Article

Influence of Bilingualism on Behavioral and Electrophysiological Parameters of Cognitive Control

No Clear Effects of Immersion, Stimulus Language, and Word Similarity

Marlene Försterling[1,2], Laura Hainke[1], Anna Redkina[1], and Paul Sauseng[1]

[1]Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany
[2]Department of Experimental Psychology, University of Oxford, UK

Abstract: Bilingualism has been associated with better cognitive control as compared to monolingualism. However, the robustness of the respective findings is subject to a recent debate, and moderators should be taken into consideration. We reasoned that groups immersed in their second language should show a greater bilingual advantage in cognitive control as compared to non-immersed participants. Further, stimulus language (first or second language), word similarity in the two languages (similar or dissimilar), as well as congruency between ink and word were varied. Forty-five participants from three different language groups (Romance, Slavic, and German) conducted a Stroop task while EEG was recorded. Higher cognitive control demand was operationalized as (1) longer reaction times, (2) higher error rates, (3) stronger N400, (4) increased Late Positive Complex (LPC), and (5) stronger Frontal Midline Theta activity. The classical Stroop interference effect was replicated for all dependent variables. Contrary to expectation, participants immersed in their second language did not exhibit any inhibition advantage in the Stroop task. Moreover, higher script similarity between first and second languages led to faster response times in general. Results are discussed in light of the current debate on the existence of a bilingual advantage in cognitive control.

Keywords: event-related potentials, frontal-midline theta, inhibition, Stroop task

Due to globalization, more than half of the global population speaks two or more languages (Grosjean & Li, 2013). During the last decades, the advantages of being bi- or multilingual have been extensively studied. It is assumed that bilingualism influences cognitive control positively (Van den Noort et al., 2019). Since both languages are simultaneously activated in bilinguals, a conflict results, which is avoided by inhibiting the language that is currently not needed (Aparicio et al., 2017; Kroll & Bialystok, 2013). This constant requirement for an inhibitory mechanism to switch between two languages constitutes training that is thought to underlie the observed advantage in conflict tasks, even in non-verbal ones (Hilchey & Klein, 2011). Beyond conflict tasks, the bilingual advantage seems to extend to other executive function components as well: according to Kroll and colleagues (2012), all three major executive functions – cognitive flexibility, working memory, and response inhibition – are positively affected by bilingualism.

The precise processes involved are subject to debate: the Bilingual Inhibitory Control Advantage hypothesis (BICA) posits that frequent use of inhibitory processes for language selection underlies bilinguals' efficient inhibition capabilities. Therefore, they, as compared to monolinguals, should display reduced interference effects, specifically in conflict trials. The Bilingual Executive Processing Advantage hypothesis (BEPA) focuses on a more general bilingual advantage across executive function domains, thus expecting performance benefits in all types of executive functions, not only conflict-related ones (Hilchey & Klein, 2011). Note that these two hypotheses are not necessarily mutually exclusive (Hannaway et al., 2019).

The claim of a bilingual advantage on cognitive control has been substantiated in empirical research, comparing bi- or multilingual to monolinguals in a wide range of tasks, such as the Simon task (e.g., Coderre & van Heuven, 2014b; Woumans et al., 2015), the Flanker task (Sorge et al., 2017), and the Stroop task (e.g., Bialystok et al., 2008; Coderre & van Heuven, 2014a). However, recent studies have failed to replicate a bilingual advantage on cognitive control (e.g., von Bastian et al., 2016; Paap & Greenberg, 2013),

Journal of Psychophysiology (2023), *37*(2), 88–100
https://doi.org/10.1027/0269-8803/a000310

and recent meta-analyses find small effects at best (Duñabeitia & Carreiras, 2015) or question the effect altogether (e.g., Lehtonen et al., 2018).

Some authors argue that the inconsistent findings can be explained by differences within the bilingual population (e.g., Bak, 2016; Duñabeitia & Carreiras, 2015). Individuals may differ with respect to immersion in their second language (L2). Living in one's second language environment should constitute frequent training in inhibiting the more automatically activated first language (L1; Coderre & van Heuven, 2014a). Accordingly, bilingual individuals should display better performances in tasks requiring inhibiting irrelevant information. Consistent with this reasoning, Yow and Li (2015) and Bonfieni and colleagues (2019) found better executive control skills in immersed bilinguals. However, others failed to find an immersion effect (e.g., Heidlmayr et al., 2014). Note that factors such as duration of immersion might need to be taken into consideration (Linck et al., 2009).

Following the same logic, an earlier age of second language acquisition should facilitate cognitive control because of more practice inhibiting the L1. Indeed, the early age of L2 acquisition has been found to relate to better executive control skills in bilinguals (Bylund et al., 2019; Tao et al., 2011; Yow & Li, 2015). Similarly, higher proficiency in L2 implies more practice in inhibiting the first language. Supporting this claim, Bonfieni and colleagues (2019) and Coderre and colleagues (2013) showed that a higher L2 proficiency predicted lower costs in an interference task. Low proficiency L2 users seem to access their L2 through their L1, while higher proficiency users may inhibit their L1 when communicating in their L2 (Luk, 2015; Tzelgov et al., 1990). The latter have more training in inhibiting, a claim supported by correlations between higher L2 proficiency and lower costs in interference tasks (Bonfieni et al., 2019; Coderre et al., 2013).

Beyond interindividual influences such as proficiency and immersion, the impact of bilingualism on cognitive control can be modulated by the similarity between two languages. Some languages use the same letters (script similarity), and some have words that sound similar (phonological similarity).

Script similarity could have opposing effects (see argumentation in Coderre & van Heuven, 2014b): on the one hand, the same script leads to greater cross-linguistic activation, and in learning to manage this, individuals acquire executive control capabilities, which underlie the bilingual advantage. This is in line with empirical results from Coderre and van Heuven (2014b). On the other hand, different-script bilinguals might experience an advantage in the Stroop task as they can use the script as a cue to help restrict lexical selection. One study showed that same-script bilinguals could experience more between-language interference during a Stroop task than different-script bilinguals (van Heuven et al., 2011).

With respect to phonological word similarity, the tendency to translate a word during a task is reduced for non-similar words (Preston & Lambert, 1969), which could lead to lower processing costs. Studies using various lexical tasks report a higher interference the more similar a foreign word is to its equivalent in participants' L1 (Allen & Conklin, 2013; Dyer, 1971; Mägiste, 1984), even if only phonologically (Sumiya & Healy, 2004). Therefore, one could expect that words that sound similar in participants' L1 and L2 lead to more interference in a lexical inhibitory control task.

In addition, when stimuli are presented in the L1, they should constitute more interference compared to stimuli presented in the L2. Since the L1 is activated more automatically and is thus more difficult to inhibit compared to an individual's L2, a higher interference of the L1 word stimulus is expected, for example, in a Stroop task (Heidlmayr et al., 2014). A higher interference has been documented empirically in various L1 tasks (Aparicio et al., 2017; Braet et al., 2011; Mägiste, 1984).

One classical inhibitory control task is the Stroop paradigm (Stroop, 1935). Participants are asked to name the ink color in which a color word is written. Ink color matches the color word (congruent trials, C) or it does not (incongruent trials, IC). Usually, as a control, neutral trials (N) in which the ink color of non-color words has to be named are included in the task. It has been consistently shown that performance in incongruent trials is worse than in neutral trials (i.e., interference effect; e.g., Kalanthroff et al., 2018). And performance in congruent trials is even better than that in neutral trials, referred to as the facilitation effect (e.g., Coderre et al., 2013). Among the different executive functions, the Stroop paradigm is usually employed to assess inhibitory control (Marian et al., 2013). The Stroop task is especially helpful for testing the effects of bilingualism on inhibitory control because it allows varying task language, as well as distinguishing between general executive or inhibition-specific advantages. Whereas interference trials specifically require inhibition, performance in all trials, regardless of congruency, does not discriminate between executive functions (Coderre & van Heuven, 2014a). Therefore, better performance collapsed across trials is an indicator of BEPA, and better performance specifically in IC-N trials of BICA (Hannaway et al., 2019).

