

# Effects of Bilingualism on Sustained Attention and Inhibition: A Bayesian Enquiry

Moulshree Rana (moulshree20@iitk.ac.in)

Department of Cognitive Science, Indian Institute of Technology  
Kanpur, India

Ark Verma (arkverma@iitk.ac.in)

Department of Cognitive Science, Indian Institute of Technology  
Kanpur, India

## Abstract

This study examines the general claim that bilingualism leads to a facilitatory effect on cognitive control. Repeatedly resolving conflict between simultaneously active representations is thought to spill over into other domains involving conflict resolution. Recent literature indicates that the effects of bilingualism on executive functions need examination with a more comprehensive characterization of bilingualism and the use of multiple measures of executive control (Backer & Bortfeld, 2021; K. R. Paap & Greenberg, 2013). Here, we operationalize bilingualism as a set of continuous variables related to language knowledge and use. Next, we employ Bayesian regression analyses to assess the evidence for the null i.e., the lack of an effect of bilingualism. We aimed to address arguments in favor of an advantage that appeal to the measurement of bilingualism, task-specificity of the effect, and the methodological issues that exist with widely used tasks such as the Simon, Stroop or Flanker (K. R. Paap, Anders-Jefferson, Zimiga, Mason, & Mikulinsky, 2020). In this study, we assess the effects of bilingualism under a newly specified mechanism of attentional control (Bialystok & Craik, 2022), specifically in sustained attention. We also administer new tasks, developed to be psychometrically sound and an improvement to existing measures of attentional control by Draheim, Tsukahara, Martin, Mashburn, and Engle. Two sustained attention tasks, along with two versions of the Flanker task were administered. The null model was the best model (with the greatest posterior probability) for all tasks. Bilingualism-related characteristics failed to show reliable influence for both sustained attention tasks. Even for "improved measures" less susceptible to methodological flaws related to RT impurity and processing confounds, the best model was the null model. The results imply that the source of null effects is not the inadequate choice of inhibition as an explanatory mechanism. We conclude that bilingualism does not have coherent and consistent effects on cognitive control (specified as either inhibition or sustained attention) and the lack of an effect is not specific to the type of conflict involved in a task or its reliance on reaction times.

**Keywords:** bilingualism; bayesian; inhibition; sustained-attention; attentional control

## Introduction

The "Bilingual Advantage in EF" hypothesis refers to the idea that bilinguals' special practice with inhibition of lexical items grants them better ability in inhibiting in non-linguistic stimuli. For a bilingual, representations are simultaneously active in both the target language and the non-target language (Marian & Spivey, 2003), where the target language is determined by the context, such as the languages the listener understands. Models explaining how bilinguals are able to produce correct utterances without many intrusion errors call upon inhibition as the mechanism. For instance, Green's In-

hibitory control model (Green, 1998), and the Adaptive Control Hypothesis model (Green & Abutalebi, 2013). The idea is that bilinguals inhibit lexical items from the non-target language. It is this capacity of inhibition which forms the link between being bilingual and performing better on commonly administered executive functioning tasks such as the Stroop, Simon or Flanker (K. R. Paap et al., 2020). Since bilinguals have almost continuous practice with inhibiting non-target lexical items, it spills over into better performance for other tasks which also involve inhibiting irrelevant information. This hypothesis of a facilitatory effect has been amply explored, with findings converging to a null effect, with effect sizes close to 0 (Lehtonen et al., 2018). A feature of this bloc of studies examining effects of bilingualism on performance of inhibition tasks is that although findings converge to a null effect, some positive findings appear (Bialystok, Craik, Klein, & Viswanathan, 2004; Prior & MacWhinney, 2010). Researchers have tried to explain the occasional positive findings as originating from possible confounds (where variables other than bilingualism are causing a difference to appear, not bilingualism itself) or are instances of Type 1 error.

