The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning

Article *in* Journal of Cognitive Psychology · March 2014

DOI: 10.1080/20445911.2014.891597

Kenneth R. Paap
San Francisco State University

Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pecp21

The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning

Kenneth R. Paap[a]

[a] Department of Psychology, San Francisco State University, 1600 Holloway Avenue, EP301, San Francisco, CA 94132, USA

Published online: 27 Feb 2014.

Journal of Cognitive Psychology, 2014
http://dx.doi.org/10.1080/20445911.2014.891597

Routledge
Taylor & Francis Group

The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning

Kenneth R. Paap

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

Kroll and Bialystok assert that managing two languages leads to a reorganisation of the neural circuits involved in language and cognitive processing and to bilingual advantages in executive functioning. This commentary documents that bilingual advantages in inhibitory control, switching and monitoring are difficult to replicate. Kroll and Bialystok argue that the use of componential analyses and categorical hypothesising are often responsible for the null results and that these research practices have created a false controversy surrounding the existence of bilingual advantages. An alternative perspective is presented suggesting that the appearance of a steady stream of published reports has been exaggerated because of the frequent use of risky small *n*'s, a confirmation bias to report positive findings and a reluctance to conduct and report exact replications.

Keywords: Bilingual advantage; Confirmation bias; Executive function; Replication; Switching.

In the lead article to a recent special issue on bilingualism, Kroll and Bialystok (K&B; 2013) make these points about the relationship between bilingualism and cognition: (1) During comprehension and production, bilinguals automatically activate both languages and this requires a mechanism to select appropriately between the competing systems so that language processing can proceed fluently in the target language without interference from the other language, (2) That mechanism is most likely found in the executive control system that is largely based on a network of component processes in the frontal cortex, (3) In the course of recruiting domain-general brain mechanisms to regulate the use of two languages, the networks themselves undergo a broad reorganisation that will, in turn, also change non-verbal processing making it more resilient and efficient. Thus, bilingualism does not simply enhance one or more of the genetically determined components of executive processing (EP; Friedman et al., 2008), but rather qualitatively alters how the brain instantiates domain-general abilities in monitoring, shifting and inhibitory control, (4) This reorganisation leads K&B to assert that the differences between bilinguals and monolinguals cannot be explained by comparing the efficacy of corresponding components in bilingual and monolingual brains, (5) Consequently, trying to explain these differences on the basis of the known components (for monolinguals) "is a reductionist error" (p. 501). Several aspects of this argument appear to be underspecified or open to counterargument.

BRINGING BALANCE TO CLAIMS ABOUT BILINGUAL ADVANTAGES

K&B state that "studies of executive function demonstrate a bilingual advantage, with bilinguals

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

outperforming their monolinguals counterparts on tasks that require ignoring irrelevant information, task switching, and resolving conflict" (p. 497). These bilingual advantages are difficult to replicate. Hilchey and Klein (2011) reviewed 31 experiments and concluded that the evidence for a bilingual advantage in inhibitory control is rare in both children and young adults. More emphatically they assert that the collective evidence "is simply inconsistent with the proposal that bilingualism has a general positive effect on inhibitory control processes" (p. 629).

Since Hilchey and Klein's sobering review, there have been many additional tests for bilingual advantages in the inhibitory control component of executive functioning (EF) in non-verbal tasks and they overwhelmingly report no bilingual advantages. Kousaie and Phillips (2012a) found no behavioural differences between groups of young adults in the Stroop, Simon or flanker tasks. Kousaie and Phillips (2012b) used both young adults and older adults and found no differences in the magnitude of Stroop interference. A similar study by Humphrey and Valian (2012) using the Simon and flanker tasks follows the same pattern. Four different groups of multilinguals show Simon and flanker effects statistically equivalent to a group of English monolinguals. Paap and Greenberg (2013) found no bilingual advantage in three Simon experiments, one anti-saccade task and one flanker experiment. In another study using young adult participants, Ryskin and Brown-Schmidt (2012) report no bilingual advantages in either the Stroop or the flanker tasks.

Early reports of bilingual advantages in switching costs (Prior & Gollan, 2011; Prior & MacWhinney, 2010) are also difficult to replicate. Tare and Linck (2011) reported no group differences when propensity scores were used to exquisitely match the two groups. Using the same colour-shape switching as Prior, we obtained no language-group differences in four experiments (Paap & Greenberg, 2013; Sawi & Paap, 2013). Similarly, Hernández, Martin, Barceló, and Costa (2013) report no bilingual differences in switching costs in three experiments, one of which was also modelled on the Prior and MacWhinney task.

The studies testing for advantages in switching costs usually conduct a concomitant test of mixing costs (the difference between mean reaction time (RT) during single-task performance compared to mean RT on the repeat trials from the mixed block). This measure is assumed to reflect the cost of monitoring for the precue and preparing for a possible switch. The studies with young adults never produce a bilingual advantage in mixing costs. Many of these new studies also test for global RT advantages in non-verbal interference tasks (e.g., Simon, flanker and Stroop) and all of them report no differences between the language groups (Humphrey & Valian, 2012; Kousaie & Phillips, 2012a, 2012b; Paap & Greenberg, 2013; Sawi & Paap, 2013). Given that, both mixing costs in switching tasks and global RT advantages in interference tasks have been used as measures of monitoring the combined results on mixing costs, and global RTs provide no evidence for bilingual advantages in the monitoring component of EF.

IMPLICATIONS OF THE NEUROSCIENCE OF THE BILINGUAL BRAIN

In general, cortical areas shown to be involved in managing two languages overlap with those shown to be involved with inhibitory control and other components of EF (Bialystok, Craik, & Luk, 2012). Furthermore, it is clear from the neuroimaging results that the neural processing of bilinguals and monolinguals differs during the performance of the Simon and flanker tasks, in part, because some of the cortical areas recruited by bilinguals are not employed by monolinguals. All of this is consistent with the view that managing two languages leads to an organisation (or reorganisation) of neural networks in cortical areas involved in EF. However, a reorganisation to accommodate bilingualism does not logically need to result in more efficient performance. Alternatively, it could lead to comparable performance or even to a compromise that results in inferior performance.

