Different bilingual experiences predict different executive functions: Evidence from mouse-tracking

Aslı Yurtsever [ID], Kaiah N. Sotebeer and John G. Grundy [ID]

Department of Psychology, Iowa State University, Ames, Iowa, USA

Abstract

There is evidence to suggest that the effects of bilingualism on executive functions (EFs) need to be examined along a continuum rather than a dichotomy. The present study addressed this need by examining the influence of different bilingual experiences on executive functioning using a Flanker and Stroop mouse-tracking task that taps into more dynamic cognitive processes than typical behavioral paradigms. We sampled 98 bilingual young adults and investigated conflict and sequential congruency effects (SCEs). We found that mouse-tracking metrics captured links that were not identified with overall reaction times. SCEs were more sensitive to detecting relations between L2 experiences and EF than simple conflict effects. Second-language age of acquisition and L1/L2 switching frequency consistently predicted EF outcomes. This association was moderated by the attentional demands of the task. These findings highlight the complexity of the effects of bilingualism on cognition, and the use of more sensitive measures to capture these effects.

Highlights

- Sensitive measures are needed to assess the effects of bilingualism on executive functions (EFs).
- Bilingualism is made up of several complex experiences that interact with EF.
- We used mouse-tracking to examine the relation between L2 experiences and EF.
- L2 age of acquisition and switching frequency predicted EF on Stroop and Flanker tasks.
- The effects of L2 experiences on EF depend on the attentional demands of the task.

1. Different bilingual experiences predict different executive functions: Evidence from mouse-tracking

Cognitive psychology relies on fundamental tools such as accuracy and reaction time (RT) to glean insights into cognitive processes, decision-making and response strategies. These variables, being both cost-effective and universally applicable across various paradigms, have been the cornerstone of cognitive research. However, their limitations become apparent when it comes to elucidating the dynamic nature of decision-making processes over time (Freeman & Ambady, 2010). When canonical measures fall short and risk overinterpretation, it becomes imperative to explore newer and more nuanced metrics. Therefore, in this study, we used a dynamic mouse-tracking paradigm to dissect bilinguals' response selection and attentional processes.

Bilingualism is one of the important experiences that increases neuroplasticity and reshapes executive functions (EFs), which are defined as higher-order cognitive functions such as working memory capacity, inhibitory control and task switching, all of which involve attentional control (Bialystok, 2017; Diamond, 2013; Friedman & Miyake, 2017). Bilinguals aptly navigate competition between two active languages in their minds (Kroll et al., 2014, 2015; Thierry & Wu, 2007; Wu & Thierry, 2010), fostering neural adaptations that exert a lasting impact on EF regions (reviews in Deluca et al., 2020; Grant et al., 2014; Grundy et al., 2017a; Pliatsikas, 2020). These neural adaptations often translate into bilinguals outperforming their monolingual counterparts on nonverbal EF tasks (e.g., Bialystok, 2009; Bialystok et al., 2004; Bialystok et al., 2014; Costa et al., 2009; Prior & MacWhinney, 2010; Singh et al., 2015). Nonetheless, these effects are not universally replicated (reviews in Gunnerud et al., 2020; Grundy, 2020; Lehtonen et al., 2018; Ware et al., 2020).

Two potential reasons for this inconsistency arise. First, the cognitive measures may lack the necessary sensitivity to delve beyond conventional metrics, thereby failing to capture the intricate aspects of decision-making over time. Second, the wide array of individual differences in bilingual experiences is often overlooked (Yurtsever et al., 2023). Despite the recent emphasis on recognizing bilingualism as a complex construct interacting with sociolinguistic variables and contexts (Gullifer & Anderson, 2023; Backer & Bortfeld, 2021; Chung-Fat-Yim et al., 2022; Leivada et al., 2021; Luk & Grundy, 2023; Pot et al., 2018; Titone & Tiv, 2023; Yamasaki et al., 2018), many

studies examining the relations between bilingualism and EFs still dichotomize groups into "monolingual" and "bilingual." This over-simplified approach can be problematic given that it overlooks crucial linguistic and contextual variations that could potentially obscure real effects (Grundy, 2020; Yurtsever et al., 2023).

In light of these considerations, the present study employed a sensitive and dynamic mouse-tracking paradigm (Freeman & Ambady, 2010) to examine how individual bilingual experiences are related to EFs. This approach is a cost-effective alternative to functional neuroimaging. Neuroimaging techniques have long been useful in examination of cognitive efficiency between monolinguals and bilinguals. We will first discuss neuroimaging evidence and then compare how mouse-tracking can complement and extend the current dialogue.

Neuroimaging techniques such as electroencephalography (EEG) and functional magnetic resonance imaging provide one path to examine cognitive processes that are not observable at the behavioral level alone and can measure the efficiency of performance on EF tasks (Luck, 2014; van Heuven & Coderre, 2015). For example, Abutalebi et al. (Abutalebi et al., 2011) compared German–Italian bilinguals and Italian monolinguals on a Flanker task and found that bilinguals used the anterior cingulate cortex (ACC) less than monolinguals in dealing with the conflicting stimuli, despite similar behavioral performance. The ACC is highly involved with monitoring and moderating conflict in the brain (Botvinick et al., 2001; Carter et al., 1998; Shenhav et al., 2013; Sheth et al., 2012); thus, Abutelabi et al.'s findings demonstrate that their bilinguals were more efficient at handling conflict resolution in the ACC than monolinguals. The finding that bilinguals use the ACC more efficiently than monolinguals has been replicated several times with different designs, populations and tasks (e.g., Anderson et al., 2021; Rodríguez-Pujadas et al., 2014).

EEG studies provide complementary findings for more efficient cognitive processing in bilinguals than monolinguals. Zunini et al. (2019) used a task-switching paradigm to demonstrate that French–English bilinguals had smaller switching and mixing costs than monolinguals, with mostly similar electrophysiological patterns in the brain, suggesting that bilinguals need the same number of resources to perform the task better than monolinguals. However, bilinguals did have larger N2 and smaller P3 amplitudes than monolinguals. These patterns suggest that bilinguals prioritize early attention to facilitate faster task-set reconfiguration at later stages in processing, consistent with more efficient processing for bilinguals than monolinguals. Recent reviews of the literature suggest that these are common findings: bilinguals demonstrate enhanced processing at early and more automatic stages to alleviate the burden from later stages of processing, which is a sign of efficiency (Antoniou, 2023; Grundy et al., 2017a). Together, neuroimaging techniques provide a useful avenue by which to examine the efficiency of processing on EF tasks, but the cost and accessibility associated with such technology limits its use. Fortunately, alternative methods, such as MouseTracker (Freeman & Ambady, 2010), provide a behavioral solution to examine multiple cognitive processes and response strategies before response execution, without the cost (Spivey et al., 2005).

The prototypical mouse-tracking paradigm involves a binary forced-choice decision task between two options represented as buttons on the computer screen (Schoemann et al., 2021). Mouse-tracking involves using a computer mouse to measure x- and y-axis coordinates as participants initiate and move toward a target over time. Motor responses, such as the movement of one's arm to select an option, are frequently updated through our perceptual and cognitive processing. These measurements of ongoing and continuous movement metrics, as measured through mouse-tracking software, give insights into the dynamics of cognition and decision-making (Freeman & Ambady, 2010). This allows for a dynamic set of measures to be calculated, including the time it takes to *initiate* a movement after the stimulus appears, and the *deviation* from the most direct path to the correct response (see Figure 1). These mouse-tracking metrics have been shown to be more reliable than overall RTs (Grundy, 2022).

Spivey et al. (2005) were the first to implement a mouse-tracking methodology to capture continuous cognitive dynamics during a spoken language task. Participants were presented with two objects on the screen, one target and one competing, displayed in the upper right or left corners. Participants were given auditory instructions on which stimulus to select with a mouse click, 500 ms after the images were presented. Competing stimuli were chosen to assess the impact of competition on lexical activation via motor output. The trials consisted of either phonologically dissimilar (control items; e.g., fountain-chair) or similar words (distractor items; e.g., sink-rink). Results showed that the distractor items yielded longer total response times than the control items. In addition, distractor items yielded greater deviations from the target compared to the control condition, indicating larger effects of spatial attraction toward the competing object in the distractor conditions than in the control conditions.