## Attentional Control replaces Inhibition

An alternative proposal is that the mix of positive and null findings is not understood under an inhibition account, but rather the mixed findings can be reconciled under a framework of "attentional control". Bialystok and Craik, as a response to null findings have proposed replacing the previous hypothesis involving inhibition with one based on a hierarchical construct of attentional control. Attentional control refers to a broad term covering several cognitive processes used to accomplish goals. These processes include goal maintenance, interference suppression, and switching of processing resources. Attentional control selects processes for current needs and goals and reallocates resources to these processes (Bialystok & Craik, 2022). This construct is proposed to be a better explanation for findings in recent studies by overcoming the limitations of the inhibition account, and argues that the conceptualization of inhibition makes two mistakes: 1. Tasks employed to measure inhibition are considered a proxy for inhibition itself (i.e. reducing inhibition to performance on tasks like Stroop, Flanker). 2. Categorizing tasks as inhibition tasks even when they involve different abilities, namely interference suppression and response inhibition. In

addition, a claim is that bilinguals would typically perform better than monolinguals on tasks involving interference suppression than those involving response suppression. Another argument calls upon studies by Costa, Hernández, Costa-Faidella, and Sebastián-Gallés which proposes a hybrid account of language control involving both selecting the target language and inhibiting the non-target language. Bialystok and Craik claim three alternate ways an advantage could appear (which an inhibition account does not explain). First, a bilingual advantage appears as a general benefit in processing speed or better efficiency of the executive system. Second, benefits would not be seen on tasks such as Simon or Flanker because these tasks have little EF demands and minimal conflict. A consequence is that bilingual advantages are instead to be expected on tasks which pose a greater demand on executive function. Finally, bilinguals supposedly show fewer atypically long RTs i.e. display fewer attentional lapses, which could be a marker for better processing speed. In this study we test this claim directly. We examine the relationship between bilingualism and performance on sustained attention tasks where the dependent variable (RT20) captures lapses of attention. The sustained attention tasks used also tentatively demand greater EF resources. The use of these tasks can also inform about other theories calling upon attention such as the hierarchical attentional model proposed by (Chun, Golomb, & Turk-Browne, 2011). Here, Limited capacity, Selection, Modulation and Vigilance are described. Limited capacity refers to focus on relevant information over other information. Selection refers to bias attention towards one candidate over another. Modulation refers to the processing of attended information. Vigilance refers to the extent of modulating information over some period of time. A sustained attention task such as the SACT, would involve demands, not only on selection (like the Flanker) but also modulation and vigilance (Draheim et al., 2021).

### **The inadequacy of Inhibition tasks**

A majority of studies examining the “Bilingual Advantage” employ tasks like the Simon, Flanker or Stroop which purport to measure inhibition (K. R. Paap et al., 2020). However, there is a growing concern regarding the validity of these tasks. This concern originates from the lack of significant or meaningful inter-task correlations, and also notes that these task showcase poor psychometric properties. The lack of established validity limits researchers who use these tasks to make strong theoretical conclusions. Another concern is that these tasks are not suited for studies examining individual differences in EF (which is what studies treating bilingualism on a continuum want to do). Researchers are divided in the reasons to explain these poor intertask-correlations and psychometric properties. One of the arguments is that these tasks do not correlate because the construct they are measuring (inhibitory control) is not unitary. The mechanism for resolving conflict is task-specific instead of a domain general inhibitory control system. Rouder and Haaf found no meaningful relationship between performance

on the Stroop and Flanker tasks despite using a robust hierarchical regression approach. They concluded that these tasks involve task-specific inhibitory mechanisms. Rey-Mermet, Gade, and Oberauer also concluded that there is no unified inhibition factor, based on the findings that when 11 tasks were loaded on a single factor, factor loadings fell below 0.2 for half of the tasks. An alternative approach posits that there is a domain general attentional control system and the poor psychometric properties are a methodological issue. Tasks typically used involve high measurement error. Tasks such as the Flanker or Stroop have small effect sizes and the ratio of trial-level noise to individual variation is high. The difference scores calculated in these tasks are a major issue. These scores have lower reliability than its components. Draheim et al. found that mean incongruent and congruent RTs had better reliability and stronger correlations with other measures compared to the difference scores. Consequently, some studies examining the bilingual advantage have also looked at overall (global) RTs instead of difference scores. Among reviewed studies which looked at global RTs in groups of monolinguals and bilinguals, 81.1% of the studies showed null effects, 16.7% showed a bilingual advantage and 2.3% reported a monolingual advantage (K. Paap, 2019). The interpretation of faster performance for bilinguals (as smaller global RTs) is murky. Notably, a dependent variable which includes only mean RT of incongruent or congruent trials is very impure since processing speed or task fluency remain confounds. This contamination argument is also made for difference scores by Hedge, Powell, Bompas, and Sumner and Draheim et al.. They argue that the performance on tasks that calculate difference scores are contaminated with irrelevant variance originating from speed-accuracy tradeoffs and processing speeds. (Draheim et al., 2021) points out that issues with the Simon and Flanker tasks are primarily due to the use of RT and difference scores in RT.

