The Brain Mechanisms Underlying the Cognitive Benefits of Bilingualism may be Extraordinarily Difficult to Discover

5 authors, including:

Kenneth R. Paap
San Francisco State University
86 PUBLICATIONS **4,075** CITATIONS

SEE PROFILE

Oliver M Sawi
University of Connecticut
22 PUBLICATIONS **977** CITATIONS

SEE PROFILE

Chirag Dalibar
San Francisco State University
5 PUBLICATIONS **55** CITATIONS

SEE PROFILE

Jack Darrow
San Francisco State University
6 PUBLICATIONS **55** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Statistics View project

Social Psychology View project

Review article

The Brain Mechanisms Underlying the Cognitive Benefits of Bilingualism may be Extraordinarily Difficult to Discover

Kenneth R. Paap [1,*], **Oliver M. Sawi** [1,2], **Chirag Dalibar**[1], **Jack Darrow** [1,3], **Hunter A. Johnson** [1]

[1] Language, Attention, & Cognitive Engineering (LACE) Lab, Department of Psychology, San Francisco State University, San Francisco, CA, USA
[2] Department of Psychology, University of Connecticut, Storrs, CT, USA
[3] Department of Philosophy, San Francisco State University, San Francisco, CA, USA

* **Correspondence:** Email: kenp@sfsu.edu; Tel: +415-338-6840.

Abstract: The hypothesis that coordinating two or more languages leads to an enhancement in executive functioning has been intensely studied for the past decade with very mixed results. The purpose of this review and analysis is to consider why it has been (and will continue to be) difficult to discover the brain mechanisms underlying any cognitive benefits to bilingualism. Six reasons are discussed: 1) the phenomenon may not actually exist; 2) the cognitive neuroscientists investigating bilingual advantages may have been studying the wrong component of executive functioning; 3) most experiments use risky small numbers of participants and are underpowered; 4) the neural differences between groups do not align with the behavioral differences; 5) neural differences sometimes suffer from valence ambiguity, that is, disagreements whether "more" implies better or worse functioning and 6) neural differences often suffer from kind ambiguity, that is, disagreements regarding what type of mental events the pattern of activation in a region-of-interest actually reflects.

1.   Introduction

Executive functions (EFs) consist of a set of general-purpose control processes believed to be central to the self-regulation of thoughts and behaviors that are instrumental to accomplishing goals. From a neuropsychological perspective the construct of EF is often viewed as a set of interrelated component processes all involving the prefrontal cortex (PFC) with each component recruiting additional areas of cortical function. This componential framework allows for the possibility that the related components have some degree of anatomical and functional independence [1]. Thus,

individuals may vary in terms of overall EF ability or with respect to specific components. If EFs are general-purpose then individuals who excel in, say, a measure of inhibitory control in one task should also show little interference (excellent inhibitory control) in a different task. That is, indices obtained in different tasks but assumed to measure the same component of EF should correlate and thus show convergent validity.

The core assumption of those who have concluded that bilingual advantages in EF exist is that coordinating two languages requires substantially more cognitive control than using only a single language. One prime example of this coordination is the need to monitor conversational partners and to switch languages as needed depending on who enters or leaves the conversation. A switch presumably requires the activation of the new target language and the inhibition of the one that no longer has a direct conversational utility. A second example of the special demands placed on bilinguals is the need to resolve the competition between translation equivalents especially during language production. If the task of managing two languages does require substantially more control than using only one language and if this control is the same (viz., involves the same neural circuits) as the cognitive control employed generally, then this ubiquitous practice of coordinating two languages should lead to bilingual advantages in nonverbal tasks that require switching or conflict resolution [2]. Many contemporary research programs seek to determine the brain mechanisms underlying the benefits of bilingualism for general EF, for example, Abutalebi et al. [3]. For multiple reasons it may be extraordinarily difficult to discover these brain mechanisms; six of those are examined in this article.

