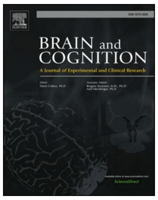
Brain and Cognition 147 (2021) 105658

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c



Increases in attentional demands are associated with language group differences in working memory performance

Kyle J. Comishen, Ellen Bialystok [*]

York University, Toronto, Ontario

ARTICLE INFO

Keywords:
Bilingualism
Working memory
Attention
Executive function

ABSTRACT

One approach to resolving the controversy over whether bilingualism affects executive function (EF) performance has been to identify the specific tasks and populations that might show these effects. The assumption is that the effect of bilingualism reliably occurs with some tasks and populations but not others and that identifying those conditions will settle outstanding contradictions. However, it is now clear that experiments using the same task (e.g., flanker, Simon, etc.) and apparently the same populations (monolingual or bilingual participants) still lead to different outcomes. Therefore, something in addition to these factors must determine performance. The present study tested the hypothesis that changes in demands for attentional control within a task is associated with performance differences for groups with different attentional resources, in this case, monolingual and bilingual participants. Sixty-four young adults who were classified as monolingual or bilingual based on a detailed questionnaire completed four increasingly difficult conditions of an n-back task while EEG was recorded. Behavioral results showed greater declines with increasing difficulty for monolinguals than bilinguals, and electrophysiological results revealed more effortful processing by monolinguals across all conditions. Our interpretation is that demands for attentional control by the task in conjunction with assessments of attentional resources in individuals or groups determines performance on executive function tasks. These results lead to a re-examination of how executive function is conceptualized and the role of bilingualism in performance on these tasks.

1. Introduction

The possibility that bilingualism is associated with neuroplastic changes in cognitive and brain systems across the lifespan has led a surge in research accompanied by vigorous debate about the reliability of the evidence (review in Bialystok, 2017). In contrast to studies supporting cognitive differences between monolingual and bilingual infants (Comishen, Bialystok, & Adler, 2019; Kovács & Mehler, 2009; Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012), children (review in Barac, Bialystok, Castro, & Sanchez, 2014) and older adults (review in Baum & Titone, 2014), other research examining similar populations has not found these effects, particularly in studies of young adults performing behavioral tasks (Bialystok, Craik et al., 2005; Bialystok, Martin, & Viswanathan, 2005; Bialystok, 2006; Paap & Greenberg, 2013; Von Bastian, Souza, & Gade, 2016). In many cases, studies that appeared to be similar in all essential respects have nonetheless led to different outcomes. Adding to the confusion, meta-analyses of these studies have ruled in favor of both the positive (Adesope, Lavin, Thompson, & Ungerleider, 2010; Grundy & Timmer, 2017; Grundy, 2020) and null effects (Donnelly, Brooks, & Homer, 2019; Lehtonen et al., 2018) as the more reliable outcome. Nonetheless, both positive and null results continue to be reported in this literature, suggesting that some further factor may be responsible for the variable outcomes.

The majority of the research investigating the cognitive effects of bilingualism has been based on the Unity and Diversity model proposed by Miyake and colleagues (Miyake et al., 2000) to define the structure of executive functioning (EF). The central assumption of the model is that the three components of EF they identify – updating (working memory), shifting, and inhibition – are largely autonomous aspects of EF despite some overlapping variance. Moreover, each of the components is associated with tasks that are assumed to be prototypical representations of that type of EF processing. Therefore, the logic of these studies is that individuals or groups who perform well on a task are expressing a general facility with that component more broadly; good performance on a flanker task is interpreted as good inhibition, so individuals (or groups) with this facility are expected to express it in all tasks identified

---

[*] Corresponding author at: Department of Psychology, York University, 4700 Keele St., Toronto, ON M3J 1P3, Canada.
E-mail address: ellenb@yorku.ca (E. Bialystok).

https://doi.org/10.1016/j.bandc.2020.105658
Received 24 July 2020; Received in revised form 17 November 2020; Accepted 23 November 2020

as inhibition in the model. Despite this being the logic that follows from the model, it has not been consistent with actual results, leading to some of the confusion and contradictions in the literature. However, rather than concluding that the failure to conform to this logic indicates that there are no language group differences in EF, it is possible that the model is insensitive to crucial factors that underlie bilingual performance. This possibility is especially important for studies of young adults where the majority of null results have been found. The current study examines the hypothesis that demands for attentional control in individual conditions constitutes a systematic source of task difficulty that supersedes the three specific EF components.

Commonly used tasks, such as the Simon task and flanker task typically reveal ceiling performance for young adult participants (Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik et al., 2005; Bialystok, Martin, et al., 2005; Costa, Hernández, & Sebastián-Gallés, 2008; Luk, Anderson, Craik, Grady, & Bialystok, 2010; Paap & Greenberg, 2013), and the lack of behavioral variability reduces the likelihood of observing group differences. However, several early studies demonstrated that manipulations in the task that created a set of scaled conditions changed the likelihood that monolingual and bilingual young adults would achieve similar outcomes. The first example is from a study in which young adults completed a Simon task under two conditions (Bialystok, 2006). The "easy" condition, low switch, presented strings of trials in which there were few sequential switches between congruent and incongruent trials, whereas the "difficult" condition, high switch, required frequent shifting between these trial types. Both language groups performed comparably on the low switch condition but the bilinguals outperformed monolinguals on the high switch condition. Confirming this pattern, a recent study using a flanker task showed that such switching between trial types was more difficult for monolinguals than bilinguals (Grundy, Chung-Fat-Yim, Friesen, Mak, & Bialystok, 2017).

A second example comes from a study by Costa, Hernández, Costa-Faidella, and Sebastián-Gallés (2009) in which a flanker task was manipulated by changing the ratio of congruent and incongruent trials, creating a low monitoring condition in which most of the trials were the same and a high monitoring condition in which congruent and incongruent trials were more evenly distributed. This manipulation is similar to the switch conditions examined by Bialystok (2006) in that the high monitoring condition also requires more frequent switching. Again, all participants performed equivalently in the low monitoring conditions, but the bilingual group outperformed the monolinguals in the high monitoring conditions. These studies indicate that manipulation of task difficulty can generate group differences even when the same participants and task are used.

The present study applies this approach to a frequently used working memory (WM) task. As with other aspects of EF, studies investigating WM have reported both better bilingual performance (Hernández, Costa, & Humphreys, 2012; Morrison, Farooq, & Taler, 2019; Morrison, Kamal, Le, & Taler, 2020; Morrison & Taler, 2020; see Grundy & Timmer, 2017 for a review) and no difference between language groups (Antón, Carreiras, & Duñabeitia, 2019; Lukasik et al., 2018; Ratiu & Azuma, 2015; see Lehtonen et al., 2018 for a review). However, some studies that reported better bilingual performance based this conclusion on electrophysiological and not behavioral evidence (e.g., Morrison et al., 2019, 2020; Morrison & Taler, 2020), suggesting that measures of reaction time and accuracy may lack sensitivity.

The n-back task is a common measure of WM and allows manipulations of difficulty without changing the essential processing that defines the task (Conway et al., 2005). The task requires participants to recognize a stimulus that was presented *n* trials previously to the current trial in a sequence. The difficulty is related to the manipulation of *n* in that it determines how many items must be retained in working memory to correctly make the judgment. The increase in *n* is associated with decreases in accuracy, increases in response time, and diminished cognitive resources for working memory (Daffner et al., 2011; Kok, 2001; Polich, 1996). The n-back, therefore, is a promising paradigm for investigating the effect of task difficulty within a task on language group differences in performance.