**The RT problem** Reaction times are contaminated with speed-accuracy tradeoffs and processing speed (Draheim et al., 2021). Although difference scores are calculated to remove influences of processing speed and task fluency, they are still prone to contamination by speed-accuracy tradeoffs. It is also possible that difference scores fail to account for processing speed (Rey-Mermet, Gade, Souza, Von Bastian, & Oberauer, 2019). Processing speed remains a confound and even if we see good inter-task correlations. It remains unclear whether the differences in performance are a result of differences in attentional control or processing speed. Even if the tasks correlate positively, inferences drawn from them remain limited. Draheim et al. suggest that the solution to the various issues with current tasks is to create and use tasks where RT is rendered irrelevant. The suggested tasks do not involve RT and are instead based on measures of accuracy and avoid difference scores altogether. Use of accuracy based measures prevents contamination from processing speed. Draheim et al. argue that the null or mixed findings prevalent in individual difference literature are due to the methodological issues with

current tasks (which are primarily based on contaminated RT reliant scores).

## This study

Based on the problems with the explanatory accounts of the executive function advantages documented in bilingualism research, and the present flaws in the commonly used tasks, in this study we not only want to ask if bilingualism variables can predict performance and which variables are important predictors, but also want to check if bilingualism variables are *good* predictors of EF performance. We aimed to test if the source of conflicting results seen in literature is the reliance on RT-based measures. We can say that the effects of bilingualism are masked because of reliance on RT measures if strong evidence for bilingualism’s effects appears on the RT-irrelevant tasks and not on the RT-based tasks. Another question this study addresses is whether bilingual “advantages” appear on tasks measuring attentional control (here, sustained attention) rather than those which measure only inhibition. To address these questions, we employ two sustained attention tasks (chosen based on recommendations by (Draheim et al., 2021)) and two versions of the flanker task. We also evaluate the inhibition versus attentional control account by comparing the predictions of bilingualism variables for performance on the Sustained-Attention-to-Cue (SACT), Psychomotor Vigilance Task (PVT), Arrow Flanker and Adaptive Flanker tasks.

## Methods

### Participants

54 subjects (11 females, 43 males) completed the experiment in two phases, each of duration between 45 minutes to 1 hour. The two sustained attention tasks were performed in one session and the two flanker tasks in another. The order of the phases was counterbalanced. The protocol for this study was approved by the institute ethics committee. Participants were bilinguals in the age group 16-35 with normal or corrected to normal vision. Participants either received monetary compensation of Rs.100/hr or 1% course credit. Mean age of participants was 21.46 (*min* : 19, *max* : 34). The native language of participants was either Marathi, Hindi, English, Tamil, Gujarati, Telugu, Kannada, Bhojpuri, Malayalam, Urdu or Marwari. Their second and third languages also included Bengali, Sindhi, Punjabi, Assamese, Odia, Russian, French and Arabic. As seen in Table 1, participants were on an average, equally proficient in their two languages, also showcasing balance in use. There is also sufficient variation in the multilingual diversity scores indicating that participants were heterogeneous in their use of multiple languages across contexts.

