Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials.

Shanna Kousaie & Natalie A. Phillips

Department of Psychology/Centre for Research in Human Development,
Concordia University
7141 Sherbrooke St. West
Montreal, Quebec
H4B 1R6 CANADA

Corresponding author:
Dr. Natalie A. Phillips
7141 Sherbrooke St. West, Montreal, Quebec, H4B 1R6
Phone: 514-848-2424 ext. 2218; Fax: 514-848-4537
Natalie.Phillips@concordia.ca

Abstract

An advantage for bilingual relative to monolingual young adults has been found for cognitive control tasks, although this finding is not consistent in the literature. The present investigation further examined this advantage using three tasks previously found to be sensitive to the effect. Furthermore, both behavioral and event-related brain potential (ERP) measures were included. Monolingual (n=25) and highly proficient bilingual (n=26) young adults completed a Stroop, Simon, and Eriksen flanker task while electrophysiological recording took place. Behaviorally there were no language group differences on any of the tasks. The ERP measures demonstrated differences between monolinguals and bilinguals with respect to conflict monitoring, resource allocation, stimulus categorization, and error-processing; however, these differences were not consistent across tasks. Given the similar behavioral performance across the groups the observed differences in brain responses may not represent an advantage for bilinguals. The results are discussed with respect to previous findings.

KEYWORDS: bilingualism; event-related brain potentials (ERP); cognitive control; N2; P3; ERN

1. Introduction

The effects of being bilingual on cognitive processes other than language *per se* have received an increasing amount of attention in the literature. Being bilingual has been associated with superior performance on tasks measuring executive function (see Bialystok, 2007, 2009), including the Simon task (Bialystok, 2006; Bialystok, Craik, Klein, & Viswanathan, 2004), the Stroop task (Bialystok, Craik, & Luk, 2008; Zied et al., 2004), and the Attention Network Test (ANT: Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Costa, Hernández, & Sebastián-Gallés, 2008). Furthermore, an advantage for bilinguals relative to monolinguals has been found in children (Bialystok & Martin, 2004; Martin-Rhee & Bialystok, 2008), young adults (e.g., (Bialystok, 2006; Costa, et al., 2009; Costa, et al., 2008), and older adults (Bialystok, et al., 2004; Bialystok, et al., 2008; Bialystok, Craik, & Ryan, 2006; Zied, et al., 2004). Until now the majority of investigations examining the bilingual advantage have used behavioral measures only; the present investigation examines the bilingual advantage in a Stroop task, a Simon task, and a modified Eriksen flanker task using both behavioral (reaction time (RT) and accuracy) and electrophysiological (event-related brain potentials; ERPs) measures. The inclusion of electrophysiological measures permit the examination of bilingualism-related differences in the neural responses associated with the performance of these tasks.

It has been hypothesized that the bilingual advantage results from the constant manipulation of two languages by bilinguals (Bialystok, 2007). The simultaneous activation of a bilingual's two languages despite being engaged in a single language has been well documented using picture identification (e.g., Blumenfeld & Marian, 2007; Marian, Spivey, & Hirsch, 2003), word identification (e.g., Dijkstra, Grainger, & van Heuven, 1999; Dijkstra, Timmermans, & Schriefers, 2000; van Heuven, Schriefers, Dijkstra, & Hagoort, 2008), translation recognition (e.g., de Groot, Delmaar, & Lupker, 2000), and semantic priming (e.g., de Bruijn, Dijkstra, Chwilla, & Schriefers, 2001; Kerkhofs, Dijkstra, Chwilla, & de Bruijn, 2006; Kousaie & Phillips, 2011; Paulmann, Elston-Güttler, Gunter, & Kotz, 2006), using stimuli that overlap languages in lexical and/or phonological features (e.g., interlingual homographs). Given the non-selective activation of languages, cognitive control processes are thought to be required to prevent interference by the non-target language. These control processes may be similar to those engaged during the performance of attentional control tasks, including selective attention to target information, inhibition of irrelevant information, and switching (Bialystok, et al., 2004). This creates a situation in which these control mechanisms are extensively practiced in bilinguals and could lead to more efficient control processes relative to monolinguals.

Evidence for this notion has been derived from several tasks, including the Stroop task, the Simon task, and the Eriksen flanker task. Stroop (1935) found a significant increase in naming time for the print color of an incongruent color word relative to naming the color of a solid square. The Stroop effect has been extensively studied since the publication of Stroop’s influential paper (see MacLeod, 1991), and for the present investigation we take the position that the Stroop effect is caused by interference resulting from competition between word reading and color naming. In order to respond correctly to an incongruent stimulus, an individual must suppress/inhibit the dominant word reading response in order to correctly name the color. This has been referred to as interference suppression (Bialystok, et al., 2008; Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002) and greater Stroop interference corresponds to less efficient interference suppression. Recent investigations have found that bilingualism is associated with smaller Stroop effects (Bialystok, et al.; Zied, et al., 2004), suggesting that bilinguals are more efficient at interference suppression relative to monolinguals.

Another task that has been investigated in relation to bilingualism is the Simon task (Simon & Rudell, 1967). Although there are variations, common to each version of the Simon task are stimuli that can vary along two dimensions (e.g., color and position); however, only one dimension is relevant for task performance (e.g., color). By manipulating the relative location of the stimulus and the response (e.g., presenting a stimulus requiring a left lateralized response on the left vs. the right side of the monitor) congruent and incongruent trials are possible. On a congruent trial both stimulus dimensions map onto the same response, whereas on an incongruent trial they map onto different responses and the irrelevant stimulus dimension must be inhibited in order to respond correctly. The Simon effect refers to the increase in RT for incongruent trials relative to congruent trials. Similar to results from the Stroop task, bilinguals have shown smaller Simon effects relative to monolinguals (Bialystok, 2006; Bialystok, et al., 2004), which has been suggested to reflect better perceptual conflict resolution in bilinguals relative to monolinguals (Bialystok).

The final task relevant to this investigation is an arrows version of the Eriksen flanker task. Eriksen and Eriksen (1974) used target letters that were flanked on either side by distractor letters which could require the same or different response as the target letter (i.e., congruent vs. incongruent conditions). RT significantly increased for incongruent trials, indicating that participants were unable to avoid processing information from the flanking stimuli.

The effect of bilingualism on performance of the Eriksen flanker task on its own has not been investigated; however, a variation of the task has shown an advantage for bilinguals relative to monolinguals (Costa et al., 2009; Costa et al., 2008). Specifically, the effect has been demonstrated using the ANT (Fan, McCandliss, Sommer, Raz, & Posner, 2002), which comprises a flanker task embedded in a cue reaction time task designed to explore three attentional networks, namely executive control, alerting and orienting. With respect to the executive control component relevant to the current investigation, congruent trials were comprised of a target and flanking arrows pointing in the same direction, whereas incongruent trials were comprised of a target arrow pointing in one direction and flanking arrows pointing in the other direction. Costa et al. (2008) found that bilinguals were faster than monolinguals overall and showed less interference from incongruent flankers than monolinguals. Furthermore, Costa et al. (2009) found that the bilingual advantage only emerged when monitoring demands were high, suggesting that the observed advantage for the bilinguals was caused by superior conflict monitoring. Given that the present investigation is concerned with executive control differences between bilinguals and monolinguals, we used a simple Eriksen flanker task.

One factor common to tasks in which bilinguals show an advantage is the need to monitor for and resolve conflict in order to maintain high accuracy. For example, in the Stroop task there is conflict between the word and the color on incongruent trials and a participant must detect and resolve this conflict by inhibiting the dominant word reading response. Given that bilinguals have been found to demonstrate superior performance than monolinguals on the Stroop, Simon, and flanker tasks, the present investigation examined all three tasks in the same sample using the same methodology. Until now differences between monolinguals and bilinguals have been examined for each task individually across studies. Although the neural systems activated by the interference in these tasks may be similar (Fan, Flombaum, McCandliss, Thomas, & Posner, 2003; Peterson et al., 2002), Fan et al. found no correlation between the behavioral interference measures. Given that there are differences in task demands between the three tasks, an examination of all three tasks in the same sample will permit us to evaluate whether there are differential effects of bilingualism on performance across the tasks.

Several theories attempt to explain how cognitive control is implemented in performing the tasks described above, and there is agreement that the anterior cingulate cortex (ACC) and dorsolateral prefrontal cortex (DLPFC) are involved (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Carter & van Veen, 2007; Liu, Banich, Jacobson, & Tanabe, 2006; Milham, Banich, Claus, & Cohen, 2003). Although different theories postulate different roles for these brain areas (e.g., the ACC as a conflict monitor vs. the ACC being involved in conflict resolution), the previously observed behavioral differences between monolinguals and bilinguals leads to the question of whether there would be differences in the strength and/or timing of brain activity. One way to address this question is using ERPs.

ERPs are extracted from the ongoing electroencephalograph and have excellent temporal resolution on the order of milliseconds allowing for the measurement of cognitive processes as they unfold in time. Different components of the ERP are associated with different cognitive processes and the amplitude and latency of the component are believed to be related to the strength and timing of the underlying cognitive process (Coles & Rugg, 1995). For the purposes of the present investigation, we were interested in the various ERP components that are related to executive control, including the N2, P3, and error-related negativity (ERN).

The N2 component that we were interested in, and that has been found using tasks most similar to our own, peaks 200-350 ms following a stimulus and has a frontocentral distribution (see Folstein & Van Petten, 2008). The N2 is thought to be related to conflict monitoring (e.g., van Veen & Carter, 2002a, 2002b; Yeung, Botvinick, & Cohen, 2004) and has been correlated with activity in the ACC as measured by functional magnetic resonance imaging (fMRI; Mathalon, Whitfield & Ford, 2003). Using an arrows version of the Eriksen flanker task, Danielmeier, Wessel, Steinhauser and Ullsperger (2009) found that the amplitude of the N2 was modulated by pre-response conflict (i.e., conflict between correct and incorrect response tendencies). Melara, Wang, Vu and Procter (2008) also found that the N2 had significantly greater amplitude for incongruent relative to congruent stimuli in a Simon task, replicating previous results indicating an association between the N2 for correct trials and conflict monitoring and/or detection (van Veen & Carter, 2002a, 2002b; Yeung, et al., 2004).

The P3 is a broad positive waveform with a centroparietal scalp distribution that peaks 300-600 ms following an eliciting stimulus. It is thought to be related to the updating of schemas (Donchin, 1981) and the allocation of resources (see Polich, 2007). P3 latency has been found to be proportional to stimulus categorization time (Kutas, McCarthy, & Donchin, 1977) and smaller in amplitude with increasing resource allocation (see Polich). Valle-Inclán (1996) found the P3 to be smaller in amplitude and delayed in latency for correct incongruent relative to correct congruent trials in a Simon task, and more recently, Melara et al. (2008) found that P3 amplitude peaked earlier for congruent relative to incongruent stimuli in a Simon task.

In contrast to these stimulus-evoked components, the ERN is a sharp negative deflection that peaks 50-100 ms following an incorrect response and is thought to reflect error-detection (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Gehring, Goss, Coles, Meyer, & Donchin, 1993). Others suggest that the ERN reflects post-response conflict resulting from a comparison of an erroneously executed response and the correct response tendency (Yeung, et al., 2004). Support for the latter comes from Danielmeier et al. (2009) who found a larger ERN for incorrect incongruent trials in a high post-response conflict condition relative to a low post-response conflict condition.

It is noteworthy that dipole modeling has found that the frontocentral N2 and the ERN can be modeled by a dipole in the same area of the ACC (van Veen & Carter, 2002a).

Furthermore, the amplitude of both components has been correlated with ACC activity measured by fMRI (Mathalon, et al., 2003). Thus, it has been suggested that the ACC is activated prior to the response in correct conflict trials (reflected by the frontocentral N2) and immediately after the response in incorrect conflict trials (reflected by the ERN; (Carter & van Veen, 2007).

The primary goal of the present investigation was to compare the neural responses of monolinguals and bilinguals when performing the Stroop, the Simon, and the Eriksen flanker tasks. Given that behavioral evidence suggests an advantage for bilinguals that is believed to be the result of well-practiced control mechanisms, it is likely that the neural correlates of these control mechanisms would differ between these two language groups.

To our knowledge there are four imaging studies that have examined this question. Bialystok et al. (2005) used magneto-encephalography (MEG) to localize differences in brain activity between monolinguals and bilinguals during the Simon task. Behaviorally there were no differences between monolinguals and French-English bilinguals. However, both Cantonese-English and French-English bilinguals showed systematic differences in MEG responses from monolinguals and both bilingual groups showed a relationship between faster responses and greater activity in areas of the left prefrontal cortex and ACC. This pattern was similar for congruent and incongruent trials and emerged in the 8-15 Hz frequency band, which is generally associated with signal processing. These results suggest that despite similar behavioral performance, monolinguals and bilinguals differed in the underlying neural processing involved in task performance, and that the management of two languages led to changes in executive function.

In another study, Luk, Anderson, Craik, Grady, and Bialystok (2010) examined language group differences in flanker task performance using fMRI. Behaviorally, Luk et al. found no language group differences. However, their neuroimaging results showed a similar relationship between brain and behavioral responses for the two language groups on congruent trials, but a different pattern for incongruent trials. Specifically, in bilinguals, facilitation and interference effects were associated with activity in different brain regions, whereas, in monolinguals these two processes were associated with similar regions of activation (and were consistent with the regions associated with the facilitation effect in the bilinguals). In addition, the version of the flanker task used by Luk et al. included a no-go condition for which the two language groups showed similar behavioral performance and similar regions of brain activation. These findings suggest that the effect of bilingualism on cognitive control is confined to interference suppression and is not present for response inhibition.