Close examination of the two studies that reveal differences in neural processing between bilinguals and monolinguals during non-verbal interference tasks shows that the brain–behaviour relationships are inconsistent with an interpretation that the identified neural differences are causing bilingual advantages in conflict resolution (namely smaller interference effects). Bialystok et al. (2005) used magnetoencephalography (MEG) imaging to investigate the relationship between brain and behaviour when participants were engaged in a Simon task. There were two groups of bilinguals, Cantonese-English and French-English and, and one group of English monolinguals. The Simon task included experimental blocks where congruent and incongruent trials were randomly presented and control blocks where the target was presented at fixation and there was never any conflict between the physical location of the target and the response required by the task rule. With respect to RTs, there was a main effect of group (the Cantonese-English bilinguals were faster than the other two groups), but no Group × Trial Type interaction. With respect to the neuroimaging results, all groups recruited similar areas but there were group differences with respect to the specific areas associated with faster responding. Both bilingual groups showed substantial overlap in terms of the loci associated with fast responding (anterior cingulate, superior frontal and inferior frontal regions) and these differ from the specific areas associated with fast responding for the monolinguals (middle frontal region).

Bialystok et al. (2005) arrive at a cautious and fair conclusion: "The evidence supports the interpretation that bilinguals perform the Simon task differently from monolinguals, even when they respond at the same speed, and that the group differences are evident for both congruent and incongruent trials" (p. 48). Note that they do not claim that their neuroimaging results support a bilingual advantage in inhibitory control. They avoid an unjustifiable leap from differences between bilingual and monolinguals in neural processing to claims about differences in inhibitory control for the simple reason that no such behavioural differences in the magnitude of the interference effect were observed in their data. Furthermore, Bialystok et al. (2005) offer no explanation for the global RT differences that were observed: "We have no explanation for the faster reaction times of the Cantonese-English bilinguals, but due to the relatively small numbers of participants in each group this result could be due to sampling variability" (p. 46).

A more recent study by Luk, Anderson, Craik, Grady, and Bialystok (2010) uses neuroimaging data to bolster claims about bilingual advantages in conflict resolution during a flanker task. There were 10 English monolinguals and 10 English-other bilinguals. With respect to the RT data, there was neither a main effect of group nor a significant Group × Trial Type interaction. Note that the two tasks differ with respect to how bilinguals neurally process congruent and incongruent trials. To review, Bialystok et al. (2005) found that bilinguals and monolinguals differ with respect to the regions of fast responding in the Simon task, but these group differences were the same for both types of trials. In contrast, Luk et al. reported that the regions associated with faster responding on *congruent* trials were the same for both monolinguals and bilinguals, but that faster responding on *incongruent* trials was associated with different areas (bilateral cerebellum, bilateral superior temporal gyri, left supramarginal gyri, bilateral post-central and bilateral precuneous)—but only for bilinguals. This additional and different pathway employed by bilinguals on incongruent flanker trials leads Luk et al. to an unjustified conclusion that bilinguals have superior inhibitory control: "these results support the proposition that bilingualism influences cognitive control of inhibition" (p. 356) and that:

> differential engagement of this more extensive set of regions during incongruent trials in the two groups suggests that bilinguals can recruit this control network for interference suppression more effectively than monolinguals, consistent with their tendency to show less interference in terms of RT. (p. 356)

The reference to showing "less interference in terms of RT" cannot refer to the concurrent behavioural performance because the interference effects for the two groups were nearly identical. More likely they are referring to the studies cited in their introduction that reported significant bilingual advantages in the magnitude of the Simon, Stroop and flanker tasks. This advantage is properly modified as a "tendency," at best, given the many null results reviewed by Hilchey and Klein (2011) and updated earlier.

Another perspective on the relationship between brain and behaviour starts with the non-controversial assumption that behaviour is caused by underlying neural activity. Thus, even if the faster responding of the Cantonese-English bilinguals is due to sampling variation, some unidentified neural circuits of this group of bilinguals must be more efficient than those in the group of French-English bilinguals. The fact that the neuroscience methods used by Bialystok et al. (2005) identify similar neural pathways for both groups of bilinguals means that these methods have failed to identify the neural cause of the behavioural difference that actually did occur.

Differences between bilinguals and monolinguals in neural processing alone cannot support the hypothesis that there are bilingual advantages in conflict monitoring tasks such as Simon or Stroop. As a first step, the neural differences must be linked to the behavioural differences of interest. At this point in time, that link has not been forged as the behavioural differences reported by Bialystok et al. (2005) for the three language groups mismatch the neural differences. Likewise, Luk et al. report neural differences between groups of bilinguals and monolinguals, but there are no behavioural differences. In summary, the current evidence linking performance in conflict monitoring tasks to neuroimaging differences does not reinforce the relatively rare reports of bilingual advantages in performance of the Simon and flanker tasks. In fact, there is a surprising disconnect between the brain differences and behaviour differences that co-occur in the same experiment.

WHY ABANDON COMPONENTIAL ANALYSIS?

The assumption that complex behaviour can be understood on the basis of its known components is often referred to as reductionism. K&B acknowledge that a componential analysis is an effective means of rendering an otherwise intractable problem manageable by reducing it to a functional combination of simpler processes. Having acknowledged the benefits of this approach, they advance the claim that it does not apply or has been misapplied to understanding the interplay between bilingualism and cognition. K&B's leitmotif is that bilingualism is so complicated and special that it must be understood differently from other aspects of cognition. Reductionism is rejected because it is:

> a counterproductive and overly simplistic approach to understanding the broadly-based reorganisation that occurs from bilingualism … Approaches based on labels applied to tasks and abilities that seek a correspondence between them fail to account for the reorganisation of whole networks that follow from bilingualism. (pp. 499–500)

As argued in the previous section, neuroimaging has clearly revealed that bilingual and monolingual brains are different, but it has made only modest progress in determining how the differences in neural processing determine the differences (or similarities) in behaviour. Thus, the current neuroscience neither obviates the need or the usefulness of a componential analysis based on latent variables associated with behavioural measures.