Morrison et al. (2019) assessed performance of monolingual and bilingual young adults on a standard n-back with ERPs included as a measure. English monolingual and English-French bilingual participants completed an n-back that included a 0-, 1-, and 2-back conditions. In addition to behavioral measures, three ERP waveforms were analyzed. The P2 is a positive-peaking centro-frontal waveform that occurs 150–300 ms post-stimulus presentation (Luck, 2014). For tasks that assess working memory, such as the n-back, the P2 is associated with the encoding of information into working memory (Dunn, Dunn, Languis, & Andrews, 1998; Finnigan, O'Connell, Cummins, Broughton, & Robertson, 2011; Lijffijt et al., 2009), so better encoding of information into working memory is associated with larger P2 amplitudes and earlier peak latencies (Finnigan et al., 2011). Recent evidence from a delayed matching-to-sample WM task found young and old bilingual adults both displayed larger P2 amplitudes than their monolinguals counterparts, indicating differences in cognitive processing (Morrison & Taler, 2020). The N2 is a negative-peaking centro-frontal waveform that occurs 200–350 ms after the presentation of a stimulus (Folstein & Van Petten, 2008; Luck, 2014) that is associated with the ability to discriminate between a presented stimulus and another stimulus stored in working memory (Bennys, Portet, Touchon, & Rondouin, 2007; Patel & Azzam, 2005). Larger (i.e., more negative) N2 amplitudes and shorter peak latencies indicate greater discriminability between a presented stimulus and one held in memory (Daffner et al., 2011). Finally, the P3 is a positive-peaking centro-parietal waveform that occurs 300–600 ms post-stimulus presentation (Luck, 2014; Polich, 2012). The P3 is a well-studied component of working memory and has been associated with the intensity of attentional processing during retrieval of a stimulus held in working memory (Kok, 2001). With increasing task difficulty, there are decreases in P3 amplitude and increases in P3 peak latency as available attentional resources required to retrieve an item from working memory are reduced (Kok, 2001; Polich, 1996). There were no behavioral or P2 and N2 differences between language groups in the Morrison et al. (2019) study, but groups did differ on P3 amplitude. Specifically, while P3 amplitude decreased with increasing task difficulty, the bilingual group exhibited larger P3 amplitudes than the monolingual group across all conditions. The researchers interpreted this difference as indicating greater attentional control for bilinguals that enabled them to perform similarly on behavioral assessments but with less recruitment of attentional resources.

Two other studies did reveal differences between monolinguals and bilinguals in the behavioral measures of an n-back task. Janus and Bialystok (2018) tested 9-year-old children on 1- and 2-back conditions in which emotional stimuli were interleaved between trials. The emotional manipulation did not affect performance for either language group, but the bilingual children were slower but more accurate than monolinguals on the 2-back, with equivalent performance on the 1-back. This paradigm was extended to young adults by Barker and Bialystok (2019) who included EEG recordings while the task was performed. Similar to the findings with children, the bilinguals were slower but more accurate than monolinguals on the 2-back and again groups were comparable on the 1-back. Analysis of P3 amplitude showed the expected task effect for the bilinguals, but not the monolinguals – the bilinguals experienced a decrease in P3 amplitude from the 1-back to the 2-back, whereas the monolinguals showed no difference. It appears that only the bilinguals adapted to the increase in task demands on the 2-back and, as a result, recruited the resources necessary to do well on the 2-back which was evident by the bilinguals behaviorally outperforming the monolinguals. While it is unclear why the monolinguals did not show the expected observation of a decrease in P3 amplitude with increasing task difficulty, they did show the predicted increase in P3 latency with increasing task difficulty suggesting they were impacted by the change in task demands. Together, these findings suggest monolinguals and bilinguals show differences in the recruitment of attentional resources.

Though the mean amplitude findings from the Morrison et al. (2019) and Barker and Bialystok (2019) appear to contradict one another, there is a possible explanation for this. In both studies, the bilingual participants displayed the expected outcome showing a decrease in mean amplitude of the P3 when task difficulty increased. The monolingual group showed the same P3 pattern in the Morrison et al. (2019) study, but not the Barker and Bialystok (2019) study – their amplitude did not differ between the 1- and 2-back. Furthermore, there were no behavioral differences between language groups in the Morrison et al. (2019) study, but there were behavioral differences in the Barker and Bialystok (2019) study where bilinguals had better performance. Together, the evidence suggests when monolinguals and bilinguals perform similarly on a behavioural task, bilinguals use their attentional resources more efficiently. When bilinguals behaviorally outperform monolinguals, however, it may be due to the bilinguals using their attentional resources more efficiently and/or they better adapt to the challenges that the task demands (as seen in the Barker and Bialystok (2019) study). It appears then that bilinguals, but not monolinguals, efficiently deploy attentional resources in a manner that translates to better behavioral performance.

Two questions remain unanswered by these studies. First, because language group differences occurred on 2-back conditions but not 1-back conditions, the assumption was that the relevant factor was the increased difficulty of the 2-back. However, confirming that interpretation requires a larger gradient of difficulty in which the only difference across conditions is attentional requirements. Second, the only previous studies that have accompanied performance on this task with EEG recordings have produced contradictory results. Therefore, it is necessary to repeat the procedure to validate which of the outcomes is reliable. The issue is important because the assumption is that electrophysiological measures provide a more sensitive index of processing resources recruited while performing these tasks than do simple behavioral measures.

The current study presented increasingly difficult conditions of an n-back task to monolingual and bilingual young adults to investigate behavioural and electrophysiological measures of performance. The central assumption was that changes in task difficulty from increasing demands for attentional control within a task are more important in determining language group differences in performance than are differences between tasks. In the n-back task used here, the same processes are recruited for all conditions but the more difficult conditions present greater challenges to attention because the sequence must be held in WM for a longer time (more trials) and encounter more distraction (intervening trials) before the judgment can be made. Participants completed 0-, 1-, 2, and 3-back conditions of an n-back task while electroencephalogram (EEG) was recorded. The hypothesis was that behavioral performance of monolingual and bilingual participants will be comparable for the easy conditions and diverge with increasingly difficult conditions. Therefore, the prediction was that monolinguals and bilinguals will perform similarly on the simple conditions (0-back and 1-back), begin to diverge on the 2-back, and will achieve significantly different levels of performance on the 3-back because the increasing difficulty will disproportionately impact performance of monolinguals.

The ERP measures reflect attentional resources recruited in the task across conditions. As in the study by Morrison et al. (2019) the relevant measures were P2, N2, and P3. P2 indexes encoding in working memory such that larger amplitude, or more positivity, reflects greater working memory capacity. Therefore, the prediction was that bilinguals will show larger P2 amplitudes than monolinguals. The N2 indexes stimulus discriminability such that larger negative amplitudes indicate less distractor noise. Since bilinguals generally perform better than monolinguals when there is distraction from misleading cues, the prediction was that bilinguals will display a larger N2 amplitude than monolinguals. Finally, P3 indexes item recognition in working memory, arguably the central process involved in this task. In this case, a more positive amplitude reflects less effortful processing, indicating that sufficient resources are available to perform the task without effort. Therefore, the prediction was that bilinguals will display a larger P3 amplitude than monolinguals. Finally, since peak latency of the P2, N2, and P3 is associated with the onset of recruiting attentional resources, the prediction was bilinguals will show earlier peak latencies than the monolinguals.