2. Bilingual Advantages May Not Exist

Paap and Greenberg [4] concluded that there was no compelling evidence for bilingual advantages in general EF and that reports of statistically significant performance advantages were likely to be artifacts. If the phenomenon does not exist, then there are no underlying brain mechanisms to discover.

One reason to doubt the existence of bilingual advantages is that the most comprehensive tests —those using multiple tasks and large numbers of participants, have shown no advantages at all. In each of four studies reported by Paap and Greenberg [4] and Paap and Sawi [5] individuals participated in either three or four tasks (antisaccade, flanker, Simon, and color-shape switching). Measures of different components of EF (e.g., monitoring, inhibitory control, switching) could be derived from each task. Across all four studies there were 28 tests for bilingual advantages and the vast majority yielded no significant differences while two yielded a monolingual advantage. Cumulatively there were 213 monolinguals and 180 bilinguals. The participants were relatively homogeneous in terms of life experiences as all were undergraduate psychology majors at the same university. Although all of the bilinguals spoke English they spoke many different other languages, varied in terms of the relative proficiency of the first language (L1) and second language (L2), age-of-acquisition of L2, and the degree of language switching regularly employed. When this composite database was used to explore these factors Paap, et al. [6,7] did not find any evidence that specific bilingual experiences potentiate the bilingual advantage.

Given the reasonable conjecture that benefits of bilingualism are likely to develop to a maximum in bilinguals who are highly proficient, acquire both languages early, and reside in language communities where most people speak the same two languages and switching is ubiquitous; the studies by Duñabeitia, et al. [8]; Antón, et al. [9]; and Gathercole, et al. [10] deserve special

attention. Duñabeitia et al. [8] compared Spanish monolinguals (n = 252) to Basque-Spanish bilinguals (n = 252) at six successive grades with respect to both a verbal Stroop task and a number-size congruency task. Bilinguals and monolinguals performed equivalently in these two tasks in terms of global RT and across all the indices of inhibitory control explored across all grade levels. Antón et al. [9] compared a group of 180 Basque-Spanish bilingual children with a group of 180 carefully matched Spanish monolinguals on a flanker task. There were no language group differences in either inhibitory control (incongruent-congruent) or in global RT. The Gathercole et al. [10] study of Welsh-English bilinguals was a lifespan study testing seven age groups (from 3 years of age through over 60). They reported no systematic language-group differences on three tasks assumed to reflect EF: dimensional card sorting (N = 650), Simon (N = 557), and a grammaticality judgment with irrelevant semantic anomalies (N = 354). All three studies share the strengths of using bilinguals immersed in a bilingual region, monolingual control groups from the same country, a very large number of participants, multiple age groups, and multiple measures of EF.

These large and comprehensive studies provide no evidence at all for bilingual advantages in EF: but it is important to address the sizeable number of published studies that show significant bilingual advantages in one measure computed from a single task. The case that many of these are artifacts is presented in Section 3. Others may involve replicable designs, but use assumed measures of EF that have little or no convergent validity as described in detail by Paap and Sawi [5].

3. Neuroscience investigations of the bilingual advantage may be investing heavily in the wrong component of EF

This discussion requires some recent history regarding a shift in thinking about the locus of the bilingual advantage. Most of the influential early investigations, Bialystok et al. [11] for example, focused on testing the hypothesis that bilinguals have superior inhibitory control. The logic underlying the hypothesis was based on evidence from psycholinguistic studies of adult language processing that showed that the two languages of a bilingual remain active even when the context strongly supports the intention to use only one of them. The joint activation of two lexicons requires a mechanism for keeping the languages separate so that fluent performance can be achieved without intrusions from the unwanted language. If this mechanism as assumed by Green's inhibitory control model [2] involves the same executive functions used for general inhibitory control, then bilinguals accrue massive practice that should make them less vulnerable (compared to monolinguals) to interference in nonlinguistic tasks.

The standard marker of inhibitory control in these tasks is the difference in mean response time between trials that require conflict resolution compared to those that do not. In the Stroop (both verbal and nonverbal versions), Simon, and flanker tasks conflict occurs on a subset of trials because a potent but task-irrelevant stimulus is paired in an incongruent manner with the task-relevant stimulus. The effectiveness of this inhibitory control can be inferred from differences in response times between congruent and incongruent trials with smaller interference effects implying superior ability.