An anonymous reviewer raised the following cogent questions in an earlier discussion. Why is bilingualism and its consequences special? Why does not the general model of cognitive psychology apply to this case? Might this not simply reflect the fact that we are in the early stages of understanding bilingualism and have not discovered the right building blocks? Reducing the richness of mental life to its possible components has proved to be a very successful strategy and, therefore, it seems to require more than what K&B present to justify that we abandon the general model.

K&B argue that "the tendency to consider bilingualism as a unitary phenomenon explained in terms of simple component processes has created a set of apparent controversies" (p. 497). Some of this controversy appears to stem from a tendency for K&B to attribute very strong conclusions to researchers who have offered quite cautious interpretations of their data. For example, consider the report by Alario, Ziegler, Massol, and DeCara (2012) of no relationship between the magnitude of Simon interference and homograph interference for a group of monolinguals. Such a relationship may have been anticipated if conflict resolution during language processing generalises and enhances conflict resolution in non-verbal processing: a process that many have conjectured underlies the bilingual advantage in EF. K&B characterise Alario et al.'s discussion as arguing that the absence of the relationship in monolinguals means that "the explanation for bilinguals is therefore incorrect. This is a reductionist error … Reducing performance to a few measurable components fails to capture the most crucial outcome of the experience, namely, the reconfiguration of these networks" (p. 501). What Alario et al. actually said was this: "These data may support the idea that the relationship between verbal and non-verbal conflict resolution is elusive and difficult to establish" (p. 556). Alario et al.'s stated conclusion comes closer to being a statement of fact than a conjecture that needlessly stirs controversy.

Also puzzling is K&B's embrace of several studies that rely on a componential analysis of the tasks. Many studies show that "bilingualism modifies not only inhibition but also monitoring (Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009), switching (Prior & Gollan, 2011), and working memory" (p. 500). On the one hand, K&B state in an unqualified way that the results of specific measures on specific tasks show that bilingualism modifies each of the core components of EP (namely inhibition, monitoring, switching and updating). For example, the bilingual advantage in switching costs reported by Prior and Gollan (2011) in a colour-shape switching task leads K&B to conclude that bilingualism modifies "switching." Prior and Gollan assume that the difference in mean RT between the switch trials and repeat trials (switching costs) is a good measure of the "switching" component of EF because the difference score compensates for individual differences in basic perceptual and motor processing, and both trial types require monitoring for the task cue and preparing for a possible switch. The point here is that, in this context, K&B appear to endorse experiments that were designed and interpreted using a very analytic and reductionist perspective. Put simply, if one believes that this type of componential task analysis is wrong and creates controversies, then one should not use such studies as evidence for a bilingual advantage.

WHO PRACTICES "CATEGORICAL HYPOTHESISING"?

K&B define categorical hypothesising as the expectation "that mutually exclusive alternatives can be compared such that supporting one invalidates the other" (p. 499), and then provide several examples of researchers who allegedly have engaged in the practice; "Hilchey and Klein (2011) assemble evidence from studies showing no bilingual advantage on simple inhibition tasks *and then use that result to discredit the entire body of work* (see also Paap & Greenberg, 2013, for a similar argument; p. 500)." This is a complete misunderstanding of the approach that Hilchey and Klein used to organise their review and of their conclusions. It certainly is true that they concluded that there was no compelling evidence for a bilingual advantage in *inhibitory control* in children or young adults. However, not only did they conclude that there was a *ubiquitous* bilingual advantage in global RT in the Simon and flanker tasks (that they associate with a bilingual advantage in monitoring rather than inhibitory control), but they also offered their own creative explanation as to why these advantages may occur on the basis of how bilinguals share resources and resolve conflict in comparison to monolinguals.

Construing Paap and Greenberg (2013) to have argued that the failure to find bilingual advantages in the inhibitory control component "discredits" the possibility that there may be advantages in other components reveals a lack of earnest effort to understand the studies they report and the research strategy that they explicitly advocated. The ensemble of tasks and measures that Paap and Greenberg report was guided by an analytic perspective that sought to isolate separable components of inhibitory control, monitoring and switching. The cumulative evidence for each component was evaluated independently from the other components. Paap and Greenberg never suggested that a preponderance of evidence against an advantage in one EF component (e.g., inhibition) should prejudice or discredit the evidence relevant to other components. This would be a silly argument if one adopts the theoretical framework that EF includes separable components. In fact, only a completely holistic view of EF that casts inhibition in an essential role might be threatened by conclusive evidence that bilingualism does not enhance inhibitory control. In summary, it is an utterly straw man argument to suggest that Hilchey and Klein, Paap and Greenberg, or anyone else has argued that differences between monolinguals and bilinguals in EF should or could be reduced to a single component.