Bartolotti and Marian (2012) investigated language competition between monolinguals and bilinguals using eye-tracking and mouse-tracking methodologies. Participants learned words in an artificial language (e.g., *shundo* → acorn) and then completed a word recognition task. In this task, they clicked on pictures corresponding to the translation of the learned words (e.g., a picture of an *acorn* as the target, or a *shovel* as the phonological competitor). Their findings revealed no group differences in accuracy and response-time metrics. However, the competitors affected the mouse movements of monolinguals and bilinguals differently. Monolinguals exhibited greater deviation along the x-coordinates toward the competitors than targets. In contrast, competitors did not affect bilinguals' x-coordinate motion but did affect their y-coordinates. This pattern suggested that bilinguals were equally drawn to both the target and competitor. In other words, they experienced language competition between responses, but they resolved it earlier and more successfully than monolinguals. Moreover, bilinguals used executive control not to eliminate interference completely, but to maintain goals.

Incera and McLennan (2016) used mouse-tracking to examine Stroop task performance between monolinguals and bilinguals. The authors demonstrated that bilinguals took longer to initiate their responses, but once movement toward the target began, they were more efficient in their path to the correct response. Incera and McLennan concluded that bilinguals behaved more like experts than monolinguals, in that they took longer to process initial demands to facilitate performance at later stages of response execution. Incera and McLennan (2018) followed this study using a diverse sample of bilinguals from different language backgrounds and examining bilingualism on a continuum. The authors found that higher bilingualism scores, as measured by the percentage of time exposed to a second language, were associated with a reduced Stroop effect (incongruent minus congruent trials), as measured by a smaller deviation toward the incorrect response (i.e., more direct path to correct response) on incongruent than congruent trials. This finding was only borne out of the Stroop task, not the Flanker task, and the authors suggested that this could be because
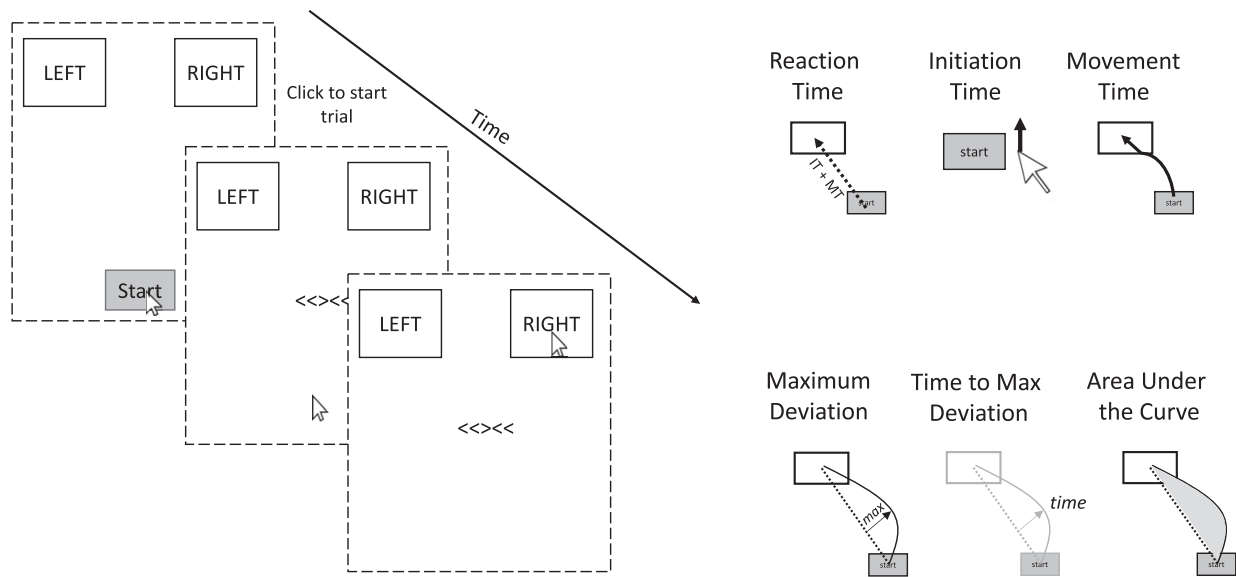
Figure 1. The mouse-tracking paradigm that allows for the examination of several behaviors that occur before response completion. Reused with permission from Grundy (2022).

bilingualism leads to better prepotent response inhibition (EF component assessed on the Stroop task) rather than resistance to distractor interference (EF component assessed on the Flanker task). Alternatively, it could be that bilingualism is more likely to influence EF on more demanding tasks (Stroop) than less demanding tasks (Flanker). Kousaie and Phillips (2017) further highlighted the importance of examining both tasks, given that bilinguals were faster and more accurate on the Stroop task than monolinguals. However, bilinguals only outperformed monolinguals on accuracy of the Flanker task, not RTs. They also demonstrated using ERPs that distinct cognitive processing between monolinguals and bilinguals emerged depending on whether the Flanker or Stroop task was used. These findings highlight the importance of examining both the Stroop and Flanker tasks within the same experiment.

The Stroop findings from these experiments complement Incera and McLennan's previous work by demonstrating that bilingualism enhances efficiency of performance on an EF task that cannot be measured with RTs alone. More recently, Damian et al. (2019) showed that Chinese–English bilinguals had smaller maximum deviations and area under the curve (AUC) for the most direct path to the correct response than English monolinguals, consistent with the previous work. Additionally, bilinguals had significantly later initiation times for both the Simon and spatial Stroop tasks than monolinguals, again consistent with the previous work. However, contrary to the previous work, bilinguals were earlier to initiate responses than monolinguals for the Flanker task. One of the reasons for this discrepancy in initiation times for the Flanker task compared to the Stroop task observed in Damian et al. (2019) could be that the Flanker task is less challenging, and more demanding tasks are required to observe an effect of bilingualism on EF. Alternatively, individual bilingual experiences could influence outcomes in different directions. Previous work has shown that age of L2 acquisition (L2 AoA; Deluca et al., 2019), L2 proficiency (Singh & Mishra, 2013; Tse & Altarriba, 2012), L2 use (Qi et al., 2023), L1/L2 balance (Yow & Li, 2015), and L1/L2 switching frequency (Liu et al., 2019; Soveri et al., 2011a, 2011b) all influence EFs in the brain. Thus, it is possible that variations in the L2 experiences of the sample populations across the studies drove the outcomes observed.

To resolve discrepancies in the literature, the present study used a dynamic mouse-tracking paradigm to examine the individual influences of different second-language experiences on initiation times and movement dynamics on conflict tasks. We tested young adult bilinguals with a Flanker and a Stroop task and examined conflict effects and sequential congruency effects (SCEs; Gratton et al., 1992). SCEs measure the extent to which individuals' responses are influenced by the preceding trial, shedding light on their ability to swiftly disengage attention—a valuable proxy for executive functioning. Specifically, when the previous trial is incongruent, the conflict effects that follow (incongruent–congruent) are smaller than when the previous trial is congruent. Smaller SCEs are typically interpreted as indicative of more effective conflict management (Gratton et al., 1992), and SCEs become smaller with practice (Egner et al., 2010). Conflict and SCEs are two EF effects that have been shown to be modified by bilingualism by some (e.g., Costa et al., 2009; Grundy et al., 2017b), but not others (e.g., Morton & Harper, 2007; von Bastian et al., 2016).

1.1. The present study

Bilingualism has been shown to lead to smaller SCEs due to bilingual experiences contributing to more rapid disengagement of attention (Antoniou, 2023; Chung-Fat-Yim et al., 2022; deMeurisse & Kaan, 2023; Grundy & Bialystok, 2018; Grundy et al., 2017b; 2017c; Grundy & Keyvani-Chahi, 2017; Grundy et al., 2020; Mishra et al., 2012). Therefore, we hypothesized that more bilingual experiences (e.g., earlier Age of acquisition [AoA], greater L2 use, proficiency and switching) would be associated with smaller Flanker and Stroop effects. Furthermore, we expect that the effects would be stronger in the Stroop task, consistent with previous work (Incera & McLennan, 2018). The individual bilingual experiences that lead to these changes in attentional disengagement are unknown, but like the conflict effects, we generally expect that more bilingual experience (e.g., L1/L2 switching frequency) will predict smaller SCEs.