The problem with this explanation is that it is did not fit the data. Hilchey and Klein [12] reviewed 31 experiments and concluded that evidence for a bilingual advantage in inhibitory control in both children and young adults was rare, but that there was an advantage in global RT. The fact that the bilingual advantage is usually equivalent for congruent and incongruent trials is perplexing. This pattern could only occur if there was an advantage that applies equally to both types of trials

and, at the same time, there was no additional advantage in inhibitory control. This gave rise to the view that bilinguals are better at managing trial-to-trial variation with respect to presence or absence of conflict under the rubric of "monitoring" [13], "coordination" [14], or "mental flexibility" [15]. From an elementary view of experimental control, a bilingual advantage in monitoring requires a control condition that experiences no conflict at all: otherwise any group differences could be attributed to differences in perceptual processing, motor processing, or general fluid intelligence rather than a component of EF. The purest test would be a difference score between the congruent trials from a mixed block and the same type of trials in a pure block that includes no conflict trials. The studies that have included this control have not shown a bilingual advantage in global RT [4–6]. Nonetheless, this is the precise contrast that should be used in a neuroimaging study if the goal is to determine the neural circuitry that is responsible for a monitoring advantage that applies equally to both congruent and incongruent trials.

For example, a widely cited article by Abutalebi, et al. [3] is titled "Bilingualism Tunes the Anterior Cingulate Cortex for Conflict Monitoring", but the neural contrast referred to as "conflict monitoring" (incongruent BOLD—congruent BOLD) actually corresponds to the behavioral contrasts typically referred to as "inhibitory control"(incongruent RT—congruent RT). Abutalebi et al. [3] report no behavioral differences in global RT and do not even run a baseline condition that would enable them to discover cortical regions involved in monitoring and preparing for conflict that would facilitate both congruent and incongruent trials. In short, there is currently a disconnect between the favored EF component in the behavioral literature (monitoring / mental flexibility) and the contrasts used most frequently in the neuroimaging work that typically compare statistical maps between congruent and incongruent trials.

4.   Small numbers of participants in each group

Another problem is that the research on the bilingual advantage has been plagued by the same questionable research practices that are, unfortunately, prevalent in both psychological science and neuroscience. As discussed by Paap [16] and Paap and Liu [17], the most frequent and serious problems combine small numbers (n) of participants with biases to confirm previous findings and to undervalue replications. The small n component of this problem can be especially troublesome when a research question is framed as a difference between two naturally-occurring populations (e.g., bilinguals and monolinguals) rather than as differences across levels of an independent variable that can be manipulated. Paap, et al. [6] tabulated 76 tests for bilingual advantages that used nonverbal interference or switching tasks and showed that bilingual advantages are in a clear minority and tend to occur when there are a small number of participants per language group whereas null results occur both with small n and large n. This is not the expected pattern if bilingualism truly does enhance EF. If the null is false, then as the sample size becomes very large the t values grow without bound, and the p values converge to zero. That is, the null will always be rejected in the large-sample limit if there is a real difference to detect. Thus, all others things being equal, one would expect significant effects (especially for a small effect size) to cluster at the higher end of sample sizes.

Small n's reduce an experimental design's power to correctly reject the null hypothesis; but as Bakker, et al. [18] demonstrate with simulations, small n's coupled with a bias against null findings also result in an inflated rate of false positives. Similarly, in an analysis of studies in the neurosciences Button et al. [19] show that the average statistical power is very low and that low power reduces the likelihood that a statistically significant result reflects a true effect. Francis [20]

strongly asserts that "Studies with unnecessarily small sample sizes should not be published" (p. 989). The meaning of small should be based on statistical power and not research tradition. Correcting the problem of underpowered experiments in neuroscience is an expensive endeavor. For example, if the effect of bilingualism on EF was generously estimated to be of medium size (Cohen's d = .5), if the effect was tested with an alpha of .05, and if a researcher was willing to accept a power of only .67, then one would need 36 and 48 participants in each of two language groups given a one-tailed and a two-tailed test, respectively.

