Sorace, 2014). However, it is difficult to disentangle the direction of causality in this case as it also plausible that individuals with better inhibition abilities are more

effective at learning a second language later in life. Furthermore, this account seems to run to counter to regression studies showing that when treated as a continuous variable, AoA is negatively associated with interference cost (Luk, De Sa, & Bialystok, 2011), or unrelated to interference cost (Paap & Sawi, 2014). However, given our results, and the plausible theoretical account, more research teasing out of the relationship between these variables seems warranted.

Multiverse analysis

The average effect sizes observed for global RT and interference cost were generally insensitive to variation in the modeling approach or dataset, although they were smaller using the multilevel model with robust variance estimation than when using the multivariate meta-analysis models. Moreover, our results are likely anti-conservative as we did not correct for multiple comparisons.

In the preferred analysis we excluded one study due to exceptionally large effect sizes that appeared to be outliers (i.e., g 's of up to 6.79). Including the outlying effect sizes strongly influenced the average effect, in some cases by large magnitudes. This study clearly had a very large influence on the average effect size while also creating a skewed pattern of residuals that likely violates the assumptions of normally distributed random effects at the levels of paper and comparisons. Although it is a landmark study without obvious methodological concerns, we are cautious about interpreting the models that included the outliers for two reasons. First, the skewed residuals likely reflect a violation of the assumptions of normally distributed random effects and sampling errors. Second, including this study in the PEESE analyses greatly influences inferences about publication bias, suggesting the true

publication-bias free estimates should be significantly lower than 0 (i.e., favoring monolinguals), which is implausible. We therefore think the effect size estimates from the models without outliers are more accurate.

Overall the results of this meta-analysis provide only very weak evidence for a bilingual advantage in interference-control tasks overall and little evidence for an advantage on global RT. For interference cost, there may be a small advantage on interference cost for late bilinguals. While future research is necessary to replicate this finding, as it is inconsistent with the null effect of AoA in Lehtonen et al.'s (2018) meta-analysis, it may suggest that later second language learners may enjoy increases in inhibitory control (Bak et al., 2014).

Limitations and methodological issues

Despite our findings casting doubt on a robust bilingual advantage on interference-control tasks, it remains possible that bilingual advantages are present, but restricted to certain bilinguals or that existing methodological and statistical practices are not sufficient to detect these sorts of effects. We consider these possibilities in turn.

One important factor to consider is that the term “bilingual” has been used in the literature quite broadly, often with little consideration of differences in participants’ language proficiency or distinct linguistic experiences. (See Surrain and Luk (2017) for a systematic review of how the term “bilingual” has been used in the literature.) It is possible that specific linguistic experiences are necessary for a bilingual advantage to emerge. For example, a potentially important moderator that was not included in the present study is the frequency of switching between languages. If bilingual advantages emerge because of frequent code switching, rather than as a consequence of lexical representation, then it is possible that bilinguals who have the experience of switching between languages more frequently would exhibit smaller interference cost than those who switch less frequently, or monolinguals. Costa et al. (2009) speculated that in socio-linguistic contexts characterized by a high degree of code switching, bilinguals must monitor for cues that determine when to switch languages; such cues may include the linguistic background of interlocutors or the broader linguistic context of the conversation. This suggests that EF advantages might only be evident for bilinguals from socio-linguistic contexts in which code switching is the norm. Unfortunately, while some studies have found a relationship between frequency of language switching and cognitive tasks (Prior & Gollan, 2011; Verreyt, Woumans, Vandelandotte, Szmalec, & Dyuck, 2016), other studies with larger samples have not (Johnson, Sawi, & Paap, 2015; Paap et al., 2017b). In a recent study, Hofweber, Marinis, and Treffers-Daller (2016) found that it was the

type of code switching engaged in by bilinguals rather than just frequency that predicted an inhibitory advantage in a high-monitoring condition of the Flanker task. Neither frequency nor type of language switching was considered in the current analysis because it could not be reliably determined from the majority of studies in the published literature; nonetheless, this may prove to be a fruitful avenue for future research.

