Psychonomic Bulletin & Review (2023) 30:1187–1226
https://doi.org/10.3758/s13423-023-02245-x

THEORETICAL/REVIEW

The ups and downs of bilingualism: A review of the literature on executive control using event-related potentials

Kyriakos Antoniou[1,2,3]

Accepted: 30 December 2022 / Published online: 26 January 2023

Abstract

Whether bilingualism enhances executive control (EC) is controversial. This article reviews 24 studies on the bilingual EC effect using event-related potentials (ERPs). It evaluates the evidence based on considerations of neural efficiency, different EC theories, and accounts regarding the locus of the bilingual effect. The review finds some evidence for a positive bilingual impact. This is more consistent for the P3 and response-locked ERPs. Moreover, when considering each component independently, evidence primarily supports a monitoring and secondarily an inhibition locus. Additionally, an N2/ERN (error-related negativity) dissociation (no bilingual N2 effect but positive ERN impact, evident as smaller ERN), coupled with the P3 results, suggest that monitoring may not be the (only) locus of a bilingual effect but (an)other post-monitoring mechanism(s). Attention disengagement also receives some support. Finally, results across studies are largely consistent with the Bilingualism Anterior to Posterior and Subcortical Shift model (BAPSS): Bilingual effects, when found, often manifest as shorter latencies, larger components or wider amplitude effects during earlier (N2, P3) but smaller components or narrower effects during later processing (stimulus-locked negativities and response-locked components). However, this evidence is not unequivocal. Many bilingual-monolingual comparisons reveal null or some suggest negative or opposite to prediction bilingual effects. Second, the scant evidence about which bilingual experiences impact EC is, generally, unclear, while some evidence indicates negative effects. Third, BAPSS is often not confirmed when multiple components are examined within subjects. Finally, this literature is challenged by confounds and small samples. Further research is required to conclude a positive bilingual effect on EC in ERPs.

Keywords Electrophysiology · ERPs and working memory/attention · Task switching or executive control · Attention and executive control

Introduction

It is estimated that bilingual speakers – individuals who regularly use more than one language or dialect – make up most of the world population (Grosjean & Li, 2013). As a result, bilingualism and its possible neuro-cognitive effects have become a topic of central interest for researchers in linguistics, education, psychology, and cognitive neuroscience (e.g., Bialystok, 2017; Bialystok et al., 2009; Costa & Sebastián-Gallés, 2014; de Houwer & Ortega, 2018; García-Pentón et al., 2016; Grundy, Anderson, & Bialystok, 2017a; Paap et al., 2015; Pliatsikas, 2019; Pliatsikas & Luk, 2016).

Within this body of work, some findings suggest that bilingualism is an experience that enhances executive control (e.g., Bialystok, 2017). Executive control (EC) refers to a domain-general, non-verbal system that regulates cognition and behavior in line with internal goals and current context (e.g., Miyake et al., 2000). There exist different accounts on the nature and structure of the EC system (see, e.g., Gratton et al., 2018). However, drawing on the influential work of Miyake et al. (2000), researchers most often examine three executive functions (Karr et al., 2018): *shifting* or *task-switching* (the ability to flexibly and rapidly switch between rules, representations, or tasks), *updating and monitoring the contents of working memory*[1] (coding and monitoring

✉ Kyriakos Antoniou
kyriakos.antoniou@cut.ac.cy

[1] Department of Rehabilitation Sciences, Cyprus University of Technology, 15 Vragadinou Street, 3041 Limassol, Cyprus

[2] Hellenic Open University, Patra, Greece

[3] Center for Applied Neuroscience, University of Cyprus, Nicosia, Cyprus

---

[1] The terms *updating* and *working memory* are often used interchangeably. I follow this practice here.

information in working memory for relevance to a given task and revising it if it is no longer relevant), and *inhibition* (inhibiting dominant responses and irrelevant information). According to Miyake et al. (2000), the EC system is characterized by unity and diversity, in that these three different and other EC processes are partly distinguishable but also moderately interrelated.

Bilingualism is thought to enhance EC because bilinguals presumably use EC on a constant basis to manage their simultaneously active languages in the mind and brain, and to monitor the interactional situation so as to select and use the right language during everyday communication (e.g., Bialystok, 2017; Costa et al., 2009). More recently, however, a surge of studies showing no bilingual effect on EC has turned the question of whether bilingualism affords EC benefits into a controversial research topic (e.g., Bialystok, 2017; Paap et al., 2015). This topic has mainly been examined using behavioral measures, such as speed and accuracy of manual responding (e.g., Lehtonen et al., 2018). Nevertheless, lately, many studies have investigated the bilingual effect on brain structure and function with methods from cognitive neuroscience (e.g., Bialystok, 2017; Cespón & Carreiras, 2020; García-Pentón et al., 2016; Grundy, Anderson, & Bialystok, 2017a; Pliatsikas & Luk, 2016).

