A behavioural and electrophysiological investigation of the effect of bilingualism on aging and cognitive control

Shanna Kousaie[a] and Natalie A. Phillips[b]

[a]Cognitive Neuroscience Unit, Montreal Neurological Institute, McGill University

[b]Department of Psychology/Centre for Research in Human Development, Concordia University

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

Given previous, but inconsistent, findings of language group differences on cognitive control tasks the current investigation examined whether such differences could be demonstrated in a sample of older bilingual adults. Monolingual and bilingual older adults performed three cognitive control tasks that have previously been used in the literature (i.e., Stroop, Simon and flanker tasks) while brain electrophysiological recordings took place. Both behavioural (response time and accuracy) and event-related brain potentials (ERPs; N2 and P3 amplitude and latency) were compared across the two language groups. Processing differences between monolinguals and bilinguals were identified for each task, although the locus differed across the tasks. Language group differences were most clear in the Stroop task, with bilinguals showing superior performance both behaviourally and electrophysiologically. In contrast, for the Simon and flanker tasks there were electrophysiological differences that indicating language group processing differences at the level of conflict monitoring (Simon task only) and stimulus categorization (Simon and flanker tasks), but no behavioural differences. These findings support suggestions that these three tasks that are often used to examine executive control processes show little convergent validity; however, there are clear language group differences for each task that are suggestive of a superior performance for bilinguals, with behavioural differences emerging only in the linguistic Stroop task. Furthermore, it is clear that behavioural measures alone do not capture the language group effects in their entirety, and perhaps processing differences between language groups are more marked in a sample of older adults who are experiencing age-related cognitive changes than in younger adults who are at the peak of their cognitive capacity.

KEYWORDS: Aging; Bilingualism; Cognitive Control; Event-Related Brain Potentials (ERPs)

1. Introduction

In recent years there has been a marked increase in interest and research in the consequences of bilingualism for cognitive function, particularly in aging. This interests stems from findings of cognitive advantages for bilinguals compared to monolinguals (e.g., see Bialystok, Abutalebi, Bak, Burke, & Kroll, 2016; Bialystok, Craik, & Luk, 2012), which has implications for research, education and policy, and health, as well as other areas of cognitive science (e.g., cognitive training). The processing differences between bilinguals and monolinguals have been found primarily using tasks that measure cognitive control, including attentional and inhibitory control. Although there is substantial controversy in the literature regarding the reliability of these findings (see Hilchey & Klein, 2011; Paap, Johnson, & Sawi, 2015), there is little debate about the fact that there are differences between monolinguals and bilinguals in terms of language processing and brain plasticity (e.g., see Costa & Sebastian-Galles, 2014). The main issue that arises with respect to the consequences of bilingualism for cognitive control is identifying the circumstances necessary for processing differences to emerge. In the current study we compare monolingual and bilingual older adults on several tasks believed to measure cognitive control, using both behavioural and electrophysiological (event-related brain potentials, ERPs) measures. Our goal is to determine whether effects are observable in older adults, whether they are consistent across three tasks requiring the resolution of conflict, and whether there is evidence of processing differences at either the level of stimulus processing (as indexed by the ERP measures) and/or response output (as indexed by behavioural measures).

In an extensive review of the literature, Hilchey and Klein (2011) suggest that the superior performance seen in bilinguals may be a general speed advantage, rather than a more specific effect limited to conditions that require inhibitory control/interference suppression. In addition, one factor that appears to influence whether or not processing differences are observed, and that is highlighted by Hilchey and Klein, is the age of the participants. That is, superior performance in bilinguals compared to monolinguals may be difficult to detect in young adults who are at the peak of their cognitive functioning; however, in older adults who are experiencing age-related changes in cognition, it has been suggested that bilingual language experience may buffer against some of these cognitive changes (Bialystok, Martin, & Viswanathan, 2005).

In general, across a variety of cognitive control tasks that manipulate stimulus-response congruency, previous research has found that older adults show larger increases in response time (RT) for incongruent compared to congruent trials than young adults, irrespective of being bilingual. Previous research comparing monolingual and bilingual older adults on these same tasks has generally found superior performance in bilinguals than monolinguals. The relevant literature is reviewed below. We start by briefly examining the evidence for age-related decline on the three cognitive control tasks used here, followed by mention of any influence of bilingualism on task performance.

Previous research that has examined the effect of aging on performance of the Stroop task has found that older adults show larger Stroop effects (i.e., decreases in performance on incongruent than congruent trials) than younger adults. Studies have found larger Stroop effects for older than younger adults that are resistant to practice (Davidson, Zacks, & Williams, 2003; Dulaney & Rogers, 1994), as well as manipulations in stimulus orientation (i.e., upside down, or upside down and backward; Weir, Bruun, & Barber, 1997). Others have found a larger Stroop effect in older than in younger adults, both in terms of RT and accuracy, when participants were required to identify the colour that the stimulus was presented in, but not when they were required to identify the word (West, 2004). Bugg, DeLosh, Davalos, and Davis (2007) also found that age was associated with slower incongruent colour naming, above what could be accounted for by general age-related slowing. In addition, in a neuroimaging study, Milham et al. (2002) suggest that there are age-related changes in the neural underpinnings of Stroop task performance. Specifically, older adults showed less brain activation in regions related to attention control (i.e., dorsolateral prefrontal and parietal cortices), increases in sensitivity in brain regions important for response level evaluatory processing (i.e., anterior cingulate cortex), and brain activity patterns consistent with a decreased ability to suppress irrelevant information. However, a previous meta-analysis concluded that the apparent age-related decline in inhibitory function demonstrated by increases in the Stroop effect are actually an artifact of general age-related slowing (Verhaeghen & De Meersman, 1998).

With respect to the hypothesized effect of bilingualism on performance, Bialystok, Craik, and Luk (2008) examined whether being bilingual had an impact on Stroop performance by comparing younger and older monolinguals and bilinguals. Their results demonstrated larger Stroop effects in older than younger participants and in monolingual compared to bilingual participants. This difference was largely replicated in another study, which showed smaller Stroop interference in both younger and older bilinguals than in their monolingual peers, with the difference being larger in the older adults (Bialystok, Poarch, Luo, & Craik, 2014). However, Kousaie, Sheppard, Lemieux, Monetta, and Taler (2014) and Kousaie and Phillips (2012a, 2012b) did not find evidence for a language group differences in Stroop task performance in older adults. Additionally, a recent study using both a verbal and a numerical Stroop task failed to find superior performance in bilingual compared to monolingual older adults, as well as no modulation of executive control functions by second language proficiency within a group of older bilinguals (Antón, Fernández García, Carreiras, & Duñabeitia, 2016).

In terms of the Simon task, previous research has found that older adults demonstrated larger Simon effects (i.e., greater increases in response time for incongruent trials compared to congruent trials), even when general age-related slowing was accounted for (Van der Lubbe & Verleger, 2002). The Simon task has also been used to examine the effect of bilingualism on the processes required to successfully inhibit information from the irrelevant dimension and respond to the relevant aspect of the stimulus. The first study to examine this found that the Simon effect was larger for older than middle-aged adults, as well as for monolinguals than bilinguals (Bialystok, Craik, Klein, & Viswanathan, 2004). It is noteworthy that Bialystok et al. found a smaller Simon effects for bilinguals compared to monolinguals in both age groups. However, in another study, Bialystok, Craik and Luk (2008) found similar performance on a Simon task for monolingual and bilingual older adults. Similarly, Kousaie et al. (2014) did not find language group differences in the Simon effect despite finding an overall larger Simon effect for older than younger adults.

Finally, the flanker task has elicited more subtle effects of aging on task performance. That is, previous research has shown similar behavioural flanker interference effects in older and younger adults[1], with more sensitive measures (i.e., electrophysiological measures) suggesting age-related differences during flanker task performance (Hsieh & Fang, 2012; Wild-Wall, Falkenstein, & Hohnsbein, 2008). Specifically, Wild-Wall et al. (2008) found similar flanker

---

[1] Note that we are referring here to studies that have used arrowhead stimuli given that this is most similar to the task that we employed. However, other studies using letter stimuli have demonstrated greater interference from incompatible flankers in older than younger participants (Zeef & Kok, 1993; Zeef, Sonke, Kok, Buiten, & Kenemans, 1996).

interference in young and older participants but greater accuracy in the older adults, which they attribute to differential target processing in the two age groups as revealed by electrophysiological measures. Similarly, Hsieh and Fang (2012) found similar performance for younger and older adults in terms of response times; however, a smaller flanker effect in older than younger adults in terms of accuracy and age-differences in the electrophysiological response suggested that older adults used compensatory strategies to attain similar performance as younger adults. To our knowledge, the flanker task has not been used to compare cognitive control processes across language groups in older adults. However, Gollan, Sandoval, and Salmon (2011) found that error rates on a non-linguistic flanker task were associated with failures in language control (i.e., cross-language intrusion errors in a category fluency task) in older but not younger bilinguals. This supports the hypothesis that bilinguals rely on general cognitive control mechanisms to manage their two languages and that these mechanisms are susceptible to age-related decline.