Garbin et al. (2010) used fMRI and a task-switching paradigm to examine the bilingual advantage in executive control. Behaviorally, only the monolinguals demonstrated a significant switch cost (i.e., increase in RT for switch trials relative to non-switch trials) and there were also language group differences in the brain regions activated on switch and non-switch trials.

Most recently, Abutalebi et al. (in press) examined both language control and cognitive control in monolinguals and bilinguals using fMRI using a language switching and a flanker task. The most relevant of their results is the finding that bilinguals demonstrated better adaptation across two testing sessions (i.e., a decrease in the conflict effect), and less activity in the ACC to achieve similar behavioral performance as the monolinguals. Furthermore, Abutalebi et al. found a significantly larger correlation between the behavioral conflict effect and grey matter density in the ACC in bilinguals relative to monolinguals.

These studies demonstrate the sensitivity of neuroimaging techniques for studying language group differences in cognitive control even in the absence of behavioural differences.

To our knowledge the current investigation is the first to use ERP methodology, a technique with the potential to reveal where differences might lie between monolinguals and bilinguals in the information processing stream. Using a Stroop, Simon, and Eriksen flanker task, we examined language group differences using both behavioral and ERP measures. In terms of the behavioral measures it was expected that all participants would show differences between all three trials types, with congruent trials having the greatest accuracy and fastest RT and incongruent trials having the lowest accuracy and longest RT. Bilinguals were expected to show faster RTs for both congruent and incongruent trials, demonstrating an overall executive control advantage as described by Hilchey and Klein (2011), as well as smaller increases in RT for incongruent relative to neutral trials (i.e., a smaller interference effect) compared to monolinguals, as has been previously described in the literature. In terms of the ERP measures, based on previous findings it was expected that all participants would show larger N2 and smaller ERN amplitude for incongruent relative to congruent trials[1] (Danielmeier, et al., 2009; Melara, et al., 2008) and that the P3 would be delayed in latency and smaller in amplitude for incongruent relative to congruent trials (Bauer, Kaplan, & Hasselbrock, 2010; Melara, et al., 2008; Valle-Inclán, 1996).

Central to the goals of the present investigation, we also expected language group differences. Specifically, we hypothesized that the bilinguals would show larger N2 amplitude for incongruent trials relative to monolinguals given that bilinguals are thought to demonstrate superior conflict monitoring abilities (Costa, et al., 2009), that is, an increase in conflict monitoring should be associated with larger N2 amplitude. Monolinguals were expected to show greater delays in P3 latency for incongruent trials than bilinguals (Bauer, et al., 2010), indicating

---

[1] This prediction may seem counterintuitive; however, it must be considered in light of Danielmeier et al.'s (2009) findings demonstrating that the amplitude of the ERN was related to the amount of post-response conflict, and in the present investigation there was more post-error conflict in incorrect congruent relative to incongruent trials.

longer stimulus categorization time. Predictions regarding the ERN are less straightforward given that this component is related to errors and the bilingual advantage has been demonstrated in RT on correct trials. Nevertheless, given that the ERN has been related to post-response conflict (Danielmeier, et al., 2009) and bilinguals are suggested to demonstrate superior conflict monitoring/resolution, it was expected that the bilinguals would show larger ERN amplitudes relative to the monolinguals, suggesting that when monitoring fails and an error is committed, there is greater post-response conflict.

2. Results

Statistical analyses were conducted using the statistical software package SPSS v. 11.5 (SPSS Inc., Chicago, IL, USA). Reported effects were significant at an alpha level of .05 (unless otherwise specified) and any significant interactions were decomposed with Bonferroni corrected simple effects analyses. Behavioral results will be reported first followed by the electrophysiological results.

2.1 Behavioral Results

We conducted a Language Group (monolingual and bilingual) x Trial Type (neutral, congruent, and incongruent) mixed ANOVA separately for the dependent variables accuracy and RT for each of the three tasks[2]. Results will be reported for each task in turn. Figure 1 shows the behavioral data for all three tasks, with accuracy on the left and RT on the right.

2.1.1 Stroop task. Accuracy. All participants demonstrated high accuracy. There was a main effect of Trial Type ($F(2,98)=15.9$, $MSE=6.1$, $p<.01$, $\eta^2_p=.25$), indicating lower accuracy

---

[2] In order to more closely replicate previous analyses that have found a bilingual advantage (e.g., Bialystok, et al., 2008) and given that the RT data can also be examined in terms of facilitation (i.e., the decrease in RT between the neutral and congruent trials) and interference (i.e., the increase in RT between neutral and incongruent trials), we also conducted a one-way ANOVA separately for the dependent variables interference and facilitation. There was no significant effect of Language Group for any of the tasks; thus, these results are not reported.

for incongruent trials relative to neutral and congruent trials (which did not differ). There was no effect of Language Group (*p*=.91), nor a Language Group x Trial Type interaction (*p*=.34).

Insert Figure 1 here

RT. There was a main effect of Trial Type (*F*(2,98)=189.8, *MSE*=226.3, *p*<.01, η^2_p =.80), indicating a significant difference between all three trial types; congruent trials had the shortest RT and incongruent trials had the longest. There was no effect of Language Group (*p*=.29), nor a Language Group x Trial Type interaction (*p*=.65).

2.1.2 Simon task[3]*. Accuracy*. There was a main effect of Trial Type (*F*(2,96)=59.3, *MSE*=8.6, *p*<.01, η^2_p =55), demonstrating that all three trial types differed with the highest accuracy for congruent trials and lowest for incongruent trials. There was no effect of Language Group (*p*=.23), nor a Language Group x Trial Type interaction (*p*=.21).

RT. There was a main effect of Trial Type (*F*(2,96)=362.4, *MSE*=53.0, *p*<.01, η^2_p =.88), indicating a significant difference between all three trial types with congruent trials having the shortest RT and incongruent trials the longest. There was no effect of Language Group (*p*=.06), nor a Language Group x Trial Type interaction (*p*=.27). It is noteworthy that the trend toward an effect of language group is suggestive of a *monolingual* advantage and not a bilingual advantage as predicted.

2.1.3 Eriksen task[4]*. Accuracy.* There was a main effect of Trial Type (*F*(2,96)=78.0, *MSE*=15.9, *p*<.01, η^2_p =.62), showing that all three trial types differed with the most accurate

---

[3] One bilingual participant was excluded from all analyses of the Simon task due to poor accuracy (i.e., 48 – 53% accuracy).

[4] One monolingual participant was excluded from all analyses of the Eriksen task due to poor accuracy (i.e., 63-73% accuracy).

responses for congruent trials and least accurate for incongruent trials. There was no effect of Language Group (*p*=.45), nor a Language Group x Trial Type interaction (*p*=.52).

RT. There was a main effect of Trial Type ($F(2,96)=292.6$, $MSE=263.0$, $p<.01$, $\eta^2_p=.86$), demonstrating faster RTs for neutral and congruent trials (which did not differ) than for incongruent trials. There was no effect of Language Group (*p*=.33), nor a Language Group x Trial Type interaction (*p*=.94).

2.2 Electrophysiological Results

Separate analyses were conducted for each component of interest (i.e., N2, P3, and ERN) for each of the tasks, and the results are presented for each task separately. ANOVAs consisted of the within-subjects factors Trial Type and Site (referring the scalp location of the electrode) and the between-subjects factor Language Group. A subset of midline electrodes were selected for each component based on previous research and inspection of the grand averaged waveforms. Sites Fz and FCz were included for analysis of the N2, and Fz, FCz, and Cz for analysis of the ERN given the frontocentral distribution of these components (Falkenstein, et al., 2000; see Folstein & Van Petten, 2008). Cz, CPz, and Pz were included for analysis of the P3 given its centroparietal scalp distribution (Falkenstein, et al., 2000; Squires, Squires, & Hillyard, 1975).

For each component, we conducted a mixed ANOVA which included the within-subjects factor Time referring to 20 ms time intervals comprising the time interval encompassing the entire component of interest (specified below for each component). The dependent variable in these analyses was mean amplitude within each 20 ms time interval. We examined the hypothesized latency shift in the P3 by analyzing the peak maximum latency within the P3 time interval for each participant in an additional ANOVA

For analyses with more than one degree of freedom in the numerator, the Huynh and Feldt (1976) correction for non-sphericity was used. The unadjusted degrees of freedom, the corrected mean square error (*MSE*), the adjusted *p*-value, and the Huynh-Feldt epsilon value (ε) are reported.

Due to poor technical quality of the EEG recording several participants were excluded from the electrophysiological analyses. In addition, one bilingual was excluded from analyses of the Simon task due to poor behavioral performance (i.e., 50.6 % accuracy)and two monolinguals were excluded from analyses of the Eriksen task, one for achieving 100 % accuracy across all three trial types and the other for poor performance (i.e., 69 % accuracy). One addition bilingual was excluded from analysis of the ERN for the Stroop and Eriksen tasks due to an insufficient number of accepted trials. Table 2 provides the sample size for each of the analyses.

Insert Table 2 here

Figures 2-10 depict the grand averaged waveforms for each task and component separately. In each figure, panel A shows the main effect of Trial Type collapsed across Language Group, panel B compares monolinguals and bilinguals for each Trial Type separately, and panel C shows the effect of Trial Type for each Language Group. We have included one representative electrode site for each component; FCz for the N2 and the ERN, and Pz for the P3. Waveforms are stimulus-locked for the N2 and P3 (Figures 2, 3, 5, 6, 8, and 9)and are response-locked for the ERN (Figures 4, 7 and 10).

2.2.1 Stroop task. The N2 was analyzed between 220 and 360 ms and the P3 between 300 and 500 ms; see Figures 2 and 3. Analysis of the N2 revealed a main effect of Language Group ($F(1,40)=6.1$, $MSE=232.3$, $p=.02$, $\eta^2_p=.13$), demonstrating larger N2 amplitude for monolinguals. There was also a Language Group x Trial Type x Site interaction (*F*(2,80)=6.2, *MSE*=1.4, *p*=.03, η^2_p =.10, ε=.78), indicating that monolinguals showed larger N2 amplitude than bilinguals for all trial types at site Fz and for neutral trials at sites Fz and FCz.

Insert Figures 2 and 3 here

Analysis of the P3 revealed a main effect of Trial Type (*F*(2,80)=14.4, *MSE*=19.3, *p*<.01, η^2_p =.27, ε=.89), demonstrating smaller P3 amplitude for incongruent trials relative to both congruent and neutral trials, which did not differ from each other. Analysis of peak P3 latency revealed a main effect of Language Group (*F*(2,80)=5.1, *MSE*=9257.7, *p*=.03, η^2_p =.11), demonstrating that the P3 peaked later in the monolinguals than in the bilinguals. There was also a main effect of Trial Type (*F*(2,80)=5.9, *MSE*=2130.8, *p*=.01, η^2_p =.13, ε=.75), demonstrating that the P3 peaked later for neutral trials relative to both congruent and incongruent trials, which did not differ.

The ERN was analyzed between 0 and 100 ms; see Figure 4. There was a Language Group x Trial Type interaction (*F*(2,78)=3.7, *MSE*=165.0, *p*=.03, η^2_p =.10, ε=.93), which revealed an effect of Trial Type in the bilinguals only, demonstrating larger ERN amplitude for neutral trials relative to both congruent and incongruent trials, which did not differ from each other. Monolinguals demonstrated larger ERN amplitude than bilinguals for congruent and incongruent (*p*=.06) trials.

Insert Figure 4 here

2.2.2 Simon task. The N2 was analyzed between 200 and 300 ms and the P3 between 240 and 460 ms; see Figures 5 and 6. Analysis of the N2 revealed a main effect of Trial Type

(*F*(2,84)=20.5, *MSE*=17.1, p<.01, η^2_p =.33, ε=1.0), indicating that N2 amplitude was larger for neutral trials than congruent and incongruent trials, which did not differ.

Insert Figures 5 and 6 here

Analysis of the P3 revealed a main effect of Language Group (*F*(1,42)=5.3, *MSE*=958.1, p=.03, η^2_p =.11), demonstrating larger P3 amplitudes for the monolinguals than the bilinguals. There was also a trend towards a main effect of Trial Type (*F*(2,84)=3.3, *MSE*=153.4, p=.06, η^2_p =.27, ε=.59), showing that incongruent trials elicited a smaller amplitude P3 than congruent (p<.01) and neutral trials (p=.07) trials, which did not differ. Analysis of P3 peak latency revealed a main effect of Trial Type (*F*(2,84)=20.2, *MSE*=875.1, p<.01, η^2_p =.33, ε=.87S), demonstrating later peak latency for incongruent trials relative to both neutral and congruent trials, which did not differ.

The ERN was analyzed between 0 and 100 ms, see Figure 7; there were no significant effects.

Insert Figure 7 here

2.2.3 Eriksen task. The N2 was analyzed between 260 and 420 ms and the P3 between 300 and 560 ms, see Figures 8 and 9. Inspection of panels A and C of Figure 8 suggests a delay in the latency of the N2 for incongruent trials; however, we have taken this to be a reflection of the delay in P3 latency. Analysis of the N2 time interval revealed a main effect of Trial Type (*F*(2,82)=28.2, *MSE*=12.3, p<.01, η^2_p =.41, ε=1.0), indicating larger N2 amplitude for incongruent relative to congruent and neutral trials, which did not differ from each other.