In summary, the hypotheses are that behavioral measures will show increasing gaps between monolinguals and bilinguals as difficulty increases but that bilinguals will demonstrate larger P2, N2, and P3 amplitudes and shorter peak latencies across all conditions reflecting greater underlying attentional resources. Put another way, behavioral data indicates the level of performance and electrophysiological data indicates the efficiency of processing. Both measures are required to understand performance on this working memory task.

2. Method

2.1. Participants and procedures

Participants were students at York University in Toronto, Canada, who participated for course credit. Potential participants underwent a preliminary screening to determine that they were monolingual or fluent in more than one language; Individuals with intermediate profiles of bilingualism were excluded at this stage. The final sample included 33 English-speaking monolingual and 31 bilingual young adults between the ages of 18 and 25 years, right-handed, with normal or corrected-to-normal vision, and no history of concussion or use of psychoactive medications. An additional 7 participants were tested but excluded from data analyses due to excessive drift and/or high frequency noise in their EEG channels during data acquisition ($n = 4$), mean behavioural responses ($n = 1$) or ERP amplitudes ($n = 1$) that were 3 standard deviations from the mean, or stopped attending to the tasks ($n = 1$). Each participant provided informed written consent at the start of the study.

To assess participants for language background, the Language Social Background Questionnaire (LSBQ; Anderson, Mak, Keyvani Chahi, & Bialystok, 2018) was administered. The questionnaire asks for proficiency estimates in all known languages and detailed questions about the languages used in specific contexts and with specific individuals. The non-English languages of the bilingual participants included: Arabic ($n = 3$), Cantonese ($n = 3$), Farsi ($n = 2$), Gujarati ($n = 3$), Hebrew ($n = 1$), Hindi ($n = 4$), Hungarian ($n = 1$), Italian ($n = 1$), Malayalam ($n = 1$), Patois ($n = 1$), Polish ($n = 1$), Spanish ($n = 3$), Swedish ($n = 1$), Tagalog ($n = 2$), Urdu ($n = 1$), and Vietnamese ($n = 3$). Eleven bilingual participants reported a non-English language as their first language; however, the age of exposure and acquisition of a second language was early for all bilingual participants ($M = 1.3$ years, $SD = 2.5$), so order of acquisition may not be relevant. All participants reported their proficiency in English and a non-English language by estimating their ability to speak and understand the language on a scale of 0 (no proficiency) to 100 (high proficiency). The questionnaire also obtained background information, including parents' education on a 5-point Likert scale in which 1 represented no high school diploma, 2 represented high school graduate, 3 represented some college or college diploma, 4 represented bachelor or first academic degree, and 5 represented graduate or professional degree. This information was used as a proxy for socioeconomic status (SES).

Participants completed the Vocabulary and Block Patterns subsets of the Shipley-2 Test (Shipley, Gruber, Martin, & Klein, 2009) to provide assessments of verbal and non-verbal intelligence, respectively. Both tests report standardized scores ($\mu = 100$, $SD = 15$).

Following the background assessments, participants completed four conditions of the n-back task in the order of progressive difficulty (i.e., 0-, 1-, 2-, and 3-back) while EEG was recorded. Each session lasted approximately 2 h. All procedures were approved by York University's

Office of Research Ethics.

2.2. N-back Task

E-prime 2.0 software (Psychology Software Tools, Inc., version 2.0.10.353) was used to present stimuli (300 × 300 pixels) on a 19-inch computer, with a refresh rate of 60 Hz. Participants sat approximately 50 cm in front of the computer, and stimuli were displayed at the center of the screen and remained until participants pressed one of two buttons or 1000 ms had elapsed. The interstimulus interval (ITI) was jittered randomly at 1400, 1500, or 1600 ms during which the screen remained empty to reduce alpha noise in the EEG signal.

There were four stimulus sets consisting of digits, letters, shapes, or symbols, with nine unique stimuli in each set. The four sets were assigned to each of the four experimental conditions using a Latin square design. Therefore, participants completed four conditions, each with a different stimulus set, and the stimulus sets were counterbalanced across conditions throughout the sample. The stimuli and design are illustrated in Fig. 1. Different stimulus sets were used to reduce proactive interference across conditions.

For all conditions, participants were asked to respond to each trial as quickly as possible by pressing the "M" key on the keyboard if the stimulus was a target or the "Z" key if it was a distractor. For the 0-back, targets were defined by a predetermined stimulus chosen from the set used in that condition; for the 1-back, targets were defined as a stimulus that matched the one on the previous trial; for the 2-back, targets were defined as a stimulus that matched one that was shown two trials previously; and for the 3-back, targets were defined as a stimulus that matched the display shown three trials previously.

There were 180 trials in each of the four n-back conditions, producing a total of 720 experimental trials. Each condition consisted of two blocks of 90 trials each with a 20 s break between blocks. In each condition, 60 of the 180 trials were targets.

2.3. EEG acquisition and preprocessing

EEG data were recorded continuously by a BioSemi Acquisition System from 64 Ag/AgCl active electrodes that were placed according to the international 10–20 system at a sampling rate of 512 Hz. Six additional electrodes were placed on both mastoids, 1 cm below the eyes, and 1 cm from the lateral canthus of each eye. These electrodes were applied for offline re-referencing, capturing eye blinks and eye-movements, respectively. All electrodes were referenced to the common mode sense electrode and grounded by the driven right leg electrode. A criterion of 20 kΩ was used for electrode impedance, with electrode sites being interpolated if this threshold was surpassed. Only 3 participants had interpolation of a site of interest.

Off-line preprocessing was conducted using EEGLAB toolbox (version 10.2.2.4b) in MATLAB (2012, Mathworks, Natick, MA, version 7.14.0.739). EEG data were filtered with a Butterworth filter of 0.01–60 Hz and re-referenced to the average of both mastoids. Stimulus-locked epochs from − 200 to 1000 ms after stimulus presentation were computed with a 200 ms pre-stimulus baseline correction. Electrode sites with impedances equal or greater than 20 kΩ or with high frequency noise were interpolated. A simple voltage threshold of 400 μV was used to detect and remove trials with drift or muscle tension. After this, Independent Component Analysis (ICA; Makeig, Bell, Jung, & Sejnowski, 1996) was conducted to detect and correct eye-movements and blinks. Finally, an additional simple voltage threshold of 150 μV was used to remove any trials with eye-related artifacts that were not corrected by ICA. The percentage of trials removed was 1.39% and 1.14% for monolinguals and bilinguals, respectively. Grand average event-related potentials were obtained for each participant by electrode site and condition.

3. Results

3.1. Behavioral measures

Mean scores for background measures by language group are reported in Table 1. One-way ANOVAs revealed no significant group differences for age, SES, verbal, or non-verbal intelligence, all $F\text{s} < 1$ There was a main effect of group for proficiency in English speaking, $F(1,62) = 13.68$, $p < .001$, $\eta_p^2 = 0.18$, and understanding, $F(1,62) = 7.98$, $p = .01$, $\eta_p^2 = 0.11$, indicating higher scores for the monolingual group than the bilingual group. In contrast, the bilingual group was more proficient in speaking, $F(1,62) = 1972.32$, $p < .001$, $\eta_p^2 = 0.97$, and understanding, $F(1,62) = 2262.69$, $p < .001$, $\eta_p^2 = 0.97$, a non-English language than the monolingual group.

The effect of stimulus set on performance was examined in a 2 × 4 ANOVA for Group and Stimulus set that was conducted for each of accuracy, false alarm rate, and reaction time. There were no interactions of group and stimulus set in any of the analyses, indicating stimulus set did not influence relevant group outcomes. The complete analyses are reported in the Appendix A.