The Flanker and Stroop tasks were chosen because they are two of the most commonly used EF tasks and they exercise attentional demands differently. The Stroop task is generally more difficult than the Flanker task given that it involves suppressing word

reading, which is a difficult automatic process to stop (Heitz & Engle, 2007; Lu & Proctor, 1995). It is important to emphasize that much of the present study was necessarily exploratory, given that there is a paucity of research on bilingualism and EF using mouse-tracking, and none have examined which mouse-tracking metrics best capture the association between bilingual experiences and conflict and SCE effects. Nonetheless, we generally predicted that greater bilingual experiences would lead to smaller conflict effects and SCEs.

2. Methods

2.1. Participants

We recruited 101 bilingual participants from undergraduate courses at Iowa State University in exchange for course credit. Data from three participants were excluded for partial completion of the study. The data analysis was conducted with a final sample of 98 (42 females, M_{age} = 19.54, SD_{age} = 1.5). Participants were assessed for language experience and proficiency using the Language and Social Background Questionnaire (LSBQ; Anderson et al., 2018; Luk & Bialystok, 2013). Language proficiencies are calculated as the average score of self-reported proficiencies in speaking, understanding, reading and writing. Frequency of L1 and L2 use, balance, and language switch frequency were computed using a scale of 0–4 by asking how frequently they used each of the languages in various domains of daily life. Parental education is indicated on a scale of 1–5. We present the descriptive statistics and data ranges in Table 1. Fifty-seven bilinguals reported English as their first language (L1). Non-English first languages included Spanish (n = 11), Arabic (n = 4), Chinese (n = 4), Korean (n = 4), Gujarati (n = 2), Japanese (n = 2), Vietnamese (n = 2), Indonesian (n = 1), Kinyarwanda (n = 1), Mizo (n = 1), Portuguese (n = 1), Tagalog (n = 1), and unreported (n = 7). A majority of participants reported L2 as Spanish (n = 39) or English (n = 30). Other second languages reported included German (n = 3), Chinese (n = 2), Arabic (n = 1), Bahasa Melayu (n = 1), Bosnian (n = 1), Burmese

(n = 1), French (n = 1), Hindi (n = 1), Italian (n = 1), Lao (n = 1), Mandarin (n = 1), Swahili (n = 1), Twi (n = 1), Urdu (n = 1), and unreported (n = 12).

Previous literature surrounding SCE with mouse-tracking using a Simon task was analyzed to have an effect size of Cohen's d = .71 for RT and d = 0.81 for AUC (Scherbaum & Kieslich, 2018). In the present study, we used the PANGEA online power analysis tool (Westfall, 2015 calculator available at https://jakewestfall.shinyapps.io/pangea/) to determine the minimum number of participants needed to achieve at least 80% statistical power. This calculation was based on an alpha level of 0.05 and a conservative effect size of d = 0.71 for the interaction term between the current trial and previous congruency in our ANOVA. The analysis revealed that 28 participants would be required to detect the SCE with 80% power, so our sample size was deemed more than sufficient.

2.2. Materials

We designed the experiment using Mousetracker software (Freeman & Ambady, 2010) and used a Dell MS116 Wired Mouse to collect data. Each trial consisted of two response boxes on the top left and right corners of the screen, and a Start button on the bottom center. When participants clicked the Start button, the stimulus appeared in the center of the screen. We asked the participants to move their mouse and click the correct response box as quickly as they could. We set the interstimulus interval to 1000 ms. The stimulus and response boxes stayed on the screen until the response. If the response time was longer than 2500 ms, an error message appeared and urged the participants to respond faster. After incorrect responses, a red X appeared in place of the stimulus for the duration of 1000 ms.

In the Flanker task, the stimuli were either congruent or incongruent arrows. We asked the participants to respond to the center flanker and ignore the flanking ones. The response boxes read "left" and "right," respectively. In the Stroop task, the stimuli consisted of color names, either congruent with the font color or incongruent (e.g., BLUE written in red ink). We asked the participants to respond to the font color and ignore the written word. The response boxes read "RED/GREEN" on the left and "ORANGE/BLUE" on the right box.

We obtained several outcome variables from the mouse-tracking paradigm. *Initiation time* refers to the time from clicking the Start button to the initiation of mouse movement. *Movement time* refers to the time between initiation and completion of the click response. *RT* refers to the total time between clicking the Start button and clicking the response box. It is calculated as the sum of initiation time and movement time. *Maximum deviation* is the distance of maximum deviation from a straight line that connects the Start point and the response box. Large deviations are observed more often on incongruent trials due to participants overcorrecting their initially incorrect response. *Time to maximum deviation* refers to the time from initiation to reaching the maximum deviation point. *AUC* refers to the area under the curve of the mouse movement trajectory relative to a straight line connecting the Start point and response box. AUC always produces a positive value regardless of the relative position to the straight line.

2.3. Procedure

Participants were tested in a computer lab. Upon signing the consent form, they were presented with the mouse-tracking task. Participants completed 96 trials of the Stroop task and 96 trials of the Flanker task. The order of the tasks was counterbalanced across

Table 1. Descriptives for demographic and language background information

Variable	M	SD	Data range	Range possible
Age	19.54	1.5	17–25	
L2 AoA	6.27	5.43	0–16	
L1 proficiency	92.72	10.59	50–100	0–100
L2 proficiency	70.48	23.46	10–100	0–100
L1 use	3.16	0.92	0.75–4	0–4
L2 use	1.93	1.06	0–4	0–4
L1/L2 switching frequency	1.69	1.23	0–4	0–4
Balance (raw)	1.24	0.88	0–3.6	0–4
Balance (rescaled)	1.08	0.71	0–2	0–2
Mother's education (in years)	3.24	1.46	1–5	1–5
Father's education (in years)	3.31	1.53	1–5	1–5

Note: L1 = first language; *L2* = second language; *AoA* = age of acquisition; *M* = mean; *SD* = standard deviation. The raw balance variable indicates 0 = L1 dominant; 2 = L1/L2 Balanced and 4 = L2 dominant. The rescaled balance variable indicates 0 = unbalanced and 2 = balanced.

Table 2. Means and standard deviations for conflict and sequential congruency effects

Variable/effect	Flanker conflict	Stroop conflict	Flanker SCE	Stroop SCE
Initiation time	7.83 (48.23)	0.45 (51.21)	−8.41 (86.89)	18.09 (98.04)
Movement time	256.43 (313.33)	241.2 (432.72)	149.14 (359.82)	60.03 (218.9)
Reaction time	264.26 (334.29)	241.65 (432.17)	140.73 (362.87)	78.13 (241.97)
Maximum deviation	0.25 (0.13)	0.14 (0.1)	0.05 (0.11)	0.04 (0.13)
Area under the curve	0.44 (0.26)	0.37 (0.3)	0.12 (0.25)	0.12 (0.34)
Time to maximum deviation	141.47 (191.53)	180.01 (273.58)	94.94 (247.49)	38.38 (258.38)

Note: SCE = sequential congruency effect.

participants. Trial types were randomized within each task. Upon completion, they completed the language and demographics survey. The total study duration was approximately 30 min.

3. Results

To facilitate transparency and reproducibility, all the raw data used for the following analyses are available on Figshare: https://doi.org/10.6084/m9.figshare.25357060. Table 2 provides the means and standard deviations (SDs) for all conflict effects and SCEs and Figure 2 presents the correlations between L2 variables for the present sample.

To investigate the independent contributions of L2 experiences on executive functioning, we conducted a series of multiple regression analyses using jamovi software (jamovi project, 2023). We report the independent contributions of each L2 experience (by controlling for the variance from the other L2 variables) on the Flanker effect, Stroop effect, Flanker SCE and Stroop SCE by adding all L2 experiences into the same model for each of the dependent variables. All variance inflation factor scores were less than 3.5 for the following analyses; thus, collinearity between L2 experiences was not an issue, despite strong correlations between experiences. In all models, we included L2 AoA, L2 proficiency, L2 use, L1/L2 balance and L1/L2 switching frequency as independent variables. We present the inferential statistics in Table 3.