This article critically reviews studies on the bilingual EC effect using the event-related potential (ERP) electrophysiological method. This review, with its focus on ERPs, is timely for several reasons. First, the prediction of a bilingual EC benefit has primarily been tested in terms of timing (e.g., bilinguals show faster response times in EC tasks; e.g., Draheim et al., 2019; Hilchey & Klein, 2011; Lehtonen et al., 2018). When measuring time, however, the ERP method is superior to reaction times (RTs): It records brain activity linked to cognitive processes directly, with millisecond temporal resolution, and without being influenced by metalinguistic and – for stimulus processing – response-related (e.g., motor) processes (Luck, 2014; Steinhauer, 2014). Thus, the ERP technique provides direct, more temporally accurate, and purer measures of the timing of cognitive processing. In turn, it may be more likely to reveal differences even in populations for whom behavioral effects are harder to detect or might not exist. A bilingual behavioral benefit, for instance, is possibly harder to find, or does not exist, in young adults. This is because young adults are presumably at the peak of cognition with less room for further improvement (e.g., Bialystok & Craik, 2010). Relatedly, ERPs enable us to determine the specific time point at which the effect occurs or the cascade of functions affected by bilingualism in the stream of processing, something not possible with behavioral measures. This, in turn, allows for testing recent proposals that posit bilingual neuro-cognitive effects at different processing stages over time (e.g., Grundy, Anderson, & Bialystok, 2017a).

Finally, there is now a growing body of work that used ERPs to examine bilingual effects in EC tasks. To date, however, reviews have focused on only a small number of ERP studies (Bialystok, 2017; Cespón & Carreiras, 2020; Grundy, Anderson, & Bialystok, 2017a; Incera, 2018), on a few ERP components (N2, P3, N450, error-related negativity (ERN)) or on specific EC tasks (Incera, 2018). It is necessary to carefully scrutinize the bulk of the evidence from this research to obtain a clear and complete picture of the ERP literature on the bilingual EC effect. Understanding whether bilingualism affects cognitive and brain functioning is important because the positive, neuroplastic effects of bilingualism have been argued to have lasting implications for the individuals involved; specifically, some evidence suggests that bilingualism contributes to cognitive reserve, protects against cognitive decline with aging, and delays the onset of neurodegenerative diseases such as dementia (Bialystok, 2017).

The goal of this review is to synthesize and critically assess the findings on whether and how bilingualism modulates various electrophysiological measures (ERP components) linked to EC. Moreover, this work aims to provide a clear theoretical framework and specific hypotheses about how a positive bilingual effect on EC may manifest at the neural, electrophysiological level. To achieve this, it draws on considerations of neural efficiency, specific EC theories, and different proposals regarding the precise neuro-cognitive locus of the bilingual EC effect (see, e.g., Cespón, 2021; de Bruin et al., 2021). It also discusses some methodological considerations for future research in order to advance the investigation of bilingual neuro-cognition through the ERP lens. Ultimately, I hope that this review will provide a useful overview of the theoretical issues, methods, and results from this literature, and a reference for researchers who are interested in using ERPs not only in the bilingualism field but also in the broader area of EC. Moreover, I wish to highlight the potential of the ERP method in providing critical evidence for research, including work on the bilingual EC effect, which relies on predictions concerning timing; and to outline methodological issues that will allow future work to take advantage of its full potential in this regard. In the next sections, I briefly review the literature on the bilingual effect on EC. I also outline different EC theories and different accounts on the locus of the bilingual EC effect, and generate predictions on how a positive bilingual effect may manifest in ERPs. I then move on to review the ERP studies.

Bilingual executive control advantage: The debate

Direct support for a bilingual EC advantage has mainly come from behavioral research showing superior – faster and/or more accurate – performance for bilinguals in EC tasks. Figures 1, 2, 3, 4, 5 and 6 present six tasks that have been employed to examine EC in bilinguals using ERPs.

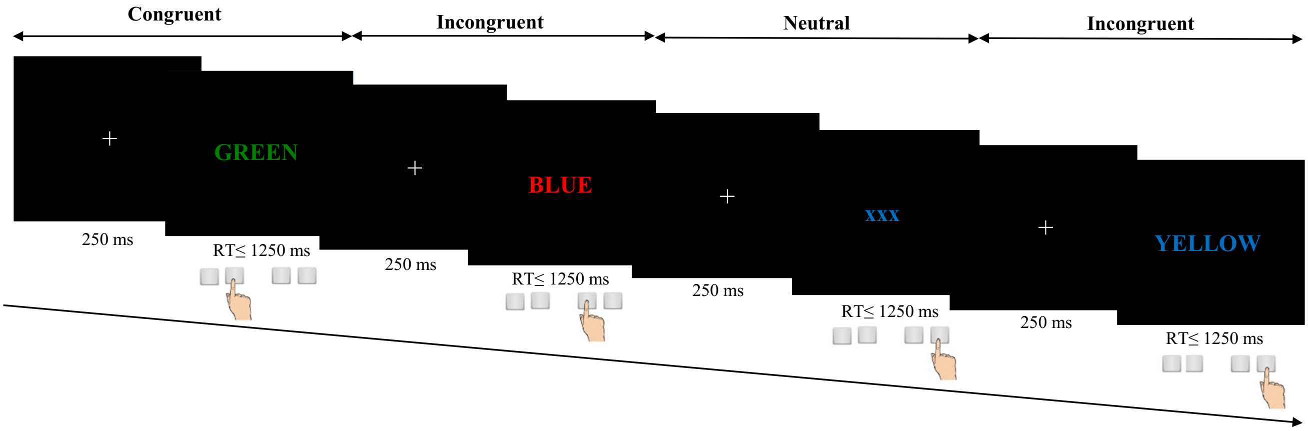

Fig. 1 The Stroop task (e.g., Kousaie & Phillips, 2012): Respond based on the font color. *RT* reaction time