Performance in the Stroop task as a measure of cognitive control on a behavioral level is assessed through reaction times and error rates. Neural responses can contribute to a better understanding of behavior during the Stroop task (Liotti et al., 2000); for instance, it can be inferred whether possible advantages arise on stimulus or response level. Electrophysiological correlates of cognitive control often reported in association with the Stroop task are, for instance, the event-related potential components N400 and Late Positive Complex (LPC), as well as rhythmical brain activity in the theta frequency range (around 5 Hz) obtained at frontal recording sites (Frontal-Midline Theta, FM-θ).

N400 represents medial frontal-central negativity between 350 and 500 ms after stimulus onset. A more negative N400 amplitude has been associated with increased interference (West & Alain, 2000). During the Stroop task, studies showed increased negative amplitude in IC compared to C or N trials (e.g., Coderre & van Heuven, 2014a; Zhao et al., 2015). N400 has been associated with response rather than stimulus conflict (e.g., Chen et al., 2011; Zhao et al., 2015).

The LPC most likely emerges between 600 and 900 ms after stimulus onset, around parietal regions (Ergen et al., 2014; Larson et al., 2009). Similar to N400, LPC has been consistently associated with behavioral measures of cognitive control with an increased positive amplitude occurring during the conflict (West et al., 2005; see also Kousaie & Phillips, 2012), but it seems to be less affected by practice than reaction times (Zhao et al., 2015).

Another electrophysiological correlate of cognitive control is FM-θ, occurring in the prefrontal cortex at around 4–8 Hz (Cooper et al., 2019; Sauseng et al., 2019). FM-θ is hypothesized to reflect the implementation of cognitive control and communication across different brain regions (Berger et al., 2019; Cavanagh & Frank, 2014). There is compelling empirical evidence relating FM-θ to cognitive control in tasks such as the Simon task (e.g., van Driel et al., 2015), Flanker task (e.g., Mückschel et al., 2016), and Stroop task (e.g., Hanslmayr et al., 2008).

Whereas N400 seems well suitable for investigating inhibition-specific processes, FM-θ does not respond selectively to inhibitory control but is involved in identifying conflict (Hanslmayr et al., 2008). FM-θ is even discussed as a rather global mechanism during cognitive control – specifically involved in several forms of proactive as well as reactive cognitive control (see, e.g., Cooper et al., 2019; Kaiser et al., 2019; McKewan et al., 2020; Sauseng et al., 2019). Therefore, in addition to behavioral parameters, these measures could be particularly helpful in dissociating BICA and BEPA.

In an EEG study using a Stroop paradigm, Hannaway and colleagues (2019) investigated some factors potentially contributing to the bilingual cognitive advantage. Instead of comparing bi- and monolinguals, they compared participants who were immersed in their second language and participants who were not. The immersed group did not outperform the non-immersed one on a behavioral level. The authors argue that the lack of an immersion effect might be due to the fact that the English native speakers inadvertently have not been more immersed than the German speakers because both groups reported no difference in their use of L1. For each participant, stimuli were presented in their mother tongue (L1) as well as in their second language (L2), with some stimuli sounding similar between those two languages and others dissimilar. Similar trials required more inhibitory control than dissimilar stimuli on a neuronal and behavioral level. In order to better understand the underlying mechanisms, namely BICA and BEPA, they compared the N400 component, which represents response conflict, and the LPC component, which rather represents stimulus conflict (Zhao et al., 2015). Part of the results were consistent with BICA, others with BEPA. Given the disparity of findings in the field, it remains an open question whether a bilingual advantage for cognitive control is robust and to what extent it is moderated by additional factors. In the present study, we aimed to reproduce and extend the findings of Hannaway and colleagues (2019) with different and more language groups and three neuronal correlates, namely N400, LPC, and FM-θ. We, too, studied the effects of bilingualism on cognitive control with a Stroop task while simultaneously recording EEG. We tested three (quasi-) experimental groups, with different mother tongues (Slavic, Romance, or German), with all participants currently living in Germany and having a second language level of at least B2. Therefore, the Slavic and Romance language groups were currently immersed in their L2, and the Germans constituted a non-immersed control group. Participants were presented congruent, incongruent, and neutral trials. Each participant was tested in their L1 and L2, with words sounding either similar (S) or dissimilar (DS) between these two languages. For each group, we measured the costs in the Stroop task as an operationalization of cognitive control. High costs are indicated by (1) longer reaction times, (2) higher error rates, (3) more negative N400 amplitudes, (4) more positive LPC amplitudes, and (5) a stronger FM-θ activity.

In our study, we aimed to replicate the classical findings of the Stroop task, that is, incongruent trials should elicit higher costs than neutral or congruent trials. Second, we assumed that the two groups immersed in their second language (Slavic and Romance) showed less costs in the Stroop task compared to the German group, which was not immersed.

Moreover, we expected that groups with the same script in L1 and L2 (Romance and German) would differ in costs from the group with different scripts (Slavic) in the Stroop task. Furthermore, we expected that similar words require more costs than dissimilar sounding words. Finally, we expected less interference in L2 than in L1 trials.

In addition, we planned exploratory correlation analyses between the costs in the Stroop task and three predictors of cognitive control that have been identified in the literature. We expected less costs in the L2 Stroop task; the longer a person was immersed in their L2, the earlier the person has acquired their second language (i.e., age of acquisition) and the more proficient a person was in their L2.

Compatible with the BEPA account, lower costs for bilinguals should be found collapsed across all congruency conditions, while BICA predicts less costs for bilinguals only when interference is given (IC-N trials).

Methods

Design

A mixed design consisting of one between-subject factor and three within-subject factors was implemented. The between-subject factor was the first language of participants (German, Romance, or Slavic). Here, we analyzed L2 immersion (Germans as a non-immersed group versus Romance and Slavic immersed) and script similarity between L1 and L2 (Slavic group with low script-similarity versus German and Romance with high script-similarity). The within-subject factors were task language (words appeared either in the mother tongue [L1] or the second language [L2]), congruency between presented words and their ink color (incongruent [IC], neutral [N] or congruent [C]); and word similarity between task languages (words sounded either similar [S] or dissimilar [DS] in the two task languages).

Participants

We calculated the necessary sample size a priori according to the effect size reported in Hannaway and colleagues (2019) of $f = .25$ with the parameters $\alpha = .05$, $1 - \beta = .8$ for a repeated measures ANOVA) using *G*Power* software (Faul et al., 2007). The analysis revealed a minimum of 14 participants per group. As a precautionary measure, we recruited one more person in each group, resulting in 45 participants in total.

Inclusion criteria for participants were age (18–45 years) and an L2 proficiency level of at least B2 (Common European Framework of Reference; Council of Europe, 2001). The proficiency level of L2 was assessed via self-report.

Two German participants were excluded because they did not fulfill the L2 proficiency criterion and one Slavic participant due to technical problems. Thus, our sample included 42 participants (13 male and 29 female, M_{age} = 24.55 years, SD = 5.06). The Romance group included 15 individuals (6 male and 9 female): 5 Portuguese, 5 Spanish, and 5 French natives. The Slavic group consisted of 14 participants (1 male and 13 female) with Russian as their L1. Both groups spoke German as L2. These two groups did not differ significantly in the duration of immersion and L2 proficiency level. The German group comprised 13 individuals (6 male and 7 female) with Spanish or French as their L2. The Slavic group was significantly older than the other groups (M = 28.29, SD = 5.66). Self-reported L2 proficiency in the German group was significantly lower than in the other two groups, as was the frequency of L2 use. For EEG analyses, four additional participants (two Slavic, one German, and one Romance) were excluded due to a large number of EEG artifacts (more than 30 out of 60 trials per block). Sample characteristics are summarized in Table 1.