Table 1: Participant scores of bilingualism-related predictors as measured by LHQ 3

Variable	Mean	S.D.	Min.	Max.
Proficiency $L_1$	0.832	0.122	0.57	1.00
Proficiency $L_2$	0.813	0.106	0.57	1.00
Immersion $L_1$	0.788	0.109	0.50	0.94
Immersion $L_2$	0.739	0.106	0.50	0.90
Dominance Ratio	1.037	0.254	0.63	2.18
MLDS	1.404	0.319	0.90	1.96

## Tasks

### Sustained Attention Tasks

#### Psychomotor Vigilance Task (Unsworth & Robison, 2016)

This task is an RT-based task and is immune to speed-accuracy tradeoffs, since an inaccurate response is not possible. This task consisted of 80 trials where participants were shown five zeros at the center of the screen. The zeros start counting up by 17 every 17ms after a variable duration of time. The wait time was either 2,4,8 or 12 seconds. Participants’s task was to press the spacebar as soon as they think the numbers have started counting up. After their response, their reaction time was displayed on the screen. The dependent variable for this task was the mean reaction time on the slowest 20% of trials.

#### Sustained Attention to Cue Task (Draheim et al., 2021)

This task is an accuracy-based analog of the Psychomotor vigilance task. First, a fixation cross was presented for 2 or 3 seconds, which was followed by a circular cue indicating where the target would appear next. The cue appeared to the left or to the right of the screen equal number of time. The target was a 3x3 array of letters and the participants were instructed to report the central letter. Once the target array appeared, the central letter was masked after 125ms, and the mask stayed for 1 second. The gap between the cue and the target was variable. (It was equally distributed among 2,4,8 and 12 seconds). Before the target appeared, a distracting asterisk flashed at the center of the screen for 300ms. A total of 64 trials were presented excluding practice trials. The dependent variable was participant’s accuracy.

### Flanker Tasks

#### Arrow Flanker (Eriksen & Eriksen, 1974)

Within this task, a string of arrows is presented to participants and the middle arrow is the target. They must press the left (‘Z’) or right (‘/’) keys to indicate the direction in which the arrow is pointing. The central arrow matches its direction with other flanker arrows on congruent trials and mismatches on incongruent trials. Participants completed 144 trials (96 congruent and 48 incongruent). The dependent variable was a difference score i.e. the Flanker Interference effect which was calculated as the difference between mean reaction times for congruent and incongruent accurate trials.

**Deadline Flanker (Draheim et al., 2021)** This is a flanker task with an adaptive procedure that does not rely on RT. A total of 324 trials were presented in 18 blocks of 18 trials (12 congruent and 6 incongruent). Participants must respond within a time limit. A beep sound indicated when the deadline was reached. For blocks 1-6, if participants were accurate on 15 out of 18 trials, the response deadline decreased by 90ms, otherwise decreased by 270ms. The starting deadline was 1050ms. For blocks 6-18, the deadline increased by 90ms or decreased by 30ms. The dependent variable was the deadline duration after 18 blocks i.e the response deadline for the hypothetical 19<sup>th</sup> block. Both types of trials had an equal bearing on the change in deadlines i.e. accuracy on incongruent and congruent trials was not counted separately.

## Demographic and Language Information

### Language History Questionnaire 3.0

The LHQ 3.0 (Li, Zhang, Yu, & Zhao, 2019) consists of 27 questions enquiring about demographic details and language use. The LHQ yields aggregate scores corresponding to proficiency in each language, dominance of each language (dominance ratios between languages), immersion in each language context and a multilingual diversity score (*MLDS*). The following aggregate scores were considered as independent variables: Proficiency L1, Proficiency L2, Immersion L1, Immersion L2, Dominance Ratio, and *MLDS*. (L1 and L2 refer to first and second languages respectively). Demographic information collected included age, gender, current educational level, and education level of parents.

### Data Preparation

A total of 57 participants registered for the study and 54 completed both sessions of the study. 3 participants were excluded from the final analysis due to their failure to complete the LHQ3.