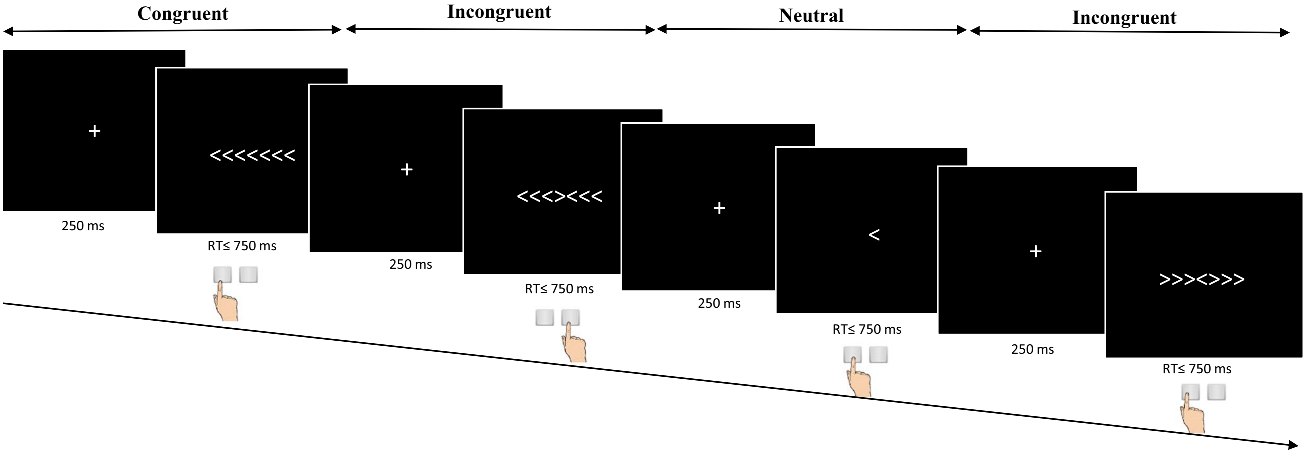

Fig. 2 The Flanker task (e.g., Kousaie & Phillips, 2012): Respond based on the direction of the middle chevron. *RT* reaction time

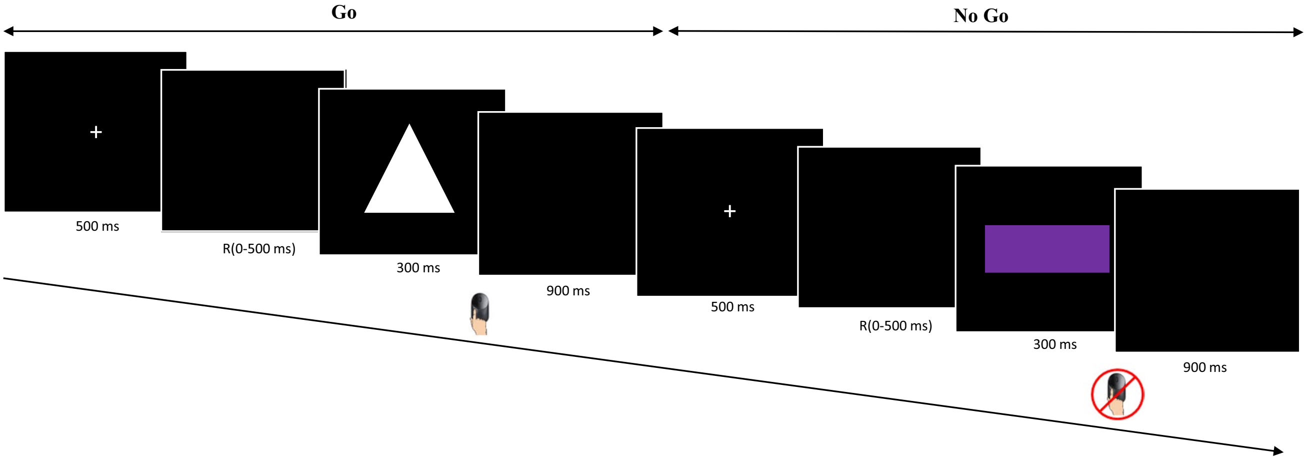

Fig. 3 The Go/No-Go task (e.g., Sullivan et al., 2014): Respond if shape color is white and withhold response if shape is of other color. *R* random number

