NeuroImage 298 (2024) 120786

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Brain correlates of attentional load processing reflect degree of bilingual engagement: Evidence from EEG

Sergio Miguel Pereira Soares[a,1,*], Yanina Prystauka[b,1], Vincent DeLuca[c], Claudia Poch[d], Jason Rothman[c,d,e]

[a] Language Development Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[b] Department of Linguistic, Literary and Aesthetic Studies, University of Bergen, Bergen, Norway
[c] Department of Language and Culture, UiT the Arctic University of Norway, Tromsø, Norway
[d] Nebrija Research Center in Cognition, University of Nebrija, Madrid, Spain
[e] Department of Linguistics and English Language, Lancaster University, Bailrigg, Lancaster LA1 4YW, UK

ARTICLE INFO

Keywords:
Bilingualism
Oscillations & ERPs
Working memory
Attention
N-back task

ABSTRACT

The present study uses electroencephalography (EEG) with an N-back task (0-, 1-, and 2-back) to investigate if and how individual bilingual experiences modulate brain activity and cognitive processes. The N-back is an especially appropriate task given recent proposals situating bilingual effects on neurocognition within the broader attentional control system (Bialystok and Craik, 2022). Beyond its working memory component, the N-Back task builds in complexity incrementally, progressively taxing the attentional system. EEG, behavioral and language/social background data were collected from 60 bilinguals. Two cognitive loads were calculated: low (1-back minus 0-back) and high (2-back minus 0-back). Behavioral performance and brain recruitment were modeled as a function of individual differences in bilingual engagement. We predicted task performance as modulated by bilingual engagement would reflect cognitive demands of increased complexity: slower reaction times and lower accuracy, and increase in theta, decrease in alpha and modulated N2/P3 amplitudes. The data show no modulation of the expected behavioral effects by degree of bilingual engagement. However, individual differences analyses reveal significant correlations between non-societal language use in Social contexts and alpha in the low cognitive load condition and age of acquisition of the L2/2L1 with theta in the high cognitive load. These findings lend some initial support to Bialystok and Craik (2022), showing how certain adaptations at the brain level take place in order to deal with the cognitive demands associated with variations in bilingual language experience and increases in attentional load. Furthermore, the present data highlight how these effects can play out differentially depending on cognitive testing/modalities – that is, effects were found at the TFR level but not behaviorally or in the ERPs, showing how the choice of analysis can be deterministic when investigating bilingual effects.

1. Introduction

Over the past two decades, much has been discussed regarding the potential for bilingualism to confer adaptations to various components of executive functions (EFs), inclusive of the underlying mechanisms (see Bialystok, 2021; Bialystok and Craik, 2022). The working hypothesis links adaptive effects in domain-general cognition to the demands/costs implicit to regulating activation and management of all the languages a multilingual knows. A majority of relevant work has tested this hypothesis by comparing monolingual and bilingual performances on EF tasks, whereby bilingualism would be expected to show some level of facilitation. While there is no shortage of individual studies providing support for this hypothesis, the literature as a whole provides conflicting results (see Paap, 2023; Donnelly et al., 2019; Grundy, 2020; Lehtonen et al., 2018; Leivada et al. 2021, 2023 for meta-analytics and discussions). Some have argued that failures of replication should be understood *a priori* to question bilingualism as a potential modulatory factor for EFs (e.g., see Paap, 2023 for review)—i.e., that the original

---

[*] Corresponding author at: Wundtlaan 1, 6525 XD Nijmegen, the Netherlands.
E-mail address: sergio-miguel.pereira-soares@mpi.nl (S.M. Pereira Soares).

[1] The first two listed authors share first authorship.

https://doi.org/10.1016/j.neuroimage.2024.120786
Received 5 April 2024; Received in revised form 9 July 2024; Accepted 12 August 2024

hypothesis is simply wrong or that the effects of bilingualism on EF are restricted to specific and undetermined circumstances. Against this point, however, recent epistemological and empirical work has shifted the focus: understanding what parameters of bilingual experience are more and less likely to give rise to adaptations (DeBruin, 2019; DeLuca et al., 2019, 2020; Green and Abutalebi, 2013; Gullifer and Titone, 2019; Luk and Bialystok, 2013; Titone and Tiv, 2022). Under such an approach, it is not assumed —if it ever was— that bilingualism *par excellence* should default to sufficient exercising within cognitive processing to warrant EF adaptations. Rather, the contemporary landscape of hypotheses maintaining a role for bilingualism as a potential mind/brain modulator claims that degree of engagement with bilingual experiences —e.g., amount and density of code-switching, proportionality of usage of the languages in various social contexts, the profiles of an individual's social network, among other factors—is ultimately what determines if, and if so the extent to which, individual bilinguals will show evidence of neurocognitive adaptations.

In a recent article, Bialystok and Craik (2022), henceforth B&C, address the elusive mechanisms of bilingualism-induced neurocognitive adaptations, situating them within the broader attentional control system. More specifically, they argue that the inconsistencies found in bilingualism studies arise, in part, from the traditional component-based view of EF, which focuses disproportionately on inhibition in the relevant bilingualism literature. Instead, they propose a holistic account based on attentional control more broadly, akin to recent updates by Miyake and Friedman (2012) to their EF proposal. Accordingly, B&C maintain adaptations to bilingual experience reflect a proportionality of deployment of relative attention needed to handle the related language/cognitive control demands of an individual, including but not exclusively involving inhibition, which would thus be modulated by the nature and degree of engagement with specific aspects of one's bilingual experience. Moreover, given that attention subsumes a broad range of cognitive subcomponents, their proposal is commensurable with empirical evidence used to support an inhibition-based mechanism for bilingual adaptations, without having to rely solely on an inhibition-based mechanism. Moreover, by appealing to the larger attentional control system, observations unaccounted for by an inhibitory mechanism can be accommodated while straightforwardly accounting for the observation that any given neurocognitive (bilingual) effect is dependent on the nature of the tasks/cognitive process being examined. As a result, the expectation is that the harder a particular task is/becomes (i.e., the more it taxes attentional resource allocation), the more likely it will be to observe either group differences (monolingual versus bilingual) or individual differences across bilinguals calibrated to degree of bilingual experiences.

Not unrelated to the above, there has been discussion pertaining to how the field, at least behaviorally, tends to assess EF performance and whether the typically used assays are (always) granular enough to reveal potentially latent effects (see e.g., Draheim et al., 2022 and Burgoyne et al., 2023 for recent work specifically on improving such task issues). This is especially plausible at particular ages of testing (around 20 to 40 years old) when cognition is generally at or around its peak (i.e., during which most individual's cognitive abilities are at their highest level of performance- Germine et al., 2011; Hertzog, 2020). If we take the position that engagement with bilingual experience is ultimately what determines (degree of) EF adaptation seriously, such discussions have important consequences because not having the most sensitive measures of EF could result in washing out weaker adaptations (especially so for studies whose subjects are (primarily) younger adults). To address this concern, not least in light of B&C's approach, the field would benefit from increased usage of EF tasks with progressively increasing attentional load. Doing so has at least two advantages. First, it would provide a more rigorous testing ground for determining if there is a bilingual EF effect at all. Second, it would permit one to specifically test B&C's mechanistic claims in accord with the position that not all bilingualism is the same in the relevant sense for EFs.

A domain for which the aforementioned is readily testable is working memory (WM), not least given the structure of the tasks commonly used to test it and the fact that many individual studies show bilingualism can contribute to WM efficiency (Anderson et al., 2021; Barker and Bialystok, 2019; Comishen and Bialystok, 2021; Morrison et al., 2019, but see Lehtonen et al., 2018 meta-analysis). The N-back task has been widely used to assess WM and permits simple manipulations to increase task difficulty/demands. Participants are required to recognize a stimulus presented *n* trials in the past and compare it—i.e., evaluate if it matches— to the current one in a sequence. Thus, the task involves continuous WM while progressively taxing attention as a function of item by item updating and the manipulation of *n* (1-back being less cognitively demanding than 2-back). As such, it is amenable to attentional resource allocation manipulation without compromising or changing (when updating is implicated, so $n = 1$ onwards) the underlying cognitive processes involved (Conway et al., 2005). Increasing *n* is associated with slower reaction times (RTs), decreases in accuracy, and lower cognitive resource availability for WM and attention (Dafner et al., 2011; Polich, 1996). This paradigm is, therefore, suitable for testing B&C's approach, especially in combination with a neuroimaging method capable of capturing attentional resource allocation in real time, such as EEG.