**Outlier removal** Reaction times smaller than 200ms that are too short to reflect task processing were removed for the Arrow flanker and PVT. For the PVT, reaction time did not exceed 10 seconds in any of the trials. For the Arrow flanker task, outliers were removed for reaction times of incongruent and congruent trials separately. Outliers were detected by fitting a t distribution to the data and points within  $\pm 3.5 \times s$  of mean were removed ( $s$  = standard deviation). Participants whose accuracy was more than 3.5 standard deviations below mean were removed for the Flanker tasks. For all tasks, outliers were removed by fitting a t distribution to the data. Data points within  $\pm 3.5 \times s$  of the mean were removed. Due to outlier removal, sample sizes for the four tasks are unequal. Descriptive statistics for the dependent variables in the four tasks are presented in Table 2.

Table 2: Descriptive Statistics for the four tasks.

Task	<i>N</i>	Mean	Std. Dev.	Min.	Max.
SACT	54	84.462	15.956	31.250	100.0
PVT	50	0.592	0.078	0.413	0.767
Arrow Flanker	48	0.044	0.021	0.003	0.083
Adaptive Flanker	45	1.172	0.477	0.470	2.300

## Data Analyses

We built four linear regression models, one for each task with the dependent variables mentioned previously. The aggregate scores obtained from the LHQ were the independent variables, namely Proficiency in L1, Proficiency in L2, Immersion in L1, Immersion in L2, Dominance Ratio and Multilingual Diversity Scores. This is how the relationship between bilingualism and "attentional control" has been traditionally examined. We believe this relationship can be better evaluated using a Bayesian approach (Rana & Verma, 2023).

**Bayesian Multiple Regression** A Bayesian multi-model analysis allows us to evaluate evidence for all possible models from selected predictors, including the null model. We measure the evidence for each model as a Bayes Factor  $BF_m$  which quantifies the change from odds of a model before seeing the data (prior odds) and after observing the data (posterior odds). In our analysis, we start with a flat prior, i.e. a  $Beta(1, 1)$  distribution. For estimates of our parameters (here, regression coefficients), we proceeded with the default *Jeffreys-Zellner-Siow* prior with scaling factor  $r = 0.354$ .

## Results

For all tasks, we evaluated the models on the evidence the data provided for that particular model ( $BF_m$ ). Our aim was also to evaluate which bilingualism characteristics are important predictors of performance on these tasks i.e., which factors have an influence on "attentional control". We can examine the posterior inclusion probabilities for each variable of interest. Inclusion probabilities refer to the probability that a particular factor would be included in a linear model. The updating factor  $BF_{incl}$  records the change from prior to posterior odds of inclusion. The posterior probability for inclusion is defined as the sum of posterior probabilities of models that contain a specific factor (such as L1 Proficiency).

### Sustained Attention tasks

For the Sustained-Attention-to-Cue task (SACT), the best model (model with the greatest posterior probability) was found to be the null model with  $BF_m = 4.31$ .  $BF_m$  for an alternate model containing Immersion in L1 and Multilingual Diversity Scores as the predictors equalled to 3.94. Prior inclusion probabilities equalled to 0.50. After combining evidence for including each predictor, we see that models perform better when bilingualism-related variables are not included (See Figure 1). For all variables, models performed better when