Morton and Harper (2007) are also criticised for categorical hypothesising:

> some researchers have argued that group effects reported for bilingualism cannot be attributed to bilingualism but instead reflect differences in SES, education, immigration, or culture. For example, a small-scale study by Morton and Harper (2007) showed no difference in performance on a Simon task between children classified as monolingual or bilingual but a significant correlation between performance and an estimate of SES. Their conclusion was that bilingualism had no effect on executive functioning …. because SES was the crucial variable. Extrapolating from this result, they assumed that previous research reporting group differences was in fact reflecting differences in SES rather than bilingualism. The error in such reasoning is to assume that categorical designs require categorical interpretations: if the effect is caused by X, then it cannot be caused by Y. (p. 6)

The last sentence succinctly presents K&B's concern about categorical hypothesising, but again misrepresents what the authors actually said. Morton and Harper did not conclude that "bilingualism had no effect on executive functioning." Rather, they modestly state that their "results are at odds with prior evidence of a bilingual advantage in the Simon task" and that "One possible reason for this discrepancy is that bilingual and monolingual children in the present study had comparable ethnic and socioeconomic backgrounds" (p. 723). Morton and Harper most certainly never appealed to the logic that if the effect (namely language-group differences in the Simon task) is caused by X [namely socio-economic status (SES)], then it cannot be caused by Y (namely the bilingual experience). Instead, they simply observe that when language-groups are matched on *objective* measures of SES (namely family income and parent's educational level) and on immigrant status, there is no bilingual advantage in the Simon task. Thus, Morton and Harper's conclusions have nothing to do with adopting a radical and simplistic assumption that: "if the effect is caused by X, then it cannot be caused by Y."

K&B extend the categorical hypothesising argument into Miyake and Friedman's (2012) *unity and diversity* model of EF. Miyake and Friedman conducted confirmatory factor analyses (CFA) using measures from three different tasks for each of the three hypothesised latent variables (namely updating, switching and inhibition). Each measure was significantly loaded on the anticipated latent variable. At the higher level, the three latent variables correlate with one another and this is consistent with the assumption that each contributes to a common EF. When the same data are reanalysed with a second-order CFA, the nine observed measures all load on common EF with two of the nested components (updating and shifting) still making unique contributions. In summary, the CFAs support the theory of a general EF ability with separable updating and switching components and an inhibition component that is not separable and that is weakly linked to the general EF ability.

K&B suggest that Miyake and Friedman provide an example of researchers who "have begun to use more general terms than those given by the standard core components to explain differences in performance between monolinguals and bilinguals" (p. 500). Since Miyake and Friedman embrace both unity and diversity, individuals (or groups) could differ on a single component, a combination of separable components or the common higher-order component. Therefore, it is true that no one component is likely to account for differences observed across populations or situations. But, it does not logically follow that the framework should be abandoned in favour of more general terms like coordination or mental flexibility. If a bilingual advantage is defined in terms that are different from and not tied to the core or common components of EF as operationally defined and validated by Miyake and Friedman (or other systematic frameworks), then these informal terms cannot inherit the construct and convergent validity that exists for validated components.

THE RELATIVE VALUE OF NULL RESULTS

K&B's critique of categorical hypothesising digresses to some dubious claims about the determinants of power in between-group designs and the relative value of positive versus negative findings:

> Finally, the overlapping distributions of two groups performing the same task, in which participants are drawn from the same population and differ by only one feature, in this case bilingualism, make it extremely difficult to obtain a reliable difference in the mean score if only one measure is being considered. Standard experimental design usually involves about 25 or 30 participants per group, and the similarity of the populations in the two groups, the simplicity of the tasks used in this research, and the tendency for regression towards the mean makes it astounding that significant group differences are ever obtained. The considerable literature that reports group differences between monolingual and bilingual participants is greatly more informative than the attempted replications that fail to find significance. (p. 502)

Why should similarity of the populations reduce power? If participants are similar to one another in all respects other than bilingualism then should not this reduce variability in the dependent measure due to other factors? As for the concern expressed that null results are more likely when only *single-measures* are considered, Paap and Greenberg advocate for and use multiple measures. Like Costa et al., we used the attentional network test (ANT) version of the flanker task and reported null effects for measures of inhibitory control (the flanker effect), monitoring (global RT), alerting, orienting and context shifting (sequential dependencies).

Do violations of normality lead to null effects?

And why are significant group differences *greatly more informative* than null results?:

> Relatedly, statistical models often assume that the variables are normally distributed, a precondition

> that is almost never tested yet leads to null effects when it is violated. Failure to obtain the standard of statistical difference between groups is often a problem of the data distribution. (p. 501)

A recent article by Schmider, Ziegler, Danay, Beyer, and Bühner (2010) examined the robustness of analysis of variance (ANOVA) when the population distributions are markedly non-normal. They use Monte Carlo methods to compare the outcomes of one-way ANOVA calculations for samples of size n = 25 per group that were randomly drawn from either normally, rectangularly or exponentially distributed populations. For each distribution, 50,000 samples were generated and only the 10% most representative of their populations were selected for further analysis. This ensured that the "normal" samples deviated only slightly from their normal parent distribution and that the "non-normal" samples were not at all normal. Across a range of effect sizes the empirical type 1 error α and type 2 error β remained constant under violations. For example, for an effect size of $f = .4$, the percentage of significant ANOVAs were 86.8, 87.5 and 86.9, for the normal, rectangular and exponential distributions, respectively. The fact that the power values for the non-normal distributions are equivalent to those obtained with normal distributions shows that K&B are incorrect when they assert that non-normality leads to null effects when violated. Schmider et al. conclude that ANOVA is, indeed, robust when the normality assumption is violated.

Does categorical hypothesising lead to null results?

K&B attribute other null results to categorical hypothesising:

> the general problem with the categorical approach is that it fails to account for the inherently non-categorical nature of the relevant constructs. *Individuals are not bilingual or not, and tasks are not measures of inhibition or not* … Because of the need to identify categorical variables, gradations in all of these dimensions are washed over, and the interactions between linguistic and cognitive systems are rarely observed. An illustration of this problem can be seen in the recent paper by Paap and Greenberg (2013). (p. 502)

Of course no one task embodies any one process. This is a description of the widely discussed task impurity problem (Burgess, 1997). Likewise, the importance of considering many different types of bilingualism is not a new concern. What is somewhat novel is using these difficult methodological problems to dismiss experiments that yield null results (e.g., Paap & Greenberg, 2013) whereas expressing no fault with categorical hypothesising that yields bilingual advantages. As an example of the latter, consider a study by Engel de Abreu, Cruz-Santos Tourinho, Martin, and Bialystok (2012) that is well designed in many respects. More than one task and one measure of EF was used and the children were carefully matched on many variables. As K&B observe: "On the tasks that involved conflict and required executive control, bilingual children outperformed the monolinguals, but on all other measures, the children in the two groups performed equivalently" (p. 502). Apparently what is sauce for the goose is a categorical error by the gander. That is, Engel de Abreu et al. do compare, with impunity, bilinguals to monolinguals and tasks that involve conflict and require EF to those that do not. This is very much in the analytic tradition used by Paap and Greenberg (2013).