The EEG cognitive neuroscience literature indicates that specific neural markers are closely associated to EF processing. Some of the most studied and better categorized EF event related potential (ERP) markers are the N2 and P3 components, and alpha and theta oscillations. The N2 component, typically observed as a negative deflection around 200–300 milliseconds post-stimulus, is linked to conflict detection and inhibitory control processes (e.g., Folstein and Van Petten, 2008). The P3 component, a positive deflection around 300–600 milliseconds post-stimulus, is associated with attention allocation, updating of working memory and its amplitude is modulated by task difficulty (Ren et al., 2023). Regarding neural oscillations, both theta (4–8 Hz) and alpha (8–12 Hz) frequencies have been shown to play a crucial role in working memory processes and can be modulated by memory load (Sauseng et al., 2005). Furthermore, (frontal) theta is involved in conflict detection and error monitoring (e.g., Brunetti et al., 2019; Hsieh and Ranganath, 2014; Pscherer et al., 2021), while alpha is generally seen as a regulator of anticipatory updating mechanisms and inhibitory control responses (Cooper et al., 2016; Suzuki et al., 2018). In the neurocognitive work of bilingualism, EEG (mostly ERP) is an often-used method and the N-back task has featured prominently, especially in behavioral studies. They have, however, seldom been combined (Bice et al., 2020; Calvo et al., 2023; Grundy et al., 2017a; Lukasik et al., 2018; Pereira Soares et al., 2022; Teubner-Rhodes et al., 2016). As such, we limit the detailed discussion to the most comparable studies that have combined the two. In Morrison et al. (2019), bilingual and monolingual young adult participants completed a standard N-back task (0-, 1-, and 2-back conditions) while behavioral performance and EEG data were recorded. They found no differences at the behavioral level between the two groups. However, they showed that monolinguals exhibited a smaller P3 amplitude in comparison to the bilinguals. Guided by argumentation from the Bilingual Anterior to Posterior and Subcortical Shift (BAPSS) framework (Grundy et al., 2017b), the findings were interpreted as the bilinguals demonstrating greater attentional control in order to perform on a similar behavioral par with monolinguals. That is, the ERP signature differences indicated a bilingual efficiency effect, recruiting less neural resources (see Morrison et al., 2020 and Morrison and Taler, 2020 for comparable results in a delayed matching-to-sample memory task). Barker and Bialystok (2019), using a variant of the N-back task (implementing a differential emotion paradigm in-between the memory trials), found young monolinguals to be faster but less accurate on the 2-back than their peer bilinguals, and also monolingual accuracy to be more impeded by the emotional stimuli. Although both groups displayed an attenuation of the P3 amplitude in the 2-back condition in comparison to the 1-back, this was more attenuated in response to distracting emotional stimuli for bilinguals than monolinguals. In general, bilinguals were less impacted by emotional distraction than monolinguals and they also exhibited better adjustments in response to task-induced cognitive demands by recruiting brain resources to outperform monolinguals behaviorally. Comishen and Bialystok (2021) tested the hypothesis that a higher bilingual performance emerges when task difficulty increases. To do so, monolingual and bilingual young adults were tested on four increasingly demanding conditions of the N-back task (0- to 3-back). Monolinguals showed higher behavioral declines with increasing difficulty than bilinguals, while at the brain level also exhibiting greater efforts in processing all conditions. Thus, performance in WM, and generally in EF paradigms, might rather be determined by a cross-combination of attentional task demands and (available) attentional resources at the individual or group level.

The present work builds on these previous studies, innovating in several dimensions. Adopting a bilingual-centric individual differences approach, we capitalize on a within group design to avoid the potential comparative fallacy of comparing bilinguals to monolinguals (see Rothman et al., 2023 for discussion). Doing so harmonizes B&C's mechanistic approach with the contemporary landscape of the literature that treats bilingualism as a dynamic spectrum of experiences of dual/multiple language engagement as the key drivers of potential EF adaptations. Furthermore, we bring together both Event-Related Potentials (ERPs) and oscillatory brain activity analyses to understand how each contributes separately and, indeed, together to the research questions articulated below. While the most comparable studies (i.e., those described immediately above) provide only ERP evidence and are restricted to comparisons between bilingual and monolingual aggregates, this is the first study, to our knowledge, that also includes a time-frequency representation (TFR) analysis of the same EEG data while focusing on individual differences in bilingual engagement. As we will see, the present approach reveals insights perhaps not capturable under an ERP analysis alone, demonstrating the utility of TFR analyses in bilingualism research more generally, the combined value of brining ERP and TFR together as well as the value of understanding bilingualism in the absence of its comparison to monolingualism (Rossi et al., 2023; Rothman et al., 2023; De Houwer, 2023).

Not least in an effort to ensure relevant variation needed to run individual difference analyses in our bilingual group, early bilinguals (German-dominant heritage speakers of Italian as a minority language, who are functional multilinguals given their English knowledge) and late bilinguals (German-dominant L2 English learners) growing up in comparable conditions in Germany were recruited. Even though these two bilingual groups could, in principle, be tested against each other to the extent one has specific questions and/or theoretical motivations that warrant such a comparison, the questions investigated in the present work do not lend themselves to such a comparison. Rather our goal is to investigate if and how individual bilingual language experiences lead to differential brain correlates in the N-back task. As such, bringing these two types of bilinguals together in a single group injects the type of variation in language experience needed to ask and address our questions adequately. In light of previous work on WM and bilingualism specifically using the n-back task with ERPs, and informed by the literature on neural oscillations in general, we ask the following research question:

(1) What, if any, is the relationship between increasing cognitive load and degree/timing of bilingual engagement, at the behavioral and brain level, in EF task performance?

Following from B&C's proposal, we predict that task performance will reflect cognitive demands of increased complexity: (i) behaviorally, slower reaction times and decreases in accuracy and/or (ii) increase in theta, decrease in alpha and smaller amplitudes in N2 and P3 at the brain level. Following proposals that expect individual bilingual engagement to be a conditioning factor, we expect such effects to load onto individual differences patterns. However, these predictions need to be qualified, at least at the behavorial level. Recall that all the present participants are bilingual and the vast majority of studies showing behavioral differences juxtapose monolinguals with bilinguals. While our bilinguals are varied in their bilingual engagement, they, in principle, could all have surpassed a minimal threshold of bilingual engagement such that they do not show behavioral differences through the 2-back level. However, we would still expect differences at the brain level where efficiency in (potentially indistinguishable) task performance can be assessed. Here, we expect the degree of engagement (duration and/or intensity) in bilingual experience, irrespective of bilingual type, should relate to increased efficiency of deployment of cognitive resources to perform the task. We predict this will manifest specifically as gradual decreases in N2 and P3 amplitudes, an increase in task-related alpha power over task-irrelevant (frontal) electrode sites and increased suppression over task-relevant (central-posterior) electrodes, and decreased reliance on theta power in task-relevant (frontocentral) electrodes (DeLuca et al., 2020; Grundy et al., 2017b). Furthermore, we predict that these bilingual experience-induced effects will manifest more clearly in conditions with higher cognitive loads.

2. Materials and methods

2.1. Participants

Sixty healthy bi-/multilingual participants participated in the study. Six participants were discarded from all analyses due to high skin artefacts. Of the final 54 participants (mean age 24.6y; SD = 3.51y), 28 were second language learners (L2ers) and 26 were early bilinguals (heritage speakers – HSs). It is perhaps worth defining HSs specifically, given that this terminology is not necessarily universally used/known. HSs are naturalistic bilinguals of a minority language –their heritage language—imparted at home or in a close, diasporic community from birth despite growing up in a society that has a distinct majority language the HS is likely to become progressively dominant in as a function of increasing age (e.g., Polinsky, 2018; Kupisch and Rothman, 2018; Rothman, 2009). For the L2ers, German was the native language whereas English was the sequentially acquired L2. The HSs had Italian as their first language and either acquired German simultaneously (2L1) or before the age of 4. German was the dominant language of all participants. Regardless of bilingual type, English was acquired as an L2/L3 under the same contexts (as a foreign language in Germany). The age that participants first became bilingual differed within our sample (mean AoA for L2 = 9.23y; SD = 1.95y, mean AoA for 2L1 = 1.73y; SD = 1.69y). Even if it seems like the first exposure to at least one additional language happened on average at a young age, such does not preclude relevant variation within individuals' bilingual language use. On the one hand, although English tends to have an important role/status in schools in non-English speaking countries (at least in the European context), timing of first exposure, but especially quantity and quality of language use can vary quite drastically even at early ages. On the other hand, language engagement and opportunity in heritage bilingualism highly depends on several parameters ("status" of the language in the host country, number of parents speaking it, size of the language community, and many more) and thus inter-individual variation is the natural outcome of diversified contexts. Indeed, within the present cohort, as we will see, there is significant variation in terms of bilingual engagement. The average duration of exposure (i.e., time being (at least) bilingual) to the L2/2L1 was 19.2 years; SD = 5.2y). The Socio-Economic Status (SES) was coded from 0 to 4 based on the participant's mother's highest level of education (0 = lower than a high school diploma, 4 = postgraduate degree). Mean SES across our entire sample was 1.07 (SD = 1.18: range 0–4).

2.2. Background measures

All participants completed a language history questionnaire, the Language and Social Background Questionnaire (LSBQ; Anderson et al., 2018), which traces language exposure and use in the participants' known languages from early childhood in various settings and activities. Three different factors can be extracted as weighted aggregate scores of a subset of relevant questions and used for analyses: language use in the home environment (Home) and social contexts (Social) and language proficiency (in our specific case in the societal majority language (Proficiency)). For both Home and Social factor scores, a higher score indicates more engagement with the non-societal language (NSL) (Italian/English for HSs and English for L2ers) and a lower score indicates more engagement with the societal language (German). Higher Proficiency scores reflect higher proficiency in German. In addition to these three, two other factors were used for analyses in predicting brain and reaction times data: age of onset of the NSL (AoA) and duration of bilingualism (Duration – computed as the difference between the biological age at time of testing and AoA of Italian/English). Mean scores were: Social (mean = 11.16, SD = 9.02), Home (mean = 2.03, SD = 9.05) and Proficiency (mean = 0.71, SD = 1.7). The LexTALE (Lemhöfer and Broersma, 2012), a quick online test which aims at assessing general English proficiency, was also administered (mean = 66.71, SD = 12.27) (see https://osf.io/m86qs/ for the participant's data).