There are, of course, individual studies that used moderately large sample sizes and obtained significant bilingual advantages in performance. For example, Engel de Abreu et al [21] compared 40 Portuguese-Luxembourgish bilinguals to 40 Portuguese monolinguals and obtained a bilingual advantage on measures of inhibitory control across multiple tasks. Many steps were taken to try to match the two language groups on a variety of demographic characteristics, but nonetheless the groups differed in terms of immigrant status, culture, and mandatory preschool. All of these factors can influence the development of cognitive control [22,23,24]. Interpreting the group differences as consequences of managing two languages is further challenged because the bilinguals were not at all proficient in their L2. Finally, these results are simply inconsistent with the studies reviewed earlier by Duñabeitia et al. [8], Antón et al. [9] and Gathercole et al. [10] that had far more participants per group, compared groups matched on immigrant status and culture, and tested bilinguals who acquired both languages early and switched between them frequently.

5.   Alignment Problem

In many influential articles the pattern of behavioral differences across language groups is different from the pattern of neural differences. The examples of this *alignment problem* that are presented below are not intended to be an exhaustive review of all neuroscience investigations of the bilingual advantage in EF; however, they do account for a large proportion of the neuroimaging evidence cited as supporting the phenomenon and identifying the underlying neural circuits.

5.1. Bialystok, Craik, Grady, Chau, Ishii, Gunji, & Pantev (2005) [25]

In general, cortical areas shown to be involved in managing two languages overlap with those shown to be involved with inhibitory control and switching [26]. 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 coordinating two languages leads to a reorganization of neural networks in cortical areas involved in EF. However, as argued by Paap [16] reorganization 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. Thus, it is imperative that the observed neural differences be linked with behavioral differences confirming bilingual advantages in actual performance—that link has not yet been forged. For example, Bialystok, et al. [25] reported that two groups of bilinguals (Cantonese-English and French-English) showed substantial overlap in terms of the loci associated with fast responding in a Simon task (ACC, superior frontal, and inferior frontal regions) and that these markedly differed from the specific areas associated with fast responding for

monolinguals (middle frontal region). However, these neural differences did not align with the differences observed in behavior.. There were no group differences at all in interference control (incongruent RT—congruent RT) and there was a global RT advantage for the Chinese-English bilinguals (n = 10) compared to both the French-English bilinguals (n = 10) and the monolingual group (n = 10). Yet, the authors conclude that *"… the combination of behavioral data and images of regional activation derived from MEG converged to show systematic differences in performance between monolingual and bilingual participants"* p. 48. Although one might quibble with the meaning of "systematic" our point is that the pattern of global RT does not match the pattern of MEG differences: only the Cantonese-English bilinguals showed a global RT advantage. Furthermore (and perhaps even more important), given that: (a) the RT advantage of the Cantonese-English bilinguals was present in the control condition (pure block of no conflict trials) as well as in the experimental block and (b) that there were no differences in the magnitude of the Simon interference effect: there is simply no evidence at all in this study for a bilingual advantage.

5.2. *Luk, Anderson, Craik, Grady, & Bialystok (2010)* [27]

A study by Luk, et al. [27] claims to have shown *"….neural correlates associated with more efficient suppression of interference…"* (p. 347). This claim is completely divorced from the actual behavioral results obtained with their flanker task. Again, the group sizes were very small: 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 x Trial Type interaction. Thus, there was no behavioral evidence for a monitoring advantage (global RT or congruent RT—baseline RT) or for an interference control advantage (incongruent RT—congruent RT). With respect to neural differences 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) only for bilinguals. This additional and different pathway employed by bilinguals on incongruent flanker trials lead Luk et al. to conclude 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 behavioral performance because the interference effects for the two groups were nearly identical. Luk et al. offer a brash defence of their conclusion: *"Equivalent performance in the two groups allows meaningful interpretation of the differences in functional neural correlates without the possible confound of behavioral differences"* p. 355; as if a strong bilingual advantage would have discredited their neural findings.