Engel de Abreu et al. characterise their results as remarkable because the bilingual children have strikingly low vocabulary scores in Luxembourgish and thus are not at all proficient in their L2. In this context, these robust differences invite consideration of alternative or additional reasons for the group differences. The matching reported by Engel de Abreu et al. is extensive, but no one study can match or hold constant all factors that could contribute to group differences on tasks assumed to measure EF. In this case, in order to hold Portuguese culture constant, the bilinguals were immigrants and the monolinguals were not. Also, two years of preschool are compulsory in Luxembourg, but that is not the case in Portugal. Although the Portuguese monolinguals did attend preschools, the quality of those programmes was not formally assessed and may have differed. More generally, it is very difficult to assure that children living in Portugal have had the same experiential advantages as those living in Luxembourg.

Given the remarkable outcomes of the Engel de Abreu et al.'s study, one would have more confidence if the bilingual advantage on measures of inhibition generalised to children in the same age range, but with groups that did not differ in immigrant status and country of schooling. In a recent study, Duñabeitia et al. (2013) compared Spanish monolinguals to Basque-Spanish bilinguals who were carefully matched on a large number of indices. Both groups were administered a verbal

Stroop task and a non-verbal version of the same task (namely the number size-congruency task). Results were unequivocal showing that bilingual and monolingual participants performed equally in these two tasks across all the indices or markers of inhibitory skills explored. The lack of differences between monolingual and bilingual children extended to all the age ranges tested (six successive grades with an age range of 8–13 years). Consistent with typical trends, the group sizes for the study finding no group differences (n = 252) were far larger than the study reporting differences (n = 40).

One possible resolution of the conflicting results is to conclude that the large group differences obtained by Engel de Abreu et al. were due to unmeasured advantages for children living in Luxembourg compared to Portugal. A contrasting resolution might attribute the conflicting results to the type of bilinguals because *individuals are not bilingual or not*. The bilingual children in Luxembourg had low proficiency in their L2, whereas the bilingual children in the Basque Country of Spain were far more balanced and fluent bilinguals. In the Basque Country, Spanish and Basque are co-official languages in everyday life, and legal regulations assure that 50% of all academic subjects are taught in each language.[1] Thus, why should the bilingual experience of the beginning bilingual lead to large bilingual advantages and those of the far more balanced bilinguals to no advantages at all?

One aspect of categorising that we might all agree on is that we should refrain from generalising beyond the specific population of monolinguals and bilinguals that we test. Another reviewer cogently observes that age of acquisition, proficiency, number of languages, language combinations, context of exposure, method of instruction and modality may all determine the extent to which differences between language groups occur. This is also consistent with many of the observations made by K&B. However, I also share this reviewer's impression that the reports of bilingual advantages too often infer that the results are likely to generalise and that anyone who has acquires a reasonable proficiency in L2 is likely to accrue these cognitive benefits.

[1] This situation is very similar to the Catalan-Spanish bilinguals and the Spanish monolinguals tested by Costa et al. Their work shows a bilingual advantage in the magnitude of the flanker effect for young adults in early blocks, but not in later blocks.

WOULD NON-NULL RESULTS BE REVEALED BY MULTIVARIATE ANALYSES?

K&B argue that the reductionist and categorical assumptions can be avoided by adopting multivariate designs:

> An alternative approach to investigating the consequences of bilingualism is to use multivariate approaches that evaluate changes on continua. In a recent example, Bialystok and Barac (2012) examined the relation between the length of time children spent in an immersion education programme and thus experienced a bilingual environment, various measures of their proficiency in both languages, and outcome measures of both nonverbal executive control and metalinguistic ability. The interesting finding is that the relations were different for the different outcomes: Bilingual experience was related to performance on executive control tasks and language ability was related to performance on metalinguistic tasks. Categorical comparisons between groups on specific tasks would not have revealed these emerging differences. (p. 503)

The Bialystok and Barac's study does not clarify why multivariate designs are a panacea that enables researchers to escape an analytic approach to a complex problem and the genuine difficulties regarding the mapping of task measures to psychological constructs. The danger of abandoning an analytic framework that embraces the testing of causal relationships framed as a priori hypotheses is that it enables a promiscuous confirmation bias where any pattern of group differences is interpreted as a complex interaction of linguistic and cognitive processing thus confirming the assumption that bilingualism enhances executive processing. For example, Bialystok and Barac conclude that: "The results of the regression analyses were remarkably consistent: metalinguistic performance improved with increased knowledge of the language of testing and executive control performance improved with increased experience in a bilingual education environment" (p. 71).

Their results are remarkably consistent only because they choose to highlight the variables and analyses that "worked" and did not discuss those that did not. For example, in Study 1 both degree of balance and time in bilingual education predicted the flanker effect (a measure of inhibitory control in EF), but neither predicted global nor local switch costs. The results of Study 1 would be more consistent if the degree of balance had also predicted switching costs (unless there is a principled argument why the bilingual educational environment should enhance inhibitory control, but not shifting). In Study 2, the full set of predictors for global switch costs results in a non-significant model for both Grades 2 and 5. However, a reduced model on the Grade 5 data using only non-verbal intelligence and time-in-program is significant. Using multiple analyses of the same data-set is a risky research practice as discussed by Bakker, van Dijk, and Wicherts (2012), Francis (2012) and others. Note that the most direct measure of bilingualism (the balance measure) is not a significant predictor in either model for Grade 5.