Costa et al. (2009) suggest that superior coordination of related cognitive skills, such as monitoring and goal maintenance, may underlie any bilingual advantage; hence, to reliably detect differences may require complex tasks. For example, Morales, Yudes, Gómez-Ariza, and Bajo (2015) used a continuous performance task to evaluate both proactive and reactive control in bilingual and monolingual young adults. With the relatively more complex task, the authors found that bilinguals demonstrated better coordination of proactive and reactive control, resulting in fewer errors and significant ERP differences (i.e., N2 and P3a following the task prompt, and greater error-related negativity following incorrect responses). Such results suggest that more complex tasks and measures may be more sensitive to bilingual advantages in skills related to cognitive control.

Another possible reason why a bilingual advantage may exist, but has been difficult to consistently identify, is the existing literature’s reliance on measures of mean RTs and differences between RTs. It has long been acknowledged that task manipulations affect components of RT distributions beyond just the mean, particularly the tail of the distribution (Luce, 1986). Indeed, some researchers have begun using the exponentially modified Gaussian distribution, which allows for estimates of the mean, variance, and tail of an RT distribution (Calabria et al., 2011). This technique is a step forward in that it could be more sensitive to experimental effects than a simple mean would be. However, one weakness of this approach is that parameters of this distribution do not have distinct theoretical interpretations (Matzke & Wagenmakers, 2009).

A more precise analysis of RT data comes from cognitive process models, such as the Ratcliff drift diffusion model of two-choice forced decision tasks (Ratcliff & McKoon, 2008). The Ratcliff diffusion model models RT and accuracy distributions for two-choice forced decision tasks using four parameters with distinct theoretical interpretations (e.g., the rate of evidence accumulation and an initial response bias). This model provides excellent fit to RT distributions and has the advantage of interpretable parameters (Ratcliff & McKoon, 2008). However, these parameters cannot be identified using the normal or exponentially modified Gaussian (ex-Gaussian) distribution. Matzke and Wagenmakers (2009)

simulated data from the diffusion model and fit the ex-Gaussian and shifted Wald distributions to the datasets. They found that diffusion model parameters did not distinctly map onto parameters of either distribution; in other words, parameters of the ex-Gaussian and shifted Wald distributions correspond to multiple cognitive processes.

A generalization of the diffusion model has been developed for interference-control tasks (White, Ratcliff & Starns, 2011). This shrinking-spotlight diffusion model contains additional parameters corresponding to the shrinking of the attentional window within a trial and provides excellent fit to the Simon, Flanker, and Stroop data (Servant, Montagnini, & Burle, 2014). However, like the simpler diffusion model, its components combine in unintuitive ways, meaning that means, mean differences, and skewnesses from RT distributions will reflect a combination of cognitive processes. Ong, Sewell, Weekes, McKague, and Abutalebi (2017) found that bilinguals exhibited smaller non-decision times than monolinguals, which they interpreted as evidence that bilinguals more efficiently filter out distracting information. An advantage in the non-decision component may not consistently manifest in mean RTs, in which case, analysis of RTs might obscure a true underlying effect.

Conclusion

The purpose of this study was to determine the magnitude of a putative bilingual advantage on interference-control tasks and determine whether effects were moderated by age, task, and age of L2 acquisition. Counter to predictions, we failed to find evidence for a significant bilingual advantage for global RT after controlling for publication bias. For interference cost, although there was little evidence of publication bias, the overall effect size was very small effect after removing a landmark study identified as an outlier. Effect sizes were not significantly moderated by age or task, but were significantly moderated by an interaction between age of L2 acquisition and the DV. This interaction was driven by an effect of AoA on interference costs, with late bilinguals exhibiting larger advantages than early bilinguals. The robustness of our findings in the multiverse analysis supports the conclusions of other recent meta-analyses (Lehtonen et al. 2018; Paap et al., 2017a, b), and suggests cognitive benefits of bilingualism for standard interference-control tasks have been overstated.