5.3. *Gold, Kim, Johnson, Kriscio, & Smith (2013)* [28]

Kroll and Bialystok [15] observed that *"… the very same tasks that fail to produce differences for young adults may produce striking results for older bilinguals (e.g., Gold et al., 2013)"*, (p. 503).

Describing the differences as *striking* provides another example of the tendency to ignore, or at least undervalue, the importance of strong behavioral evidence. In their first experiment 15 older bilinguals were compared to 15 older monolinguals in a color-shape switching task without any concurrent brain imaging. There was a significant Condition x 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). This lays the groundwork for the main experiment where both younger and older adults were recruited and the behavioural task took place when 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 hardly striking. The global RT differences for the young-adult bilinguals and young-adult monolinguals are nearly identical and clearly not significant. More problematic are the global switch costs between the two groups of older adults. The difference between the means for older bilinguals (M = 14.1%) and older monolinguals (M = 23.0%) does not meet the conventional .05 level of significance, $t(38)$ = 1.97, P = 0.056. 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. This is a situation that would benefit greatly from an exact replication.

5.4. Abutalebi, Della Rosa, Green, Hernandez, Scifo, Keim, Cappa, & Costa (2012) [3].

As described in Section 3 Abutalebi et al. derived both neural and behavioral contrasts between congruent and incongruent trials in a flanker task. There were no differences in flanker effects between a group of German-Italian bilinguals and a group of Italian monolinguals in the first of two blocks. However, in Block 2 the bilinguals reduced their flanker effect by 36 ms (95% CI = 21 to 53) compared to only 11 ms (95% CI = -12 to 41) for the monolinguals. The 36 ms reduction was significant, but the 11 ms reduction was not. This statistical evidence led Abutalebi et al. to conclude that bilinguals "…are better able to adjust to conflict, hence, to adapt to conflicting situations" (p. 2085). However, a stronger test of the hypothesis that bilinguals are "better able to adjust to conflict" would directly test if the 36 ms improvement for the bilinguals was significantly greater than the 11 ms improvement for the monolinguals. Given the specification of the number of participants (17 bilinguals, 14 monolinguals) and the two 95% CI's, one can compute the standard deviations of each group, the pooled-variance estimate, and the independent groups t-value. It is $t(29)$ = 1.48, p = 0.15. By the conventions of null hypothesis statistical testing the two groups do not statistically differ in the degree of improvement from Block 1 to Block 2. The group ns and the t value also enable the computation of a Bayes Factor, viz., 1.52 [29]. This indicates that given the actual data obtained by Abutalebi et al. the null is about 1.5 times more likely to be true than the alternative. In summary, when more appropriate statistical tests are used the behavioral and neural results are no longer in alignment because once again there is no clear evidence for a bilingual advantage in actual performance.

6. Ambiguity in the valence of neural measures: Is more better or worse?

The interpretation of neural differences is very risky in the absence of behavioral data that show the same pattern of significant differences [16]. When alignment problems occur the *inherent*

ambiguity of many neural differences is exposed. One type of ambiguity, referred to here as *valence ambiguity*, occurs if increasing neural scores are interpreted as having a positive effect on performance by one researcher and as having a negative effect by another. Another type of ambiguity arises when differences in say, the same cortical region of interest, are interpreted as reflecting different *kinds* of processing - e.g., executive functioning versus associative learning. Differences in *kind ambiguity* are the topic of Section 7.

In considering behavioral measures like speed or accuracy in choice RT tasks it seems to be inherently the case that individuals or groups who are faster and/or more accurate enjoy a performance advantage over those who are slower and/or less accurate[1]. In contrast, neural measures are often more ambiguous with respect to their interpretation. To take one example Paap and Liu [17] challenged the interpretation of the language-group differences observed by Moreno, et al. [30] in the N400 component of the ERP during sentence grammaticality judgments for sentences that were syntactically correct, but semantically anomalous. The N400 is generally assumed to index difficulty in semantic integration during sentence processing. Moreno et al interpreted the larger N400 components in bilinguals as a bilingual *advantage* in conflict resolution while Paap and Liu argued that larger N400s are indicative of a bilingual *disadvantage* because the larger N400s on the semantically anomalous sentences indicated that the bilinguals were less able to filter out the task-irrelevant semantics.