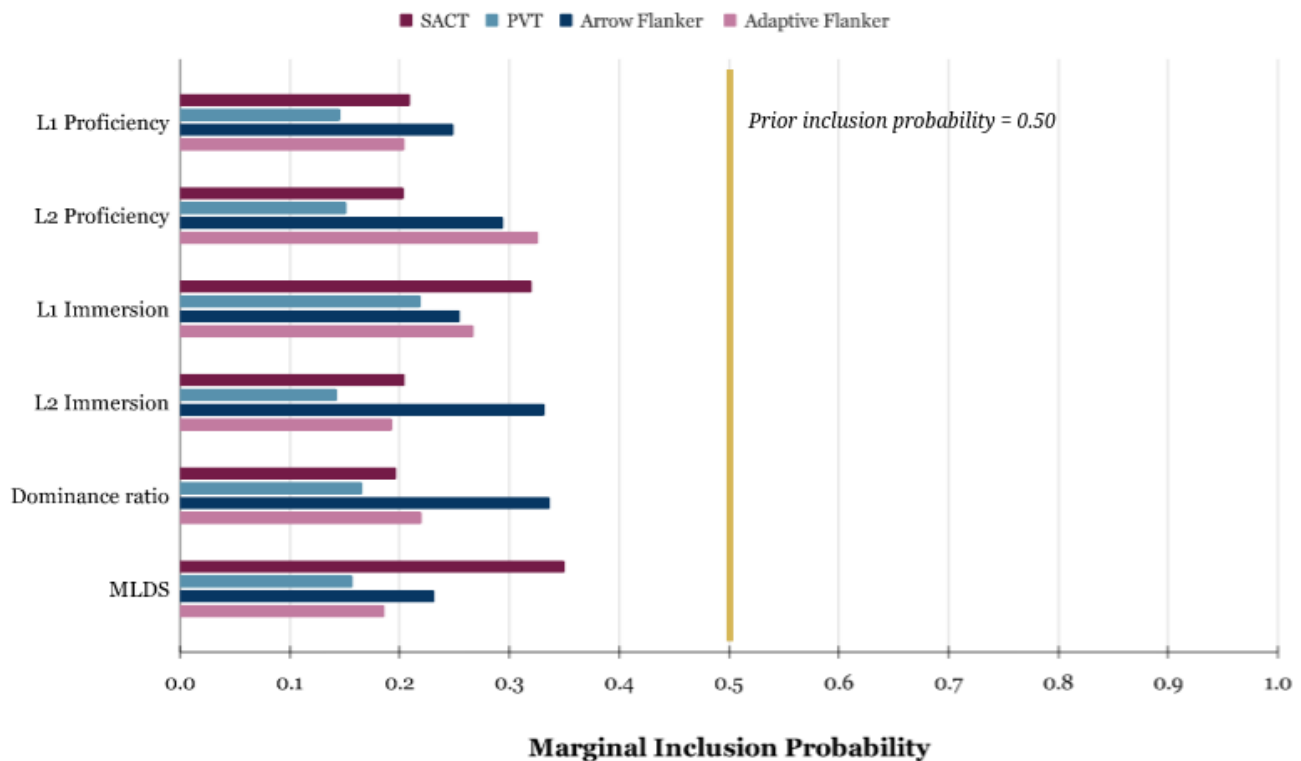


Figure 1: Updated Inclusion Probabilities. Prior inclusion probability = 0.5, represented by the yellow line.

the probability of including them was decreased. We see similar results for the Psychomotor Vigilance Task (PVT). The best model was the null model with a  $BF_m$  greater than 7. For all variables, models predicted performance on PVT better when the variables were not included.

### Flanker Tasks

For both flanker tasks, we found that the best model was the null model.  $BF_m$  for the null model for the Arrow Flanker task equalled 3.717. Compared to the alternate model containing all bilingualism characteristics,  $BF_{10} = 0.128$ . Taking the inverse (measuring evidence for the null),  $BF_{01} = 7.8$  meaning that data are 7.8 more likely if the null were true compared to the alternate model (model including all predictors). For the Adaptive flanker task,  $BF_m$  for the null is 4.2.  $BF_{10}$  comparing the null model and the next best model equalled to 1.1. The data are almost equally likely under a null model and the alternate model (containing only L2 Proficiency as predictor). The posterior probabilities of inclusion for L2 Proficiency, however, provide evidence that models perform worse when *any* of the bilingualism characteristics are inserted in the linear model (including L2 Proficiency).