All participants were based in Munich, Germany. They were recruited through personal contact, flyers, or online. The study was approved by the local ethics committee and conducted according to the Declaration of Helsinki (World Medical Association, 2013).

Materials

A pre-screening questionnaire was presented on the SoSci Survey (Leiner, 2019) website. Participants indicated their respective L1 and L2, L2 proficiency, age of L2 acquisition, and current usage of L1 and L2, as well as the duration of residence in Germany.

The Stroop task was presented via NBS presentation 0.71 software (Neurobehavioral Systems, 2019). All stimuli were presented in the center of a 22″ monitor on a gray background. Each trial started with a black fixation cross appearing for 500 ms, followed by a blank screen for 300 ms; the stimulus was then shown for 1,000 ms, with an inter-trial interval of 2,000–2,500 ms. Each participant underwent four stimulus blocks: L1-S, L1-DS, L2-S, and L2-DS. Each block contained 180 trials: 60 C, 60 IC, and 60 N (instead of a word, a string of symbols was displayed; e.g., "%%%%"), preceded by a practice block of 27 trials. Blocks with similar words used the colors orange, pink, and purple in one of the languages (German: Orange, Rosa, Violett; Portuguese: Laranja, Rosa, Violeta; Spanish: Naranja, Rosa, Violeta; French: Orange, Rose, Violet; Russian: Оранжевый, Розовый, Фиолетовый). The dissimilar words were black, green, and yellow (German: Schwarz, Grün, Gelb; Portuguese: Preto, Verde, Amarelo; Spanish: Negro, Verde, Amarillo; French: Noir, Vert, Jaune; Russian: Чёрный, Зелёный, Жёлтый). As response keys, the numbers 1–3 on the number pad were used in similar word blocks, the numbers 7–9 in dissimilar ones. Each button had a sticker displaying the respective color.

EEG data were recorded using a BrainAmp amplifier (BrainProducts, Gilching, Germany), 32-channel Ag/AgCl electrode caps (Easycap, Herrsching, Germany), and Brain Vision Recorder software (BrainProducts, Gilching,

Table 1. Sociodemographic and linguistic characteristic of participants by group

	German (*n* = 13)			Romance (*n* = 15)			Slavic (*n* = 14)		
Variable	Mean	*SD*	Range	Mean	*SD*	Range	Mean	*SD*	Range
Demographic									
Gender	1.46	0.52	1.0	1.40	0.51	1.0	1.07	0.27	1.0
Age	21.46	1.45	5.0	23.73	4.48	15.0	28.00	5.72	19.0
Handedness	0.85	0.38	1.0	0.53	0.74	2.0	0.86	0.36	1.0
L2 variables									
Immersion duration*	18.31	7.10	23.0	3.55	3.18	10.7	7.00	6.35	18.0
Age of acquisition	10.62	4.89	16.0	11.00	6.39	19.0	17.21	8.30	30.0
Proficiency	4.54	0.78	2.0	5.33	0.82	2.0	5.29	0.73	2.0
Usage	2.15	1.14	3.0	4.80	0.41	1.0	4.86	0.53	2.0

Note. Gender: 1 = female; 2 = male. Handedness: −1 = left-handed; 0 = ambidextrous; 1 = right-handed. Age, immersion duration and age of acquisition in years. Proficiency: 1 = min; 6 = max (A1–C2, CEFR). *For the German group, immersion duration refers to L1 immersion (not L2 as for the other groups), since all participants were living in Germany at the time of the experiment.

Germany). EEG and behavioral data were processed with Brain Vision Analyser 2.1 software (BrainProducts, Gilching, Germany) and statistically analyzed in RStudio (RStudio Team, 2015).

Procedure

EEG activity was recorded during the Stroop task. The experiment consisted of 4 blocks (L1-S, L1-DS, L2-S, and L2-DS) with 180 trials each. Blocks were counterbalanced across participants within each group. Before each block, participants were informed about the language and the relevant colors of the next block.

EEG data were recorded using 30 Ag/AgCl electrodes positioned according to the 10–20 system mounted in the elastic cap and two additional electrodes on the left and right ear lobes to re-reference the signal post-measurement (Easycap, Herrsching, Germany). The reference was set at FCz, the ground at AFz. Impedances were kept below 15 kΩ. Data were recorded at a 500 Hz sampling frequency and a resolution of 0.1 μV, with an online bandpass filter of 0.016–250 Hz and a notch filter at 50 Hz.

Data Analysis

Behavioral Data

We calculated median reaction times (RTs) and the percentage of correct responses (accuracy) for each condition. As suggested by Hannaway and colleagues (2019), the contrast between incongruent and neutral trials (IC-N) was calculated as an indication of an interference effect (BICA). To test for a general advantage (BEPA), trials were collapsed across congruency conditions. Since the Stroop task is not the best paradigm for dissociating inhibition from other executive functions such as shifting or updating, it can be debated whether this approach suggested by Hannaway and colleagues (2019) is optimal. To keep our study comparable with Hannaway and colleagues (2019), we nevertheless followed their suggested analysis protocol.

We calculated repeated-measures ANOVAs to test for congruency effect (IC, C, and N), and ANOVAs with within-subject factors similarity of color terms (S, DS) and task language (L1, L2), and between-subject factor L1-group (German, Romance, Slavic), for dependent variables RT and accuracy. If a Mauchly test indicated a violation of the sphericity assumption, Greenhouse-Geißer corrected p-values (p_{GG}) are reported. We calculated post hoc *t*-tests and applied a Bonferroni correction for unplanned contrasts. Spearman rank tests were used to test correlations between L2 factors (age of acquisition, usage, proficiency, and immersion duration) and dependent variables in L2 blocks for the two immersed groups.

EEG Data

EEG data were preprocessed using Brain Vision Analyzer 2.1. A zero phase-shift filter (IIR Filter) was applied to the raw EEG, with a low cutoff of 0.1 Hz and a high cutoff of 30 Hz. All channels were then re-referenced to digitally linked earlobe electrodes, A1 and A2. To control for eye-movement artifacts, the horizontal (hEOG) and vertical (vEOG) oculomotor channels were interpolated using linear derivation. The hEOG was calculated from F7 and F8 electrodes. To calculate the vEOG, we used the data from Fp1 and Fp2 electrodes. These estimates were used for automatic ocular correction Independent Component Analysis (ICA) as implemented in BrainVision Analyzer 2.1. Because visual inspection of the data showed some high-frequency noise at electrode locations Fp1 and Fp2, those were interpolated for all participants using a spline interpolation.

Raw data were inspected for artifacts based on the following criteria: (a) maximal allowed amplitude steps of 20 μV/ms, (b) maximal allowed amplitude differences of 200 μV in 200 ms, (c) a minimal allowed amplitude of

−200 μV and maximal allowed amplitude of +200 μV. EEG not fulfilling these criteria was marked as an artifact from 200 ms before the occurrence of this signal to 200 ms after the end of this diverging signal.

Each participant's preprocessed EEG was divided into segments, from 300 ms before stimulus onset to 800 ms after (for FM-θ) and 1,500 ms after (for ERP analyses). For ERP analyses, segments were baseline-corrected with respect to the 300 ms before stimulus onset. Artifact-free segments were then averaged for each participant and each level of congruency separately. On average, a similar number of artifact-free trials remained for analysis for the twelve experimental conditions (48.9, 51.1, 51.3, 47.7, 51.5, 50.5, 52.7, 45.3, 48.4, 47.8, 46.1 and 52.2 trials for conditions L1_S_C, L1_DS_C, L1_S_I, L1_DS_I, L1_S_N, L1_DS_N, L2_S_C, L2_DS_C, L2_S_I, L2_DS_I, L2_S_N, and L2_DS_N, respectively).

For N400 amplitudes, data for the time interval of 350–500 ms after stimulus onset were averaged. We used data from Cz, C3, C4, CP1, CP2, and Pz electrodes, based on previous studies (e.g., Coderre et al., 2013; Liotti et al., 2000). LPC was averaged over the time interval of 600–900 ms post-stimulus at electrode sites CP1, CP2, P3, P4, and Pz (Hannaway et al., 2019).