It is clear from the literature reviewed here that a consensus on the effects of bilingualism on the performance of tasks purported to measure cognitive control in older adults has yet to be achieved. Thus, more sensitive measures such as those provided by brain imaging may be more amenable to detecting language group differences. Two examples can be found from studies using electrophysiological (event-related brain potentials; ERPs) and functional magnetic resonance imaging (fMRI) measures. In young adults, Kousaie and Phillips (2012b) used the same three tasks used in the current investigation and found no evidence of language group differences in behavioural measures, but did find that ERP measures demonstrated differences suggestive of superior performance in bilinguals; however, these differences were not consistent across the three tasks. Specifically, differences were observed in terms of conflict monitoring

(Stroop task), resource allocation and stimulus evaluation (Stroop, Simon, and flanker tasks), and error monitoring (Stroop and flanker tasks). In older adults, Ansaldo, Ghazi-Saidi, and Adrover-Roig (2015) demonstrated language group differences in brain activation during Simon task performance in a sample of monolinguals and bilinguals who showed similar behavioural performance. In that study, older monolinguals and bilinguals performed a Simon task in the MRI scanner and demonstrated different neural correlates supporting similar behavioural performance with monolinguals showing activity in brain regions classically associated with interference control and bilinguals showing activation in regions related to visuospatial processing. These findings suggest that monolinguals and bilinguals engaged different strategies to achieve the same behavioural outcome on the Simon task. One interesting question is whether language group differences in brain imaging measures in the absence of behavioural differences constitutes superior performance. We will return to this question in the discussion.

The ERP measures used by Kousaie and Phillips (2012b) and in the current investigation are extracted from the ongoing electroencephalogram recorded during task performance. ERPs have excellent temporal resolution, thus allowing for more precise identification of the locus of language group difference in the stream of information processing. ERPs refer to characteristic waveforms elicited by underlying sensory and cognitive processes, and their amplitude and timing are believed to reflect the strength and timing of these processes (Rugg & Coles, 1995). With respect to cognitive control, two ERP components are of particular relevance: the N2 and the P3.

The N2 is a negative-going waveform that peaks 200-350 ms following a stimulus and has a frontocentral distribution (see Folstein & Van Petten, 2008). The amplitude of the N2 is thought to be modulated by the degree of conflict monitoring, with higher conflict trials eliciting a larger amplitude N2 (e.g., van Veen & Carter, 2002a, 2002b). In addition, the N2 has been correlated with activity in the anterior cingulate cortex (ACC) as measured by fMRI (Mathalon, Whitfield, & Ford, 2003), a brain area thought to be “tuned” by bilingual language experience (Abutalebi et al., 2012).

The P3 is a broad positive-going waveform that peaks 300-600 ms following the stimulus and has a centroparietal distribution. The amplitude of the P3 is thought to be related to the updating of working memory (Donchin, 1981) and the allocation of resources (see Polich, 2007), with tasks requiring fewer resource eliciting a larger amplitude P3. The latency of the P3 is believed to reflect the time it takes to categorize a stimulus (Kutas, McCarthy, & Donchin, 1977) such that factors that load stimulus processing (e.g., intensity, discriminability, etc.) result in a delayed P3 peak. However, the P3 latency may also be influenced by response-related processes when tasks are relatively simple or response times are fast (Verleger, 1997). More recent formulations comparing stimulus- and response-locked ERPs suggest that P3 latency reflects a function that is the interface between perceptual and response processing at the level of stimulus-response links, reflecting response control (Berchicci, Spinelli, & Russo, 2016).

As briefly described earlier, ERPs have been used to examine the effects of aging on flanker task performance (Hsieh & Fang, 2012; Wild-Wall et al., 2008). ERPs have also been used in previous research with monolingual and bilingual young adults to demonstrate differences in brain responses in the absence of behavioural differences (Kousaie & Phillips, 2012b). Specifically, the same sample of monolingual and bilingual young adults performed the Stroop, Simon, and flanker tasks while electrophysiological recording took place. The two language groups showed the same behavioural performance; however, they differed in terms of their ERP responses and these differences were not consistent across the three tasks. More specifically, findings were mixed with some results being suggestive of superior performance for bilinguals (e.g., earlier P3 peak latency in bilinguals than monolinguals during Stroop task performance), while others were suggestive of superior performance for monolinguals (e.g., larger P3 amplitude in monolinguals than bilinguals during Simon task performance).

In the current investigation we applied the same methodology as Kousaie and Phillips (2012b) to examine whether older monolinguals and bilinguals differ in terms of behavioural performance and/or brain responses during the performance of multiple tasks that have previously been used in the literature and that require some degree of conflict monitoring, response inhibition, and interference suppression. We used a colour Stroop task (Stroop, 1935), a spatial Simon task (Simon & Rudell, 1967), and a flanker task (Eriksen & Eriksen, 1974) using chevrons as stimuli. All three of these tasks require participants to ignore some irrelevant aspect of a stimulus in order to respond correctly, and each task is comprised of congruent and incongruent trial types. In the Stroop task, participants were presented with colour words printed in either the matching colour (i.e., congruent trial; e.g., the word *RED* printed in red) or a non-matching colour (i.e., incongruent trials; e.g., the word *RED* printed in blue) and were required to identify the colour of the print, while inhibiting the reading of the word. The Simon task was comprised of coloured squares (i.e., red or blue) presented laterally on the computer monitor and the participant was required to identify the colour of the square using lateral response keys. For some trials the response key and stimulus presentation were on the same side (i.e., congruent trials), whereas for others the stimulus and response were on opposite sides (i.e., incongruent trials). Participants were required to ignore the spatial position of the stimulus in order to correctly identify its colour. Finally, the flanker task was comprised of a central chevron flanked on either side by chevrons pointing in the same direction (i.e., congruent trials) or the opposite direction (i.e., incongruent trials) as the central target stimulus. Using lateral response keys, participants were required to identify the direction of the central chevron while ignoring the distracting flanker chevrons.

We hypothesized that if there is a clear benefit of being bilingual on task performance this would be evident in both behavioural and electrophysiological measures. Specifically, bilinguals would show faster response times and/or smaller interference effects (i.e., smaller increases in response time for incongruent compared to congruent trials) than monolinguals, as well as larger N2 and P3 amplitudes, and earlier P3 peak amplitudes. Of particular interest is whether any observed language group differences are consistent across the three tasks used here or whether the pattern of language group differences varies as a function of the task being performed. Given recent findings that report inconsistencies across different cognitive control tasks (e.g., Kousaie & Phillips, 2012b; Paap & Sawi, 2014), including those used here (see Kousaie & Phillips, 2012b), an important strength of the current investigation is the use of multiple tasks in the same group of older monolingual and bilingual participants. Thus, if a variable pattern of language group differences emerge, the use of multiple tasks of executive control that vary in the stage at which interference must be resolved could reveal interesting processing differences between the groups.

2. Material and Methods

2.1 Participants

Forty-three older adults between the ages of 60 and 83 were recruited from the Montreal, Quebec community. Twenty-one were monolingual English speakers (18 females; mean age=71.7±6.8 years) with minimal exposure to a second language, and 22 were highly proficient bilingual speakers of English and French (15 females; mean age=68.7±5.2 years). Participants self-reported no illness, health condition, or use of medication that is known to affect cognitive functioning, and all performed within normal range on the Montreal Cognitive assessment (MoCA; Nasreddine et al., 2005). Bilinguals spoke only English and French and had learned their second language before age 18 (mean age of L2 acquisition=4.9±5.1 years), self-reported high L2 proficiency and used both of their languages on a daily basis. In addition, bilinguals showed comparable performance in English and French on an animacy judgement task (Segalowitz & Frenkiel-Fishman, 2005), which we used as an objective measure of relative L2 proficiency.

Demographic information is provided in Table 1. The two language groups were matched with respect to age, education, and MoCA performance. In some case where participants were removed from an analysis due to poor quality of the electrophysiological recording the groups remained matched on these demographic variables.

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

Table 1. *Demographic information for participant groups*

	Monolinguals (n = 21; 18 females) M *(SD)*	Bilinguals (n = 22; 15 females) M *(SD)*	p^c
Age (years)	71.7 *(6.8)*	68.7 *(5.2)*	.11
Education (years)	14.8 *(3.6)*	15.6 *(3.1)*	.45
MoCA[a]	27.5 *(1.6)*	27.4 *(1.8)*	.83
L1 self-reported language proficiency[b]	5.0 *(0.0)*	4.9 *(0.3)*	.09
L2 self-reported language proficiency[b]	1.4 *(0.7)*	4.4 *(0.5)*	<.01*
Coefficient of variability L1	.26 *(.08)*	.22 *(.09)*	.14
Coefficient of variability L2	n/a	.23 *(.07)*	n/a

[a] Maximum score = 30; ≥26 normal cognitive function
[b] Self-report on a scale of 1-5: 1=no ability at all; 5=native-like ability
[c] *p*-value from independent samples t-test comparing the two language groups; the two groups differ only in terms of their self-reported L2 proficiency

2.2 Materials and Apparatus

Participants completed the MoCA (Nasreddine et al., 2005) to ensure that their cognitive function was within the range considered to be normal, as well as to ensure that the language groups were matched with respect to this measure; an animacy judgement task to assess relative L1 and L2 proficiency (monolinguals only performed this task in English, while bilinguals performed it in both English and French; Segalowitz & Frenkiel-Fishman, 2005); and three experimental tasks for which EEG recording took place, these included modified Stroop, Simon, and Flanker tasks. 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).

2.2.1 MoCA. The MoCA (Nasreddine et al., 2005) is a 10-minute cognitive assessment used to screen for mild cognitive impairment in older adults. It assesses visuospatial/executive control, memory, attention, language, and orientation. The assessment is scored out of 30, with a score of 26 or higher considered within the normal range.

2.2.2 Animacy Judgment Task. This task was used as an objective measure of relative L2 proficiency. Participants are required to categorize nouns as animate or inanimate, as quickly and accurately as possible, producing a measure of automaticity of language processing (Segalowitz & Frenkiel-Fishman, 2005). The task used here 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. Monolinguals performed the task in English only, while bilinguals performed it in both English and French in order to compare their performance across L1 and L2.

2.2.3 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 started with a fixation cross which stayed on the screen 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 Flanker tasks). The practice block was performed first and repeated if necessary until accuracy reached a minimum of 80%. A 250 Hz tone identified errors during the practice block; however, no performance feedback was provided during the experimental blocks. See Figure 1 for a sample trial of each task.