Insert Figures 8 and 9 here

Analysis of the P3 time interval showed a main effect of Trial Type ($F(2,82)=19.3$, $MSE=66.1$, $p<.01$, $\eta^2_p=.32$, $\varepsilon=.81$), demonstrating larger P3 amplitude for neutral relative to both congruent and incongruent trials, which did not differ. Analysis of P3 peak latency revealed a main effect of Trial Type ($F(2,84)=89.0$, $MSE=1881.0$, $p<.01$, $\eta^2_p=.68$, $\varepsilon=.90$), and a Language Group x Trial Type interaction ($F(2,84)=3.1$, $MSE=1881.0$, $p=.05$, $\eta^2_p=.07$, $\varepsilon=.90$), showing that P3 peak latency was delayed for incongruent relative to congruent and neutral trials in both language groups; however, the delay was longer in monolinguals than in bilinguals (mean difference between incongruent and neutral:71.2 ms vs.48.0 ms and mean difference between incongruent and congruent:68.1 ms vs.47.4 ms for monolinguals and bilinguals, respectively).

The ERN was analyzed between 0 and 100 ms; see Figure 10. The analysis revealed a trend towards a main effect of Trial Type ($F(2,80)=3.2$, $MSE=206.2$, $p=.06$, $\eta^2_p=.07$, $\varepsilon=.78$), demonstrating larger ERN amplitude for incongruent relative to neutral trials and congruent trials. There was also a Language Group x Trial Type x Site x Time interaction ($F(16,640)=2.3$, $MSE=5.5$, $p=.05$, $\eta^2_p=.05$, $\varepsilon=.30$), indicating that monolinguals showed a smaller amplitude ERN than bilinguals from 0-40 ms post-response for neutral and incongruent trials, and a larger ERN amplitude than bilinguals from 80-100 ms post-response for congruent trials. In addition , ERN amplitude did not differentiate between the trial types in monolinguals, whereas, in bilinguals, incongruent trials elicited a larger amplitude ERN than congruent and neutral trials from 40-100 ms at sites FCz and Cz. Inspection of panel C of Figure 10 suggests that the effect of trial type in the bilinguals is due to a broader ERN peak for incongruent trials.

Insert Figure 10 here
----

3. Discussion

There were two goals of the present investigation in which monolingual and bilingual participants performed a Stroop task, a Simon task, and a modified Eriksen flanker task while electrophysiological recording took place. We examined the behavioral data in an attempt to replicate previous findings of a bilingual advantage, followed by an examination of the ERPs elicited by correct and incorrect trials to determine if there were language group differences in the neural correlates of performance. Our inclusion of three tasks for which bilinguals have previously demonstrated an advantage within the same sample, and the use of both behavioral and electrophysiological methods make this a novel and thorough investigation of the bilingual advantage.

We have included a timeline in Figure 11 that shows the sequence of electrophysiological and behavioral events following the presentation of a stimulus. In addition, we have summarized the predicted and observed effects for each task in the figure.

Insert Figure 11 here

3.1 Behavioral Data

Based on previous findings we expected to find the classic effects for both language groups, namely greater accuracy and faster RTs for congruent relative to neutral trials, and for both congruent and neutral trials relative to incongruent trials. This hypothesis was largely confirmed by the main effect of Trial Type in the analysis of all three tasks. From this we can conclude that the tasks used here produced interference on incongruent trials and thus tapped the conflict monitoring and resolution processes required for performance of the task.

Of greater interest was the effect of Language Group. Analysis of the raw accuracy and RT data revealed no effect of Language Group for any of the three tasks. These findings contrast with those of others who report language group differences in young adults. Bialystok et al. (2008) found that bilinguals demonstrated smaller Stroop interference effects than monolinguals, and Bialystok (2006) found an advantage for bilinguals relative to monolinguals in an arrows version of the Simon task. However, in that study, the bilingual advantage was only present in the most demanding conditions of the Simon task. Similarly, Costa et al. (2009) found that the bilingual advantage only emerged when the proportion of congruent and incongruent trials in the task created a high demand on conflict monitoring processes. In the current investigation there were equal proportions of each trial type in each task and difficulty was not manipulated; thus it may be argued that task demands were not great enough for a bilingual advantage to be demonstrated. However, we do not believe this to be the case, as the advantage has been previously demonstrated using a blocked design which is even less demanding on conflict monitoring processes given that each block contained a single trial type (Bialystok, et al., 2008).

Given that the tasks used here comprised a large number of trials it is possible that the bilingual advantage was eliminated due to practice effects. That is, several studies have found that the bilingual advantage disappears with practice (for review see Hilchey & Klein, 2011), therefore we conducted several supplemental analyses to rule out this possibility. We examined the raw RT, as well as the interference effect relative to both neutral and congruent trials for all three of the tasks for the first block of trials only. None of these analyses yielded a significant effect of Language Group, thus we are confident that there were no behavioral differences between monolinguals and bilinguals. This is difficult to reconcile with the literature extensively reviewed by Hilchey and Klein; however, we offer the following possible explanations.

One notable difference between the current and previous investigations in the composition of the bilingual sample; that is, one of the strengths of the current study is that the bilingual sample did not include immigrants, whereas in many of the previous studies the bilingual group is comprised predominantly of immigrants[5]. It is possible that immigrant status plays an important role in the bilingual advantage and that in a group of individuals who are living in a bilingual society, as is the case in Montreal, the advantage is more difficult to detect.

The previous explanation applies to a lesser extent to the investigations conducted by Costa and colleagues, who have demonstrated an advantage for bilinguals using the ANT. However, there are important differences between the ANT and the flanker task used in the current investigation. The ANT is a flanker task that is embedded in a cue reaction time task designed to measure three attentional networks: alerting, orienting and executive control. The executive control network is measured by comparing RTs for congruent and incongruent trials that are embedded in a task that also includes cued and non-cued conditions. Thus, it is possible that this methodological difference explains the discrepancy in our findings.

3.2 Electrophysiological Data

Electrophysiological recordings were included in order to investigate language group differences in the neural responses to conflict, for both correct and incorrect trials. Given that we have been unable to replicate previous findings of a bilingual advantage in the Stroop task (Kousaie & Phillips, in press), and that previous studies have found differences in the neural

---

[5] Although many of the studies that have found a bilingual advantage have comprised predominantly immigrant samples there are several exceptions where a global advantage for bilinguals has been found using non-immigrant samples (e.g., Costa, et al., 2009; Costa, et al., 2008).

correlates of behavior in the absence of overt behavioral difference (e.g., Abutalebi, et al., in press; Bialystok, et al., 2005), we reasoned that ERPs would be a good measure of possible differences in the cognitive processes involved in performance of the Stroop, Simon, and Eriksen tasks, even in the absence of behavioral differences. We will discuss each component in turn.

3.2.1 N2. The N2 is thought to reflect conflict monitoring; thus, we predicted that all participants would show larger N2 amplitudes for incongruent relative to congruent trials due to greater demands on conflict monitoring on the former. We also predicted that bilinguals would show enhanced conflict monitoring in the form of larger N2 amplitudes relative to monolinguals.

For the Stroop task, we found that the monolinguals exhibited larger N2 amplitudes than the bilinguals overall. This result was contrary to our predictions and suggests greater conflict monitoring in the monolinguals than in the bilinguals. However, it is possible that the bilinguals required less active conflict monitoring than the monolinguals in order to perform the Stroop task. That is, if indeed bilinguals are more efficient conflict monitors as a result of their experience with two languages, then conflict monitoring in these individuals may require less activation of the ACC, thus eliciting smaller amplitude N2s. This interpretation is supported by a study demonstrating that a reduction in N2 amplitude from childhood to adolescence reflects the development of cognitive control (Lamm, Zelazo, & Lewis, 2006). Lamm et al. found that smaller N2 amplitude during a Go/Nogo task was associated with better performance on independent measures of executive function, including a color-word Stroop task. Furthermore, Abutalebi et al. (in press) find that less activity in the ACC in bilinguals relative to monolinguals is associated with the same behavioral performance in the two language groups and Garbin et al. (2010) showed ACC activity in monolinguals, but not in bilinguals during task switching. Others have suggested that a larger amplitude N2 reflects increases in the processing of target irrelevant information (e.g., flankers; Yeung & Cohen, 2006), indicating a reduction in attentional focus on target information. Following this interpretation of the N2, our findings suggest that bilinguals were better able to selectively attend to the relevant aspects of the stimulus.

In the Eriksen task incongruent trials elicited larger amplitude N2s than congruent trials and this effect did not interact with language group. This demonstrates greater conflict monitoring for incongruent relative to congruent trials, as predicted; however, there were no differences in conflict monitoring between the two language groups.

3.2.2 P3. P3 latency has been associated with stimulus categorization time; thus, we predicted that, for all participants, the P3 would be delayed in latency for incongruent relative to congruent trials. Furthermore, we expected that this delay would be greater for monolinguals relative to bilinguals given that enhanced cognitive control mechanisms should allow bilinguals to categorize stimuli more quickly. With respect to amplitude, increased resource allocation has been associated with decreased P3 amplitude; thus, we predicted that the P3 would be smaller for incongruent trials.

For the Stroop task, there was no difference in the latency of the P3 for congruent and incongruent trials, demonstrating similar stimulus categorization time for both trials types in both language groups. P3 amplitude was smaller for incongruent relative to congruent trials in both groups, confirming that there was greater resource allocation for incongruent trials. Notably, there were no group differences in the P3 amplitude, nor any interactions with group.

For the Simon task, the P3 was delayed for incongruent trials relative to congruent trials in both monolinguals and bilinguals, demonstrating that it took longer to categorize incongruent stimuli, as predicted. This delay was not larger for the monolinguals relative to the bilinguals, suggesting similar stimulus categorization time for the two groups. With respect to P3 amplitude, bilinguals demonstrated smaller amplitude P3s relative to monolinguals. This was an unexpected and surprising result given previous evidence for a bilingual advantage in the Simon task (e.g., Bialystok, 2006; Bialystok, et al., 2004). This finding suggests that monolinguals allocated fewer resources to task performance than bilinguals, which does support an advantage for bilinguals. Instead, this finding suggests that performance of the Simon task was less effortful for monolinguals than for bilinguals.This suggestion is supported by our behavioral results which demonstrated a trend toward a monolingual advantage for the Simon task. In addition, incongruent trials elicited smaller P3s than congruent trials as predicted.

As predicted, the latency of the P3 was delayed for incongruent relative to congruent trials in the Eriksen task indicating that incongruent stimuli took longer to categorize. Furthermore, the delay was larger in the monolingual group, supporting our second hypothesis concerning the P3, and suggesting even slower stimulus categorization on incongruent trials in the monolinguals compared to the bilinguals.

3.2.3 ERN. The ERN is believed to reflect error detection or post-response conflict. Following this, we predicted reduced ERN amplitudes for incongruent relative to congruent trials and that enhanced cognitive control mechanisms in bilinguals would be reflected by larger ERN amplitudes relative to monolinguals.

For the Stroop task we found similar ERN amplitudes for congruent and incongruent trials in both language groups, suggesting that there was similar post-response conflict for both trial types. However, for congruent and incongruent trials the monolinguals showed larger ERN amplitudes relative to the bilinguals. This suggests greater post-response conflict on congruent trials between the executed erroneous response and the correct response tendency in monolinguals relative to bilinguals. This was contrary to our hypothesis; however, given that we also found greater conflict monitoring (as demonstrated by the N2) in the monolinguals relative to the bilinguals, it is possible that committing an error on congruent trials was particularly salient and was thus associated with larger ERNs.

There were no significant differences in ERN amplitude between congruent and incongruent trials or between language groups for the Simon task. For the Eriksen[6] task, an effect of Trial Type that interacted with Language Group, Site and Time demonstrating larger ERN amplitude for incongruent trials in the bilingual group. Inspection of Figure 10 indicates that this is due to a broader ERN for incongruent trials, suggesting that post-response conflict lasted longer on these trials.

3.3.3 Neutral trials. To our knowledge no studies have investigated the neural responses to neutral trials in the tasks used here; thus we had no specific language group hypotheses. Nevertheless, regardless of group, we expected reduced N2 and ERN amplitudes and larger P3 amplitude on conflict-free neutral trials relative to congruent and incongruent trials.

In the Simon task we found that neutral trials elicited a larger N2 than congruent and incongruent trials, suggesting greater conflict monitoring on these trials. This does not support our prediction, but can be reconciled with the N2 as reflecting conflict monitoring. That is, neutral trials did not comprise any conflict, whereas both congruent and incongruent trials either did comprise conflict or had the potential to do so. Each trial type represented one third of the total trials; thus, on two-thirds of the trials conflict could be present, whereas on the relatively infrequent neutral trials there was no potential for conflict. This may have caused the brain to continue to monitor for conflict on neutral trials, resulting in a larger amplitude N2.

---

[6] Inspection of the left panel of Figure 4 appears to show a Trial Type difference for the monolinguals.. For this reason a supplemental within-subjects ANOVA was conducted including site separately and time intervals from 20-80 ms for the monolinguals only. The effect of Trial Type was not significant at any of the sites(Fz: *p*=.40; FCz: *p*=.28; Cz: *p*=.15).