3.1. Flanker and Stroop effects

Flanker and Stroop effects are each calculated by subtracting the congruent from the incongruent trials. Figure 3 represents the
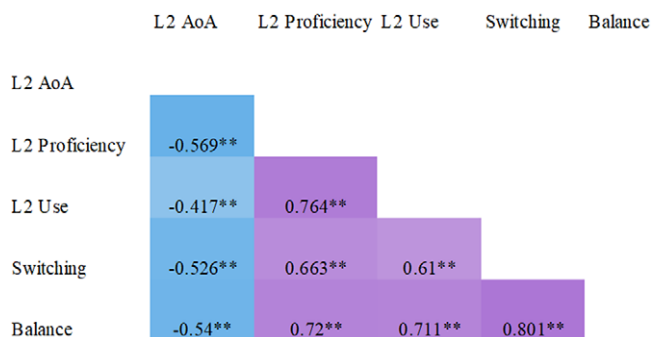
Figure 2. Heatmap for Pearson *r* correlation coefficients between second language (L2) variables.
Note: Purple values indicate positive correlations, and blue values indicate negative correlations. Darker values indicate stronger correlations, and lighter values indicate weaker correlations. All *p* < 0.01.

relation between L2 variables and Flanker (top panel) and Stroop effects (bottom panel) that reached (*p* ≤ 0.05) or approached (*p* ≤ 0.1) conventional levels of significance (see also Supplementary Figure 1 for correlations between mouse-tracking metrics and all dependent variables).

There were no significant relations between L2 experiences and Flanker effects for any of the DVs that reached conventional levels of significance. For Stroop, more L1/L2 switching frequency predicted smaller initiation time Stroop effects. No other significant associations emerged.

3.2. Flanker and Stroop SCE

The SCE (calculated as $[(cI-cC)-(iI-iC)]^1$)) is the difference between the congruency effect following congruent stimuli and the congruency effect following incongruent stimuli, and generally represents the influence of previous trial congruency on conflict (i.e., Flanker or Stroop) effects.

Table 3 and Figure 4 represent the relation between the L2 variables and Flanker SCEs that reached or approached conventional levels of significance. The Y-axis represents the size of the SCE, with higher scores representing more influence from the previous trial on current trial performance.

AoA and switching frequency consistently predicted the DVs. For initiation time, later AoA and more L1/L2 switching frequency were associated with smaller SCEs. For movement time and time to maximum deviation, the directions were reversed; earlier AoA and less L1/L2 switching frequency were associated with smaller SCEs. For RT, earlier AoA was associated with smaller RT SCEs (see Figure 4).

Table 3 and Figure 5 represent the relation between L2 variables and Stroop SCEs that reached or approached conventional levels of significance. Higher frequency of language switching was negatively associated with the SCE on initiation time, movement time and RT. Later AoA was also associated with smaller SCE RTs and movement times. No other effects reached conventional levels of significance.

4. Discussion

The present study examined the independent contributions of bilingual experiences on EFs using Flanker and Stroop mouse-tracking tasks. Several important findings emerged: 1) Mouse-tracking metrics captured the link between bilingualism and EFs

---

[1]cI = incongruent trial preceded by congruent trial, cC = congruent trial preceded by congruent trial, iI = incongruent trial preceded by incongruent trial, iC = congruent trial preceded by incongruent trial.
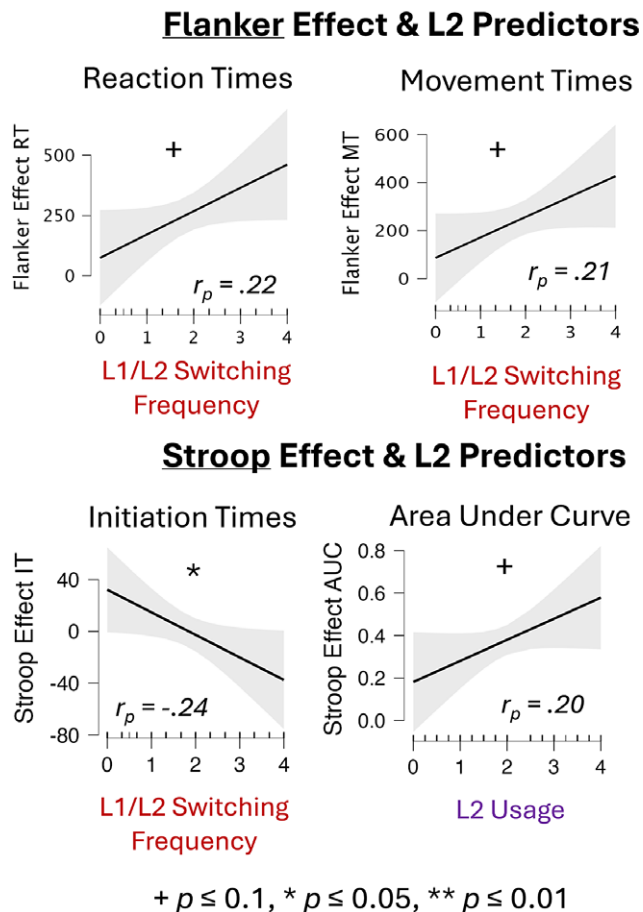
Table 3. Inferential statistics of language experiences on mouse-tracking variables

Variable/effect	Flanker conflict	Stroop conflict	Flanker SCE	Stroop SCE
Initiation time	All $p > .137$	Switching: $t(74) = -2.09$, $p = .040$	AoA: $t(74) = -2.11$, $p = .039$	Switching $t(74) = -2.00$, $p = .049$
			Switching: $t(74) = -2.31$, $p = .011$	
Movement time	All $p > .084$	all $p > .120$	AoA: $t(74) = 2.66$, $p = .010$	Switching $t(74) = -2.39$, $p = .019$
			Switching: $t(74) = 2.30$, $p = .024$	
Reaction time	All $p > .065$	all $p > .162$	AoA: $t(74) = 2.03$, $p = .046$	AoA: $t(74) = -2.01$, $p = .048$
				Switching: $t(74) = -3.03$, $p = .003$
Maximum deviation	All $p > .146$	all $p > .139$	all $p > .287$	All $p > .106$
Area under the curve	All $p > .151$	Use: $t(74) = 1.95$, $p = .055$	all $p > .129$	Use: $t(74) = 1.98$, $p = .051$
Time to maximum deviation	All $p > .115$	all $p > .230$	AoA: $t(74) = 2.47$, $p = .016$	All $p > .087$
			Switching: $t(74) = 2.57$, $p = .012$	

Note: SCE = sequential congruency effect; AoA = age of acquisition.
All models included L2 AoA, proficiency, use, balance and switching frequency as predictors in the multiple regression models.
The cell content represents the inferential statistic for the regression models. The rows represent each outcome variable for the effects.

Flanker Effect & L2 Predictors

Reaction Times

$r_p = .22$

Movement Times

$r_p = .21$

L1/L2 Switching Frequency

Stroop Effect & L2 Predictors

Initiation Times

$r_p = -.24$

Area Under Curve

$r_p = .20$

L1/L2 Switching Frequency

L2 Usage

$+ \, p \leq 0.1, \, * \, p \leq 0.05, \, ** \, p \leq 0.01$

Figure 3. Partial correlations (after controlling for all other L2 variables) for L2 variables predicting all Flanker (top panel) and Stroop (bottom panel) effects by dependent variable during mouse-tracking that reached ($p \leq 0.05$, $p \leq 0.01$) or approached ($p \leq 0.1$) conventional levels of statistical significance. Gray contours represent 95% confidence intervals.

that were not identified when solely examining overall RTs. 2) SCEs captured more relations between bilingual experiences and performance than simple conflict effects. 3) Second-language AoA and L1/L2 switching frequency consistently predicted EF outcomes. 4)

The manner in which L2 AoA and L1/L2 switching frequency predicted EF outcomes depended on the nature of the task (more demanding: Stroop, less demanding: Flanker). These findings are discussed below and highlight the importance of embracing the complexity of the effects of bilingualism on cognition and the use of more sensitive measures to capture these effects.