Accuracy to correctly identify targets across conditions is presented in Table 2. A 2 × 4 mixed-design ANOVA for Group and Condition (0-, 1-, 2-, 3-back) revealed a main effect of Group, $F(1,62) = 4.77$, $p = .03$, $\eta_p^2 = 0.07$, a main effect of Condition, $F(3,186) = 271.83$, $p < .001$, $\eta_p^2 =$

Table 1
Mean scores (standard deviation) for background measures by language group. Asterisks signify significant differences (p < .05) between groups.

	Monolingual	Bilingual
N	33	31
Age (years)	19.7 (1.3)	19.5 (1.8)
SES (parents' education)	3.2 (1.0)	3.3 (1.1)
Shipley Vocabulary	102.5 (9.8)	100.4 (8.5)
Shipley Blocks	102.6 (9.5)	101.3 (12.9)
English Proficiency (%)		
Speaking*	98.5 (6.2)	91.0 (9.8)
Understanding*	99.1 (3.8)	94.8 (7.7)
Non-English Language Proficiency (%)		
Speaking*	1.5 (5.1)	91.9 (10.5)
Understanding*	2.1 (7.0)	93.4 (8.4)

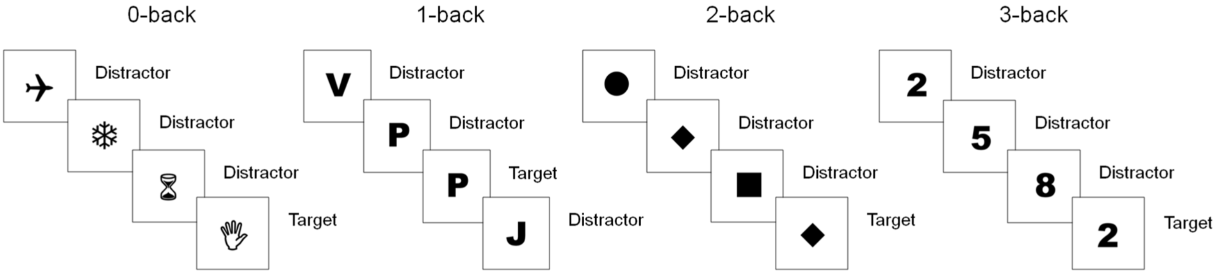


Fig. 1. Example of the stimuli sets and conditions used throughout the study.

Table 2
Mean scores (standard deviation) for behavioral measures on four conditions of the n-back task by language group.

	Monolingual	Bilingual
Accuracy (%)		
0-back	96.71 (3.42)	96.86 (2.41)
1-back	90.30 (7.34)	91.53 (5.59)
2-back	64.76 (14.30)	71.81 (16.96)
3-back	44.77 (15.56)	53.82 (18.76)
False Alarm (%)		
0-back	0.70 (0.87)	0.82 (0.91)
1-back	2.52 (2.08)	2.65 (1.98)
2-back	11.84 (7.68)	11.78 (7.44)
3-back	17.28 (9.33)	17.78 (10.00)
RT (ms)		
0-back	502 (56)	507 (70)
1-back	563 (94)	555 (80)
2-back	643 (106)	653 (71)
3-back	646 (124)	628 (112)

0.82, and a significant interaction between them, $F(3,186) = 2.70$, $p = .05$, $\eta_p^2 = 0.04$. To explore the interaction, one-way ANOVAs were conducted for each condition. There were no significant group differences on the 0-back, $F < 1$, or 1-back, $F < 1$, a non-significant trend on the 2-back, $F(1,62) = 3.24$, $p = .08$, and a significant group difference on the 3-back, $F(1,62) = 4.43$, $p = .03$, $\eta_p^2 = 0.07$. Thus, the main effect of group was driven by higher accuracy for the bilinguals in the most difficult conditions.

The rate of false alarms across conditions is reported in Table 2. False alarm rates indicate response bias and potentially inflate accuracy scores. A 2 × 4 ANOVA for Group and Condition revealed a significant main effect of Condition, $F(3,186) = 147.43$, $p < .001$, $\eta_p^2 = 0.70$, but no main effect of Group or interaction effect, *Fs* < 1. Thus, false alarm rates increased with task difficulty but did not differ by language group.

Finally, mean RT for correct responses are also reported in Table 2. A 2 × 4 ANOVA for Group and Condition revealed a significant main effect of Condition, $F(3,186) = 52.44$, $p < .001$, $\eta_p^2 = 0.46$, but no effect of Group or interaction of Group and Condition, *Fs* < 1. As expected, reaction time increased with increasing task difficulty.

3.2. *ERP analyses*

The mean amplitudes and peak latencies of the P2, N2, and P3 waveforms were analyzed. Since the P2 and N2 waveforms are largest over midline anterior scalp sites (Luck, 2014), they were analyzed by taking the average waveform recorded at electrodes: AF3, AFz, AF4, F1, Fz, F2, FC1, FCz, and FC2. Time windows of 140–170 ms and 170–210 ms were selected to analyze the P2 and N2 waveforms, respectively, corresponding to the typical onset of these waveforms (Folstein & Van Petten, 2008; Luck, 2014). The P3 waveform is largest over midline centro-parietal scalp sites (Johnson, 1986; Polich, 2012), so data were analyzed by taking the average waveform recorded at electrodes: CP1, CPz, CP2, P1, Pz, P2, PO3, POz, and PO4 with a time window of 325–425 ms, complying with the typical onset of this waveform (Luck, 2014). See Fig. 2 for a pictorial representation of the electrodes included in the anterior and parietal regions of interest.

Mean amplitude for the P2, shown in Fig. 3, was analyzed using a 2 × 4 mixed-design ANOVA for Group and Condition. There was a main effect of Group, $F(1,62) = 6.23$, $p = .02$, $\eta_p^2 = 0.09$, as the bilingual group ($M = 2.75$ μV, $SE = 0.76$) exhibited a greater mean amplitude than the monolingual group ($M = 0.75$ μV, $SE = 0.61$). There was no effect of Condition, $F(3,186) = 1.34$, *ns*, or interaction effect, $F < 1$.

Mean peak latencies of the P2 are reported in Table 3. There were no significant effects for Group, $F(1,62) = 1.07$, *ns*, Condition, $F(3,186) = 1.04$, *ns*, or their interaction, $F(3,186) = 1.68$, *ns*.

Mean amplitude of the N2 is shown in Fig. 4. A 2-way ANOVA revealed a main effect of Condition, $F(3,186) = 3.36$, $p = .02$, $\eta_p^2 = 0.05$, but no effect of Group, $F(1,62) = 1.08$, *ns*, or interaction of Group and Condition, $F(3,186) = 0.43$, *ns*. Thus, the mean amplitude of the N2 became less negative with increasing task difficulty but did not differ by language group.