For the P3, we found a delay in peak amplitude for neutral trials on the Stroop task, suggesting that stimulus categorization took longer for neutral trials. As was discussed with respect to the N2, this may be due to the relative infrequency of neutral trials, which resulted in these trials being relatively more effortful to process. For the Eriksen task, we found larger P3 amplitude for the neutral trials relative to congruent trials in both monolinguals and bilinguals, suggesting that fewer resources were allocated for these trials, as expected.

With respect to error trials, the bilinguals demonstrated larger ERN amplitude for neutral trials than both congruent and incongruent trials on the Stroop task, whereas there were no effects in the monolinguals. For the Eriksen task, the bilinguals demonstrated smaller ERN amplitude for neutral relative to incongruent trials and there were no effects in the monolinguals. Although these findings do not support our hypotheses, it is interesting to note that the differences for neutral trials emerged in the bilingual group only, demonstrating a language group difference in the processing of errors on trials that do not comprise conflict.

3.4 General Discussion

In sum, we have reported behavioral results that do not provide evidence for a bilingual advantage in the Stroop, Simon, or Eriksen flanker tasks used here. However, the electrophysiological results do reveal processing differences between monolinguals and bilinguals, although the loci of differences varied across the three tasks. This indicates that the processing differences between bilinguals and monolinguals are not uniform across the tasks.

Although previous investigations using fMRI have found that similar brain regions are activated during the performance of the tasks included here (Fan, et al., 2003; see also Peterson, et al., 2002), there is little relation between the nature of the conflict in each of the tasks (Fan et al.). In the Stroop task, a dominant word reading response is the cause of conflict, in the Simon task conflict is caused by irrelevant spatial information, and in the Eriksen flanker task flanking distractors must be inhibited. Fan et al. found that the conflict effect produced by a Stroop, a flanker, and a spatial conflict task did not correlate[7]. They further investigated this using a dual task interference paradigm, reasoning that if two tasks involve the same process, then performing them simultaneously would result in a greater increase in RT relative to when either was performed alone. Using a hybrid Stroop/flanker task (e.g., an incongruent Stroop stimulus could be flanked by a string of "x"s that were incompatible with the correct response creating a double incongruent condition) it was found that there was no additive increase in RT when both types of conflict were present. A hybrid flanker/spatial conflict task elicited the same results, suggesting that despite overlapping regions of brain activation, the cognitive processes involved in Stroop, flanker and spatial conflict differ. However, fMRI would be limited in its ability to reveal any differences in the timing of the processes associated with these neural areas.

In contrast, ERPs are well-suited to illustrating differences in stages of processing between groups on tasks; thus, it is not surprising that our three tasks yielded different results. Our findings indicate that processing differences between monolinguals and bilinguals manifest themselves differently across tasks. That is, the two language groups differed in conflict monitoring and error-related processing for the Stroop task, resource allocation in the Simon task, and stimulus categorization and error-related processing in the Eriksen flanker task.

This was the first study to examine language group differences in cognitive control using ERPs. In one study using MEG differences in the neural correlates of performance of the Simon task between monolinguals and bilinguals were found in the absence of behavioral differences (Bialystok, et al., 2005). Bialystok et al. found that faster responses were correlated with greater

---

[7] Similar to Fan et al. (2003), our behavioral data showed no correlation between the interference effects produced by the three tasks in either of the language groups, nor in the entire sample irrespective of language group.

activity in the ACC for bilinguals. More recently, Abutalebi et al. (in press) found a strong correlation between the conflict effect and activity in the ACC in bilinguals during flanker task performance using fMRI. This is interesting given that the N2 and ERN have both been correlated with ACC activity (Mathalon, et al., 2003). Bialystok et al. and Abutalebi et al.'s findings suggest that bilinguals should show larger N2 and ERN amplitudes relative to monolinguals, reflecting greater activity in the ACC. However, we found no difference between monolinguals and bilinguals in the N2 or the ERN for the Simon or Eriksen tasks. A possible reason for this is that Bialystok et al. did not report overall ACC activity; instead, they correlated ACC activity with behavioral measures. Thus, although greater ACC activity was associated with faster responding in the bilinguals, overall differences in ACC activity between monolinguals and bilinguals were not reported. In fact, Abutalebi et al. found that the ACC was more efficient in bilinguals than in monolinguals during flanker performance.

It should be noted that there were differences between our version of the Simon task and those used in previous investigations. Previous versions have either not included a neutral condition (e.g., Bialystok et al., 2004; Melara et al., 2008; Peterson et al., 2002) or included a neutral/control condition in a separate block rather than intermixed with congruent and incongruent stimuli (e.g., Bialystok, 2006; Bialystok et al., 2005; 2008). For the current investigation, we inter-mixed neutral trials with congruent and incongruent trials in order to look at RTs in terms of facilitation and interference relative to a neutral condition from within the same block. This type of design is also preferable as it controls for any differences between trial types that may result from block differences (e.g., fatigue). However, it is possible that this leads to differences in processing, as suggested with respect to the electrophysiological results.

Similarly, neutral trials comprised of a single arrowhead presented at the center of the monitor were intermixed with congruent and incongruent trials in the Eriksen flanker task. This condition is often omitted (e.g., Danielmeier et al., 2009; Gehring et al., 1993), which may help to explain differences in the processing of neutral trials, as in the Simon task.

In conclusion, the electrophysiological results have demonstrated differences between monolinguals and bilinguals in the processing of conflict that were not evident from the behavioral data alone. The electrophysiological data revealed that the two language groups differed in conflict monitoring and error-related processing for the Stroop task, resource allocation in the Simon task, and stimulus categorization and error-related processing in the Eriksen flanker task. The fact that the differences were not consistent across the three tasks suggests that the way in which bilinguals and monolinguals respond to task conflict is not universal and differs depending on the nature of the task.

These results also highlight the strength of electrophysiological methods in studies of cognitive control. Given that the observed language group differences in electrophysiological measures did not translate into behavioral differences, it is inaccurate to refer to these effects as an advantage. We tested young adult participants who were presumably at the peak of their cognitive powers. Instead, these results indicate that there are differences between young monolinguals and bilinguals in their brain responses to conflict at various stages of processing which to not lead to different behavioral outcomes. It is possible that in a population where cognitive functioning is declining (e.g., older adults), the differences in brain responses may confer an advantage in behavioral performance. Additional research is required to fully characterize the differences in cognitive control between monolinguals and bilinguals and the suggested bilingual advantage, both behaviorally and electrophysiologically. Given findings suggesting that bilingualism has a positive impact on cognitive aging (e.g., Bialystok et al., 2004; Zied et al., 2004), electrophysiological measures may be a powerful tool for elucidating the presence and/or potential mechanism underlying this positive effect.

4. Experimental Procedure

4.1 Participants

Fifty-one young adults were recruited from Concordia University and McGill University, namely 25 monolinguals (10 males) between the ages of 18 and 35 (*M* = 23.8, *SD* = 4.7), and 26 bilinguals (9 males) between the ages of 19 and 33 (*M* = 24.5, *SD* = 3.4). All participants self-reported no illness, health condition, or use of medication known to affect cognitive functioning and showed normal cognitive functioning based on the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005). The bilingual participants were native English speakers who were highly proficient in French, having learned it before age 7. All provided high self-report ratings of L2 proficiency and used French in their daily activities. In addition, they showed comparable performance across languages on an animacy judgment task (Segalowitz & Frenkiel-Fishman, 2005), which we used as an objective measure of relative L2 proficiency.

Table 1 provides demographic information for both participant groups. The groups were matched on age, education, and maternal and paternal education. When participants were excluded from an analysis due to behavioral performance or poor quality electrophysiological recordings, the groups remained matched on these demographic variables.

\---------------------------
Insert Table 1 here
\---------------------------

Ethical approval for this study was obtained from the Concordia University Human Research Ethics Committee.

4.2 Materials and Apparatus

Participants completed the MoCA (Nasreddine, et al., 2005) to assess cognitive functioning; and three experimental tasks for which EEG recording took place, including modified Stroop, Simon, and Eriksen flanker tasks. Bilingual participants also completed an animacy judgment task to assess relative L1 and L2 proficiency (Segalowitz & Frenkiel-Fishman, 2005). All computerized tasks were presented on a Dell precision 370 desktop with a Pentium 4 processor and Windows XP operating system with a 16 inch Compaq monitor using Inquisit version 2.0 (Millisecond Software, Seattle, WA).

4.2.1 MoCA. The MoCA (Nasreddine, et al., 2005) is a 10-minute cognitive screening tool used to detect mild cognitive impairment in older adults. It assesses visuospatial/executive control, memory, attention, language, and orientation. Although the MoCA is generally used in older adult samples, it was included here to allow for age group comparisons in future studies.

4.2.2 Animacy Judgment Task. Bilingual participants categorized nouns as animate or inanimate, as quickly and accurately as possible; this produced an objective measure of language proficiency (Segalowitz & Frenkiel-Fishman, 2005). The task comprised 64 nouns (preceded by 8 practice trials) in both English and French divided into separate language blocks. Stimuli were presented in yellow 20 point Arial font on a black background and participants used left and right keys (“c” and “m”) on the keyboard to categorize the noun as animate or inanimate. The different blocks contained different nouns with no translation equivalents and were matched for the number of animate and inanimate judgments and same/different responses.

4.2.3 Experimental Tasks. Each experimental task comprised 720 trials presented in 10 blocks of 72 trials and preceded by 36 practice trials. Each block included an equal number of intermixed neutral, congruent, and incongruent trials in pseudorandom order such that there was a maximum of three consecutive trials of the same type. Each trial comprised a fixation cross for 250 ms followed by the stimulus which remained on the screen until the participant responded or until the trial timed out (i.e., 1250 ms for the Stroop task; 750 ms for the Simon and Eriksen tasks). The practice block was performed first and, in the rare case when accuracy was less than 80%, the practice block was repeated until this minimum criterion was achieved. A 250 Hz tone identified errors during the practice block; however, no performance feedback was provided during the experimental blocks. See Figure 12 for a sample trial of each task.

For the Stroop task, neutral trials comprised a series of “x”s printed in green (RGB: 0, 255, 0), red (RGB: 255, 0, 0), yellow (RGB: 255, 255, 0), or blue (RGB: 0, 0, 255), with the number of “x”s corresponding to the number of letters in the color word name (e.g., “xxx” printed in red); congruent trials comprised the color words *green*, *yellow*, *red*, and *blue* printed in the corresponding color; and incongruent trials comprised the same color words printed in one of the alternate three colors (e.g., the word *red* printed in blue). Stimuli were presented at the center of the monitor in bold 27 point Arial font on a black background. Participants responded using the index and middle finger on each hand to identify the color of the print using the keyboard; the letter “z” corresponded to yellow, the letter “x” to green, the symbol “,” to red, and the symbol “.” to blue. Prior to the practice block, participants performed a key acquisition task which comprised 80 trials for which the color of green, yellow, red, and blue circles was identified. Participants could repeat the acquisition task until they felt comfortable with the response keys (most participants only completed the key acquisition task once).

The Simon task comprised red and blue squares (100 x 100 pixels) presented on a black background at the center of the monitor, or 10% to the left or right of center. Red stimuli required a left key press (i.e., the letter “x” on the keyboard) and blue stimuli required a right key press (i.e., the symbol "." on the keyboard). For neutral trials the stimulus was presented at the center of the monitor, for congruent trials the stimulus was presented on the same side of the monitor as the correct response (e.g., a red stimulus presented on the left of the monitor), and for incongruent trials the stimulus was presented on the opposite side of the monitor as the correct response (e.g., a red stimulus presented on the right of the monitor).

For the Eriksen task, stimuli comprised arrowheads presented at the center of the monitor in white, bold, 36 point Arial font on a black background. Neutral trials consisted of a single arrowhead (e.g., <); whereas congruent trials consisted of a central arrowhead flanked on either side by three arrowheads pointing in the same direction as the target (e.g., < < < < < < <); and for incongruent trials the flanking arrows pointed in the opposite direction relative to the central target (e.g., < < < > < < <). Participants responded to the direction of the central arrowhead by pressing a left key (i.e., the letter "x" on the keyboard) if the arrowhead was pointing to the left, and a right key (i.e., the symbol "." on the keyboard) if the arrowhead was pointing to the right.

4.2.4 EEG Recording. The continuous EEG was recorded from 64 scalp locations according to the international 10-20 system using sintered Ag-AgCl electrodes and an ActiveTwo nylon cap (BioSemi, Amsterdam, NL). Eight additional electrodes were used: one on each earlobe, to be used as a reference for offline processing of the data; one above and one below the left eye, to record vertical electro-oculogram (VEOG); one on the outer canthi of each eye, to record horizontal electro-oculogram (HEOG); and two corresponding to sites FT9 and FT10 according to the international 10-20 system of electrode placement. The EEG was recorded relative to Common Mode Sense and Driven Right Leg (CMS/DRL) electrodes placed at the back of the head (to the left and the right of electrode POz, respectively) and was amplified using ActiveTwo amplifiers (BioSemi, Amsterdam, NL). The EEG was acquired using ActiView version 6.05 software (BioSemi, Amsterdam, NL), time-locked to the onset of the stimulus and sampled at a rate of 512 Hz in a 104 Hz bandwidth. Polygraphic Recording Data Exchange version 1.2 (PolyRex; Kayser, 2003) software was used to convert the continuous EEG from BioSemi Data Format (.BDF) to continuous file format (.CNT) for offline processing using SCAN 4.3.1 (Compumedics USA, Charlotte, NC, USA). During conversion using PolyRex, the EEG was referenced to linked ears and a fixed gain of 0.5 was applied.