Acknowledgements We would like to thank all the authors who contributed data for this project, Ellen Bialystok, Emily Coderre, Albert Costa, Mireia Hernandez, Leah Kappa, Vera Kempe, Ana Marzecova, Paula Rubio Fernandez, and Zofia Wodniecka. We would also like to thank Jay Verkuilen for statistical advice. All data and scripts for the present study are available at: http://www.osf.io/5kmbj/view_only=bf8c91f02a544e383e2df3f45ded265.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

(Asterisks indicate studies included in the meta-analysis)

- *Abutalebi, J., Della Rosa, P., Green, D. W., Hernández, M., Scifo, P., Keim, R., ... Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, 22(9), 2076–2086. <https://doi.org/10.1093/cercor/bhr287>
- *Abutalebi, J., Guidi, L., Borsa, V., Canini, M., Rosa, P. A. Della, Parrisi, B. A., & Weekes, B. S. (2015). Bilingualism provides a neural reserve for aging populations. *Neuropsychologia*, 69, 201–210. <https://doi.org/10.1016/j.neuropsychologia.2015.01.040>
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, 80(2), 207–245. <https://doi.org/10.3102/0034654310368803>
- *Ansaldò, A. I., Ghazi-Saidi, L., & Adrover-Roig, D. (2015). Interference control in elderly bilinguals: Appearances can be misleading. *Journal of Clinical and Experimental Neuropsychology*, 37(5), 455–470. <https://doi.org/10.1080/13803395.2014.990359>
- *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. <https://doi.org/10.3389/fpsyg.2014.00398>
- *Antón, E., Fernández García, Y., Carreiras, M., & Duñabeitia, J. A. (2016). Does bilingualism shape inhibitory control in the elderly? *Journal of Memory and Language*, 90, 147–160. <https://doi.org/10.1016/j.jml.2016.04.007>
- *Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2016). The effect of childhood bilingualism and multilingualism on executive control. *Cognition*, 149, 18–30. <https://doi.org/10.1016/j.cognition.2015.12.002>
- *Arredondo, M. M., Hu, X., Satterfield, T., & Kovelman, I. (2017). Bilingualism alters children's frontal lobe functioning for attentional control. *Developmental Science*, 20(3). <https://doi.org/10.1111/desc.12377>
- Bak, T. H., Vega-Mendoza, M., & Sorace, A. (2014). Never too late? An advantage on tests of auditory attention extends to late bilinguals. *Frontiers in Psychology*, 5, 485. <https://doi.org/10.3389/fpsyg.2014.00485>
- *Barac, R., Moreno, S., & Bialystok, E. (2016). Behavioral and electrophysiological differences in executive control between monolingual and bilingual children. *Child Development*, 87(4), 1277–1290. <https://doi.org/10.1111/cdev.12538>
- Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, 144(1), 77–110. <https://doi.org/10.1037/bul0000130>
- *Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal Of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, 60(1), 68–79. <https://doi.org/10.1037/cjep2006008>
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143(3), 233–262. <https://doi.org/10.1037/bul0000099>
- *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), 859–873. <https://doi.org/10.1037/0278-7393.34.4.859>