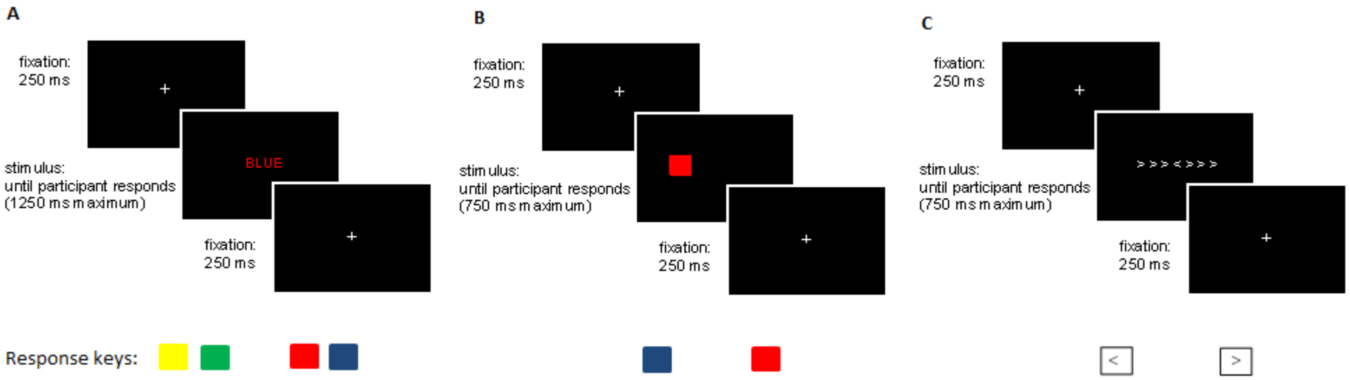

Figure 1. Sample trial for the Stroop task (A), the Simon task (B), and the flanker task (C).

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 orthographic form of the colour word (e.g., “xxx” printed in red); congruent trials comprised the colour words *green*, *yellow*, *red*, and *blue* printed in the corresponding colour; and incongruent trials comprised the same colour words printed in one of the alternate three colours (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 colour of the print using the keyboard; “z” for yellow, “x” for green, “,” for red, and “.” for blue. Prior to the practice block, participants performed a key acquisition task that comprised 80 trials for which participants were required to identify the colour of green, yellow, red, and blue circles. Participants were permitted to repeat the acquisition task until they were confident that they had learned the response keys.

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

2.2.4 EEG Recording. The continuous EEG was recorded from 64 scalp locations positioned 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 each 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 1100 ms including a 100 ms pre-stimulus baseline and stimulus-locked averages were calculated for each trial type (including only correct trials), which resulted in three averages per task for each participant (i.e., neutral, congruent, and incongruent).

2.3 Procedure

Participants were seated in a comfortable chair and informed consent was obtained. The MoCA was completed first, followed by the animacy judgement task. The electrodes 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 Flanker tasks in counterbalanced order. The testing session lasted approximately 2 hours, with approximately 60 minutes of EEG recording. Following completion of the experiment participants were debriefed and compensated $10 per hour of participation.

3. Results

Statistical analyses were conducted using the statistical software package SPSS v. 22 (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. Behavioural results will be reported first followed by the electrophysiological results. The sample size of participants varies by task due to the poor technical quality of the electrophysiological recordings for a handful of participants for specific tasks. However, with one exception, the sample size per group per task is 20 participants or higher (see Table 2). In order to facilitate a direct comparison, the behavioural data presented are from the same participants as in the ERP data. An initial set of analyses included results from the neutral trials. Results from the accuracy data, RT data, and multiple ERP components across the three tasks did not reveal differences between the language groups. Given these multiple measures, we omit report of these results in the interest of brevity.

Table 2. *Sample size for statistical analyses following exclusion of participants due to poor electrophysiological recording.*

	Monolinguals	Bilinguals
Stroop	19	20
Simon	21	21
Flanker	20	20

3.1 Behavioural Results

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

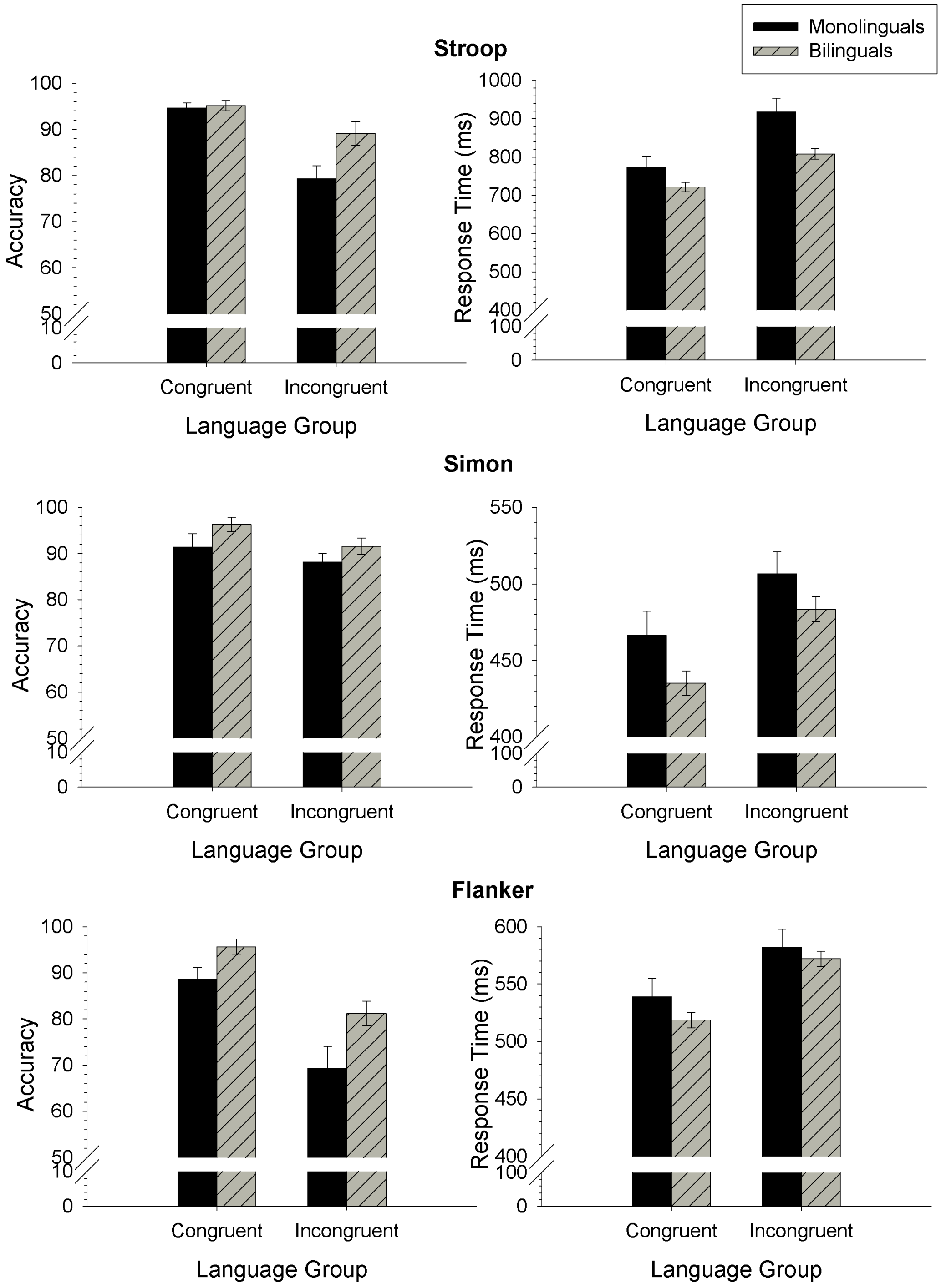

Figure 2. Accuracy (left) and response time (right) data for each task.

3.1.1 Stroop task. Accuracy. All participants demonstrated high accuracy. There was a main effect of Trial Type ($F(1,37)=57.6$, $MSE=38.49$, $p<.01$, $\eta^2_p=.61$), indicating lower accuracy for incongruent than congruent trials. There was also a main effect of Language Group ($F(1,37)=4.06$, $MSE=128.55$, $p=.05$, $\eta^2_p=.10$), demonstrating that bilinguals were more accurate than monolinguals; however, a Language Group x Trial Type interaction ($F(1,37)=11.01$, $MSE=38.49$, $p<.01$, $\eta^2_p=.23$) showed that this was only the case for incongruent trials, whereas for congruent trials the two language groups performed similarly.

RT. Results from the analysis of the RT data paralleled those from the accuracy data. Specifically, there was a main effect of Trial Type ($F(1,37)=131.4$, $MSE=1968.25$, $p<.01$, $\eta^2_p=.78$), indicating faster responses for congruent than incongruent trials. A main effect of Language Group ($F(1,37)=5.96$, $MSE=21297.05$, $p=.02$, $\eta^2_p=.14$), showed that bilinguals responded faster than monolinguals, and a Language Group x Trial Type interaction ($F(1,37)=7.97$, $MSE=1968.25$, $p<.01$, $\eta^2_p=.18$) showed that this was only the case for incongruent trials, whereas for responses times for congruent trials were similar for the two language groups.

3.1.2 Simon task. Accuracy. There was a main effect of Trial Type ($F(1,40)=14.85$, $MSE=22.62$, $p<.01$, $\eta^2_p=.27$), demonstrating higher accuracy for congruent trials than incongruent trials. There was no effect of Language Group ($F(1,40)=2.32$, $MSE=158.61$, $p=.13$), nor a Language Group x Trial Type interaction ($F(1,40)=0.49$, $MSE=22.62$, $p=.49$).