The Abutalebi et al. study presents another case of valence ambiguity in its conclusion that bilinguals adapt better to conflicting situations because "…they seem to require less ACC activity to outperform monolinguals" (p. 2084), and "…require fewer neural resources to monitor cognitive conflict" (p. 2085). One question is whether a measure of neural activity (or, in this case, the magnitude of the difference in activity between congruent and incongruent trials) reflects the amount of neural processing required in performing a neural "calculation" or the magnitude of the output of the calculation.

The description provided by Abutalebi et al. seems to imply that the measure reflects the amount of neural processing required to register and resolve the conflict and that bilinguals have a more finely tuned ACC that can accomplish the same control with less activity. Even if it is assumed that the measure reflects the amount of neural processing, the decreased activity does not inevitably lead to the conclusion that bilinguals have greater efficiency. One alternative is that the bilingual group has off-loaded the conflict resolution from the ACC to some other region. This is not unrealistic because as shown by Paap and Sawi [5] and others [22,31,32] there is very little convergent validity between the flanker effect and measures of interference from other nonverbal interference tasks. This suggests that the conflict resolution in the flanker task does not always involve general EF, but also relies on task-specific mechanisms for conflict resolution. Thus, if the ACC is involved in general conflict monitoring, then the decrease in ACC activity for bilinguals may reflect not a gain in efficiency in the ACC, but a shift to a more task-specific strategy for handling conflict in the flanker task that is regulated in a different region.

One must also consider the possibility that neural activity reflects the code (i.e., the magnitude) of what is being computed, not the amount of local processing required for the computation. If the ACC is responding to the amount of conflict detected (or the intensity of the specified control

---

[1] Speed-accuracy tradeoffs can complicate the interpretation, but the point remains that within each measure faster is better and more accurate is better.

"commanded" by the ACC as described by Shenhav, Botvinick, and Cohen [33]), then the Abutalebi et al. results would mean that bilinguals were registering less conflict and/or specifying less control than the monolinguals. This does not seem conducive to superior conflict resolution if control is needed to optimize performance (or expected value) across all types of trials. In the limiting case an individual showing no differences in ACC activity between incongruent and congruent trials should be performing poorly on the incongruent trials - unless, of course, the conflict was being efficiently resolved by a task-specific mechanism rather than general inhibitory control. A shift from controlled processing to more automatic processing is consistent with Chein and Schneider's [34] theory of skill acquisition and their experiments show diminishing activity in the ACC as practice progresses.

7. Ambiguity in kind: Is the ACC the citadel of inhibitory control or something else?

Yet another complication is that the ACC activity is modulated by many different other factors such as reward processing, pain processing, performance monitoring, value encoding, decision making, emotion, and motivation. Shenhav, et al. [33] accommodate these diverse findings by proposing a model of ACC function they call the expected value of control (EVC) model. The model assumes that the ACC integrates a wide range of information in order to estimate the EVC and then specifies both the identity and intensity of the choice that maximizes the estimated EVC. The specified intensity acts as a command that is to be implemented by the appropriate regulatory structures which for interference tasks is likely to include the lateral prefrontal cortex. For present purposes the important point is that the EVC model views conflict monitoring as just one component of a more complex control system instantiated in the ACC. This reinforces the point that it is risky to interpret differences in ACC activity as changes in the amount of conflict monitoring - especially in the absence of concurrent changes in behavior.