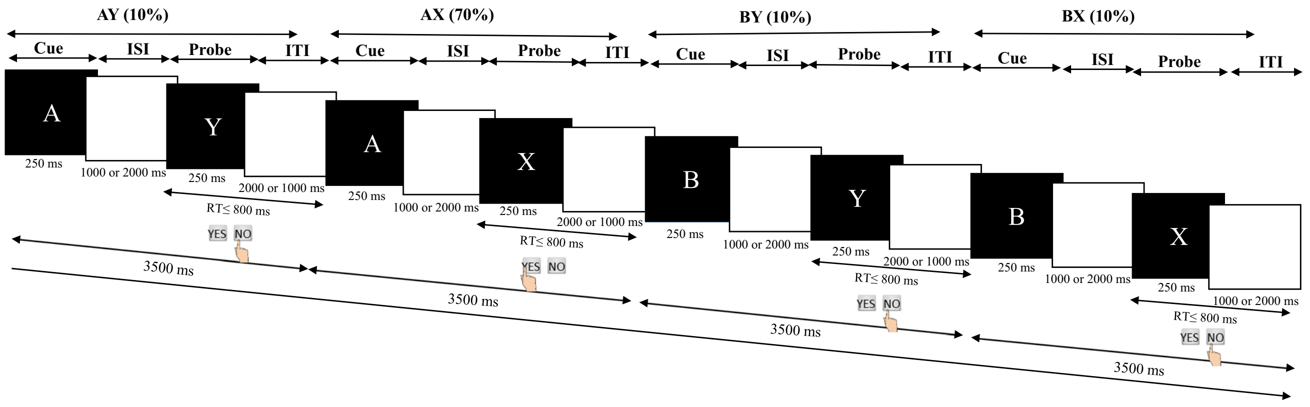

Fig. 4 The AX-Continuous Performance Task (e.g., Morales et al., 2015): Respond "Yes" to every X stimulus (probe) that is preceded by an A cue (AX trials), otherwise respond "No". *ISI* interstimulus interval, *ITI* intertrial interval, *RT* reaction time

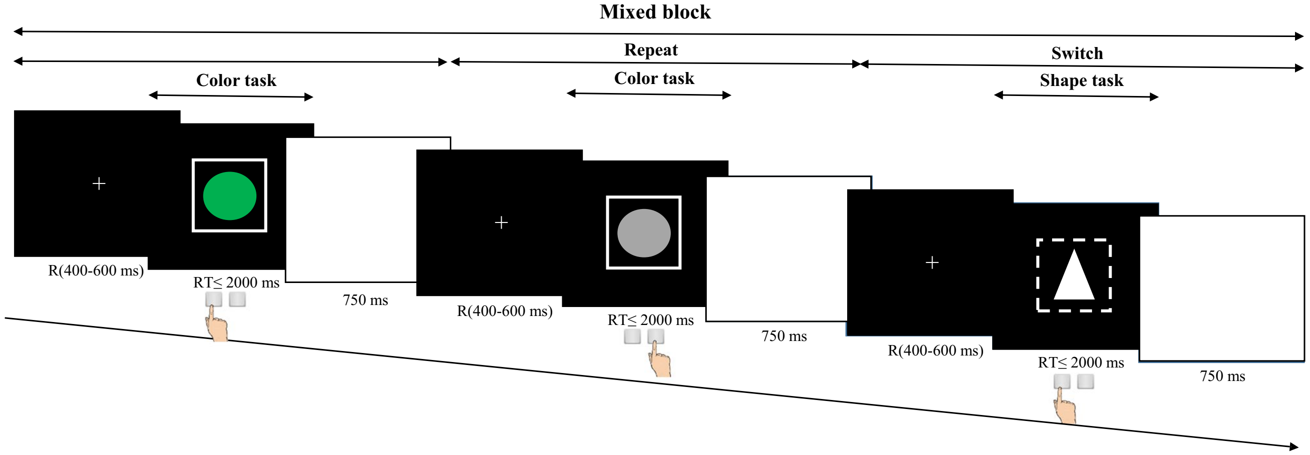

Fig. 5 The Color-Shape task (e.g., Timmer et al., 2017): Depending on the cue (solid or dashed outline), respond based on the target stimulus' color (colored or greyscale) or shape (has corners or is rounded). *RT* reaction time, *R* random number

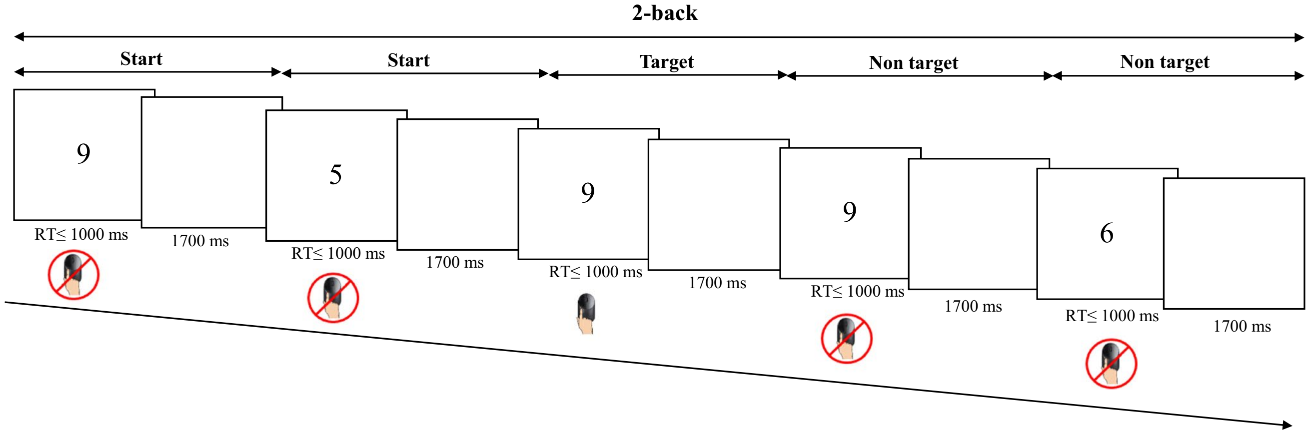

Fig. 6 The N-back task (e.g., Morrison et al., 2019). Press a button only when the current number is 0 (0-back condition) or matches the number that appeared one trial (1-back condition) or two trials (2-back condition) earlier. *RT* reaction time