2.3. Study procedure and N-back task

The research procedures in this study were approved by the ethical commission of the University of Konstanz. Before taking part in the experiment, participants provided informed consent by signing a document that contained detailed information about the study. Data were collected either in a quiet room in a designated lab (University of Konstanz, Heinrich Heine University Düsseldorf or University of Cologne) or in a quiet room in a household (mostly in Konstanz, Cologne and Düsseldorf). This was possible because we employed a portable EEG system with active shielded electrodes and some of the data were collected over the COVID period (see below for more details). Participants first completed the LSBQ. Afterwards, they were fitted with an appropriate actiCap (10–20 system - Brain Products, Inc) for the EEG recording session. The task was presented on a 17-inch screen using the experiment control program Presentation (Presentation®, Neurobehavioral Systems). Eighteen letters (B, C, D, F, G, H, K, L, M, N, P, Q, R, S, T, V, W, Z - vowels were excluded to decrease the likeliness of participants developing chunking strategies which reduce mental effort, as suggested in Grimes et al., 2008) were selected. The N-back task included three different memory load conditions/blocks. In the 0-back condition, participants were instructed to press the green button (correct) on the button box (RB-740, Cedrus®) if the letter appearing on the screen was the same one as the first of the whole sequence (e.g., if the first letter of the sequence was a C, every time C appeared on the screen green was the correct and expected answer - otherwise red/incorrect was expected). In the 1-back condition participants were instructed to press the green button if the current letter on the screen matched the previous one (e.g., a C followed by a C). Finally, in the 2-back condition, participants were instructed to press green if the current letter matched the one presented two trials prior (e.g., a C followed by a B and then another C, as in the XXXCBCXXX sequence). In each condition/block, participants were instructed to press the red button (wrong) any time the letter appearing on the screen was not a target/match. Each block, which consisted of a practice session followed by the experimental session, lasted approximately 10 minutes, for a total of approximately 30 minutes. Breaks were always permitted in-between blocks. In each block, for both the practice and the experimental part, 25 % of the trials were targets/matches (4 trials out of 16 in the practice and 60 out of 240 in the experimental part). The procedure was first orally explained to the participants. Then, the participants engaged in a brief practice session (with positive and negative feedback) to familiarize themselves with the task for a total of 16 trials. The experimental session consisted of 240 trials presented in a randomized order. Upon completion of one block, participants could take a short break or move straight to the next experimental block (the order was kept constant for each participant, i.e., always with incremental cognitive load difficulty: 0-, 1-, and 2-back). Each trial began with the appearance of a letter in the center of the screen. Trials lasted either for 2000 ms or until the participant pressed a button. In both cases, a 1700 ms inter-trial interval (ITI) blank screen followed, before the next trial was presented. The EEG was continuously recorded from 32 Ag/AgCl scalp electrodes (LiveAmp32, Brain Products, Inc). Ground and online reference electrodes were AFz and FCz, respectively. To monitor vertical and horizontal eye movements the Fp1 and Fp2 electrodes (placed on the forehead above the eyebrows) were used. Impedances were kept below 25 kΩs. The signal was amplified and continuously digitized at a 1000 Hz sampling rate using a Brain Vision LiveAmp amplifier.

2.4. Data pre-processing

Offline data processing was performed in a two-step manner. In Brain Vision Analyzer 2.0 (Brain Products, Inc), data were first visually inspected and then band-pass filtered from 0.1 to 45 Hz. The signal was then segmented from −750 ms to 1250 ms relative to stimulus onset (such long trials were defined to allow for the calculation of the time-frequency representation using a moving window approach). In order to detect and get rid of eye movements and blinks, the BVA-implemented semi-automatic independent component analysis (ICA) was employed. ICA was applied on segmented data using 512 steps and a gradient infomax restricted algorithm. Unusual looking electrodes (e.g., high noise, picking up the heartbeat signal, excessive drifting, etc.) were topographically interpolated with a spheric spline (maximally 3 electrodes per participant per condition - total number of interpolated electrodes = 0.39 % of the dataset). The signal was manually inspected for EEG artifacts, resulting in the rejection of 3.23 % of the trials. The artifact-free epochs were then baseline-corrected (−100 ms prior to stimulus onset) and re-referenced against the averaged mastoid electrodes (TP9/10). Data were then exported into Matlab, using the Fieldtrip toolbox (Oostenveld et al., 2011) for the event-related potentials (ERPs) and time-frequency analyses (TFR).

2.5. Event-related potentials (ERPs) and time-frequency representation (TFR)

Recall that one of the novelties of the present approach (within the bilingualism field) relates to our complementary use of ERP and TFR analyses. Indeed, there is value in each independently and the information they can provide are only partially overlapping. Given that a TFR approach can reveal things particularly relevant for our research questions that might not emerge in an ERP analysis, we were committed to applying this method. However, since there are no similar previous studies applying TFR, but there are a few with ERP, the present ERP analysis serves at least two methodological functions: (i) the present data could be compared to previous ERP work to assess if participants were responding in the expected way to the task and (ii) given where one might expect overlap between ERP and TFR, we could test the functionality of TFR by assessing if the TFR analyses converge with the ERP ones where correspondences are expected (i.e., expected complementary) and, by extension, have a measure to understand what, if anything more, TFR contributes independently.

Only behaviorally correct trials were used for further preprocessing and analysis. ERPs were computed between −100 and 1250 ms with a −100–0 ms baseline. The power spectrum was computed for the theta- and alpha bands traditionally defined as 4–7 and 8–12 Hz, respectively. A 500 ms long stable window and a Hanning taper were used. Steps of 50 ms and 1 Hz were computed for power changes. Because of BVA to

Fieldtrip interface export procedure properties, the resulting TFRs contained data points between 500 ms prior to and 950 ms after the stimulus onset. TFRs were averaged for each subject and separately for each of the three conditions for matched items only (0-, 1-, and 2-back). Resulting TFRs were expressed as relative change from a baseline period from −500 to −100 ms. For the remainder of the analysis at both the brain (ERPs and TFRs) and behavioral level, two cognitive loads were examined, whereby 0-back served as baseline: a low cognitive load (i.e., the difference between the 1-back and the 0-back condition) and a high cognitive load (i.e., the difference between the 2-back and the 0-back condition).

2.6. Statistical analysis

2.6.1. Behavioral analysis

All reaction times (RTs) for the condition of interest (match) below 200 ms and/or non-accurate responses were excluded from the dataset for further analysis. This led to the exclusion of 70 trials for the 0-back condition (2.16 % of data removed), 355 trials for the 1-back condition (10.96 % of data removed) and 490 trials for the 2-back condition (15.12 % of data removed). We used linear models for the RT analyses and generalized linear models from the binomial family for the accuracy data. Correlational analyses were performed between low and high cognitive load at the level of RTs and the bilingualism variables.

2.6.2. Event-related potential and time frequency analyses

1000-randomizations cluster-based permutation tests as described in Maris and Oostenveld (2007) were performed for the N2 (150 ms to 350 ms) and the P3 (250 ms to 450 ms) ERP components. In order to determine the ERP and time-frequency clusters of interest for further analysis of individual differences, we ran a 1000-randomizations cluster-based permutation approach. The tests were computed for the averaged time windows for ERPs. The tests were computed on the whole post-stimulus onset time window and averaged per power frequency band for TFRs. T-values for every single electrode-time-frequency point and cluster-values statistics were run using a statistical threshold of $p < 0.025$ per tail.

2.6.3. Comparison of working memory loads

In order to investigate cognitive loads, we first computed the power spectrum subtraction separately for both high load (2-back minus 0-back) and low cognitive load (1-back minus 0-back) for ERPs and TFRs separately. Next, we compared the resulting power spectra in both load conditions by running cluster-based permutation tests, again for ERPs and TFRs separately.

2.6.4. Event-related and time-frequency electrode regions identification

Event-related and time-frequency clusters of interest were selected via a comparison between cognitive loads as described in the previous section. Because in all instances we are performing subtractions between clusters of electrodes, depending on the power amounts being computed and used for subtraction, the resulting significant clusters can turn out being either positive or negative. Thus, the resulting measures are relative power changes. Three electrode clusters of interest were found for ERPs: within the N2 time-window, we found a positive significant difference in the low load and one positive cluster in the high load, in the P3 time-window we found a significant negative cluster only in the high load condition. For the TFRs, we found a negative theta cluster ($p < 0.001$; 50–600 ms), a second negative theta cluster ($p = 0.04$; 600–950 ms) and a negative alpha cluster ($p < 0.001$; 0–950 ms) for low cognitive load (Fig. 1) and a negative theta cluster ($p < 0.001$; 0–950 ms) and a negative alpha cluster ($p < 0.001$; 0–950 ms) in the high load condition (Fig. 2). These clusters were extracted and carried forward for individual differences analyses.

Fig. 2. High cognitive load: time-frequency representations for 2-back (column 1), 0-back (column 2) and the difference between the two conditions (2–0, column 3).

2.6.5. Interaction between bilingualism experience factors and brain outcomes

Language variables extracted from the LSBQ (non-societal language exposure and use at home -NSL Home, non-societal language use in the society or community -NSL Social, Age of L2 or 2L1 onset -AoA and duration of bilingualism - Duration) were correlated to the derived clusters described in 2.6.4 for the whole dataset as a continuum of bilingualism engagement. Socio-economic status (SES) was used as model covariate. Continuous variables included in the models were mean centered; treatment coding was applied to categorical variables. The *glm* function from the *lme4* package in R (Bates et al., 2015) was used to run the linear models.