Finally, why has the *sentence judgement task* morphed from one that definitely requires executive control to one that is now a pure measure of metalinguistic awareness? A previous article by Bialystok's group (Moreno, Bialystok, Wodniecka, & Alain, 2010) critically assumed that the sentence grammaticality task requires EF and this assumption is used to explain why bilinguals close the performance gap that was obtained with a sentence acceptability task (that is assumed to require little or no EF). In Bialystok and Barac's (2012) study, this same grammaticality task is now asserted to be a measure of metalinguistic awareness that is functionally equivalent to Berko's (1958) famous *wugs* task that was used in Study 1. Although it is certainly true that the grammaticality task requires awareness of the distinction between syntactic and semantic violations, the intent to confirm a bilingual advantage compels Bialystok and Barac to juxtapose the grammaticality task as an EF-free task that can be contrasted to a prototypical task of EF (namely switching costs). The morphing of the grammaticality task from an EF-demanding task in 2010 to an EF-free task in 2012, with neither mention nor justification, is disconcerting.

This close examination of Bialystok and Barac's study validates several of the issues raised earlier. Multivariate designs cannot replace the need to operationally define variables and carefully consider their validity. These methodological steps are part of the thinking that researchers invest in dividing complex problems into manageable components. Ironically, K&B do not eschew the use of the componential assumptions in their own practice. For example, in Bialystok and Barac's study, tasks are assumed to reflect differences in EF (flanker, switching) or differences in metalinguistic awareness (wugs, grammaticality).

Are continuous predictor variables in regression always better than discrete levels in a factorial ANOVA? Although not disagreeing that it makes sense to use degree-of-balance and English proficiency as continuous predictor variables for this particular study, we do not know whether similar differences would have been revealed by partitioning the samples into groups of say low-, medium- and high-proficiency or low-, medium- and high-balance. It might be quite informative to discover that differences in the magnitude of flanker interference were driven entirely by those children who had achieved a high level of balanced fluency.

Paap and Greenberg (2013) routinely treated bilingualism as both a categorical factor in an ANOVA and as a continuous variable in a regression analysis. Our *multilingualism* predictor is highly correlated with the *balance* variable used in Bialystok and Barac's study. In our studies, both analyses almost always converge on the same outcome.

Paap and Greenberg also explored whether bilingual advantages would surface when the level of language proficiency, age of acquisition, parent's educational level, Ravens scores, computer-gaming experience, multitasking preferences and experiences, and ability to play team sports were taken into account. Six separate regression analyses were used to predict the Simon effect, flanker effect, switching costs, mixing costs and global RT in both the Simon and the flanker tasks. The explanatory capacity of these other factors was markedly low, although two predictors (the Ravens score and excellence in team sports) were significant in three or four of the models and in the anticipated direction that greater ability should be associated with better measures of EF.

Have null results been obtained with sufficiently difficult tasks?

Another point raised by K&B in dismissing Paap and Greenberg's null results is that bilingual advantages are fragile and may not be revealed if the tasks are not sufficiently difficult in terms of their processing demands; specifically citing Costa et al.'s *Now you see it, now you don't* article. Paap and Greenberg explicitly modelled their ANT task on that used by Costa et al. and, of course, chose a ratio of congruent to incongruent trials that produces very large flanker effects. Furthermore, their task shows high test-retest reliability and a wide range of individual differences with respect to the magnitude of the flanker effect even when the sample is drawn from a highly restricted population, e.g., psychology majors at a specific university (Sawi & Paap, 2013). Likewise, their colour-shape switching task is modelled on the one used by Prior and Gollan. The mixing costs and switching costs reported by Paap and Greenberg exceed 200 ms. All their measures yield highly significant main effects of trial type that meet or exceed the magnitude of what is typically observed.

Don't bilingual advantages always occur with elderly participants?

The concern about task difficulty leads to K&B's observation that "the very same tasks that fail to produce differences for young adults may produce striking results for older bilinguals" (e.g., Gold, Kim, Johnson, Kryscio, & Smith, 2013, p. 503). The underlying premise is that most young adults have already optimised their capacity for cognitive control and therefore bilingual advantages are unlikely to appear in this population. This may well be true, but consider K&B's characterisation that there were *striking* differences for the older adults. The logic and rationale of Gold et al.'s study critically depends on the demonstration of a bilingual advantage in the behavioural measures associated with switching. In this regard, the results are not ideally clear. In their first experiment, 15 older bilinguals were compared to 15 older monolinguals in the perceptual switching task without any concurrent brain imaging. There was a significant Condition × Language Group interaction confirming a bilingual advantage in global switch costs (i.e., the difference between pure blocks of single task trials and pure blocks of switch trials), that is, the global switch costs were 50 ms greater for monolinguals than bilinguals. This sets the stage for Gold et al. to conduct the main experiment where both younger and older adults were recruited and the behavioural task took place when functional magnetic resonance imaging (fMRI) images were concurrently obtained. Thus, Experiment 2 enables a test of the Age × Condition × Language Group interaction. The behavioural results with respect to this critical interaction are ambiguous. As shown in Figure 2 of Gold et al., the global RT differences for the young adult bilinguals and monolinguals are nearly identical and clearly not significant. More problematic are the global switch costs between the two groups of older adults. A *t*-test that compares the mean for older bilinguals (M = 14.1%) to that for older monolinguals (M = 23.0%) does not meet the conventional .05 level of significance, $t(38) = 1.97$, $p = 0.056$. Because the exact p of .056 is greater than .05, a conservative interpretation would lead to the conclusion that the null hypothesis cannot be rejected. This is unfortunate because all of the sophisticated analyses conducted by Gold et al. that investigate the relationship between behavioural switch costs and the neural switch costs are restricted to the sample where the bilingual advantage on the behavioural measure is in doubt.

Summary of discussion on relative value of null results

K&B argue that the methodological issues discussed in this section render significant group differences *astounding* and almost always lead to null results. These issues have been deconstructed and found unconvincing. The logic reduces to the argument that if we believe an effect is difficult to produce, then we should discount failures to replicate.