RT. Again, analysis of the RT data paralleled the accuracy results. There was a main effect of Trial Type ($F(1,40)=315.81$, $MSE=130.19$, $p<.01$, $\eta^2_p=.89$), indicating faster responses for congruent trials than incongruent trials. There was no effect of Language Group

(*F*(1,40)=2.58, *MSE*=6034.73, *p*=.12), nor a Language Group x Trial Type interaction (*F*(1,40)=2.49, *MSE*=130.19, *p*=.12).

3.1.3 Flanker task. Accuracy. There was a main effect of Trial Type (*F*(1,38)=88.44, *MSE*=64.35, *p*<.01, η^2_p =.70), showing more accurate responses for congruent than incongruent trials. There was also a significant effect of Language Group (*F*(1,38)=5.49, *MSE*=322.84, *p*=.02, η^2_p =.13) showing that bilinguals were more accurate than monolinguals. There was no significant Language Group x Trial Type interaction (*F*(1,38)=1.87, *MSE*=64.35, *p*=.18).

RT. There was a main effect of Trial Type (*F*(1,38)=277.03, *MSE*=168.01, *p*<.01, η^2_p =.88), demonstrating faster RTs for congruent trials than for incongruent trials. There was no effect of Language Group (*F*(1,38)=0.80, *MSE*=5808.88, *p*=.38), nor a Language Group x Trial Type interaction (*F*(1,38)=3.27, *MSE*=168.01, *p*=.08.

3.2 Electrophysiological Results

Separate analyses were conducted for each component of interest (i.e., N2 and P3) 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 was 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 given its documented frontocentral distribution (e.g.,Folstein & Van Petten, 2008) and Cz, CPz, and Pz were included for analysis of the P3 given its centroparietal scalp distribution (Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Squires, Squires, & Hillyard, 1975).

For each component we conducted a series of mixed factors ANOVAs. First, we analyzed the mean amplitude of the waveform across the time window of the component, as determined by examining the grand averaged waveforms. Given that the ERP components can be quite broad we decided to analyze the mean integrated amplitude in order to capture the whole component[2]. However, in order to isolate differences in peak amplitude and peak latency, we also analyzed the amplitude and latency of the maximum/minimum peak for each component. Given that we hypothesized that there would be a Language Group x Trial Type interaction, we examined this interaction with planned simple effects comparisons for each analysis.

Figures 3 to 8 depict the grand averaged waveforms for each task and component separately. In each figure, panel A compares monolinguals and bilinguals for each Trial Type separately, and panel B 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 Pz for the P3; all waveforms are stimulus-locked.

3.2.1 Stroop task. The N2 was analyzed between 200 and 500 ms and the P3 between 350 and 830 ms; see Figures 3 and 4, respectively. Visual inspection of Figures 3 and 4 show obvious N2 and P3 components, and suggest larger N2 amplitude for monolinguals than bilinguals, as well as larger P3 amplitude for congruent than incongruent trials in both language groups.

---

[2] It is noteworthy, that in doing this the intervals that were chosen to be representative of the N2 and P3 overlap somewhat. However, importantly the two components were analyzed at different electrode locations (the N2 at sites Fz and FCz; the P3 at sites Cz, CPz, and Pz) representing their canonical topographical distribution.

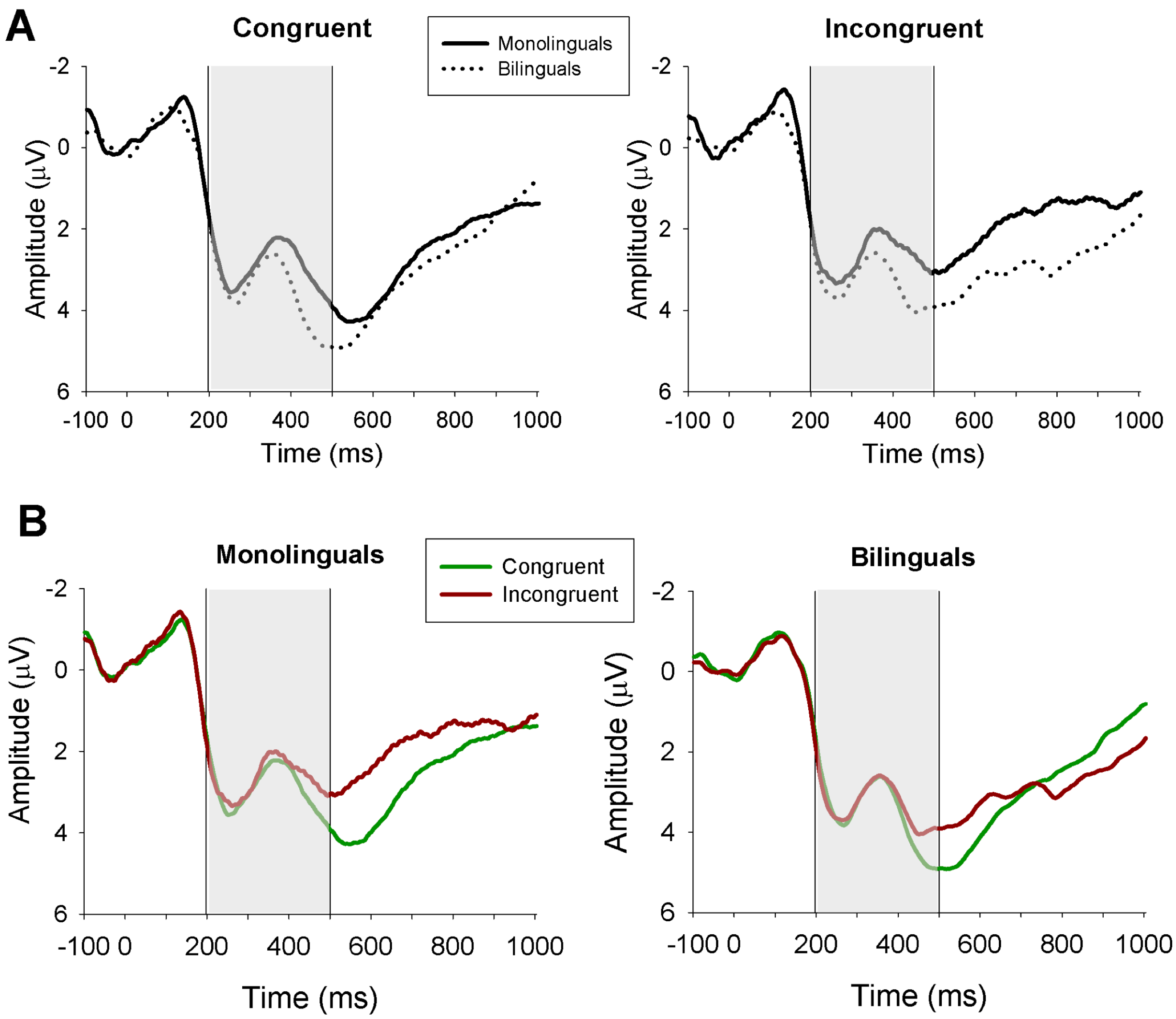

Figure 3. Stimulus-locked grand averaged waveforms depicting the N2 at site FCz for the Stroop task. Panel A compares monolinguals and bilinguals for each trial type and panel B compares congruent to incongruent trials for each language group. The shaded area indicates the interval included in the analysis.