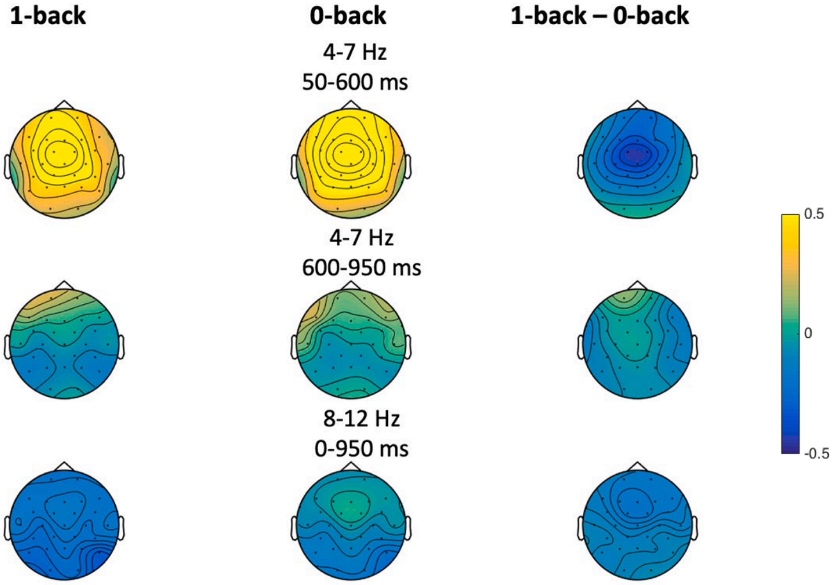

Fig. 1. Low cognitive load: time-frequency representations for 1-back (column 1), 0-back (column 2) and the difference between the two conditions (1–0, column 3).

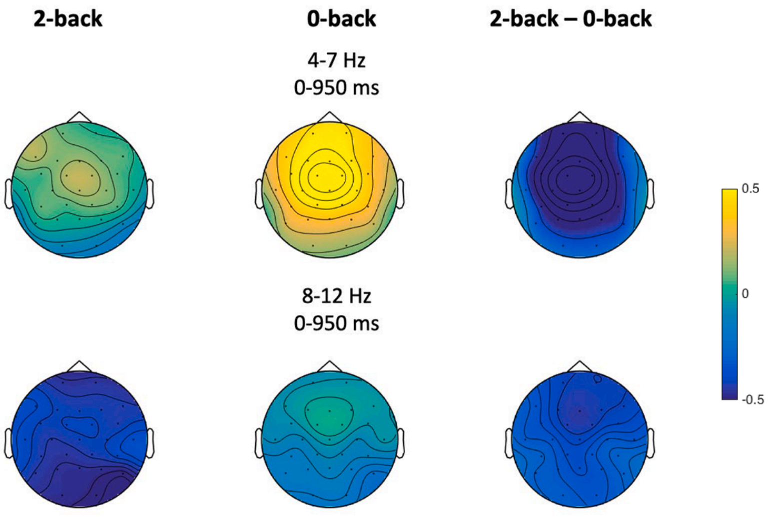

Fig. 2. High cognitive load: significant clusters for 2-back (column 1), 0-back (column 2) and the difference between the two conditions (2–0, column 3), the color scale is from −0.5 to 0.5.

3. Results

3.1. Behavioral results

A summary of the behavioral results (reaction times and mean accuracy) is shown below (Table 1). As expected, increases in task difficulties led to slower and less accurate behavioral performance as indicated by simple paired two-tailed *t*-tests both for RTs (0 vs 1, $p < 0.001$; 0 vs 2, $p < 0.001$; 1 vs 2, $p < 0.001$; p corrected at α=0.015 due to multiple *t*-tests) and accuracy (0 vs 1, $p < 0.001$; 0 vs 2, $p < 0.001$; 1 vs 2; p corrected at α=0.015 due to multiple *t*-tests). None of the correlational analyses between our bilingualism variables and the behavioral data (RT) shows statistically significant effects.

3.2. Neurophysiological results: task level

3.2.1. Event-related potentials

The ERP analyses revealed effects for the cognitive task in the typical N2 and P3 time-windows (see Fig. 3). More specifically, effects arose across the board for both low and high cognitive load [low cognitive load: N200 ($p = 0.026$), P300 ($p = 0.002$); high cognitive load: N200 ($p < 0.001$), P300 ($p < 0.001$)], indicating a stepwise task-related increase in brain/cognitive recruitment.

3.2.2. Time-Frequency analysis

The TFR results at the task level for power are shown in Fig. 4.

The results indicate a gradual decrease (see 2.6.4) in theta synchronization and a gradual increase (see 2.6.4) in alpha desynchronization. While alpha decrease as a function of task difficulty was expected, the theta effect went in the opposite direction. In order to understand this surprising observed relationship between theta and alpha, a simple correlation analysis was done between these two frequency bands. We found that the theta and alpha decreases were positively correlated ($t = 2.02$, df=52, $p = 0.048$), indicating a possible tradeoff between these two power bands (i.e., the stronger theta synchronization decreases, the stronger alpha desynchronization increases).

Table 1

Summary of the behavioral performance (mean Reaction Times and Accuracy) for the bilinguals within this study for the three conditions (0-back, 1-back, and 2-back) in the N-back task.

Condition	**meanRTs (ms)**	**sdRTs (ms)**	**meanAcc**	**sdAcc**
0-back	422	34.1	0.98	0.02
1-back	499	68.9	0.89	0.08
2-back	659	110.7	0.85	0.13

3.3. Neurophysiological results: individual bilingual differences and brain interaction

In order to examine whether bilingual engagement modulates the EEG amplitude and power levels, we ran linear models for each of the significant clusters (see 2.6.4). These were performed for ERPs and power separately.

3.3.1. ERPs

3.3.1.1. Low cognitive load (1-back – 0-back). The individual differences analyses performed on the clusters of interest did not reveal any significant effects for the relationships between language experiences and ERPs (within N200 and P300 time windows) in the low cognitive load condition.

3.3.1.2. High cognitive load (2-back – 0-back). As was true for the low cognitive load condition, the high load condition shows no significant correlations found in any ERP component.

3.3.2. Power

3.3.2.1. Low cognitive load (1-back – 0-back). As can be seen in Fig. 5, alpha negative power positively correlated with NSL Social (SE = 0.15, $t = 2.11$, $p = 0.04$).

3.3.2.2. High cognitive load (2-back – 0-back)

As can be seen in Fig. 6, we observed a positive correlation between theta negative power and AoA (SE = 0.22, $t = 3.98$, $p < 0.001$).

4. Discussion

With our research question in mind, repeated below for ease of exposition, we unpack the above presented results, seeking to understand if and how (degree of) bilingual experiences affect EF task performance manifested at distinct levels of attentional load.

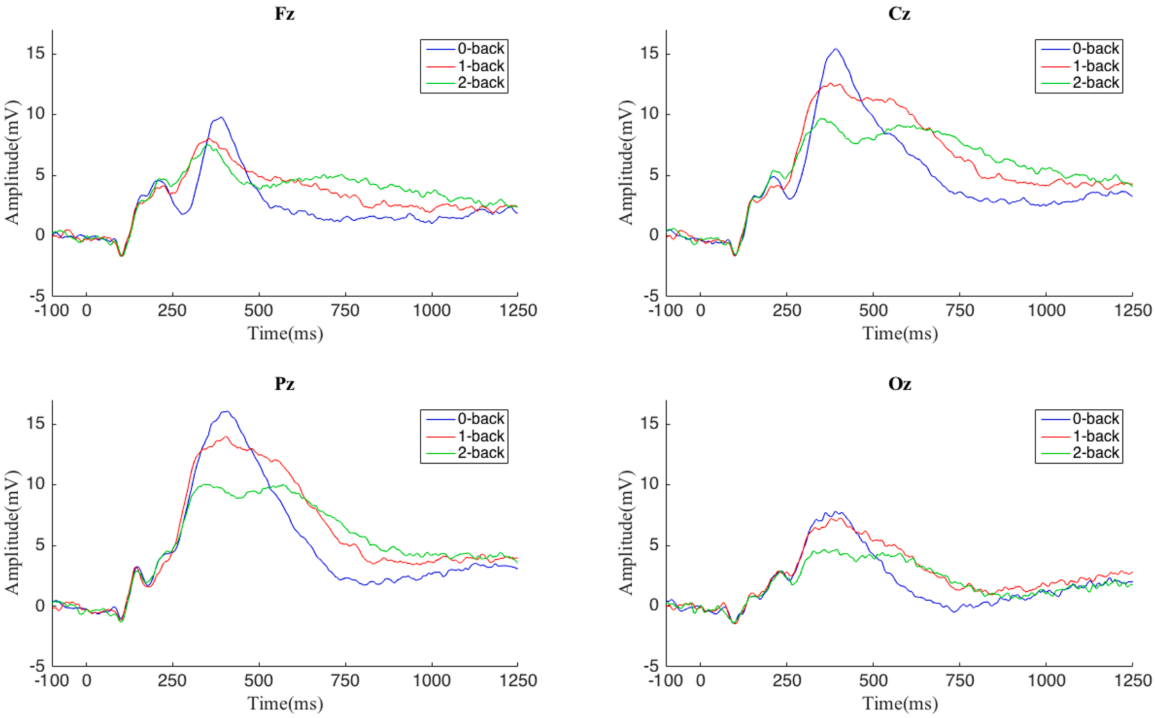

Fig. 3. ERP deflections for a flanker task (0-back in blue, 1-back in red and 2-back in green) at four different channel locations (Fz – top left; Cz – top right; Pz – bottom left; Oz – bottom right).