ARE SMALL *N*'S REALLY THAT RISKY?

Is the observation and criticism of Gold et al.'s study a nit-pick or substantial? In favour of the latter are concerns raised in the recent special issue of *Perspectives in Psychological Science* (November, 2012) on replicability and research practices. To take one example from the special issue, Bakker et al. (2012) report that 96% of research articles report significant results but typical studies are insufficiently powerful for such a track record. They explain this paradox by showing through simulations that the use of several small underpowered samples rather than the use of one larger and more powerful sample is a more efficient strategy for generating significant ($p < .05$) findings. The funnel plots of the simulations shown in the top row of Bakker et al.'s Figure 4 show the striking differences between multiple small studies and one large study when the true effect size is set to zero.

Although the significant ($p < .05$) bilingual advantages reported by Gold et al. for the older adults in Experiment 1 and the marginally significant (p = .056) differences obtained in

Experiment 2 may prove to be replicable, it is fair to point out that Experiments 1 and 2 involved *n*'s of 15 and 20 participants per group. Gold et al.'s experiments, however excellent in many other respects, are underpowered and vulnerable to the false positives as described by Bakker et al. and others. These concerns about small *n*'s also apply to Gold et al.'s fMRI-based analyses of neural switching cost. In an analysis of neuroscience rather than psychological science, Button et al. (2013) show that the average statistical power of studies in the neurosciences is very low and that low power reduces the likelihood that a statistically significant result reflects a true effect.

A second, but related, concern involves the bilingualism variable itself. The mediation analysis is interpreted to support a two-link causal chain: managing two languages enhances neural efficiency in the switching network, which, in turn, determines performance in the perceptual switching task. However, this model is only as good as the degree to which the language groups are matched on other variables that influence the efficiency of the switching network. Although acknowledging that Gold et al. have matched their groups on many characteristics, the groups do co-vary in terms of immigrant status (e.g., 75% of the older bilinguals are immigrants compared to only 15% of the older monolinguals). Older adults from immigrant families may enjoy cultural advantages (e.g., living in extended families rather than alone) that may delay cognitive decline in EF. Morton and Carlson (in press) present evidence for cultural differences in the development of EF and show how these differences can confound tests of bilingual advantages. Kirk, Scott-Brown, and Kempe (2013) report that when immigrant status and culture are matched, there are no differences in inhibitory controls between older Gaelic-English bilinguals and three monolingual control groups. Likewise, Kousaie and Phillips (2012b), using non-immigrant samples of monolinguals and bilinguals, found no differences in Stroop interference in either young or older adults. In summary, the immigrant status or cultural differences may not provide a better explanation for Gold et al.'s data, but they are at least plausible and the mounting number of null results should not be ignored.

Small *n*'s are especially risky when random assignments cannot be used because the comparison of interest is between two naturally occurring populations. Unfortunately, small *n*'s have dominated the published reports of bilingual advantages. For example, the seminal work of Bialystok, Craik, Klein, and Viswanathan (2004) showing bilingual advantages in the Simon task reported three studies using language-group sizes ranging from 10 to 32 with a mean of 15.4. In contrast, Paap and Greenberg (2013) report three Simon studies showing no bilingual advantages using group sizes ranging from 34 to 55 with a mean of 45.2. Likewise, Paap and Greenberg had more than 50 participants in each language group in their unsuccessful replication of Costa et al.'s (2009) flanker task that had 31 or 32 participants per group. A laudable exception to this pattern is that Costa, Hernández, and Sebastián-Gallés (2008) tested 100 participants in each language group.

These systematic differences in sample size are not surprising because researchers understand that unsuccessful replications are unlikely to be published, particularly when the failure results in a null result.

REPLICATION AND CONFIRMATION BIAS

K&B are correct in saying that Paap and Greenberg "call into question all evidence for bilingual advantages in EF" (p. 502). However, it is fair to note that Paap and Greenberg did so in the context of rhetorically contrasting two views:

> The two opposing views are that either there are genuine bilingual advantages that happen to be quite elusive given our current understanding of how and why they develop; or that the performance advantages, when they occur, are due to causes unrelated to bilingualism enhancing EP. The first view questions why null results occur while the second asks what may have caused performance differences favouring bilinguals. In our view the evidence points in the direction of no genuine bilingual advantage in EP. But, we are open to new and compelling evidence. (p. 255)

Paap and Greenberg go on to explicitly describe the type of evidence that they would find compelling:

- Identify the specific component(s) of EP that should be enhanced by managing two languages.
- Show a bilingual advantage in an indicator of that component across two different tasks.
- Show that the indicators correlate with one another and have some degree of convergent validity.

- Show no differences between the two groups on a pure block of easy choice-RT trials.
- Match the groups on SES and immigrant status.
- Minimise cultural differences between the groups.

The *pièce de résistance* of such as enterprise would, of course, be a successful replication in an independent laboratory. Admittedly this is a challenging agenda. Be that what it may and turning the tables, what would convince K&B that bilingualism (however special and beneficial it is in many other ways) does not enhance EP?[2]

If the asserted phenomena of bilingual advantages in EF is not falsifiable then it belongs in a category that Ferguson and Heene (2012) refer to as *undead*:

> Many theories, particularly those tied to politicized or "hot" topics, are not subjected to rigorous evaluation and, thus, are allowed to survive long past their utility. This is our use of the term "undead theory", a theory that continues in use, resisting attempts at falsification, ignoring disconfirming data, negating failed replications through the dubious use of meta-analysis or having simply maintained itself in a fluid state with shifting implicit assumptions such that falsification is not possible…. We suspect a good number of theories in popular use within psychology likely fit within this category; theories that explain better how scholars wish the world to be than how it actually is. (p. 559)