### Discussion

The current study aimed to address two sets of arguments made to argue for consequences of bilingualism in the face of observed null results. The first line of arguments called

upon looking at the effects of bilingualism by moving beyond inhibition. In reference to this, we first assessed the effects of bilingualism on Sustained Attention, a component of the hypothesized "attentional control" construct. Sustained Attention is supposed to be involved when bilinguals must continuously suppress interference from the non-target language for a given period in conversation, and is considered a crucial attentional mechanism (Chung-Fat-Yim, Calvo, & Grundy, 2022). In our study, we observe that for both sustained attention tasks, bilingualism-related characteristics failed to demonstrate any strong or reliable influence. This finding is in line with several previous studies which also failed to find a difference between monolinguals and bilinguals on tasks such as the SART (Sustained Attention to Response Task) (Bialystok, Craik, & Luk, 2008). Another study that administered the SART (Kousaie, Sheppard, Lemieux, Monetta, & Taler, 2014) found a difference in performance of French-English bilinguals and French monolinguals. The performance of the bilingual group was not different when compared to another monolingual group of English speakers. The difference seen, however, was confounded by the use of different equipment and settings. Other than looking at the comparisons of bilingual to monolingual groups, bilinguals with different ages of acquisition of the second language did not show differences in sustained attention Bak, Vega-Mendoza, and Sorace. A typical response to these null findings related

to sustained attention (Chung-Fat-Yim et al., 2022) has been that the lack of differences seen is due to the insensitivity of the measures used, arguing, instead, for the use of non-behavioral measures. Another typical response concerns the ceiling hypothesis, that is, differences do not appear with subjects in the age group (16-35) since individuals in this age group are already performing at "ceiling" efficiency. The second hypothesis is rebutted, however, in null findings seen also for different age groups across the lifespan (see (Bialystok et al., 2008; Gathercole et al., 2014)). It is common to see researchers mention null findings in literature, and continue to conclude that bilingualism enhances some faculty (here, sustained attention). While highlighting null effects seen in behavioral studies, researchers insist on positive influence of bilingualism by citing other positive findings where "newer" methods such as ex-gaussian analyses or neurophysiological measures (such as EEG) were used. Such conclusions, however, miss the fact that the hypothesis in question has been proposed for behavioral measures also, and has failed to accumulate sufficient robust evidence for it. Although the construct validity of behavioral methods remains a concern, it is inappropriate to consider all behavioral measures reporting null results as "insensitive to group differences". It has also been suggested that RT-based or difference-score based measures might be masking the true effects of bilingualism. However, even when tasks without these methodological issues are administered, no coherent relationship between bilingualism variables and attention emerges. Consider the results from the SACT and the Adaptive Flanker tasks: these tasks can be considered "improved measures" which are less susceptible to methodological flaws related to RT impurity and processing confounds (Draheim et al., 2021). For both these tasks, we observed that the best model was the null model. Even if we consider the anecdotal evidence ( $BF_{10} = 1.1$ ) for the alternate model, an examination of the posterior inclusion probabilities of each variable tells us that they were unable to predict performance. No strong evidence in favor of including any of the bilingualism characteristics as predictors in our models.

### In closing

Our findings respond to two recent developments in the Bilingual Advantage debate. One, the replacement of the inhibitory account with the attentional control account (Bialystok & Craik, 2022). In all our tasks measuring sustained attention and inhibition, the best hypothesis was the null hypothesis. The source of null effects does not seem to be the "inadequate" choice of inhibition as a mechanism. One must emphasize that the attentional control hypothesis is theoretically underspecified. Under this hypothesis, both kinds of results (null and positive findings) could further evidence for the influence of bilingualism on attention. The modulating factor is proposed to be "demands" on attentional control i.e. effects would be seen on tasks that require greater attentional resources. The sustained attention tasks used in this experi-