For example, in the incongruent condition of the Stroop task (Fig. 1), color words are shown in a conflicting font color (e.g., "RED" in blue color). Subjects need to focus on the font color and/or inhibit reading the word to resolve conflict and respond correctly. The interference effect is calculated as the RT difference between the slower incongruent and the faster, no-conflict neutral or congruent trials. A smaller score suggests better inhibition.

An early meta-analysis reported a large positive bilingual effect on a composite cognitive measure, including attentional control (Adesope et al., 2010). More recent meta-analyses, however, reveal varying but typically much smaller effects, using different methods (e.g., different ages, EC measures, publication bias correction). These effects range from null after publication bias correction (Lehtonen et al., 2018; Lowe et al., 2021) to small (e.g., de Bruin et al., 2015; Donnelly et al., 2019, only for inhibition tasks; Grundy, 2020; Gunnerud et al., 2020, only for overall EC in middle-class children; Monnier et al., 2022, for working memory) to small-to-moderate (Grundy & Timmer, 2017, for working memory; Gunnerud et al., 2020, only for switching in children), in the direction of a bilingual benefit.

More comprehensive discussions of the methodological and conceptual issues in this literature can be found elsewhere (e.g., Bialystok, 2017; Cespón, 2021; de Bruin et al., 2021; Lehtonen et al., 2018; Navarro-Torres et al., 2021; Paap et al., 2015). However, some factors that could explain the inconsistency in the findings include the following: (1) The use of small sample sizes because underpowered studies are more likely to report erroneous findings in all directions: null, significant positive or negative results, or inflated effects. (2) Failure to control for confounds, such as culture, socioeconomic status (SES), or immigration status. Research on cultural effects, for example, has reported EC differences between broader cultural groups such as Asian and Western participants (e.g., Sabbagh et al., 2006; Samuel et al., 2018). (3) Relatedly, over-control is another issue. Deliberately matching bilinguals and monolinguals on or controlling for general intelligence or attention might attenuate a potential bilingual effect because these cognitive aspects are inherently related (conceptually and empirically) with (facets of) EC. (4) The specific characteristics of bilinguals. The adaptive control hypothesis (Green & Abutalebi, 2013), for example, proposes that the context of dual-language use, and the type and degree of language switching it entails, places varying demands on different EC skills in bilinguals. Drawing on this account, larger and more widespread effects on the EC system are expected for bilinguals who frequently switch languages based on the interlocutor (dual-language context) than for bilinguals in single-language contexts who use a separate language in different situations. In turn, single-language bilinguals may exhibitit greater and more widespread EC benefits than bilinguals in dense code-switching contexts. In the latter situation, bilinguals often mix their languages in the same utterance, and language use requires minimum EC demands. Language proficiency is another bilingual experience that may affect EC. Bilinguals often exhibit lower language proficiency (e.g., smaller vocabulary, slower lexical access) when each of their languages is considered separately (e.g., Bialystok et al., 2009; Ivanova & Costa, 2008). Thus, using verbal EC tasks may mask a bilingual effect. Also, high second language (L2) proficiency possibly results in more demands for and, in turn, more training of EC during daily language use for bilinguals. (5) A final criticism is that studies often lack a theory on the nature of EC and on how bilingualism impacts this system (e.g., de Bruin et al., 2021; Jared, 2015). Theory-driven work leads to more sensitive experiments because, for example, the researcher selects the tasks and measures more likely to reveal an effect given a theory. It also results in findings more likely to be true and in more accurate conclusions (e.g., on whether effects reflect benefits or not) because data are analyzed and interpreted based only on pre-existing theory (e.g., Cespón, 2021; de Bruin et al., 2021).

Event-related potentials, bilingualism, and executive control: The present review

In the past decade, researchers have also investigated the bilingual EC benefit using the ERP neuroscientific technique (e.g., Cespón & Carreiras, 2020; Grundy, Anderson, & Bialystok, 2017a). Table 1 describes and Fig. 7 illustrates the ERP components that have been studied with different EC tasks in bilingualism research. An ERP component is a characteristic brain wave recorded from the scalp as a change in electrical brain activity or voltage and reflects a specific neuro-cognitive process (Luck, 2014). ERPs are extracted from the continuous electroencephalogram by averaging brain responses – voltage fluctuations or waveforms – time-locked to a specific task event, such as the onset of an incongruent target or of a manual response, from multiple trials. Different ERP components can vary in aspects such as polarity (positive or negative), amplitude size, latency (time in milliseconds at which they are elicited), and scalp distribution (electrodes over brain regions where they are recorded). The latency and amplitude of each ERP component provide information about the timing and strength of the associated neuro-cognitive process.