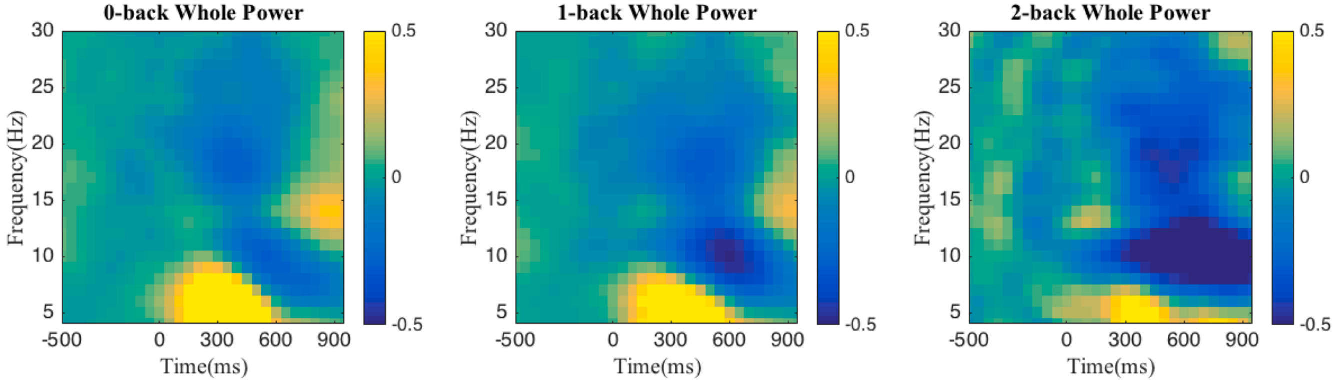

Fig. 4. Power distribution in the 0-, 1-, and 2-back conditions, electrode Cz.

(1) What, if any, is the relationship between increasing cognitive load and degree/timing of bilingual engagement, at the behavioral and brain level, in EF task performance?

4.1. Behavioral performance

Overall, we observed increased latency in response times and decreased accuracy rates with increasing task difficulty, indicating the task worked. Yet, at the behavioral level none of the bilingual measures were found to correlate with task performance expressed as the difference between cognitive load conditions. Although at first glance this might be understood as counter evidence to B&C's proposal or, in the extreme, evidence against any genuine bilingual effect on EF more generally, such a conclusion would be precipitous for several reasons. To begin with, we know from the larger literature that dissociations between behavioral and brain data are not uncommon (e.g., Abutalebi et al., 2012; Morales et al., 2015; DeLuca et al., 2020; Pereira Soares et al., 2022). And while one type of evidence is not inherently privileged over the other, an argument can be made for why brain data are more reliable for our questions. Recall that in the introduction, highlighting the absence of a monolingual comparison group, we qualified that behavioral evidence in isolation might not be particularly informative. While it certainly was possible to have shown behavioral effects, failure to show them would not suffice to reject *a priori* B&C's proposal and/or a role for bilingual engagement conditioning degree of neurocognitive outcomes. Why? It could simply be the case that the reality of our particular bilingual participants' profiles is such that each has had minimally sufficient timing (e.g., duration) and degree of bilingual engagement after which any further effect on EF is irrelevant for behavioral task performance distinctions at the 2-back level. If on the right track, we would expect that if there were a monolingual comparison group then differences might have obtained between them and our bilinguals, but pursing this question was not our goal herein. It might also be the case that behavioral differences could exist between the present bilinguals, but the cognitive load of a 2-back was not sufficient to distinguish between our bilingual experience range. For example, perhaps with our same participants behavioral differences would

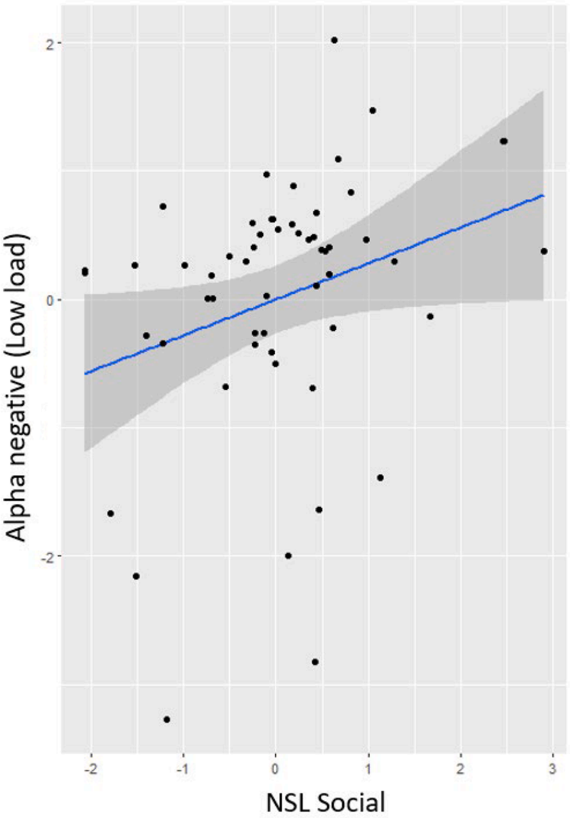

Fig. 5. Correlation between alpha negative cluster and non-societal language use in Social contexts (NSL Social) in the low load condition. Values for both axes (power for the Y axis and age for the X axes) have been scaled.

emerge with a 3-back condition. Finally, using another type of cognitive control manipulation tapping into different cognitive processes where attentional load could also be manipulated might yield individual effects at the behavioral level (in addition to the neural ones). While there is a robust history of using the N-back in the relevant bilingualism literature in the same form we have, it might be a good idea moving forward to employ a version of the N-back using entirely non-linguistic symbols or numbers. We leave the testing of these possibilities for future empirical work. What is, however, important to highlight here is that there are open questions and further considerations to keep in mind before making any conclusions on the basis of the present behavioral data alone.

Fortunately, B&C's approach makes predictions not only for behavior, but also at the brain level. On-task neuroimaging measures permit one to look beyond surface task performance to the efficiency with which any given performance is achieved. And so, before making any conclusions on what the behavioral data should be taken to mean, one wants to see which interpretations accord with the EEG evidence presented above. We now turn to doing so.

4.2. *Neurophysiological evidence: task level*

The ERP results show significant differences between conditions at all canonical timepoints (specifically N2 and P3), indicating that ERP responses were modulated by increasing task demands. Since these components reflect cognitive processes of stimulus identification/distinction and internal allocation of attentional resources and memory updating (Brower et al., 2012; Patel and Azzam, 2005; Scharinger et al., 2017), we interpret these incremental decreases of electrophysiological amplitudes (most notably in the P3 time window) as indicating a shift in the reallocation of neural resources from decision-making and memory updating operations towards maintenance, manipulation, and sustaining attention processes (Daffner et al., 2011; Ren et al., 2023).

Recall that the analysis of power over task complexity revealed a partially unexpected result. While we anticipated and found increased alpha desynchronization (e.g., Gevins et al., 1997; Gevins and Smith, 2000; Klimesch, 1999, 2012), we also found an unexpected decrease in theta power with increasing task complexity. Increases in frontal theta synchronization have been argued to reflect enhanced attention (Gevins et al., 1997) and maintenance of memory representations (Hsieh and Ranganath, 2014; Jensen and Tesche, 2002), whereas increases in alpha desynchronization (mostly across posterior electrodes) have been suggested to indicate the amount of neuronal resources allocated to the performance of the task (Gevins et al., 1997). Even though the theta effect herein presented itself in the unexpected direction, it is important to highlight that the decrease in theta follows a gradual pattern, highlighting a progressive taxation of the cognitive system to perform the task. Also important is that while increases in alpha desynchronization seem to be a robust finding, increases in theta synchronization are often less clear (Brouwer et al., 2012).

Nevertheless, in light of the observed reductions in theta power, a follow-up analysis was performed (see 3.2.2), in order to investigate the relationship between levels of theta and alpha power. Recall that the reductions in theta synchronization increase (relative to baseline) across conditions correlated positively with the suppressions in alpha. We interpret this as some sort of neural tradeoff: as theta activation decreased, alpha suppression increased with additional task demands/complexity. Thus, this theta power result can be accounted for as a shifting of neural recruitment strategies across the sample, i.e., as a redistribution of cognitive resources from memory functioning towards attention (see e.g., Missonnier et al., 2006). In fact, theta oscillatory activity in working memory can originate from different neural sources and reflect multiple cognitive processes. For example, using intracranial EEG, Brzezicka et al. (2019) examined the contribution of three areas – the dorsolateral prefrontal cortex (DLPFC), anterior cingulate cortex (ACC), and hippocampus – to working memory functioning. They found that while theta power in the ACC and hippocampus increased with increasing memory load, theta power in DLPFC decreased – a pattern compatible with what we observed in our findings. Similarly, Scharinger et al. (2017) also report a pattern where theta power decreased with increasing N-back load. They explain it by working memory overload in the high load condition. Thus, the reduction in theta activation seen herein may simply reflect contributions from different sources to handle the increasing task demands. More research is required to assess this, however. Similarly, alpha power suppression indexes a number of potential cognitive processes (Mazaheri et al., 2018) and can be interpreted here as a response to increased attentional or memory demands with increasing task complexity. Given the nature of the N-back task, both cognitive processes are likely indexed. Thus, evidence of shifting requirements between these processes, while not forming part of our original hypotheses, is not surprising.