The EVC model assumes that conflict monitoring is just one piece in the overall puzzle of ACC functioning. Other researchers [35,36] have escalated the interpretation problem by arguing that the neural differences in the ACC during the performance of the flanker task are completely unrelated to cognitive control. These arguments focus on the congruency sequence effects that are sometimes referred to as "conflict adaptation effects" or the "Gratton effect" [37]. The Gratton effect is defined as a reduction in the magnitude of the interference effect when the previous trial is incongruent rather than congruent. The most influential explanation assumes that the detection of conflict on an incongruent trial leads to increased attention to the task-relevant target and increased inhibition of the task-irrelevant flankers and that these changes in control carry over to the next trial [38]. If that trial is also incongruent there will be a facilitation, but if it is congruent there may be a cost because the influence of the supporting flankers will be suppressed.

There are alternative accounts of the Gratton effect that eschew any role of cognitive control and instead account for the pattern of results based on either: (1) feature integration and exact stimulus repetitions [39] or (2) a more complicated associative learning mechanism where event files can also be primed by flanker repetition [35] or (3) through contingency learning [36,40]. Trying to adjudicate between these opposing accounts of the Gratton effect is probably premature, but Schmidt and Weisman's [40] observation is highly relevant to the present discussion regarding the ambiguity of neural differences in the ACC between congruent and incongruent trials: all of the evidence suggesting that ACC signals presence of conflict used paradigms that contained feature integration or contingency learning confounds and consequently brain regions posited to underlie conflict processing as an explanation for Gratton effect may instead reflect learning and memory processing.

8. Conclusion.

The hypothesis that coordinating multiple languages enhances behavioral measures of EF remains contentious. Skepticism is warranted because studies using multiple measures of EF, large numbers of participants, and highly fluent bilinguals usually report no differences at all. Furthermore the commonly used nonverbal interference tasks appear to have serious psychometric problems associated with convergent validity. At first look this may appear to be a situation where converging neuroscience evidence might clarify an inconsistent behavioral database. Although measures of brain activity may often serve this important function we argue that they have a more limited role when the primary purpose is to adjudicate the presence or absence of a performance advantage between two populations. Because neuroscience data are often susceptible to contrasting interpretations with respect to valence or kind, providing consistent and compelling behavioral evidence will be a critical first step toward ascertaining the correct interpretation of the neuroscience data. As we have shown in our selective review the pattern of behavioral differences do not align with the pattern of neural differences. Consequently, the observation of neural differences (while interesting in their own right) have not contributed to resolving the basic question of whether or not there are bilingual advantages in EF.

Acknowledgments

Conflict of Interest

All authors declare no conflicts of interest in this paper.

References

1. Miyake A, Friedman NP (2012) The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Curr Dir Psychol Sci* 21: 8-14.
2. Green DW (1998) Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1: 67-81.
3. Abutalebi J, Della Rosa PA, Green DW, et al. (2012) Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cereb Cortex* 22: 2076-2086.
4. Paap KR, Greenberg ZI (2013) There is no coherent evidence for a bilingual advantage in executive processing. *Cogn Psychol* 66: 232-258.
5. Paap KR, Sawi O (2014) Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology* 5: 962.
6. Paap KR, Johnson HA, Sawi O (2014) Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology* 26: 615-639.
7. Paap KR, Darrow J, Dalibar C, et al. (2014) Effects of script similarity on bilingual advantages in executive control are likely to be negligible or null. *Frontiers in Psychology* 5:1539.

8. Dunabeitia JA, Hernandez JA, Anton E, et al. (2014) The inhibitory advantage in bilingual children revisited: myth or reality? *Exp Psychol* 61: 234-251.

9. Anton E, Dunabeitia JA, Estevez A, et al. (2014) Is there a bilingual advantage in the ANT task? Evidence from children. *Front Psychol* 5: 398.

10. Gathercole VC, Thomas EM, Kennedy I, et al. (2014) Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Front Psychol* 5: 11.

11. Bialystok E, Craik FI, Klein R, et al. (2004) Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychol Aging* 19: 290-303.

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

13. Costa A, Hernandez M, Costa-Faidella J, et al. (2009) On the bilingual advantage in conflict processing: now you see it, now you don't. *Cognition* 113: 135-149.