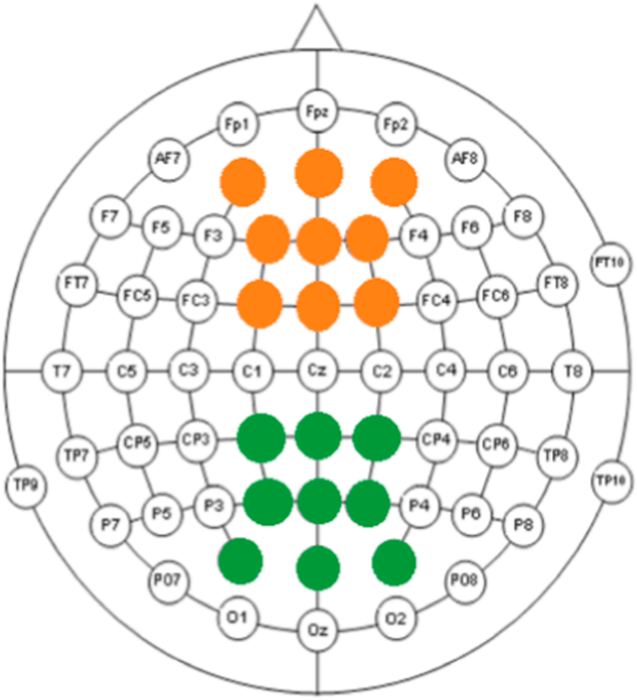


Fig. 2. The anterior (orange) and parietal (green) electrodes used for analysis of the ERP waveforms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Mean peak latencies of the N2 are reported in Table 3. Again, there were no effects of Group, $F < 1$, Condition, $F < 1$, or interaction, $F(3,186) = 1.06$, *ns*.

Mean amplitude of P3, shown in Fig. 5, was analyzed using a 2 × 4 ANOVA. There was a main effect of Group, $F(1,62) = 5.18$, $p = .03$, $\eta_p^2 = 0.08$, and a main effect of Condition, $F(3,186) = 21.10$, $p < .001$, $\eta_p^2 = 0.10$, but no interaction, $F < 1$. The bilingual group ($M = 10.34$ μV, $SE = 0.81$) exhibited larger mean amplitude than the monolingual group ($M = 8.23$ μV, $SE = 0.89$). The effect of Condition indicated equivalent amplitude for 0-back and 1-back, $F(1,63) = 0.23$, *ns*, a significant reduction for the 2-back, $F(1,63) = 12.82$, $p < .001$, $\eta_p^2 = 0.17$, and another significant reduction for the 3-back, $F(1,63) = 14.88$, $p < .001$, $\eta_p^2 = 0.19$.

To better understand the group differences in P3 amplitude, correlations were conducted between these amplitudes and the behavioral outcomes. A repeated measures correlation was applied to the data due to the violation of independent observations. There was a significant positive correlation between mean P3 amplitude and accuracy, $r_{rm}(191) = 0.41$, $p < .001$, and a significant negative correlation between mean amplitude of the P3 waveform and mean RT, $r_{rm}(191) = -0.43$, $p < .001$. Thus, larger P3 amplitudes were associated with better task performance.

Mean peak latencies of the P3 are reported in Table 3. The 2-way ANOVA indicated a main effect of Group, $F(1,62) = 5.91$, $p = .02$, $\eta_p^2 = 0.09$, a main effect of Condition, $F(3,186) = 2.84$, $p = .04$, $\eta_p^2 = 0.04$, and an interaction between them, $F(3,186) = 3.53$, $p = .02$, $\eta_p^2 = 0.05$. One-way ANOVAs for each condition indicated no group differences on the 0-back, $F < 1$, 1-back, $F < 1$, or 2-back, $F < 1.15$, but a significant group effect on the 3-back, $F(1,62) = 13.80$, $p < .001$, $\eta_p^2 = 0.18$. Thus, the bilingual group ($M = 356$ ms, $SE = 6$) displayed shorter peak latencies than the monolingual group ($M = 382$ ms, $SE = 6$) in the difficult

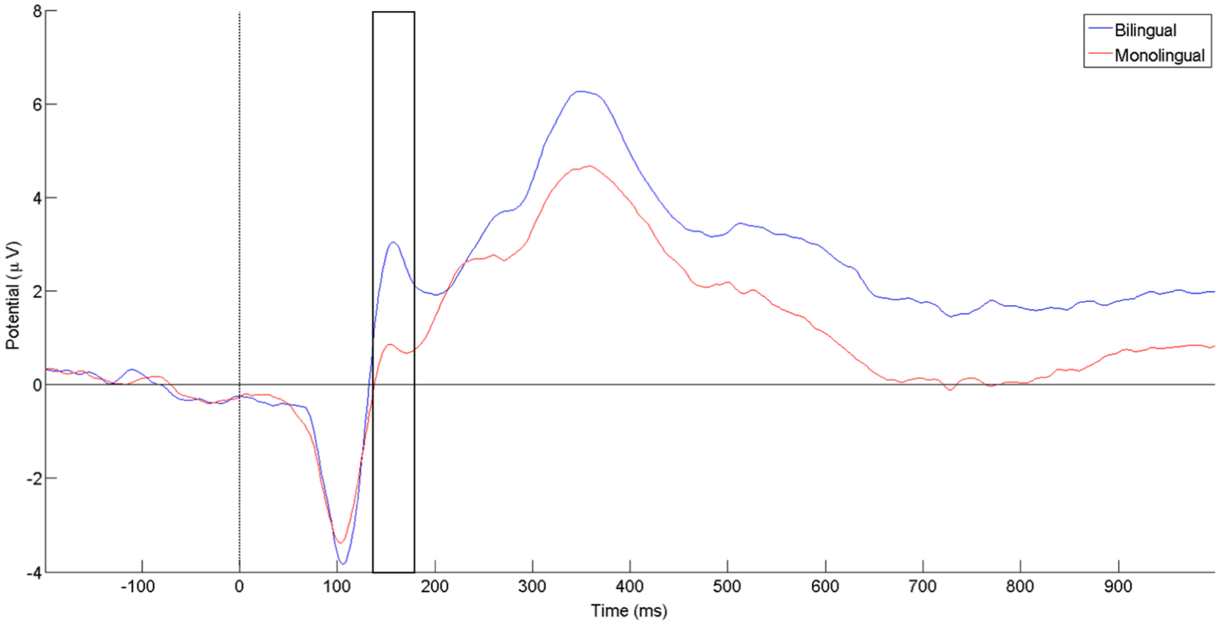


Fig. 3. Mean Amplitude for the P2 by language group collapsed across condition for correct target trials.

Table 3
Mean peak latencies (standard deviation) of the P2, N2, and P3 waveforms on four conditions of the n-back task by language group.

	Monolingual	Bilingual
P2 (ms)		
0-back	155 (10)	155 (8)
1-back	155 (10)	155 (11)
2-back	154 (10)	159 (10)
3-back	157 (11)	158 (8)
N2 (ms)		
0-back	191 (13)	189 (12)
1-back	188 (13)	192 (14)
2-back	191 (12)	192 (13)
3-back	188 (12)	191 (12)
P3 (ms)		
0-back	380 (30)	376 (27)
1-back	375 (25)	371 (20)
2-back	370 (25)	364 (22)
3-back	382 (31)	356 (25)

3-back condition.

4. Discussion

The present study examined the effect of increasing difficulty within a single task on working memory performance by monolingual and bilingual young adults. Participants completed an n-back that consisted of a 0-, 1-, 2-, and 3-back condition. Previous studies that have compared language groups across tasks, for example, administering flanker and Simon tasks to monolingual and bilingual participants, frequently found no difference between groups and no correlation across tasks (e.g., Paap & Greenberg, 2013). The conclusion was that there was no effect of bilingualism on executive functioning; the absence of inter-task correlations was used as well to reject the notion of cognitive effects of bilingualism when in fact it highlighted a problem with the Unity and Diversity model from which the hypotheses were generated. In contrast, the present study manipulated the degree of difficulty within a single task and administered the task while EEG was recorded to increase sensitivity of the measures. Both types of measures produced reliable differences between the language groups. The main results were that increasing difficulty across the conditions led to larger declines in performance for monolinguals than for bilinguals, and ERP analyses revealed greater attentional resources available for bilinguals in all conditions.