4.3. *Neurophysiological evidence: individual differences in bilingual engagement*

While no bilingual engagement effects were found to correlate with behavioral task performance, age of bilingual language acquisition (AoA) and degree of usage of the non-societal language (NSL Social) in social settings emerged as predictors of neural recruitment patterns, albeit in distinct load conditions. Given the lack of correlations in behavioral performance, the effects found herein in the neural data can be addressed as a measure of efficiency or efficacy in performing the task. Furthermore, there are no relevant ERP correlational findings to discuss. At first glance, a lack of such correlations might be considered to run in disaccord with the previous EEG-based empirical research in this domain. However, there are a few factors to keep in mind. Recall that previous studies exclusively concern group comparisons between

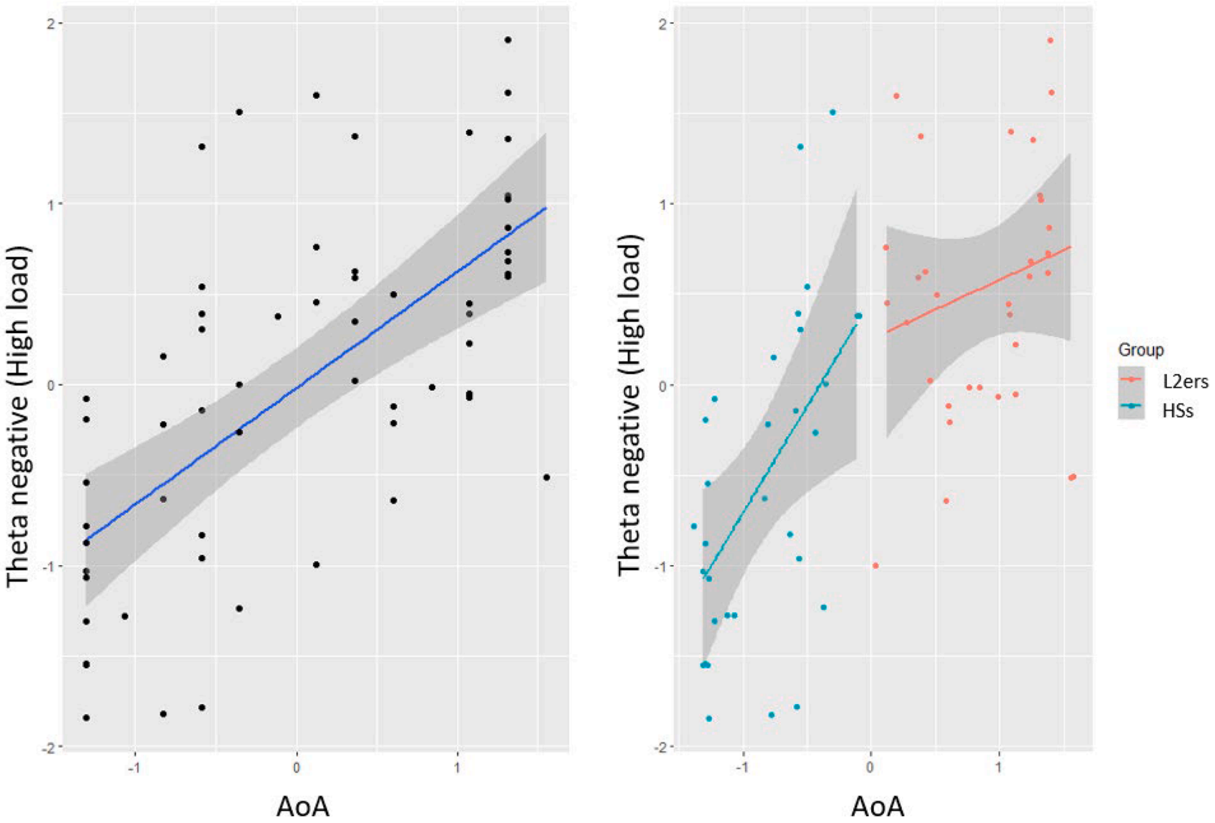

Fig. 6. (left) Correlation between theta negative cluster and age of acquisition of the L2/2L1 (AoA) in the high load condition. (Right) Same correlation as on the left panel but split by group (HSs and L2ers). Values for both axes (power for the Y axis and age for the X axes) have been scaled.

monolingual and bilingual participants, whereas the present study comprises only a bilingual group with (regressed) diverse individual experiences with bilingualism. It cannot be taken for granted, then, that what is shown as differences between monolinguals and bilinguals via ERP will manifest similarly, if at all, in an individual differences analysis across bilinguals. In fact, the lack of correlation seen for the bilingualism variables in the ERP domain, not least in light of what the TFR analysis shows, might suggest that ERPs are not the best method for a bilingual-centric approach predicated on revealing individual differences in general or for a relatively small group of participants such as the present one (Rossi et al., 2023). Conversely, TFRs by their nature are able to capture evoked activity as well as transient (induced) activity, the combination of which seems to have provided the sufficient granularity to detect the more subtle adaptation effects seen within this sample. As such, TFR analyses are perhaps generally better suited for an approach that does not compare distinct aggregates, such as monolinguals to bilinguals

The correlation observed between NSL Social and alpha desynchronization supports our predictions, but its timing (in the low- but not high cognitive load condition) was unexpected. However, upon reflection the earlier-than-expected emergence makes sense insofar as it can be due to a feature of the N-back task itself that, while amendable to differences in bilingual experience, we did not consider prior. If what we develop below is on the right track, the effect would be capturing something distinct from what underlies the basis of our prediction (i.e., not a marker of cognitive adaptation to increasing task complexity) but rather something else, a point to which we return. Recall that in the task itself, the conditions were always presented sequentially (0-, 1-, and then 2-back). As such, the effects seen for 1-back reflect a change in the cognitive demands associated with the task to that point (i.e., the emerging need for formation of memory traces). The effect in alpha can be taken then to reflect a targeted allocation of attentional/memory resources to handle this particular change in task requirements (Gevins et al., 1997). Our data indicate that how individuals deal with this initial shift in task requirement is subject to bilingual experience: those who use the non-societal language in social contexts more extensively are more apt to handle this shift more efficiently. Insofar as the underlying processes inherent to forming memory traces engage the attentional system, the apparent efficiency differential predicated on greater bilingual engagement would fall out from, and thus support, B&C's proposal.

Reasonably, one might ponder why the variable that seems to matter here is (increased) use of the NSL in social contexts, for example, as opposed to or as well as NSL use at home and/or other factors. Again, one must keep in mind that there is no monolingual comparison group as well as the reality of bilingualism that defines our participants. Given that there is no comparison group lacking bilingualism altogether, the only vehicle through which individual differentiation can manifest—indeed be investigated—is through the quantification of variables that embody a sufficient range of individual behavioral profiles. With this in mind, any given constitution of a bilingual group in terms of its diversity with relevant bilingual experiences is ultimately what drives predictions. In the present case, it makes sense that social environment would be the most reliable locus of differentiation in linguistic engagement because of the reality of how the languages distribute in Germany in general, and in particular for our participants. Recall that English is a common NSL across all participants and the opportunities for using it—and thus engaging in dual language switching—at home are essentially the same given that it is a foreign language. While roughly half our participants are Italian HSs—so they speak three languages (Italian, German and English) with high proficiency—and the other half are German-dominant speakers of English, in all cases the home context is essentially one of bilingualism with English as "the other language". Differently from other heritage language (HL) contexts, such as those described in North America (Rothman, 2009; Montrul, 2016; Polinsky, 2018), the HL, Italian, is maintained as the main home language over time, as opposed to shifting to the societal language or a balance between the two as is commonly reported in North America. And so, this means that for roughly half our participants bilingualism in the home context is between Italian and English (both parents being Italian natives), whereas for the other half it divides between German and English. There is no good reason to suggest that English would or should be more prevalent in an Italian HS's home as opposed to a German-dominant L2er's. Thus, the range of difference between participants as it relates to home use is potentially not rich enough to draw out individual differences, at least for effects on alpha in the context of the present task. Or, it could be the case that because of the ubiquity and facility of being able to engage with English as an L2 in particular in the home is high enough for all, that each individual has surpassed the threshold whereby any differences one might expect between bilingual and multilingual individuals cannot be shown. In social contexts, however, there is significantly more individual variation with respect to the balance/distribution of language usage. Choosing to use the NSL in social contexts more, not least because in principle it should rarely be needed, logically stands out as a good candidate for what differentiates individuals. This because the opportunities to engage a NSL vary considerably more than the context of "home" for our participants.

Nevertheless, it is reasonable to ponder the lack of correlational effect in alpha with any of the bilingualism experience indices in the high cognitive load. After all, we had anticipated finding such an effect, which we highlighted would serve as direct support for B&C's proposal. However, if the interpretation above regarding changes in cognitive demands for low cognitive load is on the right track, it is worth noting that the nature of the task is more consistent between the conditions that comprise high cognitive load. Of course, this does not ignore the fact that 2 back is more complex—which should engage attention more—but rather serves to highlight that the cognitive task demand itself is one and the same. And so, B&C's approach would be very suitable to explain any noted difference that could be attributable to augmented complexity for memory load, hence our original prediction. However, the fact that it did not obtain does not speak against it *per se*. What the data seem to show is that, for the present group of bilinguals, memory load complexity in what constitutes the same underlying operation was not conditioned by individual bilingual experience. This could be so either because the participant pool does not have the relevant degree of individual differences to show this, all individuals have passed a minimal engagement threshold whereby they have achieved a ceiling effect or individual differences more generally are not a good indicator of task complexity related outcome differences. Given the alpha effects noted in the lower cognitive load condition as well as other individual difference factors showing correlations to outcomes (such as AoA, which we turn to immediately below), it seems reasonable to favor the first two possibilities over the third. Ultimately, further research is needed to tease this out.