- Bialystok, E., Craik, F. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, 45(2), 459–464. <https://doi.org/10.1016/j.neuropsychologia.2006.10.009>
- *Bialystok, E., Craik, F. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005a). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *Neuroimage*, 24(1), 40–49. <https://doi.org/10.1016/j.neuroimage.2004.09.044>
- Bialystok, E., Craik, F. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, 10(3), 89–129. <https://doi.org/10.1177/1529100610387084>
- *Bialystok, E., Craik, F. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19(2), 290–303. <https://doi.org/10.1037/0882-7974.19.2.290>
- *Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 565–574. <https://doi.org/10.1037/a0012735>
- *Bialystok, E., Hawrylewicz, K., Wiseheart, M., & Toplak, M. (2017). Interaction of bilingualism and Attention-Deficit/Hyperactivity Disorder in young adults. *Bilingualism: Language and Cognition*, 20(3), 588–601. <https://doi.org/10.1017/S1366728915000887>
- Bialystok, E., Kroll, J. F., Green, D. W., MacWhinney, B., & Craik, F. I. (2015). Publication bias and the validity of evidence: What's the connection? *Psychological Science*, 26(6), 944–946. <https://doi.org/10.1177/0956797615573759>
- Bialystok, E., Martin, M. M., & Viswanathan, M. (2005b). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, 9(1), 103–119. <https://doi.org/10.1177/13670069050090010701>
- *Billig, J. D., & Scholl, A. P. (2011). The impact of bilingualism and aging on inhibitory control and working memory. *Organon*, 26(51), 39–52.
- Blom, E., Boerma, T., Bosma, E., Cornips, L. M. E. A., & Everaert, E. (2017). Cognitive advantages of bilingual children in different sociolinguistic contexts. *Frontiers in Psychology*, 8, 1–12. <https://doi.org/10.3389/fpsyg.2017.00552>
- *Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in stimulus-stimulus inhibition. *Bilingualism: Language and Cognition*, 17(3), 610–629. <https://doi.org/10.1017/S1366728913000564>
- Blumenfeld, H. K., Schroeder, S. R., Bobb, S. C., Freeman, M. R., & Marian, V. (2016). Auditory word recognition across the lifespan: Links between linguistic and nonlinguistic inhibitory control in bilinguals and monolinguals. *Linguistic approaches to Bilingualism*, 6(1), 119–146. <https://doi.org/10.1075/lab.14030.blu>
- *Bogulski, C. A., Rakoczy, M., Goodman, M., & Bialystok, E. (2015). Executive control in fluent and lapsed bilinguals. *Bilingualism: Language and Cognition*, 18(3), 561–567. <https://doi.org/10.1017/S1366728914000856>
- Calabria, M., Hernández, M., Martin, C., & Costa, A. (2011). When the tail counts: The advantage of bilingualism through ex-Gaussian distribution analysis. *Frontiers in Psychology*, 2, 250. <https://doi.org/10.3389/fpsyg.2011.00250>
- *Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, 130(3), 278–288. <https://doi.org/10.1016/j.cognition.2013.11.015>
- Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, 11(2), 282–298. <https://doi.org/10.1111/j.1467-7687.2008.00675.x>
- *Chabal, S., Schroeder, S. R., & Marian, V. (2015). Audio-visual object search is changed by bilingual experience. *Attention, Perception, & Psychophysics*, 77(8), 2684–2693. <https://doi.org/10.3758/s13414-015-0973-7>