Consider the results in more detail. False alarm rates and reaction time increased across the conditions, confirming the increasing difficulty, but there were no language group differences on these measures. This increasing difficulty was also reflected in accuracy that significantly declined across conditions, but in this case that decline interacted with language group. Although the group difference was only significant for the 3-back, the trajectory of increasing divergence in accuracy was clear across the conditions. Thus, as the conditions required greater levels of attentional control to evaluate longer sequences in WM, bilinguals could maintain better performance than monolinguals.

The accuracy level for both groups in the 3-back condition was just over (bilinguals) or just under (monolinguals) 50%, raising the possibility that the scores reflected chance responding. However, evaluating these data in conjunction with the false alarm rate rules out chance as a likely explanation. If participants were responding randomly, then not only accuracy to targets (responding YES when the correct answer was “yes”) but also rate of false alarms (responding YES when the correct answer was “no”) would be at chance. Instead, false alarm rates for both groups in the 3-back condition was less than 20%, indicating discrimination between targets and distractors. In other words, it was relatively easy to reject a target as having been seen three trials previously but difficult to confirm that the target had been displayed. This pattern rules out both chance responding and response bias in which one type of response (e.g., YES) was preferred. Together, these results indicate that the monolingual and bilingual groups performed equivalently on the n-back task up to the point of the most difficult condition, the 3-back, when the bilinguals outperformed the monolinguals.

The accuracy results indicated that bilinguals were somewhat better than monolinguals in the 2-back but the difference was not significant, corresponding to the behavioral results reported by Morrison et al. (2019), but the difference achieved significance for the 3-back, corresponding to the behavioral results reported for the 2-back in some previous research (Barker & Bialystok, 2019; Janus & Bialystok, 2018). The present P3 findings were in line with the findings of Morrison et al. (2019) but not those of Barker and Bialystok (2019). Therefore, both monolinguals and bilinguals in the present study, as in Morrison et al.

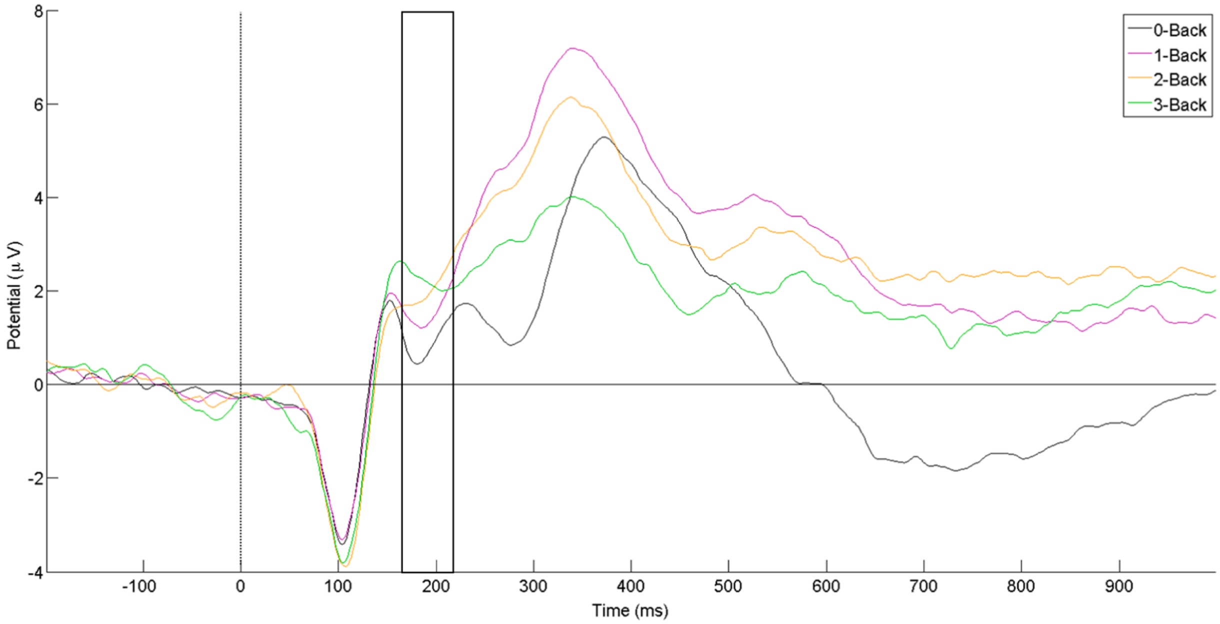


Fig. 4. Mean Amplitude for the N2 by condition collapsed across language group for correct target trials.

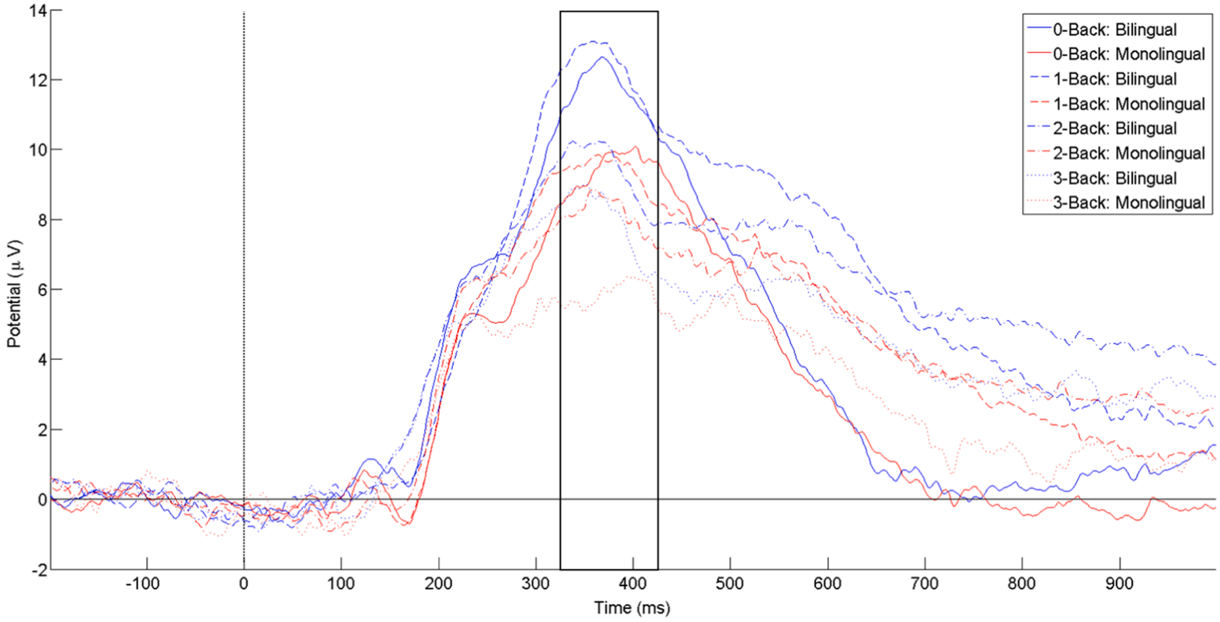


Fig. 5. Mean Amplitude for the P3 by language group and condition for correct target trials.

(2019), appropriately adapted to the demands of task. Under this condition, bilinguals are more efficient at allocating their attentional resources. Had the study used only the standard behavioral approach up to the 2-back condition, the conclusion would have been that monolingual and bilingual young adults perform comparably on an n-back task. That conclusion, however, would have been misleading. Evidence from the EEG data reveals a consistent difference between the language groups in electrophysiological signals reflecting attentional resources.