Pashler and Harris (2012) have cogently argued for the increased use of exact replications while explaining why conceptual replications (especially when such studies are numerous and low in power) may actually interact with publication bias to compound the problem. Consider the early history on the bilingual advantage in EF. Original reports that the interference effect in the Simon task are larger for monolinguals than bilinguals could not be replicated, but a global RT advantage appears … but only when overall task difficulty is high. Partial conceptual replications lead investigators to acknowledge that the underlying effect is not as robust or generalisable as initially hoped, but they do not call into question the essence of the phenomena as inferred from the original report. Successful conceptual replications are more publishable than exact replications. Thus, Pashler and Harris suggest that in fields where exact replication attempts are rare and conceptual replication attempts are common the community of researchers can easily be led to believe "beyond question in phenomena that simply do not exist" (p. 534). In their view, the key to promoting such pathological results is a pseudo result that appears both exciting and relatively easy to carry out. Pashler and Harris speculate that large numbers of low-powered conceptual replications can lead to a steady stream of those that *work* and, consequently, generate the strong impression in the community that successful confirmations are plentiful. In this context, it seems risky that K&B should insist that unless "all other explanations have been exhausted, it is misleading to call into question the reliability of the phenomena themselves" (p. 503).

Pashler and Harris also point out that one calling card of this pathology is when defenders of a controversial effect question the initial methods used to obtain the effect and rest their case upon later studies conducted using other methods. Thus, we see the proponents of a bilingual advantage in EF drift from an original explanation rooted in inhibitory control to conflict monitoring and now to mental flexibility. Likewise, the preferred task paradigm moves from Simon to Stroop to flanker to the AX-continuous performance task. Optimistically, one could view this as incremental scientific progress towards understanding the relationship between bilingualism and cognitive control. Realistically, the field needs to engage in more exact replication and articulation of a coherent theory that links specific aspects of the bilingual experience to specific enhancements in EF.

Original manuscript received June 2013
Revised manuscript received October 2013
Revised manuscript accepted February 2014
First published online February 2014

[2]This rhetorical question is not intended for those researchers who have reported bilingual advantages in non-verbal tasks, but nonetheless remains open to what those differences reflect or the possibility that the phenomena will prove to be ephemeral or constrained to very special types of bilingual experience.

REFERENCES

Alario, F.-X., Ziegler, J. C., Massol, S., & De Cara, B. (2012). Probing the link between cognitive control and lexical selection in monolingual speakers. *L'Annee Psychologique/Topics in Cognitve Psychology*, *112*, 545–559.

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

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.
Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*(1), 67–73. doi:10.1016/j.cognition.2011.08.003
Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishi, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage*, *24*, 40–49.
Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*, 290–303. doi:10.1037/0882-7974.19.2.290
Bialystok, E., Craik, I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Science*, *16*, 240–250.
Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–111). Hove: Psychology Press.
Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376. doi:10.1038/nrn3475
Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*, 135–149. doi: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. doi:10.1016/j.cognition.2006.12.013
Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2013). The inhibitory advantage in bilingual children revisited. *Experimental Psychology*. doi:10.1027/1618-3169/a000243
Engel de Abreu, P. M. J., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor: Enhanced cognitive control in low-income minority children. *Psychological Science*, *23*, 1364–1371. doi:10.1177/0956797612443836
Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*, 975–991.
Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin and Review*, *19*, 975–991. doi:10.3758/s13423-012-0322-y
Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*, 201–225. doi:10.1037/0096-3445.137.2.201
Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *Journal of Neuroscience*, *33*, 387–396. doi:10.1523/JNEUROSCI.3837-12.2013
Hernández, M., Martin, C. D., Barceló, F., & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language*, *69*, 257–276.
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*, 625–658. doi:10.3758/s13423-011-0116-7
Humphrey, A. D., & Valian, V. V. (2012, November). *Multilingualism and cognitive control: Simon and flanker task performance in monolingual and multilingual young adults*. Presentation at the 53rd annual meeting of the Psychonomic Society, Minneapolis, MN.
Kirk, N. W., Scott-Brown, K., & Kempe, V. (2013). Do older Gaelic-English bilinguals show an advantage in inhibitory control? *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Berlin, Germany, 1 August 2013.
Kousaie, S., & Phillips, N. A. (2012a). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research*, *1446*, 71–90. doi:10.1016/j.brainres.2012.01.052
Kousaie, S., & Phillips, N. A. (2012b). Ageing and bilingualism: Absence of a "bilingual advantage" in Stroop interference in a nonimmigrant sample. *The Quarterly Journal of Experimental Psychology*, *65*, 356–369. doi:10.1080/17470218.2011.604788
Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*, 497–514. doi:10.1080/20445911.2013.799170
Luk, G., Anderson, J. A. E., Craik, F. I. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and Cognition*, *74*, 347–357.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. doi:10.1177/0963721411429458
Moreno, S., Bialystok, E., Wodniecka, Z., & Alain, C. (2010). Conflict resolution in sentence processing by bilinguals. *Journal of Neurolinguistics*, *23*, 564–579. doi:10.1016/j.jneuroling.2010.05.002
Morton, J. B., & Carlson, S. (in press). The bilingual advantage: Evidence and alternative views. In M. Hoskyn, G. Iarocci, & A. Young (Eds.). Oxford University Press.
Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*, 719–726. doi:10.1111/j.1467-7687.2007.00623.x
Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232–258. doi:10.1016/j.cogpsych.2012.12.002

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

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

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

Ryskin, R. A., & Brown-Schmidt, S. (2012, November). *A bilingual disadvantage in linguistic perspective adjustment*. Presentation at the 53rd annual meeting of the Psychonomic Society, Minneapolis, MN.

Sawi, O., & Paap, K. (2013, April). *Test-retest reliability and convergent validity of measures of executive processing: Evidence from the Simon, flanker, switching, and antisaccade task*. Poster presented at the meeting of the Cognitive Neuroscience Society, San Francisco, CA.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *6*, 147–151. doi:10.1027/1614-2241/a000016

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