Offline processing of the EEG data was performed separately for each task and consisted of applying a low pass 30 Hz filter, correcting VEOG artefacts using a spatial filter (NeuroScan, EDIT4.3), and excluding trials containing HEOG artefacts exceeding ±50 μV and EEG deflections exceeding ±100 μV. The electrophysiological time window was 700 ms including a 100 ms pre-stimulus/pre-response baseline and averages were based on trial type and accuracy resulting in six averages per task for each participant (i.e., neutral correct, congruent correct, incongruent correct, neutral incorrect, congruent incorrect, incongruent incorrect). Averages were stimulus-locked for correct trials, and response-locked for incorrect trials.

4.3 Procedure

Participants were seated in a comfortable chair and informed consent was obtained. The MoCA was completed first, followed by the animacy judgement task for bilingual participants. The electrode were then applied and once set-up was complete the Stroop task was performed first due to its greater complexity relative to the other two experimental tasks (i.e., greater demands on working memory), followed by the Simon and Eriksen flanker tasks in counterbalanced order. The testing session lasted approximately 1.5 to 2 hours, with approximately 60 minutes of EEG recording. Following completion of the experiment participants were debriefed and compensated for their time in the form of course credit or $10 per hour of participation.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the Cognitive Psychophysiology Laboratory for their help with participant recruitment and data collection, special thanks go to C. McHenry for help with data collection and D. Munro for technical assistance. Thank you to Drs. N. Segalowitz and B. Woodside for helpful comments on a previous version of this manuscript. This work was supported by a Canadian Institutes for Health Research Frederick Banting and Charles Best Canada Graduate Scholarship awarded to S. Kousaie.

References

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., . . . Costa, A. (in press). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*. doi: 10.1093/cercor/bhr287

Bauer, L. O., Kaplan, R. F., & Hasselbrock, V. M. (2010). P300 and the Stroop effect in overweight minority adolescents. *Neuropsychobiology, 61*, 180-187.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60*, 68-79.

Bialystok, E. (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *The International Journal of Bilingual Education and Bilingualism, 10*, 210-223.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*, 3-11.

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control and the Simon task: evidence from MEG. *NeuroImage, 24*, 40-49.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290-303.

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 859-873.

Bialystok, E., Craik, F. I. M., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1341-1354.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibtion in bilingual children: evidence from the dimensional change card sort task. *Developmental Science, 7*, 325-339.

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes, 22*, 633-660.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (2002). Immature frontal lobe contributors to cognitive control in children: Evidence from fMRI. *Neuron, 33*, 301-311.

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 367-379.

Coles, M. G. H., & Rugg, M. D. (1995). Event-related potentials: an introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of Mind. Event-Related Brain Potentials and Cognition* (pp. 1-26). Oxford, UK: University Press.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*, 135-149.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*, 59-86.

Danielmeier, C., Wessel, J. R., Steinhauser, M., & Ullsperger, M. (2009). Modulation of the error-related negativity by response conflict. *Psychophysiology, 46*, 1288-1298.

de Bruijn, E. R. A., Dijkstra, T., Chwilla, D. J., & Schriefers, H. J. (2001). Language context effects in interlingual homograph recognition: Evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition, 4*, 155-168.

de Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology, 53A*, 397-428.

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41*, 496-518.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42*, 445-464.

Donchin, E. (1981). Surprise!...Surprise? *Psychophysiology, 18*, 493-513.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143-149.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology, 78*, 447-455.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology, 51*, 87-107.

Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. *NeuroImage, 18*, 42-57.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*, 340-347.

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

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., . . . Avila, C. (2010). Bridging language and attention: brain basis of the impact of bilingualism on cognitive control. *NeuroImage, 53*(4), 1272-1278. doi: 10.1016/j.neuroimage.2010.05.078

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*, 385-390.

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 and Review, 18*(4), 625-658. doi: 10.3758/s13423-011-0116-7

Huynh, H., & Feldt, L. S. (1976). Estimates of correction for degrees of freedom for sample data in randomised block and split block designs. *Journal of Educational Statistics, 1*, 69-82.

Kayser, J. (2003). Polygraphic Recording Data Exchange - PolyRex [Computer software]. New York State Psychiatric Institute: Department of Biopsychology. Retrieved from http://psychophysiology.cpmc.columbia.edu/PolyRex.htm

Kerkhofs, R., Dijkstra, T., Chwilla, D. J., & de Bruijn, E. R. A. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Research, 1068*, 170-183.

Kousaie, S., & Phillips, N. A. (2011). Age-related differences in interlingual priming: A behavioural and electrophysiological investigation. *Aging, Neuropsychology, and Cognition, 18*, 22-55.

Kousaie, S., & Phillips, N. A. (in press). Aging and bilingualism: Absence of a “bilingual advantage” in Stroop interference in a non-immigrant sample. *The Quarterly Journal of Experimental Psychology*. doi: 10.1080/17470218.2011.604788

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*, 792-795.

Lamm, C., Zelazo, P. D., & Lewis, M. D. (2006). Neural correlates of cognitive control in childhood and adolescence: Disentangling the contributions of age and executive function. *Neuropsychologia, 44*, 2139-2148.

Liu, X., Banich, M. T., Jacobson, B. L., & Tanabe, J. L. (2006). Functional dissociation of attentional selection within PFC: Response and non-response related aspects of attentional selection ascertained by fMRI. *Cerebral Cortex, 16*, 827-834.

Luk, G., Anderson, J. A. E., Craik, F. I. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and Cognition, 74*(3), 347-357. doi: 10.1016/j.bandc.2010.09.004

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language, 86*, 70-82.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*, 81-93.

Mathalon, D. H., Whitfield, S. L., & Ford, J. M. (2003). Anatomy of an error: ERP and fMRI. *Biological Psychology, 64*, 119-141.

Melara, R. D., Wang, H., Vu, K.-P. L., & Proctor, R. W. (2008). Attentional origins of the Simon effect: Behavioral and electrophysiological evidence. . *Brain Research, 1215*, 147-159.

Milham, M. P., Banich, M. T., Claus, E. D., & Cohen, N. J. (2003). Practice-related effects demonstrate complementary roles of anterior cingulate and prefrontal cortices in attentional control. *NeuroImage, 18*, 483-493.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment: A brief screening tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society, 53*, 695-699.

Paulmann, S., Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2006). Is bilingual lexical access influenced by language context? *NeuroReport, 17*, 727-731.

Peterson, B. S., Kane, M. J., Alexander, G. M., Lacadie, C., Skudlarski, P., Leung, H.-C., . . . Gore, J. C. (2002). An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. *Cognitive Brain Research, 13*, 427-440.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128-2148. doi: 10.1016/j.clinph.2007.04.019

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: attention and second language proficiency. *Memory and Cognition, 33*, 644-653.

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*, 300-304.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology, 38*, 387-401.

Valle-Inclán, F. (1996). The locus of interference in the Simon effect: an ERP study. *Biological Psychology, 43*, 147-162.

van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex, 18*, 2706-2716.

van Veen, V., & Carter, C. S. (2002a). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience, 14*, 593-602.

van Veen, V., & Carter, C. S. (2002b). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior, 77*, 477-482.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error-detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*, 931-959.

Zied, K. M., Phillipe, A., Karine, P., Valerie, H.-T., Ghislaine, A., Arnaud, R., & Didier, L.-G. (2004). Bilingualism and adult differences in inhibitory mechanisms: Evidence from a bilingual Stroop task. *Brain and Cognition, 54*, 254-256.

Table 1. *Demographic Information for Participant Groups*

	Monolinguals (n = 25; 10 males)	Bilinguals (n = 26; 9 males)
	M *(SD)*	M *(SD)*
Age (years)	23.8 *(4.7)*	24.5 *(3.4)*
Education (years)	15.4 *(1.5)*	15.6 *(1.1)*
MoCA*	28.3 *(1.3)*	28.4 *(1.3)*
L1 self-reported language proficiency**	5.0 *(0.0)*	4.9 *(0.2)*
L2 self-reported language proficiency**	1.5 *(0.3)*	4.6 *(0.4)*
Coefficient of variability L1	n/a	.24 *(.07)*
Coefficient of variability L2	n/a	.25 *(.09)*
Maternal Education	15.4 *(2.4)*	14.6 *(2.2)*
Paternal Education	15.0 *(3.3)*	15.3 *(2.6)*

*Maximum score = 30; ≥26 normal cognitive function
**Self-report on a scale of 1-5: 1=no ability at all; 5=native-like ability

Table 2. *Sample size for statistical analyses following exclusion of participants.*

		Monolinguals	Bilinguals
Stroop	RT	25	26
	ERP – N2 and P3	20	22
	ERP – ERN	20	21
Simon	RT	24	26
	ERP	21	23
Eriksen	RT	24	26
	ERP – N2 and P3	21	23
	ERP – ERN	21	22

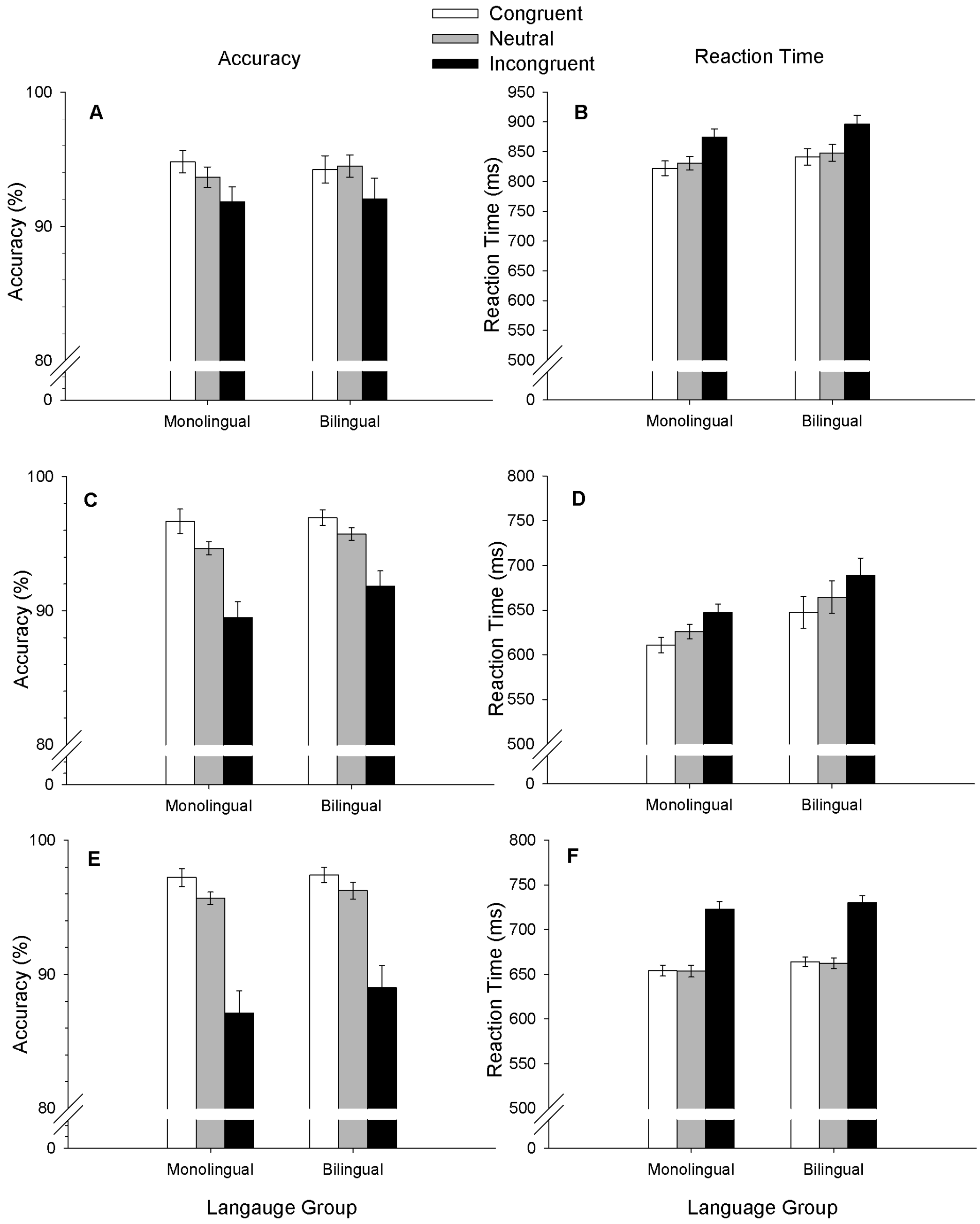


Figure 1. The left panel shows accuracy (±SE) and the right panel shows reaction time (±SE) as a function of language group for the Stroop (A, B), Simon (C, D) and Eriksen (E, F) tasks.