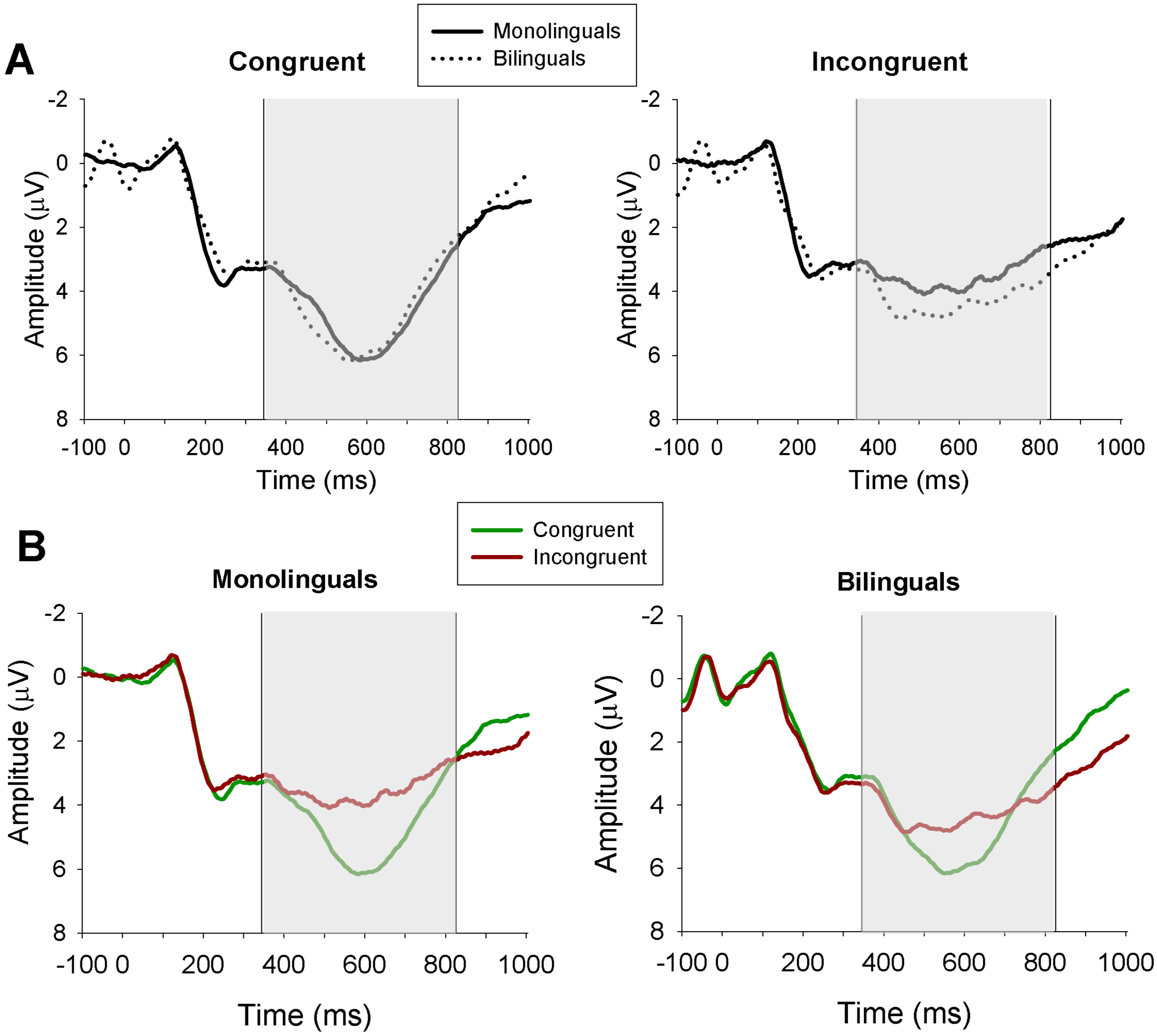

Figure 4. Stimulus-locked grand averaged waveforms depicting the P3 at site Pz for the Stroop task. Panel A compares monolinguals and bilinguals for each trial type and panel B compares congruent to incongruent trials for each language group. The shaded area indicates the interval included in the analysis.

Analysis of mean N2 amplitude and peak N2 amplitude revealed no significant effects, all *F*s < 2.4, all *p*s > .12. However, analysis of N2 peak latency revealed a main effect of Language Group (*F*(1,36)=4.56, *MSE*=7990.18, *p*=.04, η^2_p =.11), demonstrating earlier N2 peak latency in bilinguals (*M*=320.5 ms) than in monolinguals (*M*=351.5 ms). The planned Language Group x Trial Type comparison, showed that peak N2 latency was earlier in bilinguals than monolinguals for congruent trials (*p*=.04) and was a trend for incongruent trials (*p*=.06).

Analysis of mean P3 amplitude revealed a main effect of Trial Type (*F*(1,37)=18.23, *MSE*=1.86, *p*<.01, η^2_p =.33), demonstrating smaller P3 amplitude for incongruent than congruent trials. Importantly, a main effect of Language Group (*F*(1,37)=3.97, *MSE*=50.08, *p*=.05, η^2_p =.10), demonstrated larger P3 amplitude for bilinguals than monolinguals. The Language Group x Trial Type interaction was also significant (*F*(1,37)=5.87, *MSE*=1.86, *p*=.02, η^2_p =.14), which indicated the P3 amplitude was larger in bilinguals than monolinguals for incongruent trials only and only the monolingual group showed a significant effect of Trial Type (i.e., larger P3 amplitude for congruent than incongruent trials). The only significant effect that emerged from the analysis of peak P3 amplitude was a main effect of Trial Type (*F*(1,37)=22.7, *MSE*=4.69, *p*<.01, η^2_p =.38), which showed larger peak amplitudes for congruent than incongruent trials. There were no significant effects of Language Group or Trial Type on P3 peak latency all *F*s < 0.06, all *p*s > .81.

3.2.2 Simon task. The N2 was analyzed between 200 and 500 ms and the P3 between 300 and 700 ms; see Figures 5 and 6, respectively. Visual inspection of Figure 5 shows a small N2 component that is more prominent, and larger, in monolinguals than in bilinguals. Inspection of Figure 6 shows a large P3 component that appears larger in bilinguals than monolinguals, as well as earlier for congruent than incongruent trials.

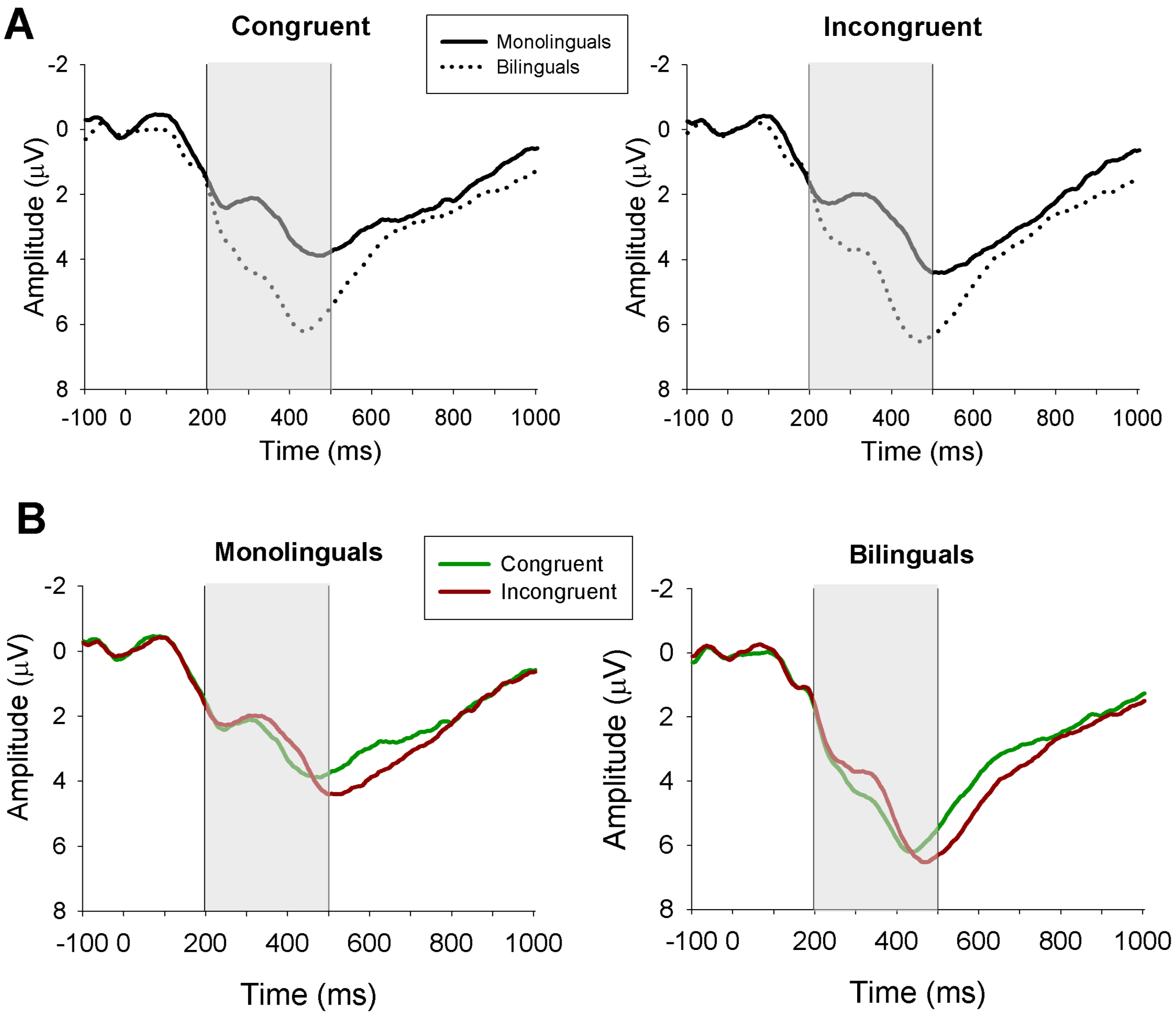

Figure 5. Stimulus-locked grand averaged waveforms depicting the N2 at site FCz for the Simon task. Panel A compares monolinguals and bilinguals for each trial type and panel B compares congruent to incongruent trials for each language group. The shaded area indicates the interval included in the analysis.