4.1. Overall RTs and conflict effects are not enough

We demonstrate here that mouse-tracking is particularly well-suited for exploring the cognitive processes linked to bilingualism, which may not be adequately captured by overall RTs (see also Bartolotti & Marian, 2012). We did not show a significant association of L2 experiences with RTs on Flanker and Stroop conflict effects; however, significant predictors emerged with initiation times. In our SCE analyses, we observed that L2 experiences significantly predicted many other mouse-tracking variables alongside RT, and in some cases, additional L2 predictors emerged only in non-RT metrics.

For conflict effects, overall RTs were not significantly predicted by any bilingual experiences on the Flanker or the Stroop task. However, initiation time conflict effects on the Stroop task (the additional time that it takes individuals to initiate a response to incongruent than congruent trials once a decision has been made), were predicted by L1/L2 switching frequency. The more an individual switches between languages, the less likely they are to wait before initiating a response on incongruent trials relative to congruent trials. This finding complements previous work showing that bilinguals have longer initiation times than monolinguals on mouse-tracking Stroop tasks (Incera & McLennan, 2016). Incera and McLennan (2018) then used a single measure of bilingualism, time exposed to L2, to extend these findings within bilinguals. They showed that greater bilingualism led to more time processing the stimulus before initiating a response. We break this down even further by including specific bilingual experiences that may contribute to this general pattern. We showed that switching between L1 and L2 is particularly indicative, beyond other L2 experiences, in predicting the duration individuals wait before initiating a response. This can be explained by a lifetime of flexibility in choosing which language needs to be spoken to any given
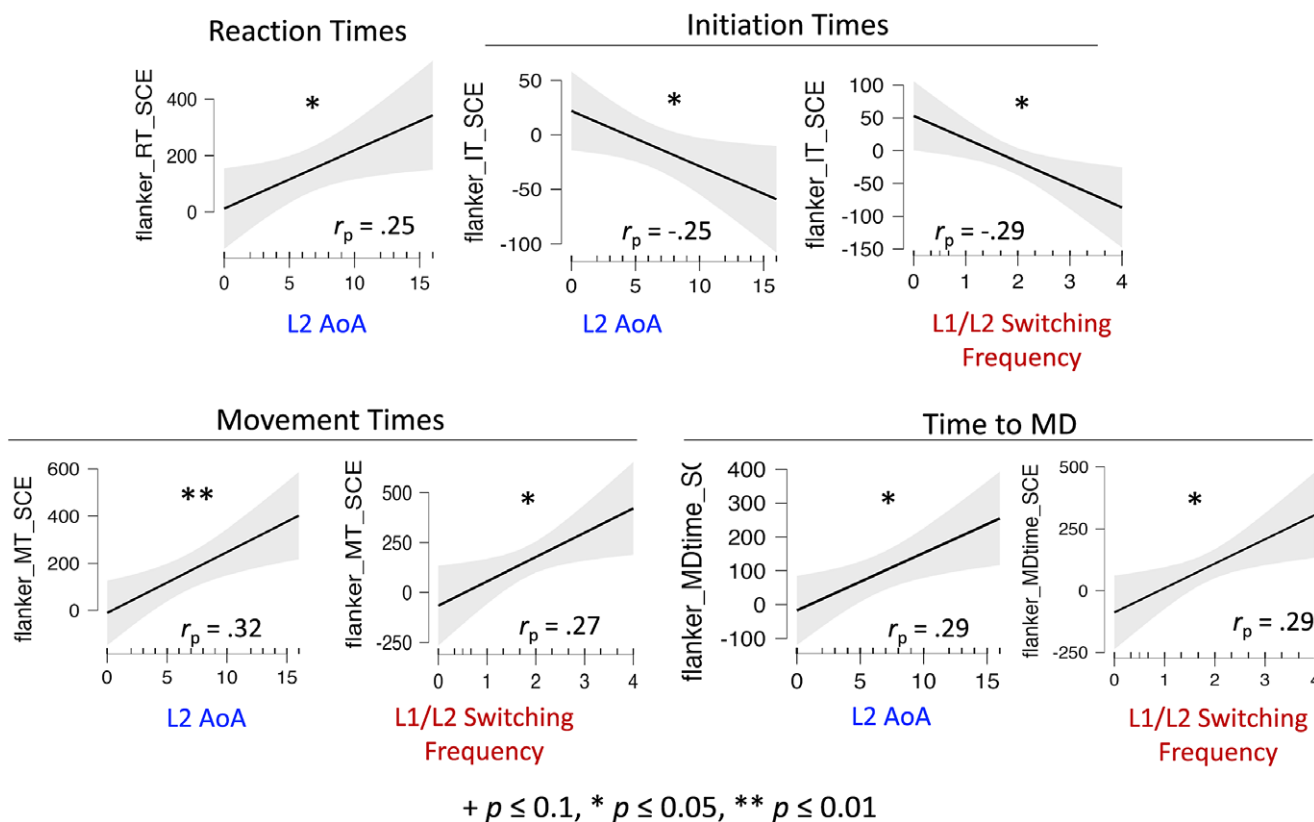
Figure 4. Partial correlations (after controlling for all other L2 variables) for L2 variables predicting SCEs on the Flanker task by dependent variable during mouse-tracking that approached ($p \leq 0.1$) or reached ($p \leq 0.05$, $p \leq 0.01$) conventional levels of statistical significance. Gray contours represent 95% confidence intervals.

interlocutor – environments in which switching between two languages is frequent, known as dense code-switching contexts (Green & Abutalebi, 2013), does not require suppressing the first word that comes to mind, given that all individuals know both languages.

The present study also explored SCEs in both Flanker and Stroop tasks and demonstrated that L2 AoA and language switching frequency consistently predicted Flanker and Stroop SCEs on several mouse-tracking variables. However, L2 AoA was the only predictor associated with overall RT SCEs during the Flanker task (see Figure 4), once again demonstrating the sensitivity of mouse-tracking metrics compared to overall RTs.

In the Flanker task, we observed that both the L2 AoA and language switching frequency predicted initiation time, movement time and time to maximum deviation SCEs, whereas no association between switching frequency and RT SCEs was observed. The positive association between the L2 AoA and SCE for the Flanker task was strongest for movement times, and in the opposite (negative) direction for initiation times. This suggests that the relation observed between the overall RT SCE and AoA is driven mainly by movement times and is weakened by initiation times given that the overall RT is the sum of initiation and movement times. For movement times and time to maximum deviation, early AoA and less switching frequency predicted smaller SCEs.

In addition to the finding that some mouse-tracking metrics are more sensitive than overall RTs, SCEs captured more relations between bilingual experiences and performance than simple conflict effects, in line with previous work (Grundy et al., 2017b). These findings are important for debates on whether bilinguals display

faster disengagement of attention than monolinguals, as several have suggested (Antoniou, 2023; Chung-Fat-Yim et al., 2022; deMeurisse & Kaan, 2023; Grundy & Bialystok, 2018; Grundy & Keyvani-Chahi, 2017; Grundy et al., 2020). Grundy et al. (2017b) first provided evidence for this position across three experiments by demonstrating that bilinguals had smaller SCEs than monolinguals. They argued that one of the reasons for failed replications in studies that examine simple conflict (e.g., Flanker and Stroop) effects is that bilinguals are faster at disengaging attention from previously irrelevant information than monolinguals. Rapid disengagement is sometimes beneficial (when the previous trial is the same as the current trial), and sometimes detrimental (when the previous trial is different from the current trial) to performance. Some studies have claimed failed replications in showing smaller SCEs for bilinguals than monolinguals (Goldsmith et al., 2023; Goldsmith & Morton, 2018; Paap et al., 2019), but the present study suggests that this may be due to collapsing L2 AoA and switching frequency variables in previous samples, effectively masking real effects that are observable at the level of continuous analyses.