In order to calculate FM-θ, the sampling rate of 500 Hz was converted into 512 Hz using spline interpolation. Thereafter, Laplacian current source density was calculated as suggested for FM-θ (Berger et al., 2019, Griesmayr et al., 2014; McKewan et al., 2020). Data were segmented 300–800 ms after stimulus onset. Fast Fourier Transformation was applied, and segments were averaged for each participant and each condition separately, resulting in amplitude spectra with a 2 Hz frequency resolution. FM-θ was obtained at electrode site Fz at a frequency range between 4 and 8 Hz (similar to Berger et al., 2019; Griesmayr et al., 2014; McKewan et al., 2020).

Results

Reaction Time

Collapsing across groups and blocks, a one-way ANOVA showed a significant congruency effect ($F[2, 82] = 98.86$, $p_{GG} < .001$, $\eta^2 = .06$). Stroop interference and facilitation effects were confirmed by paired *t*-tests, with higher RTs in IC trials ($M_{IC} = 588.28$, $SE_{IC} = 7.47$) compared to N ($M_N = 554.62$, $SE_N = 6.56$, $t[167] = 11.75$, $p < .001$) and C trials ($M_C = 538.49$, $SE_C = 6.01$, $t[167] = 17.39$, $p < .001$) and lower RTs in C compared to N trials ($t[167] = 8.47$, $p < .001$).

Collapsing across congruency levels, data were analyzed using a 3-way mixed ANOVA with the factors native language (Romance/Slavic/German), stimulus language (L1/L2), and similarity of color terms (S/DS). There was a main effect of native language ($F[2, 39] = 3.84$, $p = .03$, $\eta^2 = .15$): post hoc *t*-tests indicated that on average, the German group ($M_{Ger} = 517.34$, $SE_{Ger} = 6.46$) reacted faster than both the Romance [$M_{Rom} = 562.8$, $SE_{Rom} = 5.78$, $t[323] = -5.25$, $p < .001$) and the Slavic group [$M_{Slav} = 598.0$, $SE_{Slav} = 7.05$, $t[321] = -8.44$, $p < .001$), with lower RTs in the Romance group than in the Slavic one ($t[328] = -3.86$, $p < .001$).

A 3-way mixed ANOVA with the same factors was conducted for median RT differences between IC and N trials. Here, the interaction between the three factors reached significance ($F[2, 39] = 3.25$, $p = .049$, $\eta^2 = .02$). In the German group, IC-N differences in DS blocks were larger for L1 than L2, whereas in L1 blocks, they were larger for DS than S. After Bonferroni-corrections, the paired post hoc *t*-tests no longer revealed significant differences.

Spearman rank tests for correlations between median RTs in L2 blocks and L2-related factors (age of acquisition, usage, proficiency, and immersion duration) did not yield any significant results, neither for global RTs nor for IC-N differential values.

Task Accuracy

Collapsing across groups and blocks, a one-way ANOVA on the percentage of correct responses also showed a significant congruency effect ($F[2, 82] = 13.68$, $p < .001$, $\eta^2 = .12$). Again, interference as well as facilitation effects were found in paired t-tests, with a lower *M* percentage of correct answers in IC trials ($M_{IC} = 96.63$, $SE_{IC} = 0.32$) compared to N [$M_N = 97.83$, $SE_N = 0.18$, $t[167] = 3.64$, $p < .001$) and C trials [$M_C = 98.51$, $SE_C = 0.16$, $t[167] = 6.04$, $p < .001$) and lower percent correct values in N than C trials ($t[167] = 3.37$, $p < .001$).

A 3-way mixed ANOVA with the factors native language, stimulus language, and similarity of color terms was conducted for percent correct values collapsed across congruency. This yielded a significant similarity main effect ($F[1, 39] = 12.29$, $p < .001$, $\eta^2 = .05$) with higher percent correct values for S ($M_S = 98.11$, $SE_S = 0.17$) than DS blocks ($M_{DS} = 97.21$, $SE_{DS} = 0.22$) on average.

The equivalent ANOVA for the IC–N difference value of percent correct yielded a significant interaction between the three factors native language, stimulus language, and similarity of color terms ($F[2, 39] = 6.03$, $p = .005$, $\eta^2 = .04$). Post hoc paired *t*-tests revealed that for L1, the difference in the percentage of correct answers was higher for S than DS in the German group and higher for DS than S in the Romance group, but these effects were no longer significant after Bonferroni-correction.

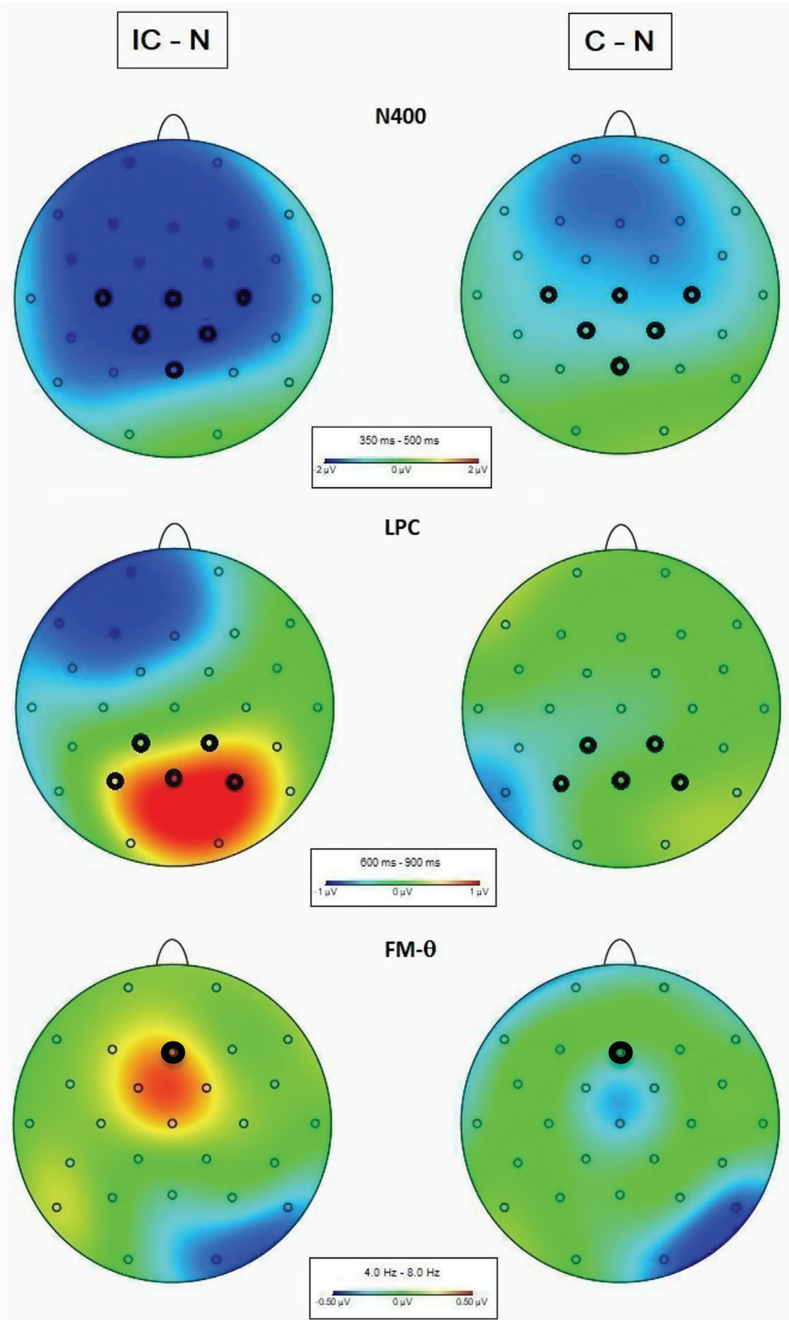

Figure 1. Congruency effects for ERP components N400 and LPC as well as FM-θ activity. All three EEG parameters show a significant interference effect (IC-N; left) but no facilitation effect (C-N; right). For statistical analysis, electrode sites highlighted in bold were averaged for N400 and LPC; for FM-θ activity, amplitude values at electrode site Fz (highlighted in bold) were used for statistical analysis.

