

Commentary

Beyond Panglossian Optimism: Larger N2 Amplitudes Probably Signal a Bilingual Disadvantage in Conflict Monitoring

Kenneth R. Paap^{1,*}, Oliver M. Sawi^{1,2}, Chirag Dalibar¹, Jack Darrow^{1,3}, Hunter A. Johnson¹

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: In this special issue on the brain mechanisms that lead to cognitive benefits of bilingualism we discussed six reasons why it will be very difficult to discover those mechanisms. Many of these problems apply to the article by Fernandez, Acosta, Douglass, Doshi, and Tartar that also appears in the special issue. These concerns include the following: 1) an overly optimistic assessment of the replicability of bilingual advantages in behavioral studies, 2) reliance on risky small samples sizes, 3) failures to match the samples on demographic characteristics such as immigrant status, and 4) language group differences that occur in neural measures (i.e., N2 amplitude), but not in the behavioral data. Furthermore the N2 amplitude measure in general suffers from valence ambiguity: larger N2 amplitudes reported for bilinguals are more likely to reflect poorer conflict resolution rather than enhanced inhibitory control.

Keywords: bilingualism; executive Function; ERP; inhibitory control; conflict resolution

1. The research database viewed through rose-colored glasses

In our contribution to this special issue [1] we discussed six reasons why it will be very difficult to discover the brain mechanisms underlying the cognitive benefits of bilingualism. Many of these problems apply to the article by Fernandez, et al. [2] that also appears in the special issue.

In their introduction Fernandez et al. present an untempered view that bilingualism enhances brain structures and function involved in EF. In the context of our discussion of the “alignment problem” ([1] Section 5) and “valence ambiguity” ([1] Section 6) in the neuroscience investigations of the bilingual advantage we feel a more circumspect perspective is warranted. Valence ambiguity

refers to the surprisingly common disagreements regarding whether a larger neural measure is “good” or “bad” with respect to its influence on actual performance. The alignment problem refers to the fact that in many tests for bilingual advantages the language group differences in the neural data do not align with the differences in performance. We argued that alignment between the neural and behavioral results is especially critical in situations of valence ambiguity. As discussed below, there is considerable ambiguity regarding whether larger N2 amplitudes that are the focus of the Fernandez et al. study reflect more or less effective cognitive control.

In contrast to the unequivocal perspective on the relevant cognitive neuroscience Fernandez et al. do acknowledge inconsistencies in the behavioral results. However, these inconsistencies have accumulated to levels that catch many observers by surprise and the analysis presented in our Section 2 and elaborated upon in Paap and Greenberg [3] and Paap, Johnson, and Sawi [4] lead us to question if the phenomenon actually exists. This skepticism rests heavily on the fact that large-n studies overwhelmingly yield null results.

2. Implications if bilingual advantages were restricted to auditory tasks

Before turning to those aspects of the Fernandez et al. experiment that concern us we agree that directly comparing visual and auditory tests of executive functioning (EF) in the same study is innovative and potentially fertile. If bilingual advantages occur in preschoolers, then those advantages must be the product of managing the production and comprehension of speech as they have yet to learn to read and write. Similarly, older children and adult bilinguals are likely to produce more spoken language than written and to do so at a faster speed—a rate that should require more active coordination of the two languages. Thus, we find this research question very worthwhile. If bilingual advantages in performance were larger or more consistently observed in auditory compared to visual tasks, this may lead to an additional inference that Fernandez et al. do not discuss: namely, that the advantages may not reflect an enhancement in general EF. The reason that researchers use non-linguistic tasks to test for bilingual advantages in EF is to make sure that any group differences in performance reflect domain-general differences in ability and not linguistic differences that may have been honed by the unique experiences associated with managing two languages. Applying a similar logic, if the components of EF (e.g., updating, inhibitory control, switching) are assumed to be general and not modality-specific, then one might be reluctant to attribute language-group differences restricted to the auditory modality to differences in a modality-free EF. Differences between bilinguals and monolinguals in auditory versus visual processing may, of course, be very interesting in their own right.

3. The risks of testing small numbers of participants

One of the concerns we have raised repeatedly [3–7] is the use of small numbers (n) of participants in the language groups used to test for bilingual advantages. This is especially prevalent in studies that include both behavioral and neuroscience measures. In the RT data Fernandez et al. have only 6 monolinguals and 11 bilinguals. If one generously assumes that the effects of bilingualism are of medium size and a standard alpha of 0.05 (two-tailed), then the power for an independent-groups t-test is only 0.11. Thus, the estimated probability of rejecting a medium effect size in the RT data was only 11%.