- *Clare, L., Whitaker, C. J., Martyr, A., Martin-Forbes, P. A., Bastable, A. J. M., Pye, K. L., ... Hindle, J. V. (2016). Executive control in older Welsh monolinguals and bilinguals. *Journal of Cognitive Psychology*, 28(4), 412–426. <https://doi.org/10.1080/20445911.2016.1148041>
- Coderre, E. L., Van Heuven, W. J., & Conklin, K. (2013). The timing and magnitude of Stroop interference and facilitation in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 16(2), 420–441. <https://doi.org/10.1017/S1366728912000405>
- *Coderre, E. L., & van Heuven, W. J. (2014). The effect of script similarity on executive control in bilinguals. *Frontiers in Psychology*, 5, 1070. <https://doi.org/10.3389/fpsyg.2014.01070>
- *Coderre, E. L., Smith, J. F., Van Heuven, W. J., & Horwitz, B. (2016). The functional overlap of executive control and language processing in bilinguals. *Bilingualism*, 19(3), 471–488. <https://doi.org/10.1017/S1366728915000188>
- *Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, 113(2), 135–149. <https://doi.org/10.1016/j.cognition.2009.08.001>
- *Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, 106(1), 59–86. <https://doi.org/10.1016/j.cognition.2006.12.013>
- *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. doi:<https://doi.org/10.1016/j.neuropsychologia.2016.08.029>
- *de Bruin, A., Bak, T. H., & Sala, S. D. (2015a). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *Journal of Memory and Language*, 85, 15–26. <https://doi.org/10.1016/j.jml.2015.07.001>
- de Bruin, A., Treccani, B., & Della Sala, S. (2015b). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, 26(1), 99–107. <https://doi.org/10.1177/0956797614557866>
- de Bruin, A., & Della Sala, S. (2017). Effects of age on inhibitory control are affected by task-specific features. *The Quarterly Journal of Experimental Psychology*. <https://doi.org/10.1080/17470218.2017.1311352>
- *De Leeuw, E., & Bogulski, C. A. (2016). Frequent L2 language use enhances executive control in bilinguals. *Bilingualism: Language and Cognition*, 19(5), 907–913. <https://doi.org/10.1017/S1366728916000201>
- *Desideri, L., & Bonifacci, P. (2018). Verbal and nonverbal anticipatory mechanisms in bilinguals. *Journal of Psycholinguistic Research*, 47(3), 719–739. <https://doi.org/10.1007/s10936-017-9556-1>
- Donnelly, S., Brooks, P. J., & Homer, B. (2015). Examining the bilingual advantage on conflict resolution tasks: A meta-analysis. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.) Proceedings of the 37th Annual Meeting of the Cognitive Science Society (pp. 596–601). Austin: 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: Myth or reality? *Experimental Psychology*, 61(3), 234–251. <https://doi.org/10.1027/1618-3169/a000243>
- *Duncan, H. D., Segalowitz, N., & Phillips, N. A. (2016). Differences in L1 linguistic attention control between monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 19(1), 106–121. <https://doi.org/10.1017/S136672891400025X>
- *Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science*, 19(12), 1201–1206. <https://doi.org/10.1111/j.1467-9280.2008.02224.x>