4.2. Task demands influence the effects of bilingualism on EF

For both conflict effects and SCEs, the Stroop task reversed the relation seen between certain bilingual experiences and outcomes on the Flanker task, and these changes may be due to attentional demands of the tasks. Past studies have shown that the Stroop task generally requires more attentional processing than the Flanker task (Heitz & Engle, 2007; Lu & Proctor, 1995). We confirmed this

Stroop – Sequential Congruency Effects & L2 Predictors
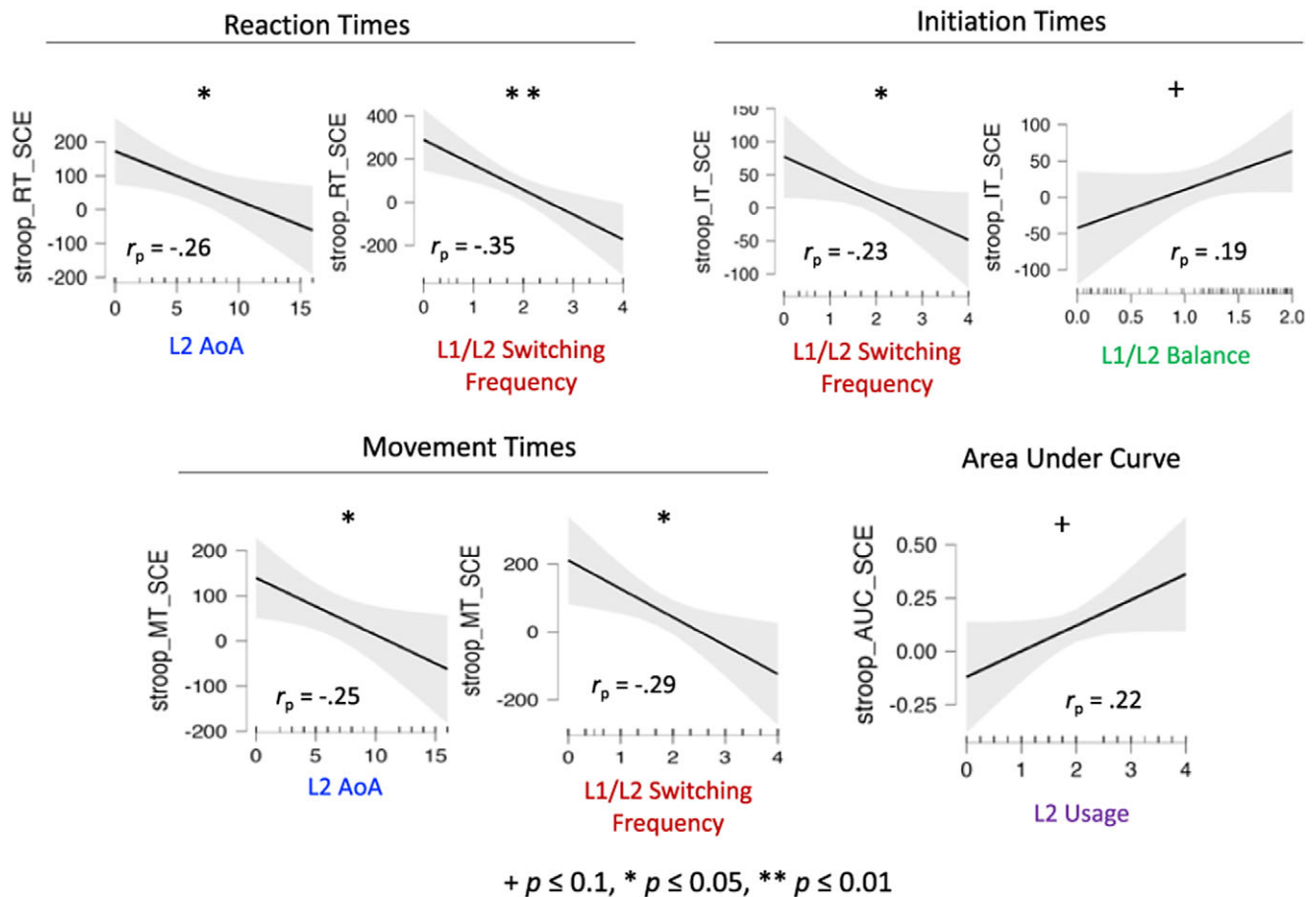
Figure 5. Partial correlations (after controlling for all other L2 variables) for L2 variables predicting SCEs on the Stroop task by dependent variable during mouse-tracking that approached ($p ≤ 0.1$) or reached ($p ≤ 0.05$, $p ≤ 0.01$) conventional levels of statistical significance. Gray contours represent 95% confidence intervals.

pattern in the present experiment by directly examining RTs: Stroop RTs ($M = 1383$ ms, $SD = 373$) were significantly longer than Flanker RTs ($M = 1196$ ms, $SD = 351$), $t(97) = −4.92$, $p < 0.001$, Cohen's $d = 0.497$. This is important because work in the field demonstrates that EF effects between monolinguals and bilinguals are more likely to appear on more attentionally demanding than less attentionally demanding tasks (e.g., Costa et al., 2009; Comishen & Bialystok, 2021).

More L2 switching frequency led to larger RT conflict effects ($r_p = 0.22$) on the Flanker task but led to smaller conflict effects on the Stroop ($r_p = −0.24$) task. A comparison of the coefficients using a Fisher's r to Z transformation (Fisher, 1915) reveals that they are statistically different from each other and in the opposite direction (Z-score = 2.51, $p = 0.006$). This pattern might be explained by the idea that rapidly switching between languages leads to adaptations to the underlying attentional system required for demanding, but not simple, tasks. The adaptation that results from switching between languages might not be efficient for simpler EF tasks, in which case, adaptations from avoiding intrusions from the unwanted language (i.e., conditions where less switching is preferred) could prove beneficial.

For movement times and RTs on the Flanker task, greater L2 AoA and L2 switching frequency led to larger SCEs, whereas the opposite was true for Stroop SCEs. Like the conflict effects, these

findings highlight the importance of considering the attentional demands of the task. Not considering this task characteristic could lead to "failed" conceptual replications for the effects of bilingualism on EFs. Among bilinguals, Mishra et al. (2012) demonstrated that high L2 proficiency bilinguals disengaged attention more rapidly than low L2 proficiency bilinguals on an inhibition of return paradigm, but later did not replicate this effect with a different sample (Saint-Aubin et al., 2018). Their failed replication could reflect the possibility that L2 AoA and L2 switching frequency are better predictors of disengagement of attention than L2 proficiency, and that their second study did not have equivalent levels of L2 AoA and L2 switching frequency, masking the effects of L2 proficiency on this EF task.

Disengagement of attention is also heavily implicated in task switching. Bilinguals who switch more often between languages show better performance on nonlinguistic task switching (Prior & Gollan, 2011; Sanchez-Azanza et al., 2020; Verreyt et al., 2015) than those who switch less. Some researchers have argued that the Flanker task exercises more reactive control than the Stroop task, which, in turn, exercises more proactive control (Gonthier et al., 2016; Kalanthroff et al., 2017). The latter is deemed similar to the nature of language switching in which bilinguals engage. This difference in the underlying control modes of the tasks may partially account for the observed

reversal in patterns. These findings underscore the significance of considering task demands, along with the timing and nature of language acquisition, in understanding how attentional control is exerted.

4.3. L2 AoA and language switching frequency

The present study delved into the complex relation between bilingual experiences and conflict effects using various mouse-tracking metrics. According to the adaptive control hypothesis, different bilingual experiences place varying demands on the EF system (Green & Abutalebi, 2013). Our findings revealed that L2 AoA and language switching frequency consistently predicted Flanker and Stroop SCEs. We provide some explanations for these emerging patterns.

The more consistent relation between earlier L2 AoA and smaller SCEs found in our data suggests that individuals who acquire a second language *earlier* in life are better able to disengage from past trials. This efficiency in response selection may stem from prolonged exposure and experience in managing two languages (Bialystok, 2017; Luk et al., 2011). This idea is further supported by findings that early bilinguals exhibit reduced prepotent responses on a number–letter task-switching paradigm (Yow & Li, 2015) and reduced mixing costs (Soveri et al., 2011a) compared to later bilinguals, who have less experience navigating language conflict.