For the ERP analyses of the auditory task there were 13 monolinguals and 13 bilinguals yielding a power of 0.23 for the same scenario (medium effect size, $\alpha = 0.05$, two-tailed). Are the significant differences between the group means of the N2 amplitudes on the auditory NoGo trials all the more compelling under these circumstances? Beyond mere optimism, we believe that additional analyses probing additional questions are warranted for a better understanding. A Bayes Factor ratio [8] is the probability of the alternative hypothesis given the data over the probability of the null given the data. The BF ratio for the audio task is 3.83. Based on Jeffreys' (1961) guidelines for interpreting BFs: the obtained BF escapes the 1 to 3 range ("worth no more than a brief mention"), and lands on the 3 to 10 side of the fence that the guidelines suggest indicates "substantial evidence" for the alternative hypothesis.

Accepting that the N2 amplitudes are greater for the bilinguals on the auditory NoGo trials, what may have caused those differences? Fernandez et al. conclude that they are caused by bilingual experiences in spoken language that do not generalize to the visual task. However, small sample sizes in tests between two naturally occurring populations are difficult to match with respect to other characteristics [9,10]. To their credit Fernandez et al. showed that their groups did not differ on measures of SES and general fluid intelligence, but the groups were not matched on immigrant status with 11 of the 18 bilinguals emigrating from Central or South America.

4. Valence ambiguity in N2 amplitude

Are larger N2 amplitudes on the NoGo trials really indicative of better cognitive control? Fernandez et al. do not justify the assumption that bigger is better in their 2014 article, but in an earlier article [11] they cite a study by Falkenstein et al. [12] who concluded that larger N2 amplitudes reflect better inhibitory control because participants with high false-alarm rates on the NoGo trials also exhibited smaller N2 amplitudes.

Although this evidence and logic should be given just consideration it is fair to say that it is not the consensus view. Some of the most compelling evidence comes from developmental studies showing that N2 amplitude in the Go/NoGo task declines over the span of 7 to 16 years even when potential physical artifacts are taken into account [13]. Furthermore, regression analyses showed that N2 amplitudes on NoGo trials significantly predicted the magnitude of Stroop interference- a common behavioral test of inhibitory control. Similarly, Espinet, Anderson, and Zelazo [14] showed that 3 to 4.5 year-old-children who can pass the dimensional change card sort (DCCS) task have significantly smaller N2 amplitudes during the post-switch phase of the task compared to those who perseverate and fail the test. The last results are compelling because the neural results (N2 amplitude differences between two groups) align with the behavioral performance results (success versus failure on the DCCS) and as argued in Paap et al. [1] this is critical to resolving the "valence ambiguity" often associated with neural measures. In this case it strongly suggests that a smaller N2 amplitude reflects superior performance. The valence ambiguity of the N2 amplitude component is strikingly exhibited by Kousaie and Phillips [15] who predicted a bilingual advantage in the form of greater N2 amplitudes in their introduction only to reverse their interpretation when bilinguals showed smaller N2 amplitudes in their Stroop task. Whatever the interpretation of the N2 amplitude differences in the Stroop task there were no differences in performance in either the flanker or Simon task showing that the language-group difference itself is inconsistent across tasks.

5. Selecting a task/measure

It is customary, but somewhat puzzling, that researchers testing for bilingual advantages usually provide only a cursory rationale for the selection of their tasks and measures. Paap and Greenberg [3] suggested that researchers should start with a theoretical framework that specifies the aspects of EF that they assume are extensively exercised in managing two languages and consequently those components of EF that should be enhanced and yield bilingual advantages. Having selected the critical constructs *a priori* one should then select tasks/measures that have demonstrated convergent and discriminant validity. Furthermore, given that most of the standard tasks for measuring the monitoring and inhibitory control components of EF have low levels of convergent validity [7] it is preferable for a study design to include two measures of the same component of EF (derived from two different tasks) so that convergent validity can be demonstrated within the study.

How should the choice of the Go/NoGo task be evaluated? Across a variety of taxonomies [16–19] the NoGo task is assumed to involve the inhibition of a prepotent response, but task analyses will not necessarily identify the tasks that are affected by common components of EF. A more empirical approach is to use latent variable analyses to identify measures that load on the same psychological construct. In their seminal study of inhibition Friedman and Miyake [20] considered three categories of interference tasks, but the best model did not empirically separate response inhibition (*viz.*, stop signal, antisaccade, Stroop) from resistance to distractor interference (*e.g.*, flanker task). Unfortunately the Go/NoGo task was not included in Friedman and Miyake's design.

In a very large meta-analysis of self control measures Duckworth and Kern [21] report an average correlation (based on 131 correlations and a total N of 4,855 participants) between the Go/No-Go task and other tasks assumed to reflect EF of $r = 0.16$. The Go/NoGo task appears in the middle of the pack of the EF tasks with, for example, the flanker task having a higher average correlation ($r = 0.19$) and the Stop-signal a lower correlation ($r = 0.11$). Thus in the grand scheme of things performance on the Go/No-Go task is associated other “measures” of EF, but the magnitude of the association is unimpressive and glosses over the various ways that performance is measured in the Go/No-Go task.