- *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(11), 1364–1371. [10.1177/0956797612443836](https://doi.org/10.1177/0956797612443836)
- *Freeman, M. R., Blumenfeld, H. K., & Marian, V. (2017). Cross-linguistic phonotactic competition and cognitive control in bilinguals. *Journal of Cognitive Psychology*, 29(7), 783–794. <https://doi.org/10.1080/20445911.2017.1321553>
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133(1), 101–135. <https://doi.org/10.1037/0096-3445.133.1.101>
- *Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., ... 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. <https://doi.org/10.3389/fpsyg.2014.00011>
- *Grady, C. L., Luk, G., Craik, F. I. M., & Bialystok, E. (2015). Brain network activity in monolingual and bilingual older adults. *Neuropsychologia*, 66(1), 170–181. <https://doi.org/10.1016/j.neuropsychologia.2014.10.042>
- *Grant, A., & Dennis, N. (2017). Increased processing speed in young adult bilinguals: Evidence from source memory judgments. *Bilingualism: Language and Cognition*, 20(2), 327–336. <https://doi.org/10.1017/S1366728916000729>
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2), 67–81. <https://doi.org/10.1017/S1366728998000133>
- *Grundy, J., Chung-Fat-Yim, A., Friesen, D., Lorinda, M., Bialystok, E. (2017). Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults. *Cognition*, 163, 42–55. <https://doi.org/10.1016/j.cognition.2017.02.010>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- *Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition*, 13(3), 315–325. <https://doi.org/10.1017/S1366728909990010>
- Higgins, J. P., Deeks, J. J., & Altman, D. G. (2008) Special topics in statistics. In J. P. Higgins & S. Green (Eds.) *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series* (pp. 481–529). Chichester: Wiley. <https://doi.org/10.1002/9780470712184.ch16>
- 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 & Review*, 18(4), 625–658. <https://doi.org/10.3758/s13423-011-0116-7>
- Hilchey, M. D., Saint-Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance cognitive fitness in nonlinguistic executive processing tasks. In J. W. Schwieter (Ed.). *The Cambridge Handbook of Bilingual Processing* (pp. 586–613). Cambridge: Cambridge University Press.
- Hofweber, J., Marinis, T., & Treffers-Daller, J. (2016). Effects of dense code-switching on executive control. *Linguistic Approaches to Bilingualism*, 6(5), 648–668. <https://doi.org/10.1075/lab.15052.hof>
- *Jiao, L., Liu, C., Wang, R., & Chen, B. (2017). Working memory demand of a task modulates bilingual advantage in executive functions. *International Journal of Bilingualism*. <https://doi.org/10.1177/1367006917709097>
- Johnson, H., Sawi, O., & Paap, K. (2015). Language switching frequency in bilinguals is inconsistently linked to executive functioning. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society, San Francisco.
- *Kapa, L. L., & Colombo, J. (2013). Attentional control in early and later bilingual children. *Cognitive Development*, 28(3), 233–246. <https://doi.org/10.1016/j.cogdev.2013.01.011>
- *Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology*, 26(6), 640–648. <https://doi.org/10.1080/20445911.2014.929580>
- *Kousaie, S., Chai, X. J., Sander, K. M., & Klein, D. (2017). Simultaneous learning of two languages from birth positively impacts intrinsic functional connectivity and cognitive control. *Brain and Cognition*, 117, 49–56. <https://doi.org/10.1016/j.bandc.2017.06.003>
- *Kousaie, S., & Phillips, N. A. (2012). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research*, 1446, 71–90. <https://doi.org/10.1016/j.brainres.2012.01.052>
- *Kousaie, S., Sheppard, C., Lemieux, M., Monetta, L., & Taler, V. (2014). Executive function and bilingualism in young and older adults. *Frontiers in Behavioral Neuroscience*, 8, 250. <https://doi.org/10.3389/fnbeh.2014.00250>
- *Kramer, R., & Mota, M. B. (2015). Effects of bilingualism on inhibitory control and working memory: A study with early and late bilinguals. *Gragoatá*, 20(38), 309–331.
- *Ladas, A. I., Carroll, D. J., & Vivas, A. B. (2015). Attentional processes in low-socioeconomic status bilingual children: Are they modulated by the amount of bilingual experience? *Child Development*, 86(2), 557–578. <https://doi.org/10.1111/cdev.12332>
- 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*, 144(4), 394–425. <https://doi.org/10.1037/bul0000142>
- Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31(4), 581–612. <https://doi.org/10.1080/15326900701399905>
- Linck, J. A. (2015). Methods matter for critical reviews too. *Cortex*, 73, 354–355. <https://doi.org/10.1016/j.cortex.2015.07.011>
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- *Luk, G., Anderson, J. E., Craik, F. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain & Cognition*, 74(3), 347–357. <https://doi.org/10.1016/j.bandc.2010.09.004>
- *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(4), 488–495. <https://doi.org/10.1017/S1366728911000010>
- *Marian, V., Chabal, S., Bartolotti, J., Bradley, K., & Hernandez, A. E. (2014). Differential recruitment of executive control regions during phonological competition in monolinguals and bilinguals. *Brain and Language*, 139, 108–117. <https://doi.org/10.1016/j.bandl.2014.10.005>
- *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. <https://doi.org/10.1017/S1366728907003227>
- *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(3), 608–623. <https://doi.org/10.1017/S1366728912000569>
- Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, 16(5), 798–817. <https://doi.org/10.3758/PBR.16.5.798>