Spearman rank tests for correlations between *M* percentages of correct responses in L2 blocks and the four L2-related factors did not yield any significant results, neither for global percent correct nor for IC-N differential values.

N400

A one-way ANOVA analyzing the N400 amplitude collapsed across all blocks and groups showed a difference in amplitude depending on congruency levels ($F[2, 74]$ = 32.43, $p_{GG} < .001$, $\eta^2 = .01$). Post hoc paired *t*-tests showed a significant interference effect ($t[37] = 7.48$, $p < .001$). However, we did not find a facilitation effect: on the contrary, the average amplitude was more negative for C than for N trials ($t[37] = 4.28$, $p < .001$) (Figures 1 and 2A).

Regarding the 3-way mixed ANOVA for N400 amplitudes collapsed across congruency conditions with factors of native language, stimulus language, and similarity of color terms, it revealed no significant effects. As for the analogue 3-way ANOVA for the IC-N amplitude difference, a significant interaction between the native language and stimulus language was found ($F[2, 35] = 3.76$, p = .03, $\eta^2 = .04$). Post hoc *t*-tests showed that the IC-N

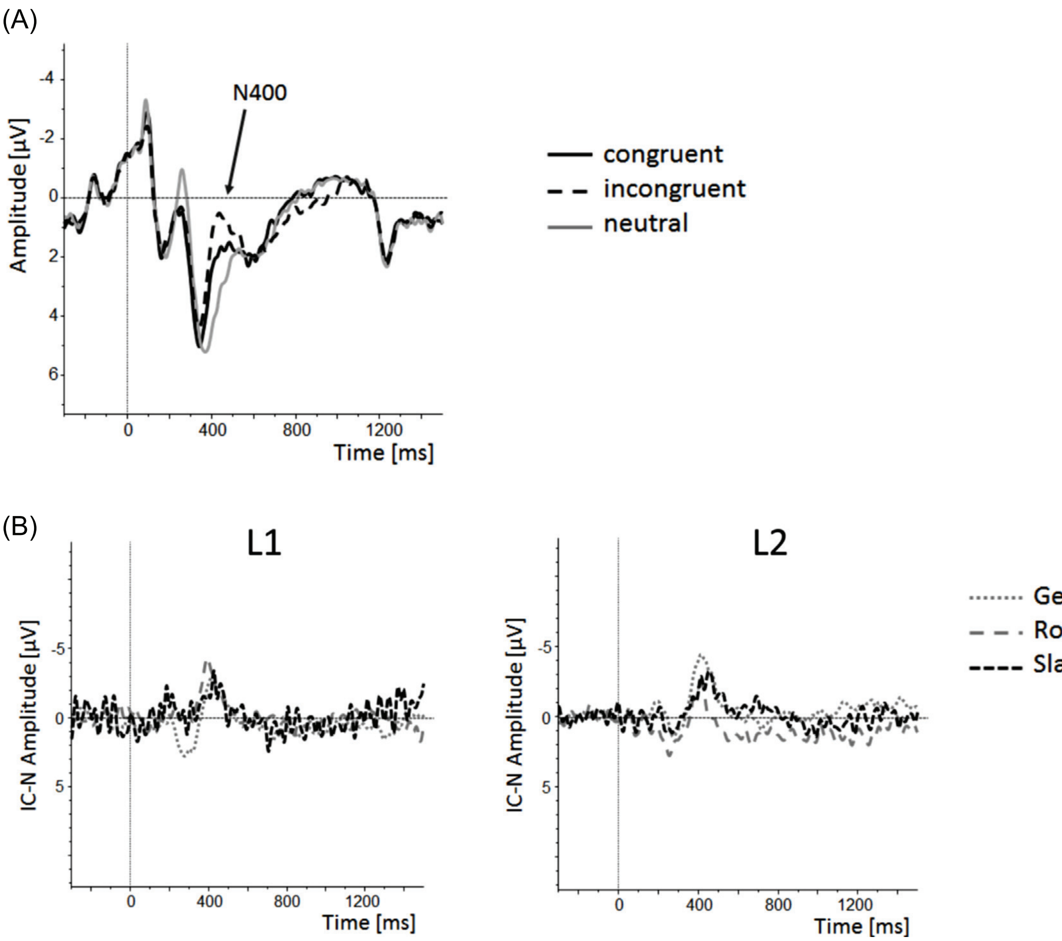

Figure 2. (A) ERP traces for congruent, incongruent, and neutral conditions across all participants averaged for electrode sites C3, Cz, C4, CP1, CP2, and Cz. Note the interference effect for the N400 component between 350 and 500 ms post-stimulus. (B) Difference waves between incongruent and neutral conditions in L1 and L2 for German, Romance, and Slavic groups. Note the reduced N400 congruency effect in L1 and the enlarged congruency effect in L2 for the German group.

difference for L2 was larger for the German group than for the other two groups; also, in the German group, the difference was larger for L2 than for L1 (Figure 2B). These effects were no longer significant after Bonferroni-correction.

Spearman rank tests for correlations between N400 amplitudes in L2 blocks and the four L2-related factors did not yield any significant results, neither for global N400 amplitudes nor for IC-N differential values.

LPC

A congruency effect was also found for LPC in a one-way ANOVA ($F[2, 74] = 3.8$, $p_{GG} = .034$, $\eta^2 = .005$) with paired *t*-tests confirming the Stroop interference effect in IC trials compared to N ($t[37] = 2.16$, $p = .02$) and C ($t[37] = 2.2$, $p = .02$) but not a facilitation effect ($p > .05$) (Figures 1 and 3A).

In the 3-way mixed ANOVA with LPC amplitudes collapsed across congruency as the dependent variable, an interaction between the three factors native language, stimulus language, and similarity of color terms was significant ($F[2, 35] = 3.4$, $p = .045$, $\eta^2 = .012$). Post hoc paired *t*-tests showed that for S, amplitudes were higher in L1 for the German group and higher in L2 for the Slavic group; after applying a Bonferroni-correction, these comparisons were no longer significant. For the equivalent ANOVA with the IC-N amplitude difference as the dependent variable, no effects were found.

Spearman rank tests for correlations between LPC amplitudes in L2 blocks and the four L2-related factors did not yield significant results for IC-N differential values. Global LPC amplitudes in L2 blocks correlated positively only with L2 proficiency ($r_s = 0.31$, $p = .03$) and L2 usage ($r_s = 0.318$, $p = .03$); however, these correlations were no longer significant after Bonferroni correction.