Timing of bilingual engagement (AoA) correlated with theta power, but the directionality of this effect was not in line with our predictions. Taking into consideration B&C's proposal and other theoretical proposals on bilingual adaptation (e.g., DeLuca et al., 2020), we predicted a negative correlation with AoA, which would indicate prolonged duration of bilingual experience to correspond to adaptations towards increased efficiency and thus lower requirements on theta activity (lower cognitive control requirements), particularly at higher cognitive loads. In our sample, we were able to tease apart time of bilingual duration (Duration) and AoA, controlling for the former across a spectrum of bilingual individuals who could be matched for Duration despite being early versus late bilinguals. And yet, Duration did not emerge as a predictor, but rather timing of onset of bilingualism did. In other words, this effect is arguably driven by the cohort of bilinguals who were early childhood acquirers, that is, the heritage bilinguals. Fig. 6 (right panel) above, which demarcates the two types of bilinguals within our sample by color, visualizes the following: while the effect goes in the same direction for both groups, one can note the dramatic differences in slope that seemingly drives the effect.

And so, AoA here is a true proxy for timing as opposed to the durative length of being bilingual itself. As such, the present correlation does not support our prediction that duration is what ultimately matters. What we observe, then, could be a true age-effect, suggesting some fundamental differences in how bilingualism experience matters more generally by AoA. Alternatively, it might not be age *per se*—we note that many of the German-dominant L2 learners report an age of onset to English in later childhood— but rather a difference that pertains to simultaneous over sequential bilingualism *par excellence*, meaning effect differences related to whether or not two languages are acquired at the same time in the absence of previous linguistic experience overall. Finally, given the reality of our participant pool whereby the HSs are highly proficient in three languages, we would not want to preclude the possibility that this effect reflects a difference between bilingualism (2 languages) versus true multilingualism (3 or more languages). Future research designed to tease these possibilities out is called for.

4.4. Limitations

While the present study contributes valuable insights into the neurocognition of bilingualism by complementing existing findings employing EEG as a methodology, it is worth acknowledging and addressing certain limitations that may impact the interpretation and generalizability of the results. A first limitation of this study is the relatively small sample size. Higher participant numbers could have given us the possibility to run more complex statistical methods (e.g., structural equation models), which would permit us to analyze more in depth the complex relationships among bilingual variables and the brain data (e.g., Carter et al., 2023). A second important limitation regards the task itself. While the majority of the studies using an N-back task stop at 2-back, it is not rare to find (mostly behaviorally) studies with even higher memory loads (e.g., 3-back). Within the context of the present study and the B&C's proposal, a further increase in cognitive load might/could have highlighted effects not entirely captured herein, as for example behavioral, ERP or language usage ones. Future studies that seek to investigate the neurocognitive effects of bilingualism within the broader attentional/cognitive networks and with incrementing attentional loads should strongly consider the limitations mentioned herein.

4.5. Conclusion

The present study tested B&C's (2022) mechanistic proposal for the cognitive effects of bilingualism, couching it within the wider attentional control network. We did so using an N-back task, which taxes attentional resource allocation incrementally, examining behavioral as well as two types of EEG data (ERP and TFR). In line with contemporary proposals regarding the determinism of relative degree of bilingual language engagement as a conditioning factor for (calibrated) EF adaptations at the individual level, we measured and regressed various indices of bilingual experience across a diverse cohort of bilingual individuals. Taken together, the results indicate two overarching trends. Firstly, the data showed effects of bilingual experience/timing in the low cognitive and high cognitive load conditions, lending some support to B&C's proposal and the general approach to treating bilingualism as a reflection of the dynamic set of bilingual experiences that define each individual's journey. Secondly, we show how methodological considerations matter. For instance, the TFR analysis captured bilingual experience effects unattested in the behavorial and ERP analyses. As a result, we offered some reasoning as to why TFR might be the most suitable EEG methodology specifically for investigating within (bilingual) group differences and can confidently recommend the increased usage of TFR for similar research in our field (see Rossi et al., 2023).

Ethics statement

This study involved human participants. It was reviewed and approved by the University of Konstanz Ethical Committee. The participants provided their written informed consent to participate in this study.

CRediT authorship contribution statement

Sergio Miguel Pereira Soares: Writing – review & editing, Writing – original draft, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yanina Prystauka:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Vincent DeLuca:** Writing – review & editing, Writing – original draft, Supervision, Formal analysis. **Claudia Poch:** Writing – review & editing, Supervision, Methodology. **Jason Rothman:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

All scripts (Fieldtrip and R) and data can be found on the following OSF link: https://osf.io/m86qs/.

Acknowledgments

This article was supported by generous funding to Sergio Miguel Pereira Soares by the European Union's Horizon 2020 researchand innovation programme under the Marie Skłodowska Curie Grant Agreement No 765556. Jason Rothman, Yanina Prystauka, and Vincent DeLuca were funded by the AcqVA Auora Center grant. Jason Rothman and Yanina Prystauka received funding from the Tromsø Forskningsstiftelse (Tromsø Research Foundation) Grant No. A43484 and the Heritage-bilingual Linguistic Proficiency in their Native Grammar (HeLPiNG) (2019–2023). Claudia Poch was supported by the PID2022-143111NB-I00 grant from the Spanish Ministry of Science.

References

Abutalebi, J., Della Rosa, P.A., Green, D.W., Hernandez, M., Scifo, P., Keim, R., Cappa, S. F., Costa, A., 2012. Bilingualism tunes the anterior cingulate cortex for conflict monitoring. Cereb. Cortex 22 (9), 2076–2086.

Anderson, J.A., Mak, L., Keyvani Chahi, A., Bialystok, E., 2018. The language and social background questionnaire: assessing degree of bilingualism in a diverse population. Behav. Res. Methods 50, 250–263.

Anderson, J.A., Grundy, J.G., Grady, C.L., Craik, F.I., Bialystok, E., 2021. Bilingualism contributes to reserve and working memory efficiency: evidence from structural and functional neuroimaging. Neuropsychologia 163, 108071.

Barker, R.M., Bialystok, E., 2019. Processing differences between monolingual and bilingual young adults on an emotion n-back task. Brain Cogn. 134, 29–43.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bialystok, E., 2021. Bilingualism: pathway to cognitive reserve. Trends Cogn. Sci. (Regul. Ed.) 25 (5), 355–364.

Bialystok, E., Craik, F.I., 2022. How does bilingualism modify cognitive function? Attention to the mechanism. Psychon. Bull. Rev. 29 (4), 1246–1269.

Bice, K., Yamasaki, B.L., Prat, C.S., 2020. Bilingual language experience shapes resting-state brain rhythms. Neurobiol. Lang. 1 (3), 288–318.

Brouwer, A.M., Hogervorst, M.A., Van Erp, J.B., Heffelaar, T., Zimmerman, P.H., Oostenveld, R., 2012. Estimating workload using EEG spectral power and ERPs in the n-back task. J. Neural Eng. 9 (4), 045008.

Brunetti, M., Zappasodi, F., Croce, P., Di Matteo, R., 2019. Parsing the Flanker task to reveal behavioral and oscillatory correlates of unattended conflict interference. Sci. Rep. 9 (1), 13883.

Brzezicka, A., Kamiński, J., Reed, C.M., Chung, J.M., Mamelak, A.N., Rutishauser, U., 2019. Working memory load-related theta power decreases in dorsolateral prefrontal cortex predict individual differences in performance. J. Cogn. Neurosci. 31 (9), 1290–1307.

Burgoyne, A.P., Mashburn, C.A., Tsukahara, J.S., Hambrick, D.Z., Engle, R.W., 2023. Understanding the relationship between rationality and intelligence: a latent-variable approach. Think Reason. 29 (1), 1–42.

Calvo, N., Grundy, J.G., Bialystok, E., 2023. Bilingualism modulates neural efficiency at rest through alpha reactivity. Neuropsychologia 180, 108486.

Carter, F., DeLuca, V., Segaert, K., Mazaheri, A., Krott, A., 2023. Functional neural architecture of cognitive control mediates the relationship between individual differences in bilingual experience and behaviour. Neuroimage, 120085.

Comishen, K.J., Bialystok, E., 2021. Increases in attentional demands are associated with language group differences in working memory performance. Brain Cogn. 147, 105658.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., Engle, R.W., 2005. Working memory span tasks: a methodological review and user's guide. Psychon. Bull. Rev. 12 (5), 769–786.

Cooper, P.S., Darriba, Á., Karayanidis, F., Barceló, F., 2016. Contextually sensitive power changes across multiple frequency bands underpin cognitive control. Neuroimage 132, 499–511.

Daffner, K.R., Chong, H., Sun, X., Tarbi, E.C., Riis, J.L., McGinnis, S.M., Holcomb, P.J., 2011. Mechanisms underlying age-and performance-related differences in working memory. J. Cogn. Neurosci. 23 (6), 1298–1314.