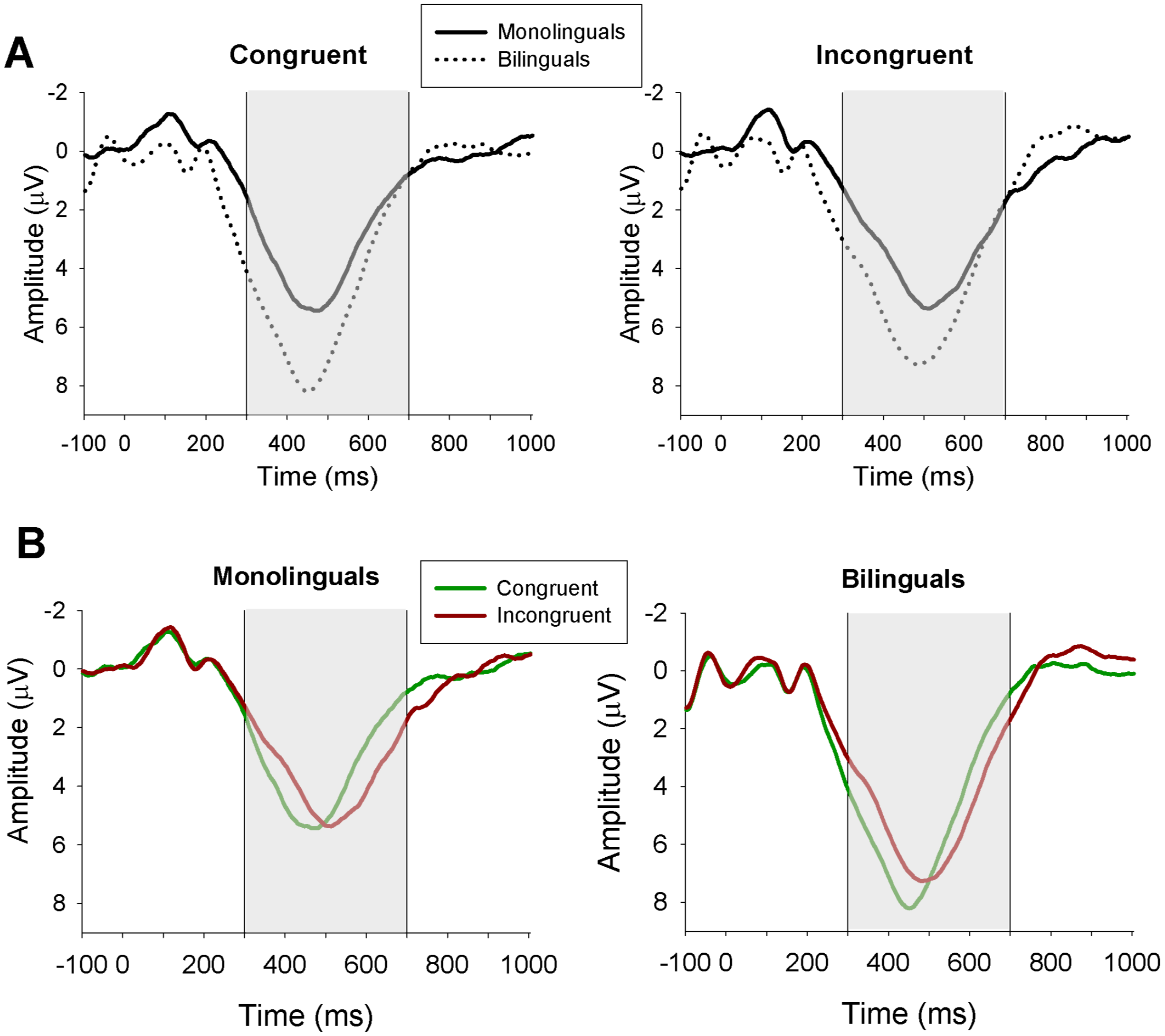

Figure 6. Stimulus-locked grand averaged waveforms depicting the P3 at site Pz for the Simon task. Panel A compares monolinguals and bilinguals for each trial type and panel B compares congruent to incongruent trials for each language group. The shaded area indicates the interval included in the analysis.

Analysis of the mean N2 amplitude revealed a main effect of Language Group ($F(1,40)=4.88$, $MSE=19.17$, $p=.03$, $\eta^2_p=.11$), demonstrating larger N2 amplitude for monolinguals than bilinguals. Analysis of peak N2 amplitude showed a main effect of Trial Type ($F(1,40)=8.72$, $MSE=1.18$, $p<.01$, $\eta^2_p=.18$), showing larger peak amplitude for incongruent than congruent trials. The planned comparison of the Language Group x Trial Type interaction ($p=.47$) showed that only the bilinguals showed larger peak amplitudes for incongruent than congruent trials. With respect to N2 peak latency, there was a main effect of Trial Type ($F(1,40)=6.43$, $MSE=1483.72$, $p=.02$, $\eta^2_p=.14$), showing earlier peak latency for congruent than incongruent trials. The planned comparison of the Language Group x Trial Type interaction revealed that only the monolinguals demonstrated a significant effect of Trial Type.

Analysis of mean P3 amplitude revealed a main effect of Language Group ($F(1,40)=3.97$, $MSE=41.55$, $p=.05$, $\eta^2_p=.09$), demonstrating larger P3 amplitudes for bilinguals than monolinguals. There were no other significant effects, all Fs < 3.34, all ps > .08. Analysis of peak P3 amplitude revealed no significant effects, all Fs < 3.10, all ps > .09; however, peak P3 latency was earlier for congruent than incongruent trials (main effect of Trial Type: $F(1,40)=17.39$, $MSE=311.82$, $p<.01$, $\eta^2_p=.30$) and for bilinguals than monolinguals (main effect of Language Group: $F(1,40)=8.35$, $MSE=17563.00$, $p<.01$, $\eta^2_p=.17$).

3.2.3 Flanker task. The N2 was analyzed between 220 and 420 ms and the P3 between 320 and 820 ms, see Figures 7 and 8, respectively. Visual inspection of Figure 7 shows no obvious N2 component, although the waveform in the time window encompassing a traditional N2 appears larger in monolinguals than bilinguals, and for incongruent than congruent trials. Inspection of Figure 8 shows a large P3 component that appears larger for bilinguals than monolinguals, and earlier for congruent than incongruent trials.

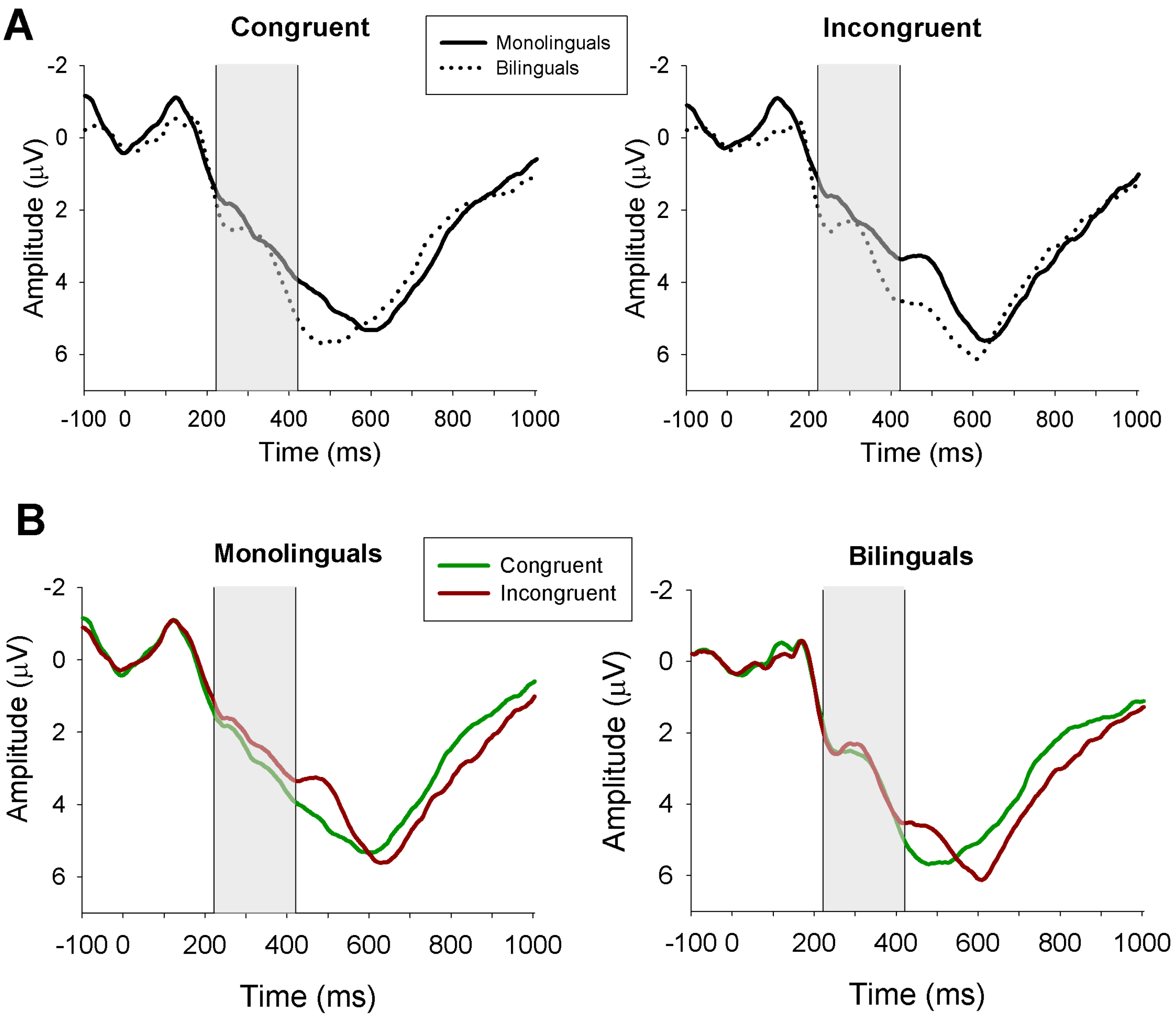

Figure 7. Stimulus-locked grand averaged waveforms depicting the N2 at site FCz for the flanker task. Panel A compares monolinguals and bilinguals for each trial type and panel B compares congruent to incongruent trials for each language group. The shaded area indicates the interval included in the analysis.