- 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. <https://doi.org/10.1177/0963721411429458>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogn.1999.0734>
- *Mohades, S. G., Struys, E., Van Schuerbeek, P., Baeken, C., Van De Craen, P., & Luypaert, R. (2014). Age of second language acquisition affects nonverbal conflict processing in children: An fMRI study. *Brain and Behavior*, 4(5), 626–642. <https://doi.org/10.1002/brb3.246>
- *Mor, B., Yitzhaki-Amsalem, S., & Prior, A. (2015). The joint effect of bilingualism and ADHD on executive functions. *Journal of Attention Disorders*, 19(6), 527–541. <https://doi.org/10.1177/1087054714527790>
- *Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114(2), 187–202. <https://doi.org/10.1016/j.jecp.2012.09.002>
- Morales, J., Yudes, C., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. *Neuropsychologia*, 66, 157–169. <https://doi.org/10.1016/j.neuropsychologia.2014.11.014>
- *Morton, J., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, 10(6), 719–726. <https://doi.org/10.1111/j.1467-7687.2007.00623.x>
- Nair, V. K., Biedermann, B., & Nickels, L. (2017). Effect of socio-economic status on cognitive control in non-literate bilingual speakers. *Bilingualism*, 20(5), 999–1009. <https://doi.org/10.1017/S1366728916000778>
- *Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism*, 13(5), 597–616. <https://doi.org/10.1080/13670050.2010.488288>
- 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(A), 28–38. <https://doi.org/10.1016/j.jneuroling.2016.08.002>
- Paap, K. R., Anders, R., Mikulinsky, R., Mason, L., & Alvarado, K. (2017a). More than 100 studies and 150 tests of the hypothesis that the differences between congruent and incongruent trials are smaller for bilinguals compared to monolinguals. Paper presented at the 58th Annual Meeting of the Psychonomic Society, Vancouver.
- *Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–258. <https://doi.org/10.1016/j.cogpsych.2012.12.002>
- Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265–278. <https://doi.org/10.1016/j.cortex.2015.04>
- Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017b). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, 29(2), 89–112.
- *Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems with convergent validity, discriminant validity and the identification of theoretical constructs. *Frontiers in Psychology*, 5, 962. <https://doi.org/10.3389/fpsyg.2014.00962>
- *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), 313–325. <https://doi.org/10.1037/a0035224>
- *Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a model for multitasking. *Developmental Review*, 35, 113–124. <https://doi.org/10.1016/j.dr.2014.12.003>
- *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(4), 535–551. <https://doi.org/10.1016/j.jecp.2012.06.013>
- Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A., & Alfieri, L. (2013). Effects of video-game play on information processing: A meta-analytic investigation. *Psychonomic Bulletin & Review*, 20(6), 1055–1079. <https://doi.org/10.3758/s13423-013-0418-z>
- 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(4), 682–691. <https://doi.org/10.1017/S1355617711000580>
- *Rainey, V. R., Davidson, D., & Li-Grinning, C. (2016). Executive functions as predictors of syntactic awareness in English monolingual and English-Spanish bilingual language brokers and nonbrokers. *Applied Psycholinguistics*, 37(4), 963–995. <https://doi.org/10.1017/S0142716415000326>
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922. <https://doi.org/10.1162/neco.2008.12-06-420>
- *Rodrigues, L. R., & Zimmer, M. C. (2016). Inhibitory and attentional control: The interaction between “professional activity” and bilingualism. *Psicologia: Reflexão e Crítica*, 29(1), 1–10. <https://doi.org/10.1186/s41155-016-0034-8>
- *Ross, J., & Melinger, A. (2017). Bilingual advantage, bidialectal advantage or neither? Comparing performance across three tests of executive function in middle childhood. *Developmental Science*, 20(4), e12405. <https://doi.org/10.1111/desc.12405>
- Rosselli, M., Ardila, A., Lalwani, A. N., & Vélez-Urbe, I. (2016). The effect of language proficiency on executive functions in balanced and unbalanced Spanish–English bilinguals. *Bilingualism: Language and Cognition*, 19(193), 489–503. <https://doi.org/10.1017/S1366728915000309>
- *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. <https://doi.org/10.1016/j.jml.2014.04.003>
- Sabourin, L., Brien, C., & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition*, 17(3), 542–555. <https://doi.org/10.1017/S1366728913000643>
- Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin*, 144(2), 111–139. <https://doi.org/10.1037/bul0000139>
- 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. <https://doi.org/10.1016/j.bandc.2010.02.003>
- *Salvatierra, J., & Rosselli, M. (2011). The effect of bilingualism and age on inhibitory control. *International Journal of Bilingualism*, 15(1), 26–37. <https://doi.org/10.1177/1367006910371021>
- *Schroeder, S. R., & Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology*, 24(5), 591–601. <https://doi.org/10.1080/20445911.2012.669367>
- *Schroeder, S. R., Marian, V., Shook, A., & Bartolotti, J. (2016). Bilingualism and musicianship enhance cognitive control. *Neural Plasticity*, 2016, Article ID 4058620. <https://doi.org/10.1155/2016/4058620>