The general interpretation is that the bilinguals have greater resources for selection and attention that enable them to maintain accuracy levels as difficulty increases. The P2 reflects encoding in WM and shows a greater positive amplitude for groups or individuals with larger working memory capacity (Dunn et al., 1998; Finnigan et al., 2011; Lijffijt et al., 2009). As predicted, bilinguals demonstrated a significantly larger mean amplitude than monolinguals on the P2 across all conditions. The N2 indexes stimulus discriminability in which a larger negative amplitude is observed when there is less distractor noise. The hypothesis was that bilinguals will have a larger N2 amplitude because of their ability to ignore distraction, but there were no language group differences on this measure. Instead, there was a significant effect of condition in which each subsequent condition reduced the amplitude of this waveform, reflecting increased interference from the intervening trials. Thus, the measure did not capture individual or group differences in stimulus selection but rather reflected task differences in which conditions became increasingly challenging, similar to the results reported by Morrison et al. (2019). Finally, the P3 indexes item recognition in working memory, with larger amplitudes associated with less effortful processing (Donchin, 1981; Mertens & Polich, 1997; Polich, 2007). The prediction was that bilinguals will demonstrate larger P3 amplitude than monolinguals. The results revealed two significant main effects with no interaction: easier conditions were associated with a larger amplitude than harder conditions, and bilinguals demonstrated a larger amplitude than monolinguals. To summarize, the N2 and P3 reflected task difficulty and the P2 and P3 reflected resource differences between monolingual and bilingual groups.

Peak latency of the P2, N2, and P3 ERP waveforms is associated with the onset of recruiting attentional resources. While it was predicted that bilinguals will show earlier peak latencies than the monolinguals across all conditions for each ERP waveform, bilingualism only influenced peak latency of the P3 waveform, consistent with the effect reported by Barker and Bialystok (2019) on the 2-back condition. The difference in peak latency suggests bilinguals were able to retrieve information from working memory faster than the monolingual group and had more attentional resources available for stimulus recognition (Kok, 2001; Polich, 1996).

The controversy around whether bilingualism leads to improvements in cognitive performance, usually operationalized as performance on standard EF tasks, has typically been based on studies that compare monolinguals and bilinguals performing a single task, such as a flanker task, or two (or more) tasks considered to reflect the same EF component, such as inhibition or updating. However, the results frequently show no evidence of behavioral differences between groups and no correlation between tasks from the same EF component. Both findings are problematic because if (a) the Unity and Diversity model is correct, and (b) bilingualism improves EF performance, then these are the logical implications from those premises. The failure to obtain supporting evidence has typically been interpreted as grounds for rejecting premise (b), namely, the claim that bilingualism modifies EF ability. Equally, however, the failure of these studies to support the expected outcomes may instead be grounds for rejecting premise (a), namely, the validity of the Unity and Diversity model.

An alternative conceptualization of EF abandons the components of the Unity and Diversity model and considers the common features across all the tasks used in this research (Bialystok, 2017). What emerges most clearly is that all these tasks rely on varying degrees of effortful attention and selection; the additional processes associated with each of the components, such as avoid interference for inhibition tasks and hold items in mind for updating tasks, may be more distraction than explanation. Put this way, performance on an EF task depends on adequate attentional resources to manage specific task demands, regardless of those demands. This approach leads to a different set of predictions than those from the componential model because group performance is not tied to entire tasks that contain multiple processes but rather to a continuous evaluation of intensity, in this case, the degree to which effortful processing is required in a specific version of a task. This approach, and the current results, were anticipated by the studies by Bialystok (2006) and Costa et al. (2009) in which manipulating complexity within a task affected the relationship between language group outcomes. However, in those studies, the notion of what was contributing to the complexity manipulation was vague. In the current study, evidence from the ERP waveforms, particularly P3, supports the interpretation that the groups differ in attentional resources. Increasing demands for attention across conditions was reflected in smaller amplitudes for these waveforms, but the amplitudes for the bilinguals was always at a higher level than those of the monolinguals, indicating greater reserve of resources. From this perspective, performance is determined by the fit between individual resources and task demands. For conditions that are relatively simple, most individuals will be able to perform the task to a high level, but as conditions require increasingly large resources, only participants whose capacity is commensurate with those demands will continue to perform the task. Our primary claim is that bilingualism improves attentional resource capacity, not necessarily by increases in volume, but rather by increases in efficiency.

Three studies using fMRI have supported the notion of greater efficiency in bilingual performance on EF tasks, despite comparable behavioral outcomes. Abutalebi and colleagues (Abutalebi et al., 2013) administered a flanker task to monolingual and bilingual young adults. In this case, bilinguals outperformed the monolinguals on the behavioural measures but also crucially displayed less activation of the anterior cingulate cortex, the region most responsible for performing this task. Similarly, Gold and colleagues (Gold, Kim, Johnson, Kryscio, & Smith, 2013) administered a task switching paradigm to younger and older adults and reported better behavioral outcomes and less activation in the cingulate cortex and other frontal regions for the bilinguals than for their monolingual counterparts. Finally, Berroir and colleagues (Berroir et al., 2017) administered a Simon task to older adult monolingual and bilingual participants and applied a network perspective to examine brain recruitment in the two groups. Bilinguals performed the task by allocating fewer resources than monolinguals, a difference they interpreted as demonstrating increased efficiency in the bilingual group. Although preliminary, these studies support the interpretation that bilinguals use attentional resources more efficiently than monolinguals when performing these tasks. That sometimes leads to differences in behavioral outcomes and sometimes it does not, but the evidence is clear that these tasks are less effortful for bilinguals.

The conclusion from this study is that the effect of bilingualism is to increase efficiency with which attentional resources are used, sometimes resulting in better performance by bilinguals than monolinguals on these tasks. It is unknown at this time how attentional efficiency is increased through bilingual experience, but it is likely related to intense experience dealing with complex linguistic environments where selection is a constant necessity. It is well established that both languages are jointly activated in the bilingual mind, yet intrusion errors are rare (review in Kroll, Dussias, Bice, & Perrotti, 2015). Our suggestion is that constant recruitment of attention to select the target language potentially increases the efficiency of those systems in a way that benefits their application to all situations. Further evidence for that speculation is required but the possibility sets out a direction for investigating bilingual effects on cognition that is qualitatively different from the current approaches motivated by the Unity and Diversity model. The interpretation relies on a distinction between the type of information contributed by the two measures used in this study. Behavioral measures reflect achievement levels whereas electrophysiological measures reflect resources. To the extent that resources are adequate, all individuals can achieve the same levels. The important information comes from situations where task demands exceed resources of an individual or group. Put this way, the determination of performance on these tasks requires an assessment of the relation between task demands and individual resources. Our current speculation is that bilingual experience has modulated the efficiency with which bilinguals can engage attentional control. Ongoing research with other tasks and other populations, currently in progress, will hopefully add clarity to this position.

CRediT authorship contribution statement

Kyle J. Comishen: Conceptualization. **Ellen Bialystok:** Conceptualization.

Acknowledgements

The research reported in this paper was funded by grant A2559 from the Natural Sciences and Engineering Research Council, Canada, to EB. We thank David Jesin for his assistance with data collection.

Appendix A

Supplementary analyses were conducted to ensure there was no confound with stimulus set and task difficulty. 2 × 4 ANOVAs with Group (monolingual, bilingual) as a between-participant factor and Stimulus (digits, letters, shapes, symbols) as a within-participant factor were conducted on accuracy, false alarm rate, and reaction time. Condition was not included in the model because doing so would reduce statistical power (i.e., approximately eight participants within each Group, Condition, and Stimulus interaction).