14. Bialystok E (2011) Coordination of executive functions in monolingual and bilingual children. *J Exp Child Psychol* 110: 461-468.

15. Kroll JF, Bialystok E (2013) Understanding the Consequences of Bilingualism for Language Processing and Cognition. *Journal of Cognitive Psychology* 25: 497-514.

16. Paap KR (2014) The role of componential analysis, categorical hypothesizing, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology* 26: 242-255.

17. Paap KR, Liu Y (2014) Conflict resolution in sentence processing is the same for bilinguals and monolinguals: The role of confirmation bias in testing for bilingual advantages. *Journal of Neurolinguistics* 27: 50-74.

18. Bakker M, van Dijk A, Wicherts JM (2012) The rules of the game called psychological science. *Perspectives in Psychological Science* 7: 543-554.

19. Button KS, Ioannidis JP, Mokrysz C, et al. (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14: 365-376.

20. Francis G (2012) Publication bias and the failure of replication in experimental psychology. *Psychon Bull Rev* 19: 975-991.

21. Engel de Abreu PM, Cruz-Santos A, Tourinho CJ, et al. (2012) Bilingualism enriches the poor: enhanced cognitive control in low-income minority children. *Psychol Sci* 23: 1364-1371.

22. Kousaie S, Phillips NA (2012) Conflict monitoring and resolution: Are two languages better than one? *Brain Research* 1446: 71-90.

23. Morton J, Harper S (2007) What did Simon say? Revisiting the bilingual advantage. *Developmental Science* 10: 719-726.

24. Carlson SM, Choi HP (2009, April) Bilingual and biculural: Executive function in Korean and American children. Biennial meeting of the society for research in child development. Denver, Colorado.

25. Bialystok E, Craik FI, Grady C, et al. (2005) Effect of bilingualism on cognitive control in the Simon task: evidence from MEG. *Neuroimage* 24: 40-49.

26. Bialystok E, Craik FI, Luk G (2012) Bilingualism: consequences for mind and brain. *Trends in Cognitive Science* 16: 240-250.

27. Luk G, Anderson JA, Craik FI, et al. (2010) Distinct neural correlates for two types of inhibition in bilinguals: response inhibition versus interference suppression. *Brain Cogn* 74: 347-357.

28. Gold BT, Kim C, Johnson NF, et al. (2013) Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *J Neurosci* 33: 387-396.

29. Rouder JN, Speckman PL, Sun D, et al. (2009) Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review* 16: 225-237.

30. Moreno S, Bialystok E, Wodniecka Z, et al. (2010) Conflict Resolution in Sentence Processing by Bilinguals. *J Neurolinguistics* 23: 564-579.

31. Salthouse TA (2010) Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. *Brain Cogn* 73: 51-61.

32. Keye D, Wilhelm O, Oberauer K, et al. (2009) Individual differences in conflict-monitoring: testing means and covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychol Res* 73: 762-776.

33. Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* 79: 217-240.

34. Chein JM, Schneider W (2005) Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Brain Res Cogn Brain Res* 25: 607-623.

35. Davelaar EJ (2013) When the ignored gets bound: sequential effects in the flanker task. *Front Psychol* 3: 552.

36. Schmidt JR, De Houwer J (2011) Now you see it, now you don't: controlling for contingencies and stimulus repetitions eliminates the Gratton effect. *Acta Psychol (Amst)* 138: 176-186.

37. Gratton G, Coles MG, Donchin E (1992) Optimizing the use of information: strategic control of activation of responses. *J Exp Psychol Gen* 121: 480-506.

38. Botvinick MM, Braver TS, Barch DM, et al. (2001) Conflict monitoring and cognitive control. *Psychol Rev* 108: 624-652.

39. Mayr U, Awh E, Laurey P (2003) Conflict adaptation effects in the absence of executive control. *Nat Neurosci* 6: 450-452.

40. Schmidt JR, Weissman DH (2014) Congruency sequence effects without feature integration or contingency learning confounds. *PLoS One* 9: e102337.