ment can be described as tasks with high attentional demands, and the null results seen cannot be simply discarded as arising due to low resource demands. The failure to find an advantage was not driven by the lack of task difficulty. An "easy" task such as the Flanker cannot be expected to bring out any real effects as all subjects would be good at it, but the same is not true for the sustained attention tasks. The second development relates to the poor psychometric properties of tasks typically used in bilingualism research (K. R. Paap, Johnson, & Sawi, 2016; Draheim et al., 2021). We tested for the effects of bilingualism and found no consistent evidence for it, despite the use of improved and more robust tasks. The null findings were not limited to the Arrow Flanker task which was a task reliant on RT, with poor reliability, and relied on the calculation of difference scores. A reason that bilingualism may not have a specific facilitatory influence on other cognitive domains could be that bilingualism does not confer any additional practice to these domains of inhibition or attention which is *exclusive* to the use of multiple languages. In summary, we examined the effects of bilingualism on sustained attention - in an attempt to go beyond inhibition. Since sustained attention is only one aspect of the broadly specified Attentional Control construct, our comments on the newer version of the BA hypothesis are limited. We also note that past studies also failed to find a bilingual advantage on sustained attention, and argue that this version of the hypothesis falls short in explaining these null results. This work is subject to some limitations - we looked at only sustained attention and used the flanker task (a task for which there is considerable disagreement regarding what it really measures (K. R. Paap, Anders-Jefferson, Mikulinsky, Masuda, & Mason, 2019)). We also limited ourselves to individual differences for behavioral measures, excluding any neurophysiological measures. Finally, our sample of interest was limited to one part of the lifespan- that of adulthood. This population is the one where the lack of converging findings in behavioral measures is especially stark. In addition, although the bilingualism-related variables were on a continuum, their measures were restricted to self-reports and did not include any objective measures.

### References

- Backer, K. C., & Bortfeld, H. (2021). Characterizing bilingual effects on cognition: the search for meaningful individual differences. *Brain Sciences*, *11*(1), 81.
- Bak, T. H., Vega-Mendoza, M., & Sorace, A. (2014). Never too late? an advantage on tests of auditory attention extends to late bilinguals. *Frontiers in psychology*, *5*, 485.
- Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, memory, and cognition*, *34*(4), 859.
- Bialystok, E., & Craik, F. I. (2022). How does bilingualism modify cognitive function? attention to the mechanism. *Psychonomic Bulletin & Review*, 1–24.
- Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M.

- (2004). Bilingualism, aging, and cognitive control: evidence from the simon task. *Psychology and aging*, 19(2), 290.
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual review of psychology*, 62, 73–101.
- Chung-Fat-Yim, A., Calvo, N., & Grundy, J. G. (2022). The multifaceted nature of bilingualism and attention. *Frontiers in Psychology*, 13.
- Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, 113(2), 135–149.
- Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, 150(2), 242.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, 16(1), 143–149.
- Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., . . . Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? lifespan evidence from preschoolers through older adults on card sorting, simon, and metalinguistic tasks. *Frontiers in Psychology*, 5, 11.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, 1(2), 67–81.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 515–530.
- Hedge, C., Powell, G., Bompas, A., & Sumner, P. (2022). Strategy and processing speed eclipse individual differences in control ability in conflict tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(10), 1448.
- Kousaie, S., Sheppard, C., Lemieux, M., Monetta, L., & Taler, V. (2014). Executive function and bilingualism in young and older adults. *Frontiers in Behavioral Neuroscience*, 8, 250.
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., De Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? a meta-analytic review. *Psychological bulletin*, 144(4), 394.
- Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and cognition*, 6(2), 97–115.
- Paap, K. (2019). The bilingual advantage debate: Quantity and quality of the evidence. *The handbook of the neuroscience of multilingualism*, 701–735.
- Paap, K. R., Anders-Jefferson, R., Mikulinsky, R., Masuda, S., & Mason, L. (2019). On the encapsulation of bilingual language control. *Journal of Memory and Language*, 105, 76–92.
- Paap, K. R., Anders-Jefferson, R., Zimiga, B., Mason, L., & Mikulinsky, R. (2020). Interference scores have inadequate concurrent and convergent validity: Should we stop using the flanker, simon, and spatial stroop tasks? *Cognitive Research: Principles and Implications*, 5(1), 1–27.
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive psychology*, 66(2), 232–258.
- Paap, K. R., Johnson, H. A., & Sawi, O. (2016). Should the search for bilingual advantages in executive functioning continue. *Cortex*, 74(4), 305–314.
- Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and cognition*, 13(2), 253–262.
- Rana, M., & Verma, A. (2023). Effects of bilingualism on inhibition unlikely-evidence from a bayesian inquiry. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).
- Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 501.
- Rey-Mermet, A., Gade, M., Souza, A. S., Von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General*, 148(8), 1335.
- Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic bulletin & review*, 26(2), 452–467.
- Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, 16, 601–615.