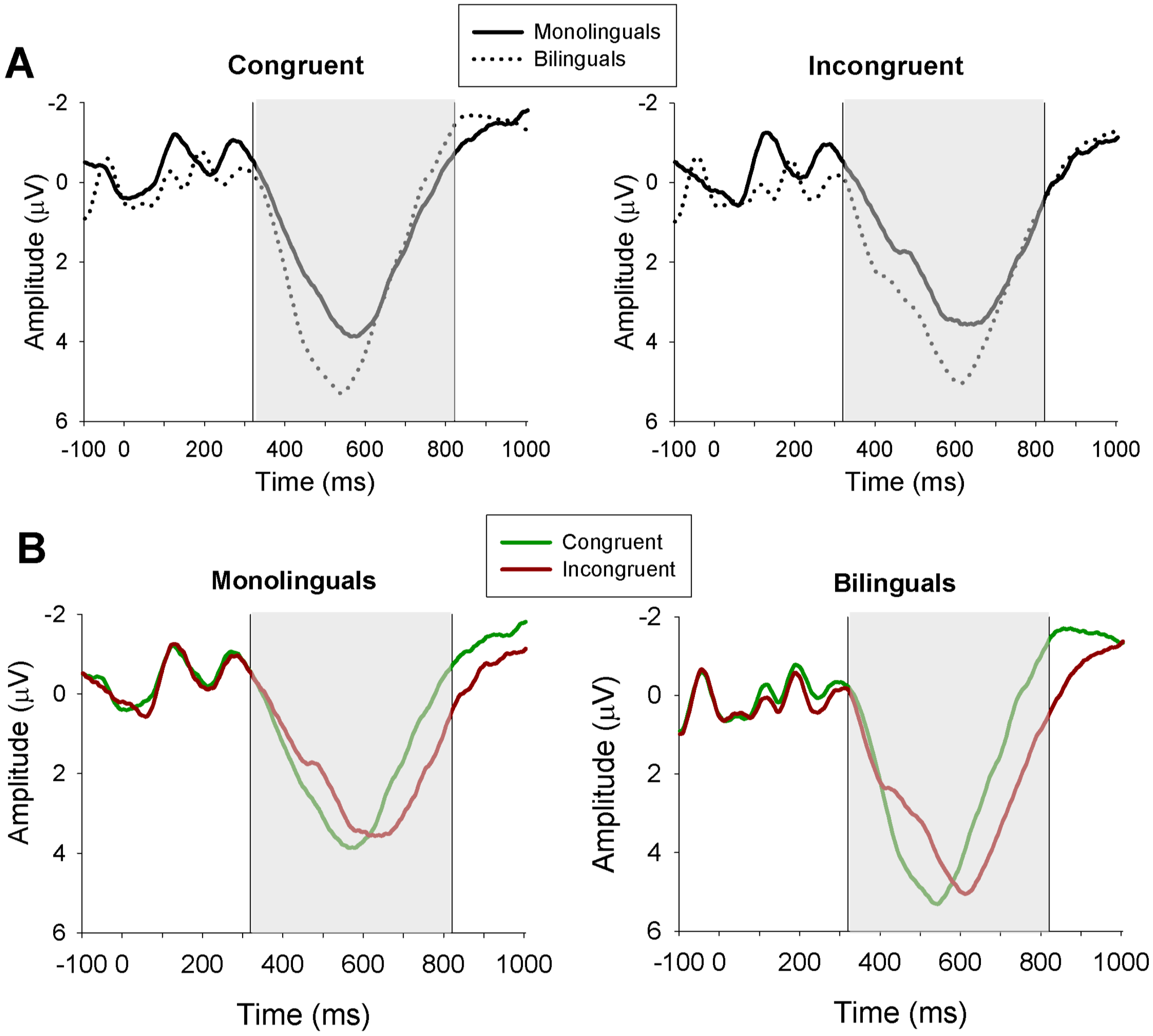

Figure 8. Stimulus-locked grand averaged waveforms depicting the P3 at site Pz for the flanker task. Panel A compares monolinguals and bilinguals for each trial type and panel B compares congruent to incongruent trials for each language group. The shaded area indicates the interval included in the analysis.

Analysis of the mean N2 amplitude revealed a main effect of Trial Type ($F(1,38)=4.71$, $MSE=0.63$, $p=.04$, $\eta^2_p=.04$), demonstrating larger mean amplitudes for incongruent than congruent trials. The planned comparison of the Language Group x Trial Type interaction showed a significant effect of Trial Type in the monolinguals only. The main effect of Trial Type was replicated in the analysis of peak N2 amplitude ($F(1,38)=8.6$, $MSE=0.79$, $p<.01$, $\eta^2_p=.19$), and the planned comparison of the Language Group x Trial Type interaction again showed a significant effect of Trial Type in the monolinguals only. Analysis of peak N2 latency showed that the N2 peaked earlier in bilinguals than in monolinguals (main effect of Language Group: ($F(1,38)=4.23$, $MSE=11166.4$, $p=.05$, $\eta^2_p=.10$), and the planned comparison of the Language Group x Trial Type interaction showed that this was only the case for incongruent trials.

There were no significant effects of Language Group or Trial Type on mean or peak P3 amplitude, all Fs < 2.3, all ps > .11. However, the P3 peaked earlier for congruent than incongruent trials (main effect of Trial Type: ($F(1,38)=34.52$, $MSE=5743.18$, $p<.01$, $\eta^2_p=.48$) and in bilinguals than monolinguals (main effect of Language Group: ($F(1,38)=4.42$, $MSE=25187.10$, $p=.04$, $\eta^2_p=.10$). In addition, the planned comparison of the Language Group x Trial Type interaction ($p=.30$) revealed that the P3 peaked earlier in bilinguals than monolinguals for the congruent condition only.

4. Discussion

The goal of the current investigation was to determine whether previous findings demonstrating superior performance in bilinguals than monolinguals on tasks measuring cognitive control could be replicated in a sample of healthy older monolingual and bilingual adults using three different cognitive control tasks in the same participants. In addition, we included both behavioural (RT and accuracy) and electrophysiological (ERP) measures given that previous work has shown that in some cases behavioural measures alone may not detect language group differences (e.g., Ansaldo et al., 2015; Kousaie & Phillips, 2012b). Furthermore, the temporal sensitivity of ERPs permits more precise identification of the locus of language group differences, when they exist, in the information processing pipeline.

In terms of behavioural results, we found a reliable effect of Trial Type for all three tasks (i.e., greater accuracy and faster RT for congruent than incongruent trials), demonstrating that the tasks themselves were effective at introducing interference on incongruent trials. In addition, the size of the interference effects reported here are similar to those previously reported for the Stroop task (e.g., Verhaeghen & De Meersman, 1998), slightly smaller for the Simon task (e.g., Ansaldo et al., 2015)[3], and slightly larger for the Flanker task (e.g., Gollan et al., 2011). Thus, we can be encouraged that our tasks were well-suited to detect any Language Group differences if they exist.

With respect to Language Group effects, the findings were mixed and not consistent across the three tasks. Behaviourally, bilinguals showed superior performance on the Stroop task, but not for the Simon or flanker tasks. Specifically, bilinguals showed more accurate and faster response times than monolinguals. Importantly, the significant interaction between Language Group and Trial Type showed that this was specific to incongruent trials, demonstrating that the monolinguals were showing a larger interference effect than the bilinguals. Bilinguals showed greater accuracy than monolinguals on the flanker task with no reliable differences in RT[4], and there were no Language Group effects for the Simon task. This is in contrast to our previous

---

[3] This may be due to differences in methodology. That is, a large number of trials are required for ERP designs, whereas relatively few trials were included in the fMRI paradigm used by Ansaldo et al..

[4] It is noteworthy that monolinguals showed a trend towards a smaller interference effect than bilinguals

study examining these same tasks in young adults where we found no reliable Language Group differences on behavioural measures (Kousaie & Phillips, 2012b). The emergence of behavioural Language Group differences in older adults in the present study supports the suggestion that the effects of bilingualism on cognition are more evident in this age group, possibly as a result of age-related decline in cognition.

The electrophysiological findings were less straightforward and will be discussed for each task separately. As was briefly mentioned earlier, ERP differences in the absence of behavioural differences can be more difficult to interpret, and whether these electrophysiological differences can be interpreted as an advantage is a matter of opinion. However, given that the ERP components of interest in this investigation have been well studied and their modulation by different task condition well characterized we are confident in our interpretation of specific language group differences being indicative of superior performance in bilinguals compared to monolinguals. Recall that the N2 is being taken as a measure of conflict monitoring and P3 amplitude and latency as measures of resource allocation and stimulus evaluation, respectively.

For the Stroop task, the electrophysiological results were entirely consistent with the behavioural findings and support the interpretation of superior performance in bilinguals than monolinguals. That is, the N2 peaked earlier and P3 amplitude was larger for the bilinguals, suggesting earlier conflict detection and allocation of fewer resources in the bilinguals than in the monolinguals. In addition, in both language groups the P3 showed a classic Trial Type effect – smaller amplitude for incongruent than congruent trials, implying greater resource allocation for the more difficult incongruent trials.

The electrophysiological results from the Simon task also provide evidence for superior performance in bilinguals. That is, N2 amplitude was larger for incongruent than congruent trials in the bilinguals only, while overall N2 amplitude was larger for the monolinguals compared to the bilinguals. Taken together, this suggests that monolinguals were monitoring for conflict to a greater extent than bilinguals, and that monitoring demands were equally high for both congruent and incongruent trials. Larger N2 amplitude for monolinguals than bilinguals has previously been reported in young adults using a Stroop task (Kousaie & Phillips, 2012b), where it was suggested that this may be the result of more efficient conflict monitoring in bilinguals as a consequence of constantly managing two languages. As suggested by Abutalebi et al. (2012), bilingualism may result in more efficient conflict processing by the anterior cingulate cortex resulting in less activation and a smaller amplitude N2 in bilinguals despite similar behavioural performance as monolinguals. For the P3, the bilinguals showed larger amplitude and earlier P3 peaks than monolinguals, again suggesting faster categorization and allocation of fewer resources in bilinguals than monolinguals. There was also a Trial Type effect in terms of P3 peak latency in both language groups, demonstrating faster categorization for congruent than interference inducing incongruent trials, as expected.

Finally, results from the flanker task provided some support for superior performance in bilinguals. For the N2, monolinguals showed a trial type effect with respect to amplitude (larger amplitude for incongruent than congruent trials) but there was no overall effect of Language Group. However, bilinguals did show earlier peak N2 amplitudes than monolinguals, and this was the case for incongruent trials only. Taken together, these findings suggest that the bilinguals were faster at monitoring for conflict than the monolinguals on incongruent trials, and perhaps the failure to observe a Trial Type effect in bilinguals indicates that both types of trials were equally easy; however, this is difficult to interpret in the absence of a main effect of Language Group. Results from the P3 showed earlier peaks in bilinguals than monolinguals on congruent trials, and for congruent than incongruent trials in both groups, suggestive of subtle processing differences between the language groups favouring the bilinguals.