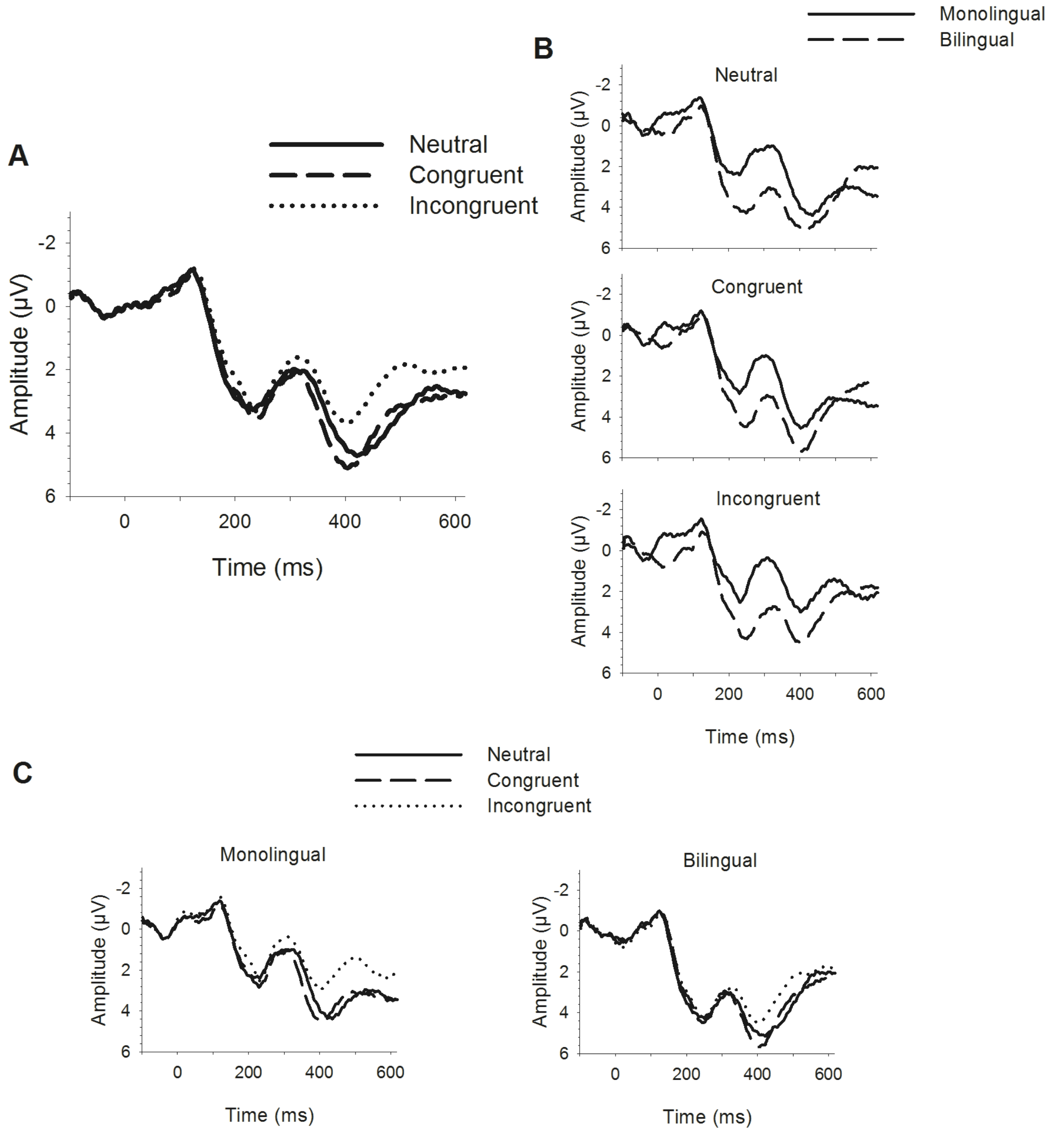


Figure 2. Stimulus-locked grand averaged waveforms for the N2 for the Stroop task depicted at site FCz. Panel A: main effect of Trial Type collapsed across Language Group; Panel B: effect of Language Group for each Trial Type; Panel C: effect of Trial Type for each Language Group.

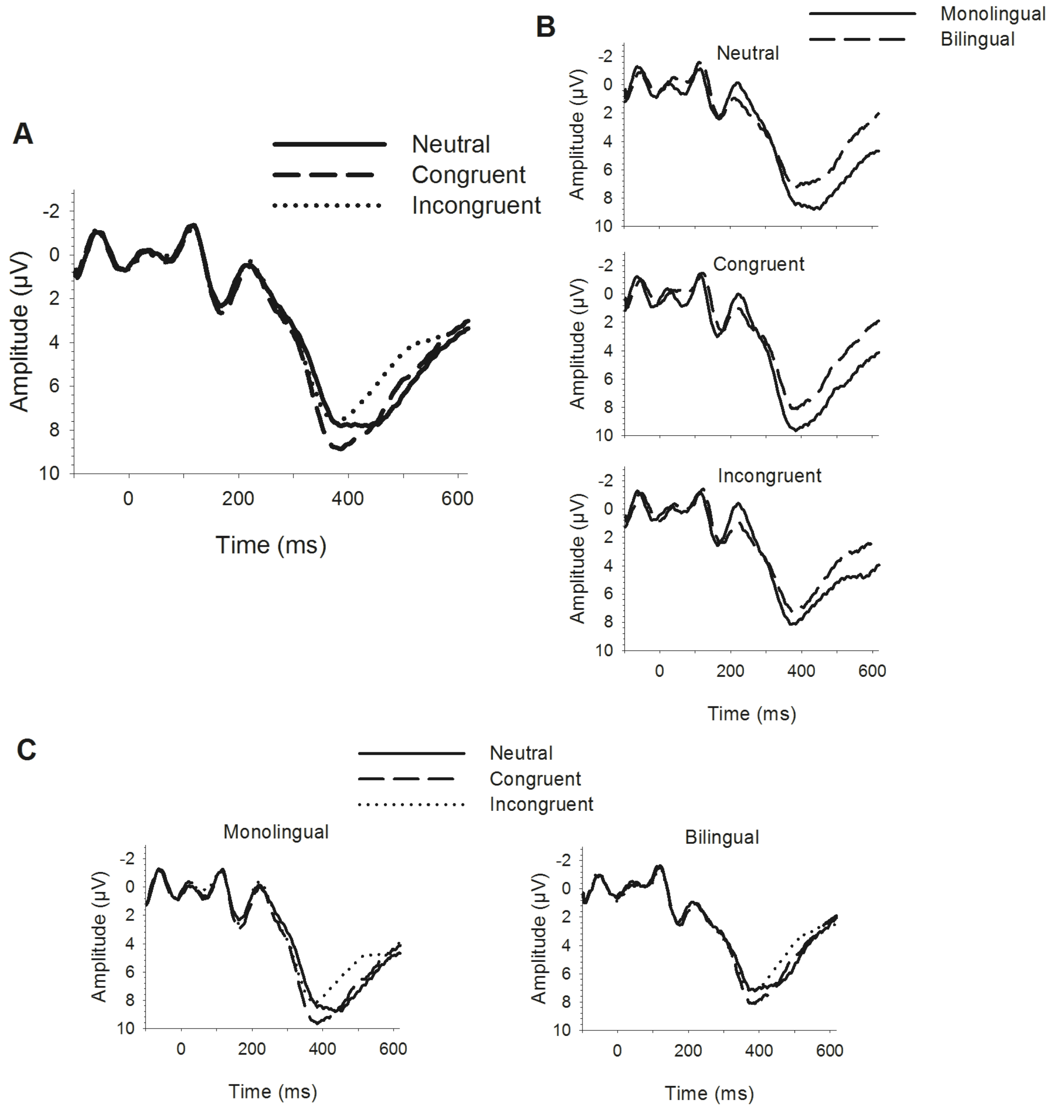


Figure 3. Stimulus-locked grand averaged waveforms for the P3 for the Stroop task depicted at site Pz. Panel A: main effect of Trial Type collapsed across Language Group; Panel B: effect of Language Group for each Trial Type; Panel C: effect of Trial Type for each Language Group.

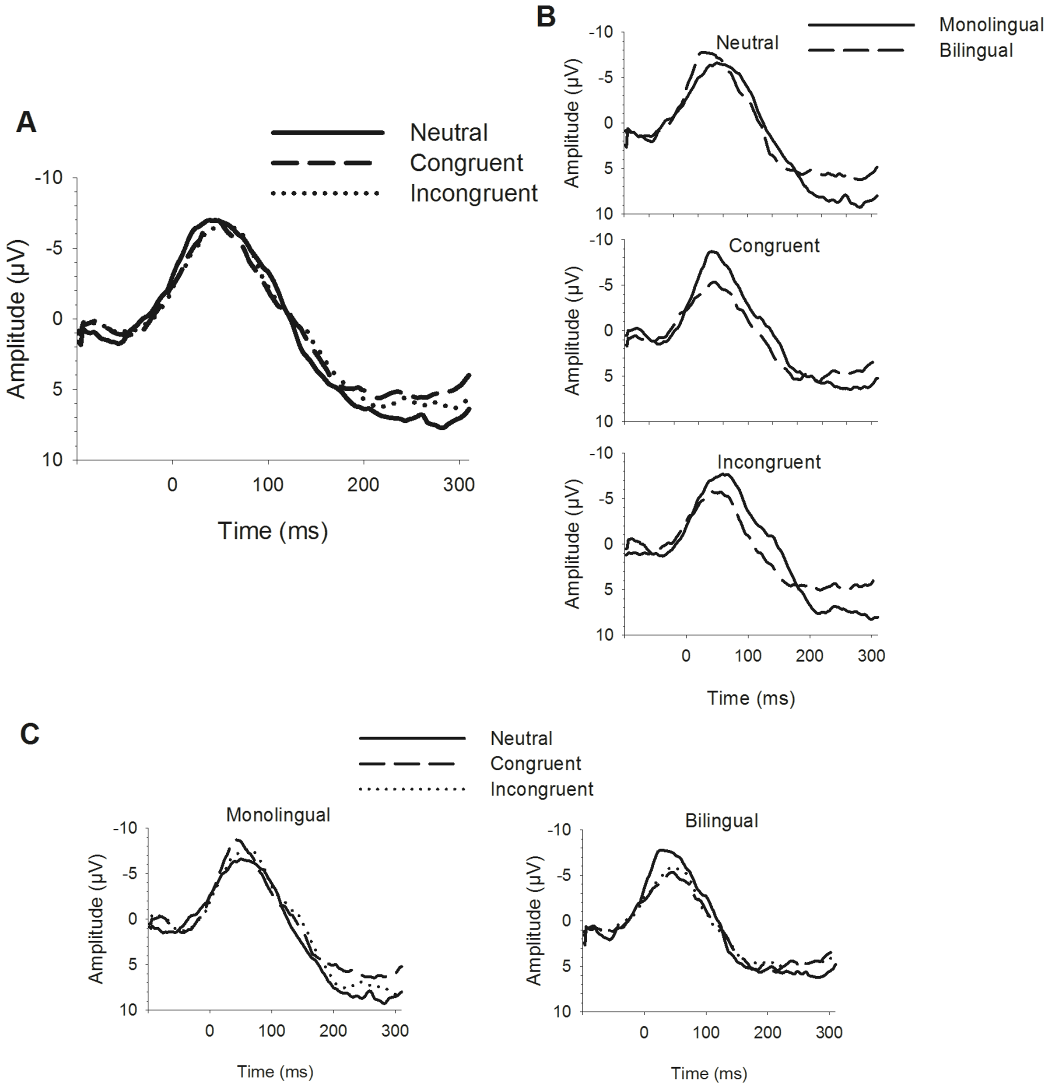


Figure 4. Response-locked grand averaged waveforms for the ERN for the Stroop task depicted at site FCz. Panel A: main effect of Trial Type collapsed across Language Group; Panel B: effect of Language Group for each Trial Type; Panel C: effect of Trial Type for each Language Group.