FM-θ

FM-θ amplitudes differed between congruency conditions, indicated by the results of a one-way ANOVA ($F[2, 74] = 6.01$, $p = .004$, $\eta^2 = .003$). Paired *t*-tests revealed a significant interference effect. The average amplitude was higher in IC trials than in N ($t[38] = 2.53$, $p = .016$) and

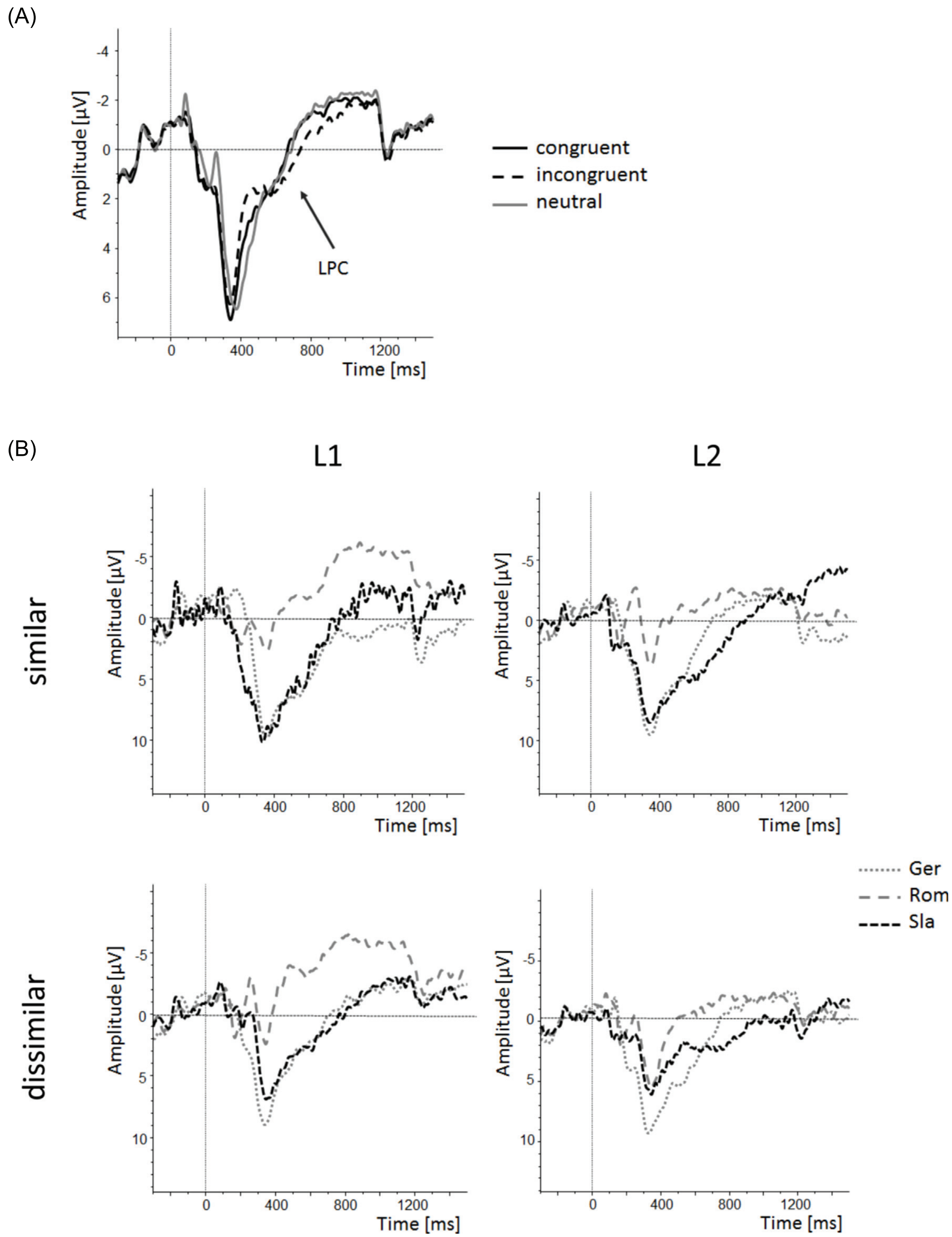

Figure 3. (A) ERP traces for congruent, incongruent, and neutral conditions across all participants averaged for electrode sites CP1, CP2, P3, Pz, and P4. Note the congruency effect for the LPC component between 600 and 900 ms post-stimulus. (B) LPC effect between German, Romance, and Slavic groups as a function of script language and similarity of color words. Note that for similar color words, the German group showed a stronger LPC in L1, whereas this was the case for the Slavic group in L2.

C trials (t[38] = 3.30, p = .002). There was no facilitation effect: the amplitudes did not significantly differ between C and N trials (p > .05) (Figure 1).

Collapsing across congruency conditions, a 3-way mixed ANOVA with factors of native language, stimulus language, and similarity of color terms did not reveal any significant effects. The same applies to the equivalent ANOVA for IC-N differential values.

Spearman rank tests for correlations between FM-θ power in L2 blocks and the four L2-related factors did not yield any significant results, neither for global FM-θ amplitudes nor IC-N differential values.

Discussion

The present study investigated the effects of factors associated with bilingualism on cognitive control, using a Stroop task while simultaneously recording EEG. We were able to replicate the classical Stroop effect. We found a robust interference effect, i.e., incongruent trials produced more costs than congruent and neutral trials for all dependent variables. We found a facilitation effect on a behavioral level. This pattern of results constitutes a plausibility check of our task and our dependent variables. Note, though, that we did not find evidence for a facilitation effect on a neural level. In contrast, the N400 amplitude for C trials was even more negative than for N trials, which is consistent with Hannaway and colleagues (2019).

We predicted that groups that are immersed in their L2 perform better in the Stroop task. Therefore, the Romance and Slavic groups should outperform the German group. There was an effect of L1 on RTs when trials collapsed across congruency conditions. However, contrary to our expectations, the German group was faster than the Romance and Slavic groups. We did not find an effect of immersion on RTs calculated for IC-N trials, error rates, or neuronal activity. The lack of an effect for IC-N trials suggests that the L1 effect on RTs is not specific to inhibition but rather constitutes a broader advantage in RTs (for the German group). This might be due to group characteristics that were unintended. On average, the German group was the youngest, followed by the Romance and the Slavic group. In an exploratory ANCOVA analysis, including age as a covariate, the effect of the group is indeed no longer significant ($p = .24$). Thus, it is conceivable that age slowed down overall RTs. This seems likely because age-related RT differences in executive control tasks have been empirically documented (e.g., Bialystok et al., 2008). Note that in the study by Hannaway and colleagues (2019), immersed participants also showed higher costs, even though their groups did not differ in age.

Moreover, there were no significant correlations between the duration of immersion and any dependent variables. Therefore, our study challenges the assumption of a positive effect of immersion on the bilingual advantage and is in line with the results obtained by Hannaway and colleagues (2019), who also did not find a positive behavioral effect of immersion in L2. Hannaway and colleagues explained their findings with a lack of difference in the degree of L2 usage between their immersed and their non-immersed group. In our study, the immersed and non-immersed groups differed in their degree of L2 usage ($p < .001$), and we still did not find any positive immersion effect.

Note that participants recruited in the study by Hannaway and colleagues (2019) and our study have been exposed to their second language for a long time. Heidlmayr and colleagues (2014) argued that cognitive control is only necessary at the beginning of an immersion period, which might explain the lack of an immersion effect here. That is, there might be a curvilinear rather than a linear effect of immersion. Moreover, immersion in a second language might strongly vary in time. For instance, one might feel much more immersed when writing a paper in L2 than when having family or friends over for some time and mainly using L1. These rather state-like than trait-like immersion effects would need to be addressed in future research.

There was an effect of word similarity on the percentage of correct answers collapsed across congruency levels; that is, when words sounded similar between the two languages, there were more correct answers than when words sounded dissimilar. This pattern is contrary to our expectation that similar color words would lead to more interference between the two languages, resulting in higher costs. One possible explanation for this pattern is that similar items require less translation time. Consistent with this reasoning, similarity sped up response times most in congruent trials, i.e., trials with minimal conflict.

A special feature of the present study is the inclusion of a language group (Slavic) with a different scripture. This group displayed the highest reaction times collapsed across congruency conditions, as was also observed by Coderre and van Heuven (2014b). If individuals with a lower orthographic overlap trained inhibitory control less, the effect should have manifested in the IC-N analyses. It, therefore, seems plausible that it occurred due to a higher mean age. The unexpected effect of script similarity must also be interpreted with caution, as a German (L1) Slavic (L2) group could not be recruited.

Finally, we expected L1 stimuli to elicit more costs than L2 stimuli because L2 stimuli should constitute less interference. However, there was no effect of task language on any of the dependent variables. This contradicts the results found by Aparicio and colleagues (2017), Braet and colleagues (2011), and Liu (2007). One possible explanation for this discrepancy is that L2 proficiency was rather high in our sample, with its mean level slightly above C1 (Council of Europe, 2001), and L2, therefore, might approach the L1 level in its automaticity. In line with this contention, Mägiste (1984) showed that balanced bilinguals did not differ in their inhibition costs for any of their two spoken languages.