Regarding language switching frequency, our results varied across tasks and mouse-tracking metrics. For the Stroop task, we observed that *higher* switching frequency was associated with smaller SCEs. This finding suggests that frequent switchers were less influenced by the context (i.e., trial type). Frequent switching between multiple languages requires rapid attentional disengagement, and these domain-general adaptations may lead to better disengagement of attention over time, resulting in smaller SCEs.

Interestingly, in the Flanker task, the pattern was reversed. Bilinguals with *less* frequent language switching demonstrated smaller SCEs. This discrepancy might be explained by the different cognitive demands of the two tasks. The Flanker task involved lower attentional demands, and individuals who switch languages less frequently might employ different response strategies. Less frequent switchers may have optimized attentional control for less variable environments (i.e., single-language contexts), which predicts more efficient performance in simpler tasks like the Flanker. The low reported switching frequency ($M = 1.49$ out of 4) among our participants suggests that they may predominantly engage in single-language contexts (e.g., L2 at school vs. L1 at home). In these environments, selective attentional control may not be exercised as vigorously as in dual-language contexts (Green & Abutalebi, 2013). Therefore, the cognitive strategies developed by less frequent switchers could be more effective for tasks with lower demands, such as the Flanker task. These findings are complex, and indicate a need for a more nuanced examination of how specific bilingual experiences influence SCEs. Our study aims to initiate this exploration, and by sharing our data, we encourage further analysis and theoretically driven predictions to unravel these complexities.

4.4. Limitations

It is important to consider that mouse-tracking is a relatively new methodology. We urge future research to undertake theoretical work essential for understanding what mouse-tracking metrics reveal about bilinguals' cognitive processes and different response strategies, as these mechanisms have yet to be consistently identified. Therefore, inferences about mechanisms and nuances between different metrics beget caution. Furthermore, the design features of the mouse-tracking paradigm have a considerable effect on the cognitive effects observed (Schoemann et al., 2021). There is also much heterogeneity in the design of the paradigms and inconsistent reporting of design choices. For example, the use of click versus hover responses, static versus dynamic starts, or issuing a warning if initiation time is long, allow the researchers to infer different parts of the decision-making process. We have opted for a static start and click response because they yield larger conflict effects (Grage et al., 2019). Therefore, our results might not generalize to other mouse-tracking paradigms.

In the present study, we have described our sample's bilingualism characteristics in detail. The specificity and complexity of our sample's experiences warrant consideration in assuming generalizability of our results. However, rather than deem this a limitation, we urge researchers to take on this complexity by measuring and reporting bilingual experiences, as have others (Anderson et al., 2024; Dass et al., 2024; De Bruin, 2019; Yurtsever et al., 2023). In addition to individual differences in bilingual experiences, environmental factors like the community and context in which bilinguals use their languages may also influence attentional control. Due to the study location and individual and linguistic diversity in our sample, we were not able to address these factors. Incorporating detailed information regarding bilingual experience can enhance our understanding of cognitive processes revealed by mouse-tracking metrics. In this study, we have also taken an exploratory data-driven approach to observe the effect of bilingualism predictors on EF tasks. Future research should incorporate theory-driven approaches to enhance the robustness and generalizability of our findings.

5. Conclusion

The present study used Flanker and Stroop mouse-tracking paradigms to examine the effects of bilingualism on EFs and found that second-language AoA and first-/second-language switching frequency predicted both conflict and SCEs. These relations were limited when overall RTs were examined rather than the additional metrics that the mouse-tracking paradigms provide, including the time that it takes to initiate a response, complete a response, and maximum deviation from the most efficient path to the correct response. More bilingualism and EF associations were observed for SCEs than for simple conflict effects, and the attentional demands of the task (Stroop: High vs. Flanker: Low) reversed the direction of many of these patterns. These findings provide strong evidence that different bilingual experiences predict different EF outcomes, and that more sensitive measures are needed to capture these dynamics.

References

Anderson, J. A., **Mak, L.**, **Keyvani Chahi, A.**, & **Bialystok, E.** (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, **50**, 250–263.

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

Anderson, J. A., Yurtsever, A., Fisher-Skau, O., Cherep, L. A., MacPhee, I., Luk, G., & Grundy, J. G. (2024). Examining the consistency in bilingualism and white matter research: A meta-analysis. *Neuropsychologia*, **195**, 108801.

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., Cappa, S. F., & Costa, A. (2011). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, **22**, 2076–2086.

Antoniou, K. (2023). The ups and downs of bilingualism: A review of the literature on executive control using event-related potentials. *Psychonomic Bulletin & Review*, **30**, 1187–1226.

Backer, K. C., & Bortfeld, H. (2021). Characterizing bilingual effects on cognition: The search for meaningful individual differences. *Brain Sciences*, **11**, 81.

Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science*, **36**, 1129–1147.

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

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

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**, 696–705.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, **143**, 233–262.

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

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, **280**, 747–749.

Chung-Fat-Yim, A., Calvo, N., & Grundy, J. G. (2022). The multifaceted nature of bilingualism and attention. *Frontiers in Psychology*, **13**, 910382.

Comishen, K. J., & Bialystok, E. (2021). Increases in attentional demands are associated with language group differences in working memory performance. *Brain and Cognition*, **147**, Article 105658. https://doi.org/10.1016/j.bandc.2020.105658

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

Damian, M. F., Ye, W., Oh, M., & Yang, S. (2019). Bilinguals as "experts"? Comparing performance of mono-to bilingual individuals via a mousetracking paradigm. *Bilingualism: Language and Cognition*, **22**, 1176–1193.

Dass, R., Smirnova-Godoy, I., McColl, O., Grundy, J. G., Luk, G., & Anderson, J. A. (2024). A content overlap analysis of bilingualism questionnaires: Considering diversity. *Bilingualism: Language and Cognition*, 1–7.

De Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, **9**, 33.

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, **116**, 7565–7574.

DeLuca, V., Segaert, K., Mazaheri, A., & Krott, A. (2020). Understanding bilingual brain function and structure changes? U bet! A unified bilingual experience trajectory model. *Journal of Neurolinguistics*, **56**, 100930.

deMeurisse, G., & Kaan, E. (2023). Bilingual attentional control: Evidence from the Partial Repetition Cost paradigm. *Bilingualism: Language and Cognition*, FirstView, 1–11.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, **64**, 135–168.

Egner, T., Ely, S., & Grinband, J. (2010). Going, going, gone: Characterizing the time-course of congruency sequence effects. *Frontiers in Psychology*, **1**, 6922.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*. **10**, 507–521.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, **42**, 226–241.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, **86**, 186–204.

Goldsmith, S. F., & Morton, J. B. (2018). Sequential congruency effects in monolingual and bilingual adults: A failure to replicate Grundy et al. (2017). *Frontiers in Psychology*, **9**, 2476.

Goldsmith, S. F., El-Baba, M., He, X., Lewis, D. J., Dirani, L. A., Liu, J., & Morton, J. B. (2023). No bilingual advantage in children's attentional disengagement: Congruency and sequential congruency effects in a large sample of monolingual and bilingual children. *Journal of Experimental Child Psychology*, **233**, 105692.

Grant, A., Dennis, N. A., & Li, P. (2014). Cognitive control, cognitive reserve, and memory in the aging bilingual brain. *Frontiers in Psychology*, **5**, 1401.

Grage, T., Schoemann, M., Kieslich, P. J., & Scherbaum, S. (2019). Lost to translation: How design factors of the mouse-tracking procedure impact the inference from action to cognition. *Attention, Perception, & Psychophysics*, **81**, 2538–2557.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, **121**, 480–506.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, **25**, 515–530.

Grundy, J. G., Anderson, J. A., & Bialystok, E. (2017a). Neural correlates of cognitive processing in monolinguals and bilinguals. *Annals of the New York Academy of Sciences*, **1396**, 183–201.

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

Grundy, J. G., & Keyvani Chahi, A. (2017). Post-conflict slowing effects in monolingual and bilingual children. *Developmental Science*, **20**, e12488.