For accuracy, the ANOVA indicated a main effect of Group, $F(1,62) = 4.77$, $p = .03$, $\eta_p^2 = 0.07$, and Stimulus, $F(3,186) = 2.71$, $p = .04$, $\eta_p^2 = 0.04$, but no interaction, $F < 1$. Additional analyses of Stimulus revealed participants' accuracy was lower for symbols than digits, $t(63) = 2.39$, $p = .02$, $\eta_p^2 = 0.08$, and letters, $t(63) = 2.18$, $p = .03$, $\eta_p^2 = 0.07$. All other comparisons were not significant, all $ps > 0.10$.

For false alarm rates, the ANOVA indicated a main effect of Stimulus, $F(3,186) = 2.95$, $p = .03$, $\eta_p^2 = 0.05$, but no effect of Group, $F < 1$, or interaction of Group and Stimulus, $F < 1$. Additional analyses of Stimulus indicated false alarm rates were greater for symbols than digits, $t(63) = 2.33$, $p = .02$, $\eta_p^2 = 0.08$, letters, $t(63) = 2.41$, $p = .02$, $\eta_p^2 = 0.08$, and shapes, $t(63) = 2.10$, $p = .04$, $\eta_p^2 = 0.07$. There was no difference in false alarm rate between digits, letters, and shapes, all $ps > 0.73$.

The ANOVA on reaction time indicated a main effect of Stimulus, $F(3,186) = 3.77$, $p = .01$, $\eta_p^2 = 0.06$, with no effect of Group or interaction effect, $Fs < 1$. Additional analyses indicated reaction times were significantly faster for letters than shapes, $t(63) = 2.98$, $p < .01$, $\eta_p^2 = 0.12$, and symbols, $t(63) = 2.90$, $p < .01$, $\eta_p^2 = 0.12$. All other comparisons were not significant, all $ps > 0.10$.

In summary, these analyses indicate that while stimulus type influenced accuracy, false alarm rates, and reaction times, these effects were equivalent across groups because of the counterbalancing of stimulus set and condition throughout the sample.

References

Abutalebi, J., Della Rosa, P. A., Ding, G., Weekes, B., Costa, A., & Green, D. W. (2013). Language proficiency modulates the engagement of cognitive control areas in multilinguals. *Cortex, 49*(3), 905–911.

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207–245.

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods, 50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9.

Antón, E., Carreiras, M., & Duñabeitia, J. A. (2019). The impact of bilingualism on executive functions and working memory in young adults. *PLoS One, 14*(2), Article e0206770.

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly, 29*(4), 699–714. https://doi.org/10.1016/j.ecresq.2014.02.003.

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

Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics, 35*, 857–894. https://doi.org/10.1017/S0142716414000174.

Bennys, K., Portet, F., Touchon, J., & Rondouin, G. (2007). Diagnostic value of event-related evoked potentials N200 and P300 subcomponents in early diagnosis of Alzheimer's disease and mild cognitive impairment. *Journal of Clinical Neurophysiology, 24*(5), 405–412.

Berroir, P., Ghazi-Saidi, L., Dash, T., Adrover-Roig, D., Benali, H., & Ansaldo, A. I. (2017). Interference control at the response level: Functional networks reveal higher efficiency in the bilingual brain. *Journal of Neurolinguistics, 43*, 4–16.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 60*(1), 68–79.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin, 143*(3), 233–262. https://doi.org/10.1037/bul0000099.

Bialystok, E., Craik, F. I., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage, 24*(1), 40–49.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*(2), 290–303. https://doi.org/10.1037/0882-7974.19.2.290.

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism, 9*(1), 103–119. https://doi.org/10.1177/13670069050090010701.

Comishen, K. J., Bialystok, E., & Adler, S. A. (2019). The impact of bilingual environments on selective attention in infancy. *Developmental Science, e12797*. https://doi.org/10.1111/desc.12797.

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

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*(2), 135–149.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*(1), 59–86.

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

, … Donchin, E. (1981). Surprise!… surprise? *Psychophysiology, 18*(5), 493–513.

Donnelly, S., Brooks, P. J., & Homer, B. D. (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychonomic Bulletin & Review, 26*(4), 1122–1147.

Dunn, B. R., Dunn, D. A., Languis, M., & Andrews, D. (1998). The relation of ERP components to complex memory processing. *Brain and Cognition, 36*(3), 355–376.

Finnigan, S., O'Connell, R. G., Cummins, T. D., Broughton, M., & Robertson, I. H. (2011). ERP measures indicate both attention and working memory encoding decrements in aging. *Psychophysiology, 48*(5), 601–611.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x.

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *Journal of Neuroscience, 33*(2), 387–396.

Grundy, J. G. (2020). The effects of bilingualism on executive functions: An updated quantitative analysis. *Journal of Cultural Cognitive Science*. https://doi.org/10.1007/s41809-020-00062-5.

Grundy, J. G., Chung-Fat-Yim, A., Friesen, D. C., Mak, L., & Bialystok, E. (2017). Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults. *Cognition, 163*, 42–55.

Grundy, J. G., & Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research, 33*(3), 325–340.

Hernández, M., Costa, A., & Humphreys, G. W. (2012). Escaping capture: Bilingualism modulates distraction from working memory. *Cognition, 122*(1), 37–50.

Janus, M., & Bialystok, E. (2018). Working memory with emotional distraction in monolingual and bilingual children. *Frontiers in Psychology, 9*, 1582.

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology, 23*, 367–384.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557–577.

Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences, 106*(16), 6556–6560. https://doi.org/10.1073/pnas.0811323106.

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, Mind, and Brain. *Annual Review of Linguistics, 1*, 377–394. https://doi.org/10.1146/annurev-linguist-030514-124937.

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

Lijffijt, M., Lane, S. D., Meier, S. L., Boutros, N. N., Burroughs, S., Steinberg, J. L., Moeller, F. G., & Swann, A. C. (2009). P50, N100, and P200 sensory gating: Relationships with behavioral inhibition, attention, and working memory. *Psychophysiology, 46*(5), 1059–1068.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luk, G., Anderson, J. A., Craik, F. I., 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.

Lukasik, K. M., Lehtonen, M., Soveri, A., Waris, O., Jylkkä, J., & Laine, M. (2018). Bilingualism and working memory performance: Evidence from a large-scale online study. *PloS one, 13*(11), Article e0205916.

Makeig, S., Bell, A. J., Jung, T. P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In Advances in neural information processing systems (pp. 145–151).

Mertens, R., & Polich, J. (1997). P300 from a single-stimulus paradigm: Passive versus active tasks and stimulus modality. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 104*(6), 488–497.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100.

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

Morrison, C., Kamal, F., Le, K., & Taler, V. (2020). Monolinguals and bilinguals respond differently to a delayed matching-to-sample task: an ERP study. Bilingualism: Language and Cognition, 1–11.

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

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*(2), 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002.

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the event-related potential. *International Journal of Medical Sciences, 2*(4), 147–154.

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

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148.

Polich, J. (2012). Neuropsychology of P300. *Oxford Handbook of Event-Related Potential Components*, 159–188.

Ratiu, I., & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology, 27*(1), 1–11.

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W., & Werker, J. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science, 23*(9), 994–999. https://doi.org/10.1177/0956797612436817.

Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). *Shipley-2 manual*. Los Angeles, CA: Western Psychological Services.

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