On an exploratory basis, we asked participants to report their age of L2 acquisition, L2 proficiency, and L2 immersion duration. We did not find any correlation between these factors with performance measures in the L2 Stroop task. This is consistent with the study of von Bastian and colleagues (2016), which did not find any effect of these factors with a stronger bilingual advantage.

It is surprising that, in contrast to Hannaway and colleagues (2019), no main effects were found when the IC-N difference was the dependent variable. It seems that our results fit better into the BEPA framework, which expects general executive divergences, not interference-specific ones. This also supports the general trend found in the literature (see, e.g., Hilchey & Klein, 2011).

There were some noteworthy differences between the present study and the one by Hannaway and colleagues (2019). First, our sample size was larger. Lehtonen and colleagues (2018) argued in their review that the bilingual advantage was typically found only in studies with small sample sizes, possibly because of a publication bias. Nonetheless, a larger sample size might be desirable to reduce the possibility of false negatives accounting for null effects in the present data. Second, we included a wider range of languages. Recruiting a Slavic language group for our experiment made it possible to test for a difference in the script. It is possible that the English and German groups that were tested in Hannaway and colleagues' (2019) study were more homogenous than our samples. Our participant groups differed in age and gender, and it is also possible that there are more cultural differences between Slavic and Romance groups and a German group as compared to an English-speaking group. It is noteworthy that the Romance group consisted of three different language groups, which were French, Spanish, and Portuguese. Testing more heterogeneous groups might mask some of the previously reported effects. A limiting factor might be that in the present study, participants achieved a percentage of correct responses close to the ceiling (M = 97.67%). Our task might have been too easy to stress participants' cognitive capacities, possibly due to a large number of trials. Moreover, a wider range of paradigms beyond the Stroop task might be necessary to properly investigate the impact of bilingualism on inhibition, specifically, or more generally, on executive functions (such as shifting and updating in addition).

Conclusion

In light of the ongoing debate about the robustness of a bilingual advantage, the present study was designed based on Hannaway and colleagues (2019) while analyzing different language groups (German, Slavic, Romance) as well as further linguistic and neural variables.

Stroop effects were present on a neural and behavioral level. The postulated inhibitory advantage of L2-immersed bilinguals was not confirmed: as in Hannaway and colleagues' (2019) study, RTs showed the opposite pattern. The same applies to word similarity, which was expected to increase interference but led to a higher percentage of correct answers, as in Hannaway and colleagues (2019). Participants with higher script similarity between their L1 and L2 had faster RTs than participants with different scripts. Lastly, neither task language nor L2-related factors affected costs in the Stroop task.

In sum, while some explored factors did not influence costs in the Stroop task, others produced unexpected patterns. These mixed results highlight the complexity of the effect that bilingualism is claimed to have on executive inhibitory functions.

References

Allen, D. B., & Conklin, K. (2013). Cross-linguistic similarity and task demands in Japanese-English bilingual processing. *PLoS One, 8*(8), Article e72631. https://doi.org/10.1371/journal.pone.0072631

Aparicio, X., Heidlmayr, K., & Isel, F. (2017). Inhibition efficiency in highly proficient bilinguals and simultaneous interpreters: Evidence from language switching and Stroop tasks. *Journal of Psycholinguistic Research, 46*(6), 1427–1451. https://doi.org/10.1007/s10936-017-9501-3

Bak, T. H. (2016). Cooking pasta in La Paz: Bilingualism, bias and the replication crisis. *Linguistic Approaches to Bilingualism, 6*(5), 699–717. https://doi.org/10.1075/lab.16002.bak

Berger, B., Griesmayr, B., Minarik, T., Biel, A. L., Pinal, D., Sterr, A., & Sauseng, P. (2019). Dynamic regulation of interregional cortical communication by slow brain oscillations during working memory. *Nature Communications, 10*(1), Article 4242. https://doi.org/10.1038/s41467-019-12057-0

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), 859–873. https://doi.org/10.1037/0278-7393.34.4.859

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta Psychologica, 193*, 160–170. https://doi.org/10.1016/j.actpsy.2018.11.004

Braet, W., Noppe, N., Wagemans, J., & Op de Beeck, H. (2011). Increased Stroop interference with better second-language reading skill. *Quarterly Journal of Experimental Psychology, 64*(3), 596–607. https://doi.org/10.1080/17470218.2010.513735

Bylund, E., Abrahamsson, N., Hyltenstam, K., & Norrman, G. (2019). Revisiting the bilingual lexical deficit: The impact of age of acquisition. *Cognition, 182*, 45–49. https://doi.org/10.1016/j.cognition.2018.08.020

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences, 18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Chen, A., Bailey, K., Tiernan, B. N., & West, R. (2011). Neural correlates of stimulus and response interference in a 2–1 mapping stroop task. *International Journal of Psychophysiology, 80*(2), 129–138. https://doi.org/10.1016/j.ijpsycho.2011.02.012

Coderre, E. L., & van Heuven, W. J. B. (2014a). Electrophysiological explorations of the bilingual advantage: Evidence from a Stroop task. *PLoS One, 9*(7), Article e103424. https://doi.org/10.1371/journal.pone.0103424

Coderre, E. L., & van Heuven, W. J. B. (2014b). The effect of script similarity on executive control in bilinguals. *Frontiers in Psychology, 5*, Article 1070. https://doi.org/10.3389/fpsyg.2014.01070

Coderre, E. L., van Heuven, W. J. B., & Conklin, K. (2013). The timing and magnitude of Stroop interference and facilitation in monolinguals and bilinguals. *Bilingualism: Language and Cognition, 16*(2), 420–441. https://doi.org/10.1017/S1366728912000405

Cooper, P. S., Karayanidis, F., McKewen, M., McLellan-Hall, S., Wong, A. S. W., Skippen, P., & Cavanagh, J. F. (2019). Frontal theta predicts specific cognitive control-induced behavioural changes beyond general reaction time slowing. *NeuroImage, 189*, 130–140. https://doi.org/10.1016/j.neuroimage.2019.01.022

Duñabeitia, J. A., & Carreiras, M. (2015). The bilingual advantage: Acta est fabula? *Cortex, 73*, 371–372. https://doi.org/10.1016/j.cortex.2015.06.009

Dyer, F. N. (1971). Color-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior, 10*(3), 297–302. https://doi.org/10.1016/S0022-5371(71)80057-9

Ergen, M., Saban, S., Kirmizi-Alsan, E., Uslu, A., Keskin-Ergen, Y., & Demiralp, T. (2014). Time–frequency analysis of the event-related potentials associated with the Stroop test. *International Journal of Psychophysiology, 94*(3), 463–472. https://doi.org/10.1016/j.ijpsycho.2014.08.177

Council of Europe, Council for Cultural Co-operation, Education Committee, Modern Languages Division. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Griesmayr, B., Berger, B., Stelzig-Schoeler, R., Aichhorn, W., Bergmann, J., & Sauseng, P. (2014). EEG theta phase coupling during executive control of visual working memory investigated in individuals with schizophrenia and in healthy controls. *Cognitive, Affective and Behavioral Neuroscience, 14*, 1340–1355. https://doi.org/10.3758/s13415-014-0272-0

Grosjean, F., & Li, P. (2013). *The psycholinguistics of bilingualism*. Wiley.

Hannaway, N., Opitz, B., & Sauseng, P. (2019). Exploring the bilingual advantage: Manipulations of similarity and second language immersion in a Stroop task. *Cognitive Neuroscience, 10*(1), 1–12. https://doi.org/10.1080/17588928.2017.1389874

Hanslmayr, S., Pastötter, B., Bäuml, K. H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *Journal of Cognitive Neuroscience, 20*(2), 215–225.