- Servant, M., Montagnini, A., & Burle, B. (2014). Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws. *Cognitive Psychology*, *72*, 162–195. <https://doi.org/10.1016/j.cogpsych.2014.03.002>
- Shilling, V. M., Chetwynd, A., & Rabbitt, P. A. (2002). Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologia*, *40*(6), 605–619. [https://doi.org/10.1016/S0028-3932\(01\)00157-9](https://doi.org/10.1016/S0028-3932(01)00157-9)
- *Shulley, L. J., & Shake, M. C. (2016). Investigating the relationship between bilingualism, cognitive control, and mind wandering. *Journal of Cognitive Psychology*, *28*(3), 257–274. <https://doi.org/10.1080/20445911.2015.1128438>
- Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. <https://doi.org/10.1177/1745691616658637>
- Surrain, S. & Luk, G. (2017). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*. <https://doi.org/10.1017/S1366728917000682>.
- *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. <https://doi.org/10.3389/fpsyg.2011.00123>
- Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, *18*(1), 3–24. <https://doi.org/10.1017/S1366728914000522>
- Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism*, *19*(1), 181–190. <https://doi.org/10.1017/S1366728914000352>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. <http://www.jstatsoft.org/v36/i03/>
- *Vivas, A., Ladas, A., Salvani, V., & Chrysochoou, E. (2017). Revisiting the bilingual advantage in attention in low SES Greek-Albanians: Does the level of bilingual experience matter? *Language Cognition and Neuroscience*, *32*(6), 743–756. <https://doi.org/10.1080/23273798.2016.1271442>
- Waszak, F., Li, S., & Hommel, B. (2010). The development of attentional networks: Cross-sectional findings from a life span sample. *Developmental Psychology*, *46*(2), 337–349. <https://doi.org/10.1037/a0018541>
- White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the Flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, *63*(4), 210–238. <https://doi.org/10.1016/j.cogpsych.2011.08.001>
- *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), 1579–1586. <https://doi.org/10.1037/xlm0000107>
- *Xie, Z., & Dong, Y. (2015). Contributions of bilingualism and public speaking training to cognitive control differences among young adults. *Bilingualism: Language and Cognition*, *20*(1), 55–68. <https://doi.org/10.1017/S1366728915000474>
- *Yang, H., & Yang, S. (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. <https://doi.org/10.1016/j.jecp.2016.01.011>
- *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. <https://doi.org/10.1017/S1366728910000611>
- *Yudes, C., Macizo, P., & Bajo, T. (2011). The influence of expertise in simultaneous interpreting on non-verbal executive processes. *Frontiers in Psychology*, *2*, 309. <https://doi.org/10.3389/fpsyg.2011.00309>
- Zelazo, P. D., & Lee, W. S. C. (2010). Brain development. In R. M. Lerner & W. F. Overton (Eds.) *The handbook of life-span development*, Volume 1, Cognition, biology, and methods (pp. 89–114). Hoboken: Wiley.