Naturally, a shorter latency is interpreted positively – as faster or less effortful processing – at least when linked to (known) evidence indicating a better or no effect on behavioral performance; specifically, it suggests that the brain reacts faster or devotes less resources to achieve a better or same behavioral result (e.g., Cespón, 2021; Cespón & Carreiras, 2020; Grundy, Anderson, & Bialystok, 2017a).[2] However, a shorter latency may also be interpreted negatively if linked to worse behavioral results; in particular, it reflects inadequate processing.[3] For amplitude, interpretation is more complicated because there is often no intuitive way to decide whether a larger or smaller size[4] reflects better or worse functioning (e.g., Cespón, 2021; Cespón & Carreiras, 2020). Thus, additional information, behavioral or theoretical, is required to facilitate interpretation (Cespón, 2021; Cespón & Carreiras, 2020; García-Pentón et al., 2016; Paap et al., 2015; but see Bialystok, 2017: pp. 18-19; Grundy, Anderson, & Bialystok, 2017a; Kappenman & Luck, 2016: p. 114; Luck, 2014: pp. 142-143; Yeung et al., 2007: p. 354). For some ERPs, larger amplitudes reflect greater brain activity or effort (e.g., for the N2, N450, and ERN, see Botvinick et al., 2001; Carter & van Veen, 2007; see also Table 1). Thus, in the lack of evidence for behavioral effects or in presence of better behavioral results, smaller amplitudes show greater neural efficiency. This is because less neural resources are presumably used for the same or better behavioral outcome (e.g., Barulli & Stern, 2013; Bialystok, 2017; Gray et al., 2005). Second, for these same ERPs, larger amplitudes may also be interpreted positively if linked to better behavior. In this case, larger amplitudes suggest that more on-task neural effort likely underlies better behavior. Alternatively, they may show greater capacity (e.g., Barulli & Stern, 2013); that is, while one group is still able to devote neural effort to a difficult task, another group is overwhelmed by task demands and does not allocate enough resources. Of course, the latter interpretation applies only when there are behavioral differences because not allocating enough neural resources means that the task is not adequately performed (Gray et al., 2005).

For other components, however, greater difficulty is reflected in smaller amplitudes (amplitude suppression), suggesting that smaller amplitudes reflect more neural effort. Thus, smaller and larger amplitudes for these ERPs may be interpreted in the opposite manner, in the two scenarios above. In any case, commonly for both component types, a narrower[5] amplitude difference (i.e., smaller differentiation) between EC- (e.g., incongruent) and less-demanding (e.g., neutral) conditions, reflects greater efficiency, if there is evidence for better or no effect on behavior. This is because both amplitude enhancement and suppression in an EC condition result in a wider amplitude difference with the less-demanding condition (e.g., Grundy & Bialystok, 2018; Heidlmayr et al., 2015). In general, greater efficiency has been argued to provide another neural basis for reserve against brain deterioration (e.g., Barulli & Stern, 2013; Grundy, Anderson, & Bialystok, 2017a).

The N2 and P3, for instance, are two of the most-studied ERP components linked to EC. The N2 is a negative component that peaks about 200–350 ms after target stimulus, has a fronto-central distribution, and has been linked to EC processes such as response inhibition and conflict monitoring (e.g., Folstein & Van Petten, 2008). In EC tasks, the target-locked frontocentral N2 is typically larger for trials that require more EC, such as for incongruent compared to congruent, No-Go relative to Go, and switch compared to repeat trials (e.g., Folstein & Van Petten, 2008; Karayanidis & Jamadar, 2014). The frontocentral N2 (henceforth, N2) is thought to be generated in the anterior cingulate cortex (ACC) brain area (e.g., Folstein & Van Petten, 2008).

Moreover, the P3 is a positivity that peaks within 250–500 ms post-stimulus (Pires et al., 2014). It is typically divided into two sub-components: the frontal P3a that has been linked to response inhibition and the orienting of attention to novel or rare events; and the parietal P3b that has been linked to the updating of working memory and task complexity (Folstein & Van Petten, 2008; Gratton et al., 2018). The P3a may appear in tasks that require withholding a response, as, for example, in Go/No-Go or Stop-signal tests. It is larger for (more) EC-demanding targets, such as for No-Go/successful Stop compared to Go trials (e.g., Cespón & Carreiras, 2020; Pires et al., 2014). Other EC (e.g., switching, working memory, inhibition) tasks may elicit a P3b. The P3b is often smaller for EC targets, such as for switch compared to repeat, high relative to low working memory, and incongruent compared to no-conflict trials (Cespón & Carreiras, 2020; Gratton

---

[2] Theoretically, a delayed latency may also indicate a positive effect if linked to better behavioral results: More neural effort or cautious processing likely underlies better behavior. This possibility is not considered here to keep the discussion simpler and because no reviewed study reported this pattern.

[3] The same processing route may be efficient and effective in some cases – leading to a desired, accurate behavioral outcome with less effort – but ineffective – resulting in an unwanted or erroneous outcome – in other situations. For example, shallow, underspecified, or good-enough language processing may often result in fast, effortless, and successful communication. However, in other cases, the same processing route may lead to miscommunication (e.g., Christianson, 2016; Clahsen & Felser, 2018). Crucially, it is the end-result (successful communication or not) that determines whether the processing strategy may be interpreted as efficient and effective or as ineffective. Finally, insufficient or lack of neural processing may manifest as a complete absence of an ERP, in which case latency is irrelevant. However, insufficient neural processing should also be evident as a lower behavioral result.