Heidlmayr, K., Moutier, S., Hemforth, B., Courtin, C., Tanzmeister, R., & Isel, F. (2014). Successive bilingualism and executive functions: The effect of second language use on inhibitory control in a behavioural Stroop Colour Word task. *Bilingualism: Language and Cognition, 17*(03), 630–645. https://doi.org/10.1017/S1366728913000539

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review, 18*(4), 625–658. https://doi.org/10.3758/s13423-011-0116-7

Kaiser, J., Simon, N. A., Sauseng, P., & Schütz-Bosbach, S. (2019). Midfrontal neural dynamics distinguish between general control and inhibition-specific processes in the stopping of motor actions. *Scientific Reports, 9*, Article 13054. https://doi.org/10.1038/s41598-019-49476-4

Kalanthroff, E., Davelaar, E. J., Henik, A., Goldfarb, L., & Usher, M. (2018). Task conflict and proactive control: A computational theory of the Stroop task. *Psychological Review, 125*(1), 59–82. https://doi.org/10.1037/rev0000083

Kousaie, S., & Phillips, N. A. (2012). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research, 1446*, 71–90. https://doi.org/10.1016/j.brainres.2012.01.052

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology, 25*(5), 497–514. https://doi.org/10.1080/20445911.2013.799170

Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. R. V. (2012). Chapter seven–Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 56, pp. 229–262). Academic Press. https://doi.org/10.1016/B978-0-12-394393-4.00007-8

Larson, M. J., Kaufman, D. A. S., & Perlstein, W. M. (2009). Neural time course of conflict adaptation effects on the Stroop task. *Neuropsychologia, 47*(3), 663–670. https://doi.org/10.1016/j.neuropsychologia.2008.11.013

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., De, A. B., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin, 144*(4), 394–425. https://doi.org/10.1037/bul0000142

Leiner, D. J. (2019). *SoSci survey*. (Version 3.1.06) [Computer software]. https://www.soscisurvey.de

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507–1515. https://doi.org/10.1111/j.1467-9280.2009.02480.x

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia, 38*(5), 701–711. https://doi.org/10.1016/S0028-3932(99)00106-2

Liu, X. (2007). Event-related potentials associated with Chinese and English color-word Stroop tasks in Chinese bilinguals. *Journal of Medical Colleges of PLA, 22*(4), 201–208. https://doi.org/10.1016/S1000-1948(07)60041-0

Luk, G. (2015). Who are the bilinguals (and monolinguals)? *Bilingualism: Language and Cognition, 18*(1), 35–36. https://doi.org/10.1017/S1366728914000625

Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, memory, and cognition, 10*(2), 304–315. https://doi.org/10.1037/0278-7393.10.2.304

Marian, V., Blumenfeld, H. K., Mizrahi, E., Kania, U., & Cordes, A.-K. (2013). Multilingual Stroop performance: Effects of trilingualism and proficiency on inhibitory control. *International Journal of Multilingualism, 10*(1), 82–104. https://doi.org/10.1080/14790718.2012.708037

McKewan, M., Cooper, P. S., Wong, A. S. W., Michie, P. T., Sauseng, P., & Karayanidis, F. (2020). Task switching costs have distinct phase-locked and nonphase-locked EEG power effects. *Psychophysiology, 57*, Article e13533. https://doi.org/10.1111/psyp.13533

Mückschel, M., Stock, A.-K., Dippel, G., Chmielewski, W., & Beste, C. (2016). Interacting sources of interference during sensorimotor integration processes. *NeuroImage, 125*, 342–349. https://doi.org/10.1016/j.neuroimage.2015.09.075

Neurobehavioral Systems. (2019). *Presentation* (0.71) [Computer software]. Neurobehavioral Systems, Inc. https://www.neurobs.com/

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. https://doi.org/10.1016/j.cogpsych.2012.12.002

Preston, M. S., & Lambert, W. E. (1969). Interlingual interference in a bilingual version of the stroop color-word task. *Journal of Verbal Learning and Verbal Behavior, 8*(2), 295–301. https://doi.org/10.1016/S0022-5371(69)80079-4

RStudio Team. (2015). *RStudio: Integrated development for R* (Version 3.6) [Computer software]. Rstudio Inc. http://www.rstudio.com/

Sauseng, P., Tschentscher, N., & Biel, A. L. (2019). Be prepared: Tune to FM-theta for cognitive control. *Trends in Neurosciences, 42*(5), 307–309. https://doi.org/10.1016/j.tins.2019.02.006

Sorge, G. B., Toplak, M. E., & Bialystok, E. (2017). Interactions between Levels of attention ability and levels of bilingualism in children's executive functioning. *Developmental Science, 20*(1), Article e12408. https://doi.org/10.1111/desc.12408

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Sumiya, H., & Healy, A. F. (2004). Phonology in the bilingual Stroop effect. *Memory & Cognition, 32*(5), 752–758. https://doi.org/10.3758/BF03195865

Tao, L., Marzecova, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: The role of age of acquisition. *Frontiers in Psychology, 2*, Article 123. https://doi.org/10.3389/fpsyg.2011.00123

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 760–771. https://doi.org/10.1037/0278-7393.16.5.760

Van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., Vermeire, K., Lee, S.-H., & Lim, S. (2019). Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? A systematic review. *Behavioral Sciences, 9*(3), Article 27. https://doi.org/10.3390/bs9030027

Van Driel, J., Swart, J. C., Egner, T., Ridderinkhof, K. R., & Cohen, M. X. (2015). (No) time for control: Frontal theta dynamics reveal the cost of temporally guided conflict anticipation. *Cognitive, Affective, & Behavioral Neuroscience, 15*(4), 787–807. https://doi.org/10.3758/s13415-015-0367-2

Van Heuven, W. J. B., Conklin, K., Coderre, E. L., Guo, T., & Dijkstra, T. (2011). The influence of cross-language dimilarity on within- and between-language Stroop effects in trilinguals. *Frontiers in Psychology, 2*, Article 374. https://doi.org/10.3389/fpsyg.2011.00374

Von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General, 145*(2), 246–258. https://doi.org/10.1037/xge0000120

West, R., & Alain, C. (2000). Effects of task context and fluctuations of attention on neural activity supporting performance of the Stroop task. *Brain Research, 873*(1), 102–111. https://doi.org/10.1016/S0006-8993(00)02530-0

West, R., Jakubek, K., Wymbs, N., Perry, M., & Moore, K. (2005). Neural correlates of conflict processing. *Experimental Brain Research, 167*(1), 38–48. https://doi.org/10.1007/s00221-005-2366-y

World Medical Association. (2013). World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association, 310*(20), Article 2191. https://doi.org/10.1001/jama.2013.281053

Woumans, E., Ceuleers, E., van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(5), Article 1579. https://doi.org/10.1037/xlm0000107

Yow, W. Q., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: Why variations in bilingual experiences matter. *Frontiers in Psychology, 6*, Article 164. https://doi.org/10.3389/fpsyg.2015.00164

Zhao, J., Liang, W.-K., Juan, C.-H., Wang, L., Wang, S., & Zhu, Z. (2015). Dissociated stimulus and response conflict effect in the Stroop task: Evidence from evoked brain potentials and brain oscillations. *Biological Psychology, 104*, 130–138. https://doi.org/10.1016/j.biopsycho.2014.12.001

History
Received August 19, 2021
Revision received July 12, 2022
Accepted July 12, 2022
Published online October 4, 2022

Publication Ethics
The study was approved by the local ethics committee and conducted according to the Declaration of Helsinki (World Medical Association, 2013).

Authorship
Marlene Försterling, Laura Hainke, and Anna Redkina contributed equally.

Funding
Open access publication enabled by Ludwig-Maximilians-Universität München.

Paul Sauseng
Department of Psychology
Ludwig-Maximilians-Universität München
Leopoldstr. 13
80802 Munich
Germany
paul.sauseng@lmu.de