Grundy, J. G., & Bialystok, E. (2018). Monolinguals and bilinguals disengage attention differently following conflict and errors: Evidence from ERPs. *Brain and Cognition*, **128**, 28–36.

Grundy, J. G., Pavlenko, E., & Bialystok, E. (2020). Bilingualism modifies disengagement of attention networks across the scalp: A multivariate ERP investigation of the IOR paradigm. *Journal of Neurolinguistics*, **56**, 100933.

Grundy, J. G. (2020). The effects of bilingualism on executive functions: An updated quantitative analysis. *Journal of Cultural Cognitive Science*, **4**, 177–199.

Grundy, J. G. (2022). The specificity and reliability of conflict adaptation: A mouse-tracking study. *Frontiers in Psychology*, **12**, 770509.

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, **44**, 778–788.

Gullifer, J., & Anderson, J. A. (2023). Challenges of complexity, and possible solutions: A commentary on rethinking multilingual experience through a Systems Framework of Bilingualism by Titone and Tiv. *Bilingualism: Language and Cognition*, **26**, 22–24.

Gunnerud, H. L., Ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, **146**, 1059–1083.

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, **136**, 217–240.

Incera, S., & McLennan, C. T. (2016). Mouse tracking reveals that bilinguals behave like experts. *Bilingualism: Language and Cognition*, **19**, 610–620.

Incera, S., & McLennan, C. T. (2018). Bilingualism and age are continuous variables that influence executive function. *Aging, Neuropsychology, and Cognition*, **25**(3), 443–463.

The jamovi project (2023). jamovi (Version 2.3) [Computer Software]. https://www.jamovi.org

Kalanthroff, E., Davelaar, E. J., Henik, A., Goldfarb, L., & Usher, M. (2017). Task conflict and proactive control: A computational theory of the Stroop task. *Psychological Review*, **125**, 59–82.

Kousaie, S., & Phillips, N. A. (2017). A behavioural and electrophysiological investigation of the effect of bilingualism on aging and cognitive control. *Neuropsychologia*, **94**, 23–35.

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science*, **23**, 159–163.

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics*, **1**, 377–394.

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**, 394–425.

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

Liu, C., Yang, C. L., Jiao, L., Schwieter, J. W., Sun, X., & Wang, R. (2019). Training in language switching facilitates bilinguals' monitoring and inhibitory control. *Frontiers in Psychology*, **10**, 1839.

Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant information on performance: A review of the Stroop effect and related phenomena. *Psychonomic Bulletin & Review*, **2**, 174–207.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, MA: MIT Press.

Luk, G., Bialystok, E., Craik, F. I., & Grady, C. L. (2011). Lifelong bilingualism maintains white matter integrity in older adults. *The Journal of Neuroscience*, **31**, 16808–16813.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, **25**, 605–621.

Luk, G., & Grundy, J. G. (2023). The importance of recognizing social contexts in research on bilingualism. *Bilingualism: Language and Cognition*, **26**, 25–27.

Mishra, R. K., Hilchey, M. D., Singh, N., & Klein, R. M. (2012). On the time course of exogenous cueing effects in bilinguals: Higher proficiency in a second language is associated with more rapid endogenous disengagement. *Quarterly Journal of Experimental Psychology*, **65**, 1502–1510.

Morton, J. B. & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, **10**(6), 719–726.

Paap, K. R., Myuz, H., Anders-Jefferson, R., Mason, L., & Zimiga, B. (2019). On the ambiguity regarding the relationship between sequential congruency effects, bilingual advantages in cognitive control, and the disengagement of attention. *AIMS Neuroscience*, **6**, 282–298.

Pot, A., Keijzer, M., & De Bot, K. (2018). Intensity of multilingual language use predicts cognitive performance in some multilingual older adults. *Brain Sciences*, **8**, 92.

Pliatsikas, C. (2020). Understanding structural plasticity in the bilingual brain: The Dynamic Restructuring Model. *Bilingualism: Language and Cognition*, **23**, 459–471.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, **13**, 253–262.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society : JINS*, **17**, 682–691.

Qi, G., Qin, F., Zhang, X., & Yun, T. (2023, May). Re-exploring the effect of bilingualism on inhibitory control: Application of survival analysis in flanker task. In *International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)* (Vol. **12604**, pp. 524–529). SPIE.

Rodríguez-Pujadas, A., Sanjuán, A., Fuentes, P., Ventura-Campos, N., Barrós-Loscertales, A., & Ávila, C. (2014). Differential neural control in early bilinguals and monolinguals during response inhibition. *Brain and Language*, **132**, 43–51.

Saint-Aubin, J., Hilchey, M. D., Mishra, R., Singh, N., Savoie, D., Guitard, D., & Klein, R. M. (2018). Does the relation between the control of attention and second language proficiency generalize from India to Canada?. *Canadian Journal of Experimental Psychology*, **72**, 208–218.

Sanchez-Azanza, V. A., López-Penadés, R., Aguilar-Mediavilla, E., & Adrover-Roig, D. (2020). Latent variable evidence on the interplay between language switching frequency and executive control in Spanish-Catalan bilinguals. *International Journal of Bilingualism*, **24**(5–6), 912–930.

Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods*, **50**, 2097–2110.

Schoemann, M., O'Hora, D., Dale, R., & Scherbaum, S. (2021). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin & Review*, **28**, 766–787.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, **79**, 217–240.

Sheth, S. A., Mian, M. K., Patel, S. R., Asaad, W. F., Williams, Z. M., Dougherty, D. D., … & Eskandar, E. N. (2012). Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature*, **488**, 218–221.

Schoemann, M., O'Hora, D., Dale, R., & Scherbaum, S. (2021). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin & Review*, **28**, 766–787.

Singh, N., & Mishra, R. K. (2013). Second language proficiency modulates conflict-monitoring in an oculomotor Stroop task: Evidence from Hindi–English bilinguals. *Frontiers in Psychology*, **4**, 322.

Singh, L., Fu, C. S., Rahman, A. A., Hameed, W. B., Sanmugam, S., Agarwal, P., … & GUSTO Research Team. (2015). Back to basics: A bilingual advantage in infant visual habituation. *Child Development*, **86**, 294–302.

Soveri, A., Laine, M., Hämäläinen, H., & Hugdahl, K. (2011a). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism: Language and Cognition*, **14**, 371–378.

Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011b). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. *Frontiers in Psychology*, **2**, 183.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 10393–10398.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, **104**, 12530–12535.

Titone, D. A., & Tiv, M. (2023). Rethinking multilingual experience through a Systems Framework of Bilingualism. *Bilingualism: Language and Cognition*, **26**, 1–16.

Tse, C. S., & Altarriba, J. (2012). The effects of first-and second-language proficiency on conflict resolution and goal maintenance in bilinguals: Evidence from reaction time distributional analyses in a Stroop task. *Bilingualism: Language and Cognition*, **15**, 663–676.

van Heuven, W. J., & Coderre, E. L. (2015). A call for sophisticated statistical approaches and neuroimaging techniques to study the bilingual advantage. *Cortex*, **73**, 330–331.

Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2015). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, **19**, 181–190.

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.

Ware, A. T., Kirkovski, M., & Lum, J. A. (2020). Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Frontiers in Psychology*, **11**, 1458.

Wu, Y. J., & Thierry, G. (2010). Chinese–English bilinguals reading English hear Chinese. *Journal of Neuroscience*, **30**, 7646–7651.

Yamasaki, B. L., **Stocco, A.**, & **Prat, C. S.** (2018). Relating individual differences in bilingual language experiences to executive attention. *Language, Cognition and Neuroscience*, **33**, 1128–1151.

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

Yurtsever, A., **Anderson, J. A.**, & **Grundy, J. G.** (2023). Bilingual children outperform monolingual children on executive function tasks far more often than chance: An updated quantitative analysis. *Developmental Review*, **69**, 101084.

Zunini, R. a. L., **Morrison, C.**, **Kousaie, S.**, & **Taler, V.** (2019). Task switching and bilingualism in young and older adults: A behavioral and electrophysiological investigation. *Neuropsychologia*, **133**, 107186.