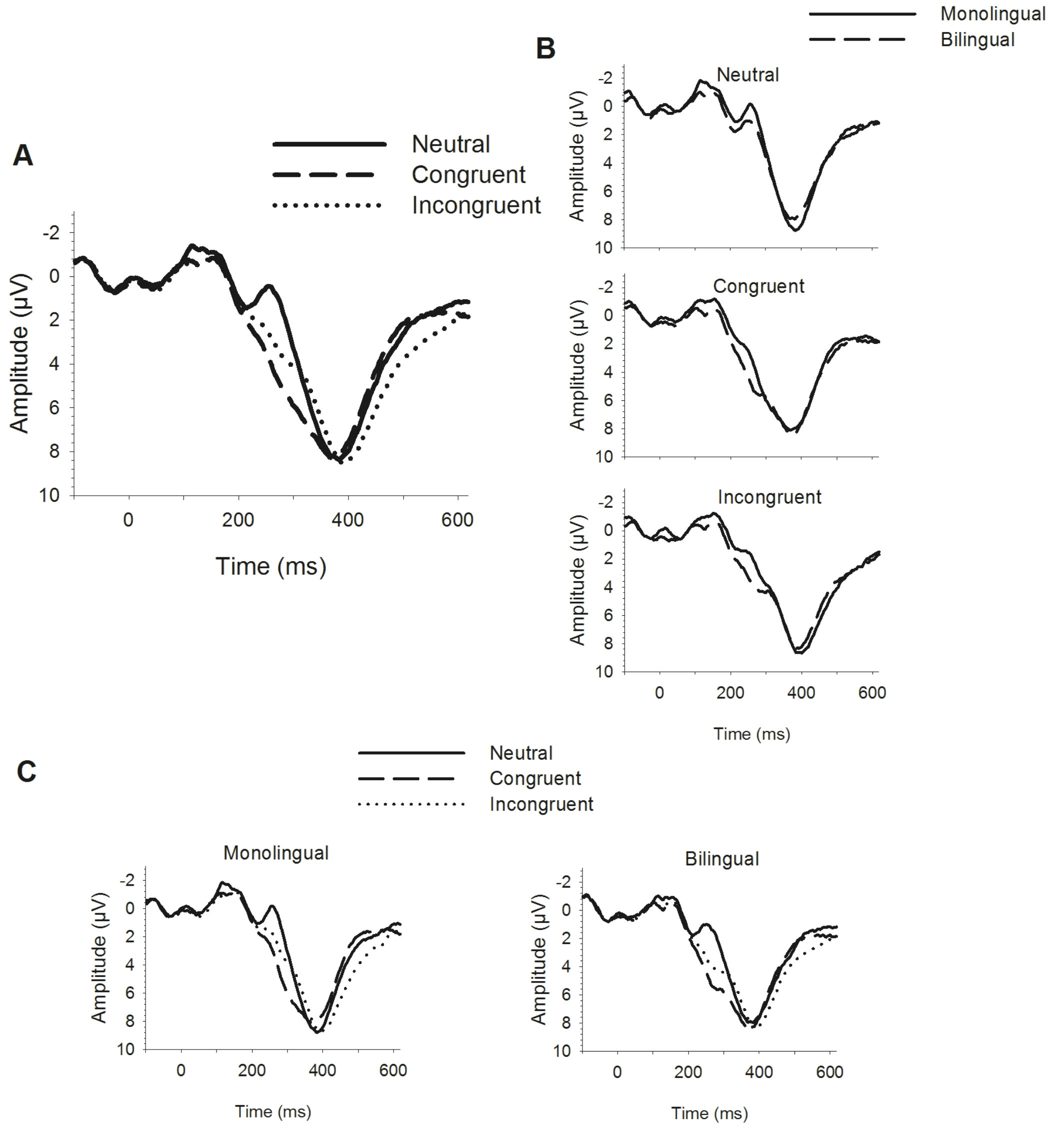


Figure 5. Stimulus-locked grand averaged waveforms for the N2 for the Simon task depicted at site FCz. Panel A: main effect of Trial Type collapsed across Language Group; Panel B: effect of Language Group for each Trial Type; Panel C: effect of Trial Type for each Language Group.

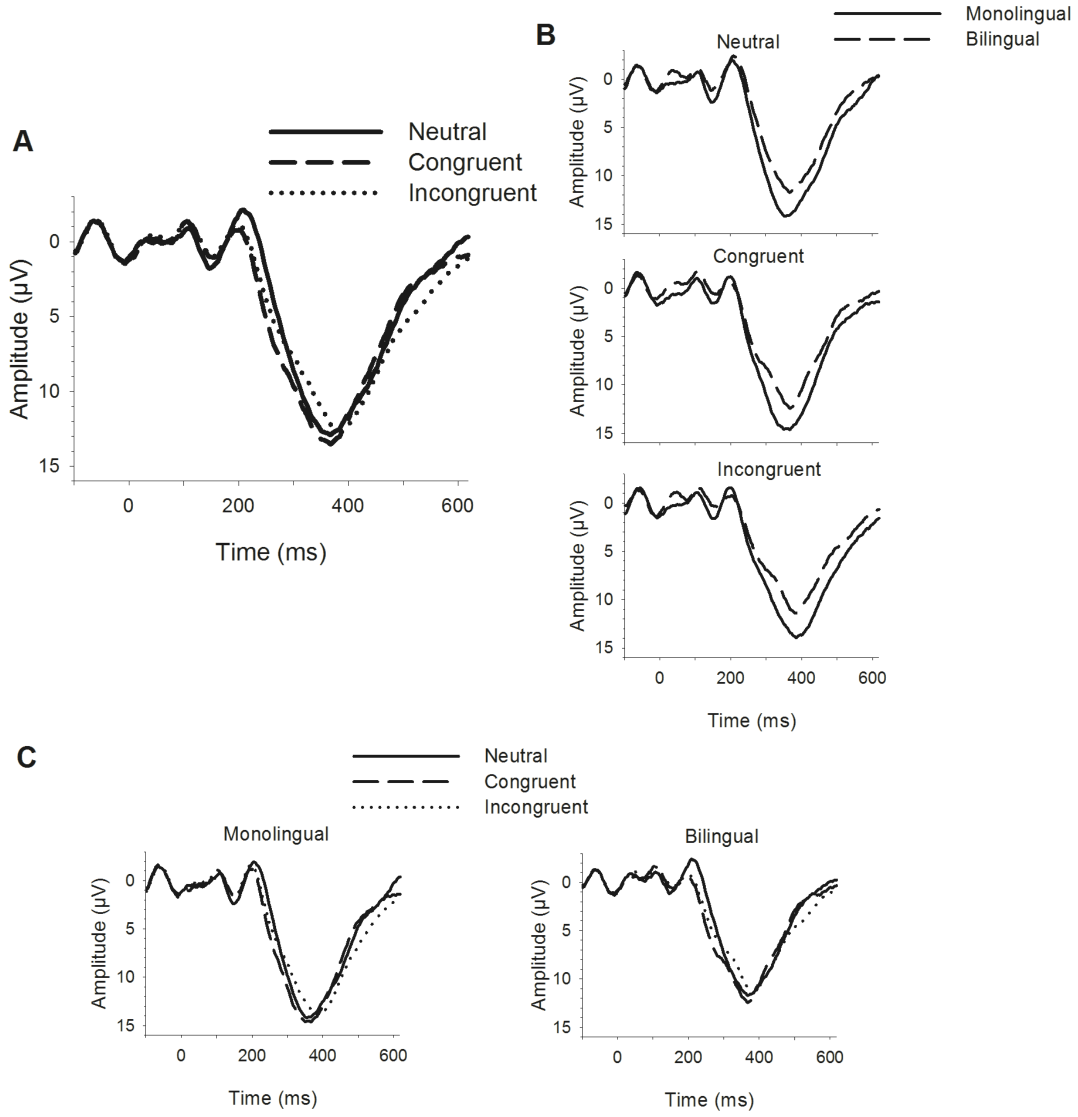


Figure 6. Stimulus-locked grand averaged waveforms for the P3 for the Simon task depicted at site Pz. Panel A: main effect of Trial Type collapsed across Language Group; Panel B: effect of Language Group for each Trial Type; Panel C: effect of Trial Type for each Language Group.

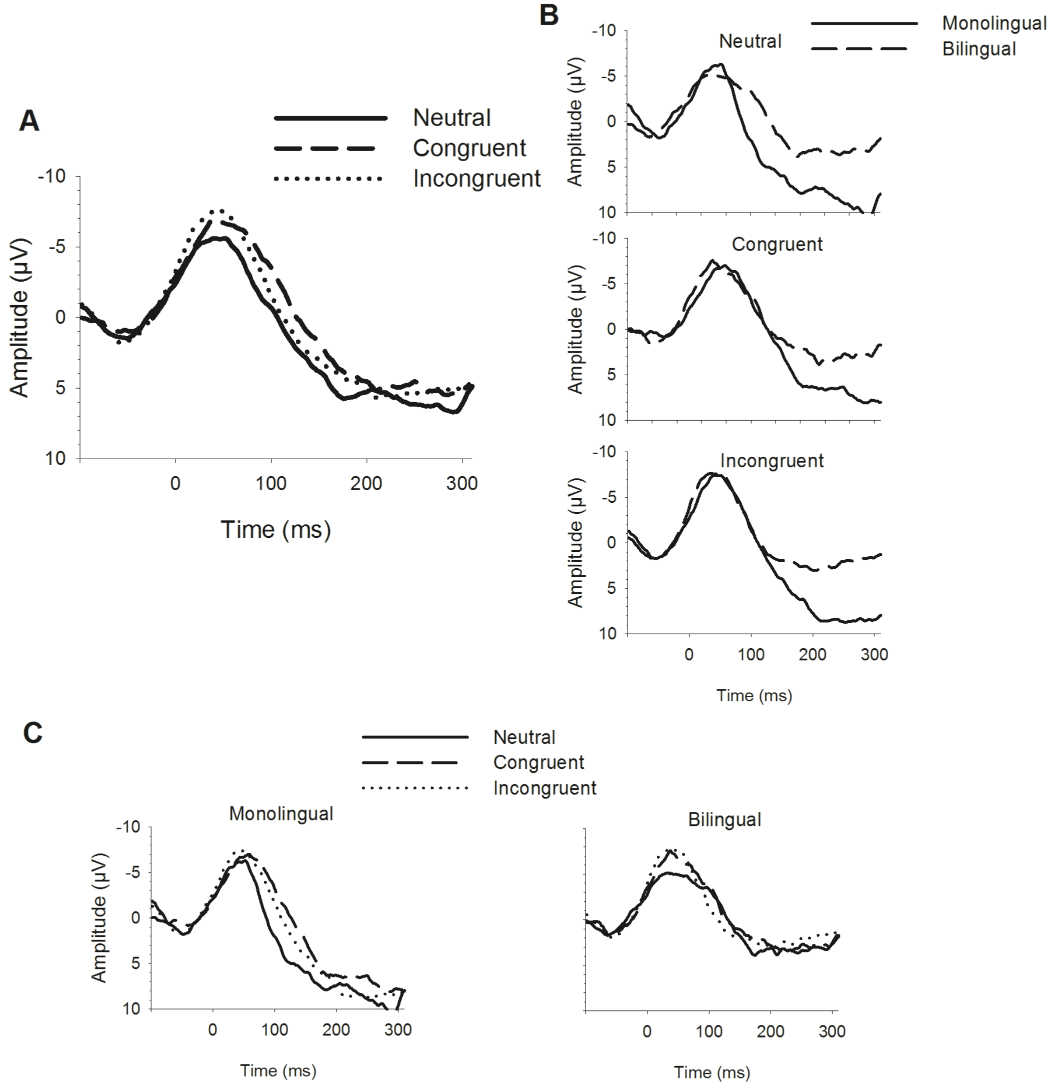


Figure 7. Response-locked grand averaged waveforms for the ERN for the Simon task depicted at site FCz. Panel A: main effect of Trial Type collapsed across Language Group; Panel B: effect of Language Group for each Trial Type; Panel C: effect of Trial Type for each Language Group.

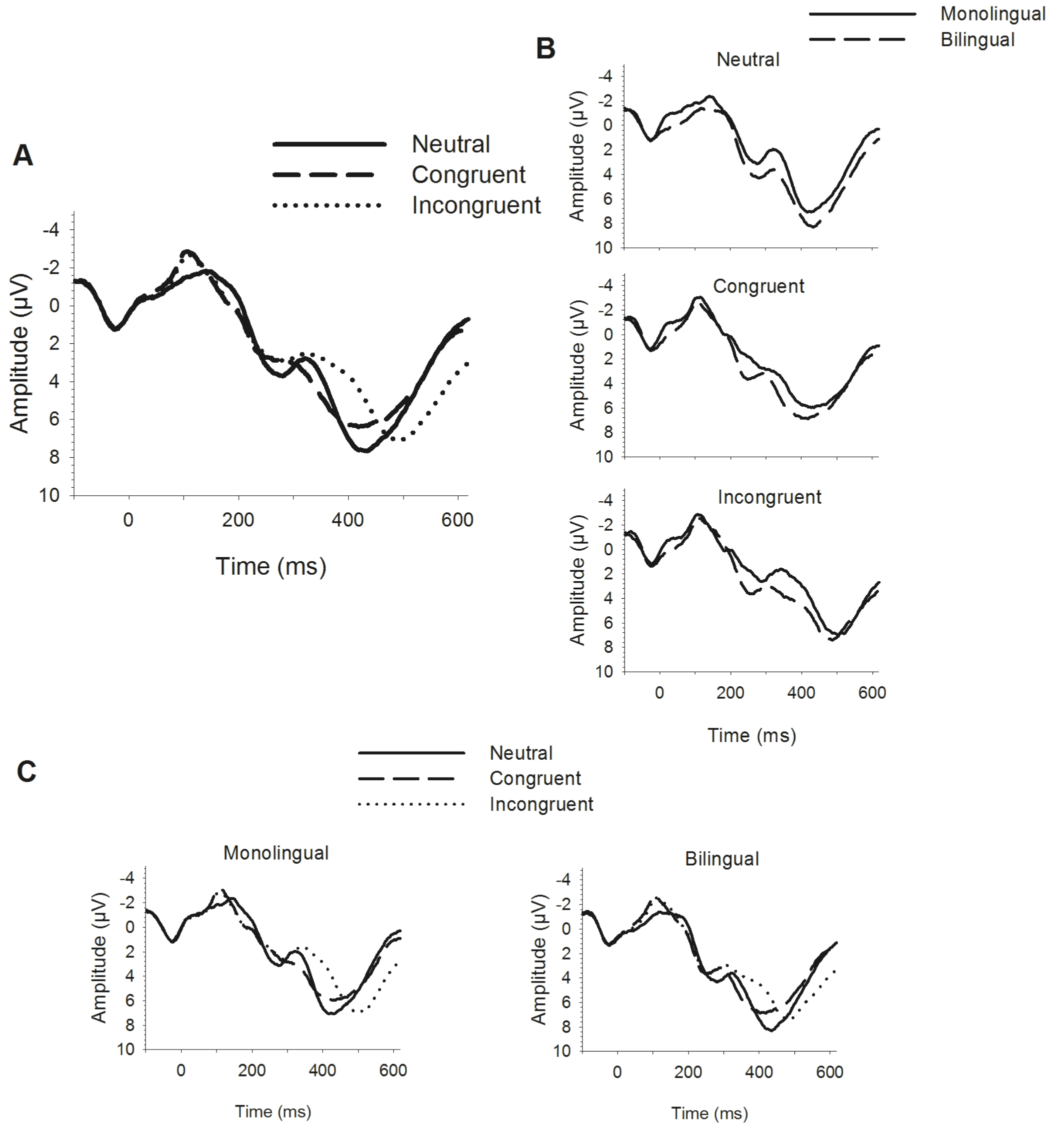


Figure 8. Stimulus-locked grand averaged waveforms for the N2 for the Eriksen task depicted at site FCz. Panel A: main effect of Trial Type collapsed across Language Group; Panel B: effect of Language Group for each Trial Type; Panel C: effect of Trial Type for each Language Group.