A more detailed analysis of performance on Go versus NoGo trials is presented by Votruba and Langenecker [22]. Their Parametric Go/No-Go (PGNG) task is similar to one strain of Go/NoGo tasks (those where the NoGo is contingent upon whether a target is repeated or alternated), but differs in many ways (modality, number of targets) from the one used by Fernandez et al. Nonetheless, the Votruba and Langenecker study provides an interesting analysis of how proportion correct (PC) on the Go trials and the NoGo trials correlate with measures of EF derived from other tasks. A typical outcome is that the correlation between PC on the Go trials and Stroop interference ($r = + 0.29$) is as strong as or stronger than the correlation with the NoGo trials ($r = + 0.19$). The implication is that if the bilinguals in the Fernandez et al. (2014) do have better inhibitory control than the monolinguals, then those advantages might have appeared on both Go trial performance (where there were no differences in performance) and No Go trials (where no data is reported on the grounds that there were very few false alarms).

6. Conclusion

Fernandez et al. are pursuing an interesting research question in asking if the benefits of bilingualism on nonlinguistic tasks might be limited to auditory tasks. However, we recommend that this work, and work on the bilingual advantage in general, embrace the need for adequately powerful sample sizes, avoid obvious confounds between the language groups, and select tasks that enable any valence ambiguity in the neural measure (N2 amplitude in this case) to be adjudicated by differences in performance that are faster, more accurate, or both.

References

1. Paap KR, Sawi OM, Dalibar C, et al. (2014) The Brain Mechanisms Underlying the Cognitive Benefits of Bilingualism may be Extraordinarily Difficult to Discover. *AIMS Neuroscience* 1: 245-256.
2. Fernandez M, Acosta J, Douglass K, et al. (2014) Speaking Two Languages Enhances an Auditory but not a Visual Neural Marker of Cognitive Inhibition. *AIMS Neuroscience* 1: 145-157.
3. Paap KR, Greenberg ZI (2013) There is no coherent evidence for a bilingual advantage in executive processing. *Cogn Psychol* 66: 232-258.
4. Paap KR, Johnson HA, Sawi O (2014) Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *J Cogn Psychol* 26: 615-639.
5. Paap KR (2014) The role of componential analysis, categorical hypothesizing, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *J CognPsychol* 26: 242-255.
6. 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. *J Neurolinguistics* 27: 50-74.
7. 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.
8. Rouder JN, Speckman PL, Sun D, et al. (2009) Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bull Rev* 16: 225-237.
9. Valian V (2014) Bilingualism and cognition. *Biling Lang Cogn* 18: 3-24.
10. Paap KR (2014) Do many hones dull the bilingual whetstone? *Biling Lang Cogn* 18: 41-42.
11. Fernandez M, Tartar JL, Padron D, et al. (2013) Neurophysiological marker of inhibition distinguishes language groups on a non-linguistic executive function test. *Brain Cognition* 83: 330-336.
12. Falkenstein M, Hoormann J, Johnsein J (1999) ERP components in Go/NoGo tasks and their relation to inhibition. *Acta Psychol* 101.
13. Lamm C, Zelazo PD, Lewis MD (2006) Neural correlates of cognitive control in childhood and adolescence: disentangling the contributions of age and executive function. *Neuropsychologia* 44: 2139-2148.

14. Espinet SD, Anderson JE, Zelazo PD (2012) N2 amplitude as a neural marker of executive function in young children: an ERP study of children who switch versus perseverate on the Dimensional Change Card Sort. *Dev Cogn Neurosci* 2 Suppl 1: S49-58.
15. Kousaie S, Phillips NA (2012) Conflict monitoring and resolution: Are two languages better than one? *Brain Res* 1446: 71-90.
16. Kornblum S, Hasbroucq T, Osman A (1990) Dimensional Overlap: Cognitive Basis for Stimulus-Response Compatibility - A Model and Taxonomy. *Psychol Rev* 97: 253-270.
17. Nigg JT (2000) On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychol Bulletin* 126: 220-246.
18. Bunge SA, Dudukovic NM, Thomason ME, et al. (2002) Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. *Neuron* 33: 301-311.
19. 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.
20. Friedman NP, Miyake A (2004) The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *J Exp Psychol Gen* 133: 101-135.
21. Duckworth L, Kern ML (2011) A meta-analysis of the convergent validity of self-control measures. *J Res Personal* 45: 259-268.
22. Votruba KL, Langenecker SA (2013) Factor structure, construct validity, and age- and education-based normative data for the Parametric Go/No-Go Test. *J Clin Exp Neuropsychol* 35: 132-146.

© 2015, Kenneth R. Paap, et al., licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)