De Bruin, A., 2019. Not all bilinguals are the same: a call for more detailed assessments and descriptions of bilingual experiences. Behav. Sci. 9 (3), 33.

De Houwer, A., 2023. The danger of bilingual–monolingual comparisons in applied psycholinguistic research. Appl. Psycholinguist. 1–15.

DeLuca, V., Rothman, J., Bialystok, E., Pliatsikas, C., 2019. Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. Proc. Natl. Acad. Sci. 116 (15), 7565–7574.

DeLuca, V., Segaert, K., Mazaheri, A., Krott, A., 2020. Understanding bilingual brain function and structure changes? U bet! A unified bilingual experience trajectory model. J. Neurolinguist. 56, 100930.

Donnelly, S., Brooks, P.J., Homer, B.D., 2019. Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. Psychon. Bull. Rev. 26, 1122–1147.

Draheim, C., Pak, R., Draheim, A.A., Engle, R.W., 2022. The role of attention control in complex real-world tasks. Psychon. Bull. Rev. 29 (4), 1143–1197.

Folstein, J.R., Van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology 45 (1), 152–170.

Germine, L.T., Duchaine, B., Nakayama, K., 2011. Where cognitive development and aging meet: face learning ability peaks after age 30. Cognition 118 (2), 201–210.

Gevins, A., Smith, M.E., McEvoy, L., Yu, D., 1997. High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. Cereb. Cortex 7 (4), 374–385.

Gevins, A., Smith, M.E., 2000. Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. Cereb. Cortex 10 (9), 829–839.

Green, D.W., Abutalebi, J., 2013. Language control in bilinguals: the adaptive control hypothesis. J. Cognit. Psychol. 25 (5), 515–530.

Grimes, D., Tan, D.S., Hudson, S.E., Shenoy, P., Rao, R.P., 2008. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 835–844.

Grundy, J.G., 2020. The effects of bilingualism on executive functions: an updated quantitative analysis. J. Cult. Cogn. Sci. 4 (2), 177–199.

Grundy, J.G., Anderson, J.A., Bialystok, E., 2017a. Bilinguals have more complex EEG brain signals in occipital regions than monolinguals. Neuroimage 159, 280–288.

Grundy, J.G., Anderson, J.A., Bialystok, E., 2017b. Neural correlates of cognitive processing in monolinguals and bilinguals. Ann. N. Y. Acad. Sci. 1396 (1), 183–201.

Gullifer, J.W., Titone, D., 2019. Characterizing the social diversity of bilingualism using language entropy. Bilingualism Lang. Cognit. 23 (2), 283–294.

Hertzog, C., 2020. Aging and peak human performance: a glance back. J. Gerontol. Ser. B 75 (8), 1621–1624.

Hsieh, L.T., Ranganath, C., 2014. Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. Neuroimage 85, 721–729.

Jensen, O., Tesche, C.D., 2002. Frontal theta activity in humans increases with memory load in a working memory task. Eur. J. Neurosci. 15 (8), 1395–1399.

Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. 29 (2–3), 169–195.

Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn. Sci. (Regul. Ed.) 16 (12), 606–617.

Kupisch, T., Rothman, J., 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. Int. J. Bilingualism 22 (5), 564–582.

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. Psychol. Bull. 144 (4), 394.

Leivada, E., Westergaard, M., Duñabeitia, J.A., Rothman, J., 2021. On the phantom-like appearance of bilingualism effects on neurocognition:(How) should we proceed? Bilingualism Lang. Cognit. 24 (1), 197–210.

Leivada, E., Dentella, V., Masullo, C., Rothman, J., 2023. On trade-offs in bilingualism and moving beyond the Stacking the Deck fallacy. Bilingualism Lang. Cognit. 26 (3), 550–555. https://doi.org/10.1017/S1366728922000761.

Lemhöfer, K., Broersma, M., 2012. Introducing LexTALE: a quick and valid lexical test for advanced learners of English. Behav. Res. Methods 44, 325–343.

Luk, G., Bialystok, E., 2013. Bilingualism is not a categorical variable: interaction between language proficiency and usage. J. Cognit. Psychol. 25 (5), 605–621.

Lukasik, K.M., Lehtonen, M., Soveri, A., Waris, O., Jylkkä, J., Laine, M., 2018. Bilingualism and working memory performance: evidence from a large-scale online study. PLoS ONE 13 (11), e0205916.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG-and MEG-data. J. Neurosci. Methods 164 (1), 177–190.

Mazaheri, A., Slagter, H.A., Thut, G., Foxe, J.J., 2018. Orchestration of brain oscillations: principles and functions. Eur. J. Neurosci. 48 (7), 2385–2388.

Missonnier, P., Deiber, M.P., Gold, G., Millet, P., Gex-Fabry Pun, M., Fazio-Costa, L., Giannakopoulos, P., Ibáñez, V., 2006. Frontal theta event-related synchronization: comparison of directed attention and working memory load effects. J. Neural Transm. 113, 1477–1486.

Miyake, A., Friedman, N.P., 2012. The nature and organization of individual differences in executive functions: four general conclusions. Curr. Dir. Psychol. Sci. 21 (1), 8–14.

Montrul, S., 2016. The Acquisition of Heritage Languages. Cambridge University Press.

Morales, J., Yudes, C., Gómez-Ariza, C.J., Bajo, M.T., 2015. Bilingualism modulates dual mechanisms of cognitive control: evidence from ERPs. Neuropsychologia 66, 157–169.

Morrison, C., Farooq, K., Taler, V., 2019. The influence of bilingualism on working memory event-related potentials. Bilingualism Lang. Cognit. 22 (1), 191–199.

Morrison, C., Kamal, F., Le, K., Taler, V., 2020. Monolinguals and bilinguals respond differently to a delayed matching-to-sample task: an ERP study. Bilingualism Lang. Cognit. 23 (4), 858–868.

Morrison, C., Taler, V., 2020. ERP measures of the effects of age and bilingualism on working memory performance. Neuropsychologia 143, 107468.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1–9.

Paap, K., 2023. The Bilingual Advantage in Executive Functioning Hypothesis: How the Debate Provides Insight into Psychology's Replication Crisis. Taylor and Francis.

Patel, S.H., Azzam, P.N., 2005. Characterization of N200 and P300: selected studies of the event-related potential. Int. J. Med. Sci. 2 (4), 147.

Pereira Soares, S.M., Prystauka, Y., DeLuca, V., Rothman, J, 2022. Type of bilingualism conditions individual differences in the oscillatory dynamics of inhibitory control. Front. Hum. Neurosci. 16, 910910.

Polich, J., 1996. Meta-analysis of P300 normative aging studies. Psychophysiology 33 (4), 334–353.

Polinsky, M., 2018. Heritage Languages and Their Speakers. Cambridge University Press.

Pscherer, C., Bluschke, A., Mückschel, M., Beste, C., 2021. The interplay of resting and inhibitory control-related theta-band activity depends on age. Hum. Brain Mapp. 42 (12), 3845–3857.

Ren, M., Xu, J., Li, Y., Wang, M., Georgiev, G., Shen, L., Zhao, J., Cao, Z., Zhang, S., Wang, W., Xu, S., Zhou, Z., Chen, S., Chen, X., Shi, X., Shan, C., 2023. Neural signatures for the n-back task with different loads: an event-related potential study. Biol. Psychol., 108485

Rossi, E., Pereira Soares, S.M., Prystauka, Y., Nakamura, M., Rothman, J., 2023. Riding the (brain) waves! Using neural oscillations to inform bilingualism research. Bilingualism Lang. Cognit. 26 (1), 202–215.

Rothman, J., 2009. Understanding the nature and outcomes of early bilingualism: romance languages as heritage languages. Int. J. Bilingualism 13 (2), 155–163.

Rothman, J., Bayram, F., DeLuca, V., Di Pisa, G., Dunabeitia, J.A., Gharibi, K., Hau, J., Kolb, N., Kubota, M., Kupisch, T., Lamérism, T., Luque, A., van Osch, B., Pereira Soares, S.M., Prystauka, Y., Tat, D., Tomić, A., Voits, T., Wulff, S., 2023. Monolingual comparative normativity in bilingualism research is out of "control": arguments and alternatives. Appl. Psycholinguist. 1–14.

Sauseng, P., Klimesch, W., Doppelmayr, M., Pecherstorfer, T., Freunberger, R., Hanslmayr, S., 2005. EEG alpha synchronization and functional coupling during top-down processing in a working memory task. Hum. Brain Mapp. 26 (2), 148–155.

Scharinger, C., Soutschek, A., Schubert, T., Gerjets, P., 2017. Comparison of the working memory load in n-back and working memory span tasks by means of EEG frequency band power and P300 amplitude. Front Hum. Neurosci. 11, 6.

Suzuki, K., Okumura, Y., Kita, Y., Oi, Y., Shinoda, H., Inagaki, M., 2018. The relationship between the superior frontal cortex and alpha oscillation in a flanker task: simultaneous recording of electroencephalogram (EEG) and near infrared spectroscopy (NIRS). Neurosci. Res. 131, 30–35.

Teubner-Rhodes, S.E., Mishler, A., Corbett, R., Andreu, L., Sanz-Torrent, M., Trueswell, J. C., Novick, J.M., 2016. The effects of bilingualism on conflict monitoring, cognitive control, and garden-path recovery. Cognition 150, 213–231.

Titone, D.A., Tiv, M., 2022. Rethinking multilingual experience through a Systems Framework of Bilingualism. Bilingualism Lang. Cognit. 26 (1), 1–16.