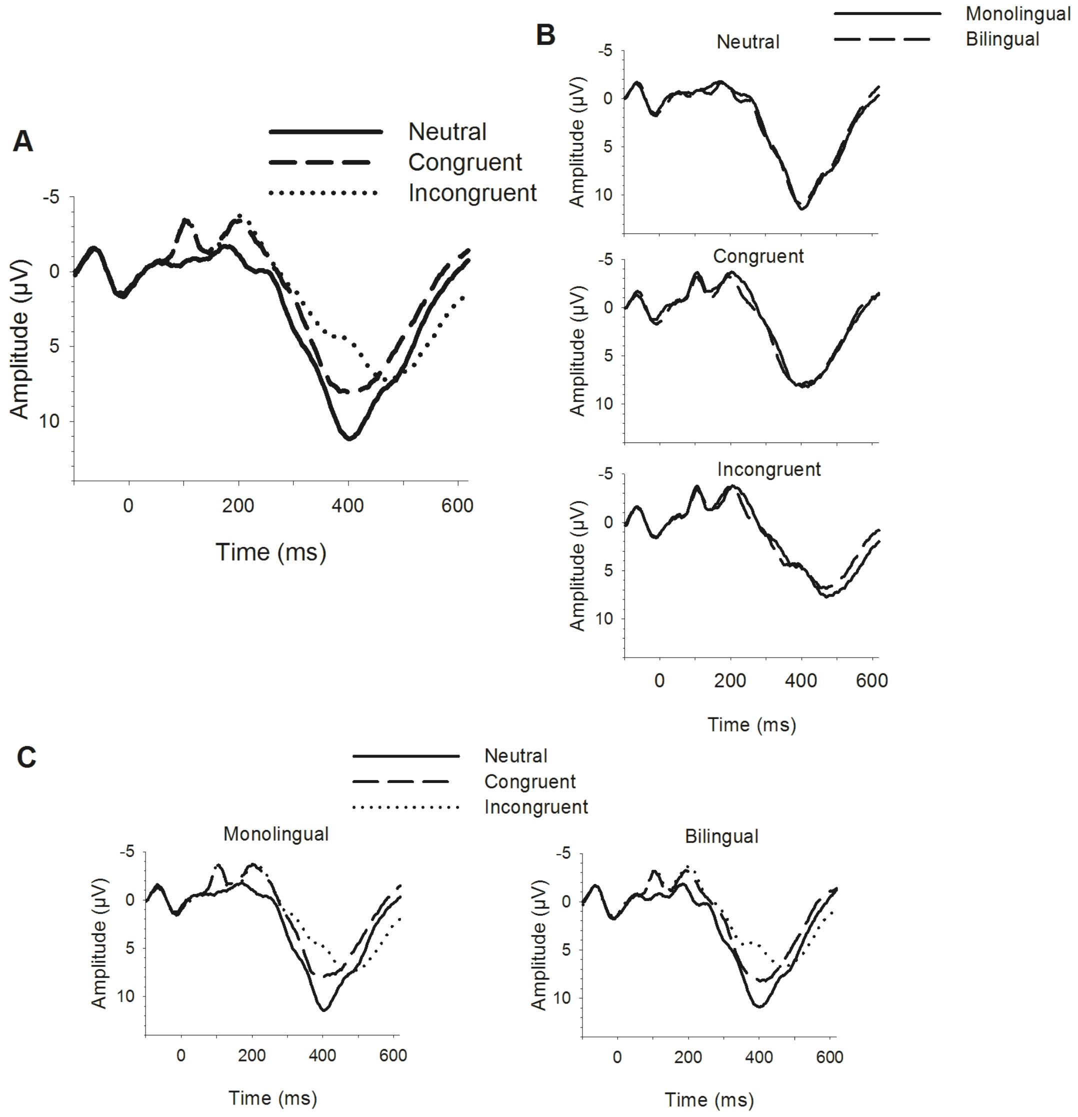


Figure 9. Stimulus-locked grand averaged waveforms for the P3 for the Eriksen task depicted at site Pz. Panel A: main effect of Trial Type collapsed across Language Group; Panel B: effect of Language Group for each Trial Type; Panel C: effect of Trial Type for each Language Group.

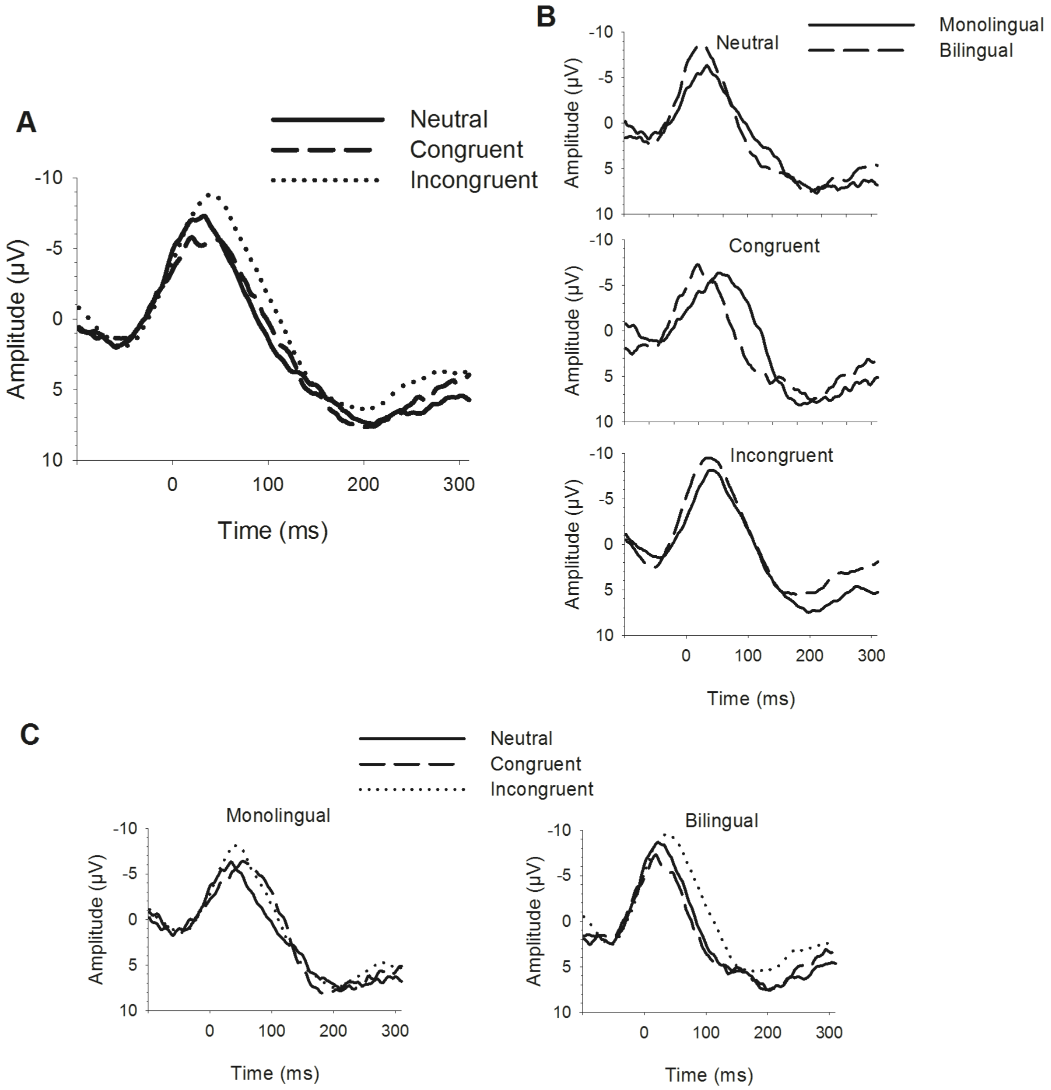
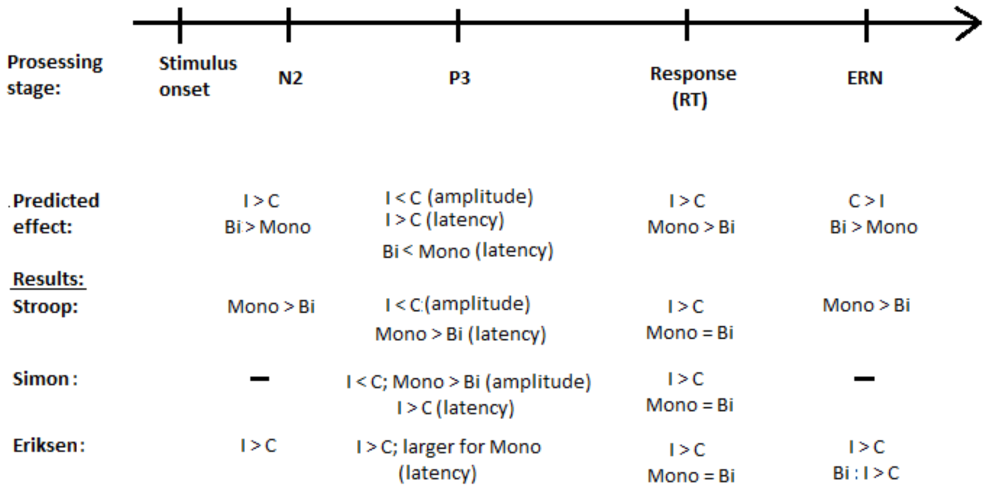


Figure 10. Response-locked grand averaged waveforms for the ERN for the Eriksen task depicted at site FCz. Panel A: main effect of Trial Type collapsed across Language Group; Panel B: effect of Language Group for each Trial Type; Panel C: effect of Trial Type for each Language Group.



Prosessing stage:	**Stimulus onset**	**N2**	**P3**	**Response (RT)**	**ERN**
Predicted effect:		I > C Bi > Mono	I < C (amplitude) I > C (latency) Bi < Mono (latency)	I > C Mono > Bi	C > I Bi > Mono
Results:					
Stroop:		Mono > Bi	I < C (amplitude) Mono > Bi (latency)	I > C Mono = Bi	Mono > Bi
Simon:		—	I < C; Mono > Bi (amplitude) I > C (latency)	I > C Mono = Bi	—
Eriksen:		I > C	I > C; larger for Mono (latency)	I > C Mono = Bi	I > C Bi : I > C

Figure 11. Timeline of events following stimulus presentation with predicted and observed results for congruent (C) and incongruent (I) trials in monolinguals (mono) and bilinguals (bi). Main effects are indicated first (e.g., I < C indicates a main effect of Trial Type; mono > bi indicates a main effect of Language Group; — indicates no significant effects), followed by significant interactions when present (e.g., Bi: I > C indicates an effect in bilinguals only).

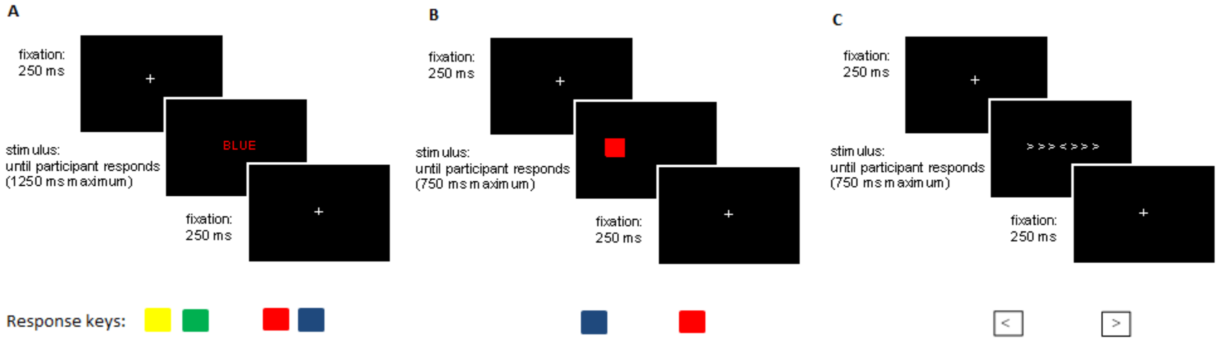


Figure 12. Sample incongruent trial for each task. The Stroop task is represented in panel A, the Simon task in panel B, and the Eriksen task in panel C.