[4] In this review, by larger and smaller amplitude, I mean more and less negative amplitudes for negative components, respectively; and more and less positive amplitudes for positive components, respectively.

[5] By narrower, I mean that the amplitude distance (differentiation) between two conditions is smaller, not that the difference number resulting from subtracting amplitude in one condition from another is smaller. For negative ERPs, for example, a narrower distance between a more negative incongruent and less negative neutral condition results in a larger difference number.

Table 1 Description (polarity, timing, scalp topography) and functional significance of the event-related potential (ERP) components investigated in the reviewed studies

ERP component	Description and function
Stimulus-locked components	
N1	• Negative deflection that occurs approximately within 100–200 ms post-stimulus. • The visual N1 is usually largest over the occipital region or inferior temporal sites (e.g., Key et al., 2005). • More negative in certain EC conditions, such as in No-Go compared to Go, incongruent Flanker compared to congruent/neutral trials, and in successful Stop-signal compared to failed stop trials (Hsieh & Fang, 2012; Mahé et al., 2014; Pires et al., 2014). However, these effects are not always reported (e.g., Johnstone et al., 2009; Key et al., 2005; Pires et al., 2014; Wild-Wall et al., 2008). • Linked to early attentional processing; particularly, discriminative processing and attentional orientation (Key et al., 2005; Luck, 2014; Pires et al., 2014). • Associated with EC-related processes such as facilitation or enhancement of relevant sensory information (Pires et al., 2014).
P2	• Positive deflection that appears around 150–250 ms post-stimulus at fronto-central sites (Kałamała, Szewczyk, et al., 2018b; Key et al., 2005; Luck, 2014). • Larger for incongruent compared to congruent trials and for switch compared to repeat trials, even though these effects are not always found (Gajewski et al., 2018;Kałamała, Szewczyk, et al., 2018b ; Rey-Mermet et al., 2019). • Thought to reflect a selective attention process engaged in the evaluation of a task-relevant stimulus (Kałamała, Szewczyk, et al., 2018b; Rey-Mermet et al., 2019). • Linked to EC-(switching-, working memory-, inhibition-)related processes (Gajewski et al., 2018; Kałamała, Szewczyk, et al., 2018b; Lijffijt et al., 2009; Rey-Mermet et al., 2019).
FCP	• Cue-locked positivity that appears around 200–300 ms post-cue, with a frontal distribution, in the Antisaccade task. • Smaller in anti- than prosaccade trials. • Considered to reflect preparatory processes linked to the decision to inhibit a response (Heidlmayr et al., 2016; Mueller et al., 2009).
N2	• Negative component that peaks around 200–350 ms post-stimulus (e.g., Folstein & Van Petten, 2008). • Comprises three different sub-components, but only one of them is sensitive to EC. • The control N2 sub-component has a fronto-central distribution (Folstein & Van Petten, 2008). • Typically, larger for EC-demanding trials, such as for switch compared to repeat trials in mixed switching blocks, incongruent compared to congruent/neutral trials; and for trials that involve withholding a response compared to trials that require responding, as in the Go/No-Go task (Folstein & Van Petten, 2008; Jamadar et al., 2015; Karayanidis & Jamadar, 2014; see also in Gajewski et al., 2018). • In switching tasks, may be larger for repeat trials in pure blocks compared to the more-demanding repeat trials in mixed blocks (see in Gajewski et al., 2018). • Linked to EC processes, such as response inhibition, conflict resolution, conflict and error monitoring (Folstein & Van Petten, 2008; Larson et al., 2014).
P3	• Positivity (larger than the N2) that peaks approximately 250–500 ms after stimulus onset, although its latency can vary depending on the task, stimuli, and participant characteristics (Pires et al., 2014; Polich, 2007). • Typically divided into two sub-components that have different scalp topographies, latencies, and functional correlates: The frontally maximal and shorter in latency P3a that may reflect response inhibition and the allocation of attention to novel or rare events; and the later in latency P3b that has a parietal maximum and may reflect the updating of working memory and task complexity (Folstein & Van Petten, 2008; Luck, 2014; Polich, 2012). • The P3a appears in tasks that require withholding a response, such as in Go/No-Go or Stop-signal tests. • The P3a is larger for (more) EC-demanding trials, such as for No-Go/Stop compared to Go trials (e.g., Pires et al., 2014). • The P3b may appear in other EC (switching, working memory, inhibition) tasks. • The P3b is often smaller for EC-demanding target stimuli, such as for switch compared to repeat trials, mixed compared to pure blocks, high compared to low working memory; and incongruent compared to no-conflict trials (e.g., Cespón & Carreiras, 2020; Gajewski et al., 2018; Gratton et al., 2018; Karayanidis & Jamadar, 2014). • The centro-parietal Flanker P3 is larger for incongruent compared to congruent trials and may reflect response inhibition or the recruitment of attentional control (Clayson & Larson, 2011a, 2011b; see also Larson et al., 2016). • Modulations of a parietal positivity in EC tasks may not always reflect a P3b (Gratton et al., 2018). • P3-like positivities also appear at cue onset in cued EC tasks, such as in switching tasks where a cue indicates the task for the upcoming trial or whether a switch or repeat trial is upcoming. These have been linked to proactive control processes. • In switching tasks, cue-locked P3-like responses are larger for switch than repeat trials in mixed blocks; and for repeat trials in mixed blocks than for repeat trials in pure blocks (Jamadar et al., 2015).