As with our behavioural findings, the electrophysiological results from the older adults are not consistent with our previous findings from young adults (Kousaie & Phillips, 2012b). That is, in young adults we found no strong evidence for superior performance in bilinguals[5], although there were some P3 latency effects. Specifically, young bilinguals showed earlier P3 peaks than monolinguals for the Stroop task, and monolinguals showed a greater delay in P3 peak latency for incongruent compared to congruent trials than bilinguals. Although these specific effects do not parallel those in our older adults, they are consistent in that they involve the latency of the P3 ERP component, which was indicative of superior performance in bilingual older adults across all tasks.

The current results make two important contributions to the current debate in the literature regarding the existence of, or necessary conditions for, effects of bilingualism to be observed. First, the only clear behavioural difference that we see is for the Stroop task, with no such behavioural difference on the Simon or flanker tasks. Moreover, the interference effects (i.e., increases in RT for incongruent relative to congruent trials) on these tasks do not correlate with each other in either the monolinguals or the bilinguals[6]. These results from our older adults

---

[5] Although there were some findings from Kousaie and Phillips (2012b) that could be interpreted as superior performance in bilinguals compared to monolinguals, there were also findings suggesting the opposite (e.g., larger P3 amplitudes for monolinguals than bilinguals on the Simon task). This influenced our conclusion that there was a lack of strong evidence for a positive effect of bilingualism on cognitive control in young adults.

[6] These analyses are not reported in the results section, as they are not relevant to the primary aims of the study. Pearson correlations were calculated between the interference effect obtained for each of the three tasks across the entire group of participants, as well as for each language group separately. There were no significant correlations between the interference effects elicited by any of the tasks (all *p*’s > .09).

are consistent with conclusions drawn from samples of younger adults (Paap & Greenberg, 2013; Paap & Sawi, 2014), suggesting little convergent validity between these tasks. Thus, these tasks should not be taken as interchangeable measures of executive control. It is interesting that the clearest evidence for superior performance in bilinguals comes from the Stroop task, a task that requires the control of interference between linguistic/semantic information. Given that the superior performance on cognitive control tasks seen in bilinguals is thought to result from the constant management of two languages, perhaps the stronger finding in the Stroop task is due to the linguistic nature of the task. Alternatively, it has been suggested that, although the flanker task and the Stroop task share similarities, they differ in important ways (Nee, Wager, & Jonides, 2007). For instance, in the flanker task, the distractors are adjacent to the target stimulus. However, in the Stroop task, the target and the distractor information are different attributes of the same stimulus. Moreover, the word information elicits an automatic and pre-potent response on incongruent trials. Thus, it is possible that in the Stroop task greater demands are placed on selective attention to filter or suppress distracting information.

Second, we see consistent evidence for superior performance in bilinguals in terms of P3 latency in the Simon and flanker tasks, with earlier P3 latencies in the bilinguals compared to the monolinguals. This indicates a language group difference in the speed of stimulus evaluation that is not specific to conflict management. This finding is consistent with the more generalized bilingual executive processing advantage suggested by Hilchey and Klein (2011), albeit one that does not percolate through to the behavioural level. The question is, why are we detecting differences in the brain responses that are thought to be reflective of stimulus processing time, but this difference is not seen in the behavioural results? Of interest here is that the P3 effect is observed for the two non-linguistics tasks where no behavioural differences were detected, whereas on the Stroop task we observed a Language Group effect on behaviour but not on P3 latency. This suggests that the two measures (behavioural and electrophysiological) reflect different aspects of task performance, with the RT measures reflecting later aspects of cognitive processing (e.g., semantic processing required for Stroop task performance) and P3 latency reflecting earlier processing (e.g., basic stimulus characteristics).

To conclude, the current investigation finds some support for language group differences, with bilinguals demonstrating superior performance in both behavioural and electrophysiological measures. The findings presented here are consistent with previous reports that show different effects of bilingualism across different tasks (Kousaie & Phillips, 2012b) as well as those that demonstrate a greater sensitivity of brain-based measures to detecting language group effects (Ansaldo et al., 2015; Kousaie & Phillips, 2012b). That is, although we did not find superior behavioural performance in bilinguals across the three tasks used here, we did find some electrophysiological evidence indicative of enhanced cognitive processing for all three tasks. The current investigation also supports the notion that language group differences may emerge more strongly in a sample of older adults than younger adults.

Acknowledgements

The authors are grateful to C. McHenry and J. Pinheiro Carvalho, as well as members of the Cognitive Aging and Psychophysiology laboratory for help with participant recruitment and data collection.

Funding: This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [grant number 203751].

References

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., . . . Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex, 22*(9), 2076-2086. doi:10.1093/cercor/bhr287

Ansaldo, A. I., Ghazi-Saidi, L., & Adrover-Roig, D. (2015). Interference control in elderly bilinguals: Appearances can be misleading. *Journal of Clinical and Experimental Neuropsychology*, 1-16. doi:10.1080/13803395.2014.990359

Antón, E., Fernández García, Y., Carreiras, M., & Duñabeitia, J. A. (2016). Does bilingualism shape inhibitory control in the elderly? *Journal of Memory and Language, 90*, 147-160. doi:http://dx.doi.org/10.1016/j.jml.2016.04.007

Berchicci, M., Spinelli, D., & Russo, F. D. (2016). New insights into old waves. Matching stimulus- and response-locked ERPs on the same time-window. *Biological Psychology, 117*, 202-215. doi:http://dx.doi.org/10.1016/j.biopsycho.2016.04.007

Bialystok, E., Abutalebi, J., Bak, T. H., Burke, D. M., & Kroll, J. F. (2016). Aging in two languages: Implications for public health. *Ageing Research Reviews, 27*, 56-60. doi:10.1016/j.arr.2016.03.003

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences, 16*(4), 240-250. doi:10.1016/j.tics.2012.03.001

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., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism, 9*, 103-119.

Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging, 29*(3), 696-705. doi:10.1037/a0037254

Bugg, J. M., DeLosh, E. L., Davalos, D. B., & Davis, H. P. (2007). Age differences in Stroop interference: Contributions of general slowing and task-specific deficits. *Aging, Neuropsychology, and Cognition, 14*, 155-167.

Costa, A., & Sebastian-Galles, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews. Neuroscience, 15*(5), 336-345. doi:10.1038/nrn3709

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop interference, practice, and aging. *Aging, Neuropsychology, and Cognition, 10*, 85-98.

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

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(2), 470-484. doi:10.1037/0278-7393.20.2.470

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

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.

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science, 22*(9), 1155-1164. doi:10.1177/0956797611417002

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

Hsieh, S., & Fang, W. (2012). Elderly adults through compensatory responses can be just as capable as young adults in inhibiting the flanker influence. *Biological Psychology, 90*(2), 113-126. doi:10.1016/j.biopsycho.2012.03.006

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

Kousaie, S., & Phillips, N. A. (2012a). Ageing and bilingualism: Absence of a "bilingual advantage" in Stroop interference in a nonimmigrant sample. *The Quarterly Journal of Experimental Psychology, 65*(2), 356-369. doi:10.1080/17470218.2011.604788

Kousaie, S., & Phillips, N. A. (2012b). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research, 1446*, 71-90.

Kousaie, S., & Phillips, N. A. (2012b). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research, 1446*, 71-90.

Kousaie, S., Sheppard, C., Lemieux, M., Monetta, L., & Taler, V. (2014). Executive function and bilingualism in young and older adults. *Frontiers in Behavioral Neuroscience, 8*. doi:10.3389/fnbeh.2014.00250

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

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

Milham, M. P., Erickson, K. I., Banich, M. T., Kramer, A. F., Webb, A., Wszalek, T., & Cohen, N. J. (2002). Attentional control in the aging brain: Insights from an fMRI study of the Stroop task. *Brain and Cognition, 49*, 277-296.

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.

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective & Behavioral Neuroscience, 7*(1), 1-17.

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. doi:10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex, 69*, 265-278. doi:10.1016/j.cortex.2015.04.014

Paap, K. R., & Sawi, O. (2014). Bilingual avantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00962

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

Rugg, M. D., & Coles, M. G. H. (1995). The ERP and cognitive psychology: conceptual issues. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition.* (pp. 27-39). Oxford, UK: University Press.

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.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Van der Lubbe, R. H. J., & Verleger, R. (2002). Aging and the Simon task. *Psychophysiology, 39*, 100-110.

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.

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology and Aging, 13*, 120-126.

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology, 34*, 131-156.

Weir, C., Bruun, C., & Barber, T. (1997). Are backward words special for older adults? *Psychology and Aging, 12*(12), 145-149.

West, R. (2004). The effects of aging on controlled attention and conflict processing in the Stroop task. *Journal of Cognitive Neuroscience, 16*, 103-113.

Wild-Wall, N., Falkenstein, M., & Hohnsbein, J. (2008). Flanker interference in young and older participants as reflected in event-related potentials. *Brain Research, 1211*, 72-84. doi:10.1016/j.brainres.2008.03.025

Zeef, E. J., & Kok, A. (1993). Age-related differences in the timing of stimulus and response processes during visual selective attention: performance and psychophysiological analyses. *Psychophysiology, 30*(2), 138-151.

Zeef, E. J., Sonke, C. J., Kok, A., Buiten, M. M., & Kenemans, J. L. (1996). Perceptual factors affecting age-related differences in focused attention: performance and psychophysiological analyses. *Psychophysiology, 33*(5), 555-565.