Table 1 (continued)

ERP component	Description and function
N450 (or N_{inc} or N400)	• Negative deflection that peaks approximately 450 ms post-stimulus, with a fronto-central distribution (but extending to parietal sites) in the Stroop task. • Larger for incongruent relative to no-conflict trials (Folstein & Van Petten, 2008; Pires et al., 2014). • Has been interpreted as an index of inhibition processes related to suppressing word information; and as a measure of response conflict or conflict detection (Pires et al., 2014).
LPC	• The LPC is a positivity that occurs within 500–800 ms post-stimulus, with a centro-parietal distribution (Davidson & Pitts, 2014; Ergen et al., 2014; Luck, 2014; Pires et al., 2014). • Often described as an extension of the P3 (Davidson & Pitts, 2014; Polich, 2012). May have the same onset as the P3 and extend for several milliseconds (Gevins & Smith, 2000; Luck, 2014). • In the Stroop task, it is typically larger for incongruent compared to congruent trials, while in proactive interference paradigms (Sternberg's working memory task; Sternberg, 1966) it reduces in amplitude for the more difficult recent negative probe trials (Pires et al., 2014; Zhang et al., 2010). • Linked to EC-(working memory-, inhibition-)related processes. Pires et al. (2014) suggest that it corresponds to a process involved in proactive interference resolution and, during the Stroop task, reflects perceptual conflict and semantic processing of word meaning.
LPP	• In switching tasks, the LPP appears approximately 500–600 ms after target. • Larger for switch than repeat trials. • Thought to indicate anticipatory reconfiguration of task set, attentional shifting to or the biasing of attention to the task set that is currently relevant (Mueller et al., 2009).
LNP	• The LNP is a negative deflection that follows the N450 and appears approximately 500 ms post-stimulus, with a fronto-central distribution in the Stroop task (e.g., Heidlmayr et al., 2015; Larson et al., 2009; Naylor et al., 2012; West & Alain, 1999). • More negative for incongruent compared to congruent trials. • This may reverse in polarity over the centro-parietal scalp, in that amplitude is more positive for incongruent compared to congruent trials (e.g., Heidlmayr et al., 2015; Larson et al., 2009; Larson et al., 2014); and is called the conflict slow potential (Larson et al., 2014). • Associated with response selection, conflict resolution, conflict monitoring, or a process that signals the need for EC adjustment to improve task performance (Larson et al., 2014; Naylor et al., 2012).
Response or saccade-locked components	
PSP	• In the Antisaccade task, the PSP is an inhibition-related component occurring approximately 250-50 ms before a saccade. • Has central scalp topography. • Reduced in the antisaccade than the prosaccade condition (e.g., Evdokimidis et al., 1996).
ERN	• The ERN or Ne has an onset at or shortly before an erroneous response; and a peak approximately 100 ms later. • Has a fronto-central distribution (Gehring et al., 2012). • Larger for erroneous than for correct responses. • May be larger for congruent than for incongruent trials (e.g., Botvinick et al., 2001; Gehring et al., 2012). • Thought to index processes such as (a) error detection – for example, the process of using the error signal to adjust and improve performance or the process or the output of the process of comparing the actual ongoing erroneous motor response with the intended correct response; (b) conflict monitoring or response conflict – that is, activation of or detection of two conflicting responses that signals poor performance and the need for increased control; (c) reinforcement learning – that is, mismatch between learned, expected, correct values and an incorrect value of stimulus-response combinations that results in an error, negative-reinforcement signal to improve performance; (d) more generally, a process that evaluates the need for or that is involved in implementing EC (Folstein & Van Petten, 2008; Gehring et al., 2012; Larson et al., 2014).
CRN	• Negative deflection associated with correct responses. • Appears at the same time window and has the same scalp topography to the ERN. • Smaller than ERN; that is, errors linked to larger negativity (ERN) than correct responses (CRN; e.g., Gehring et al., 2012). • May reflect that participants are uncertain for their accuracy during correct responses or that they mistakenly perceive correct responses as errors or (a similar to the ERN) conflict-monitoring process (Gajewski et al., 2018; Gehring et al., 2012; Gratton et al., 2018).
Pe	• Positive component that follows the ERN, approximately 200–400 ms after an erroneous response. • Has a centro-parietal scalp topography. • Larger for errors compared to correct responses (e.g., Clawson et al., 2017). • May indicate participants' awareness of their errors, an affective response to an error, the detection or evaluation of an erroneous response; or it may be involved in post-error response strategy adaptation (Gehring et al., 2012).

EC executive control, *FCP* frontal cue-locked positivity, *LPC* late positive component or late positive complex, *LPP* late parietal positivity, *LNP* late sustained negative-going potential, *PSP* presaccadic positivity, *ERN* error-related negativity, *Ne* error negativity, *CRN* correct-related negativity, *Pe* error positivity

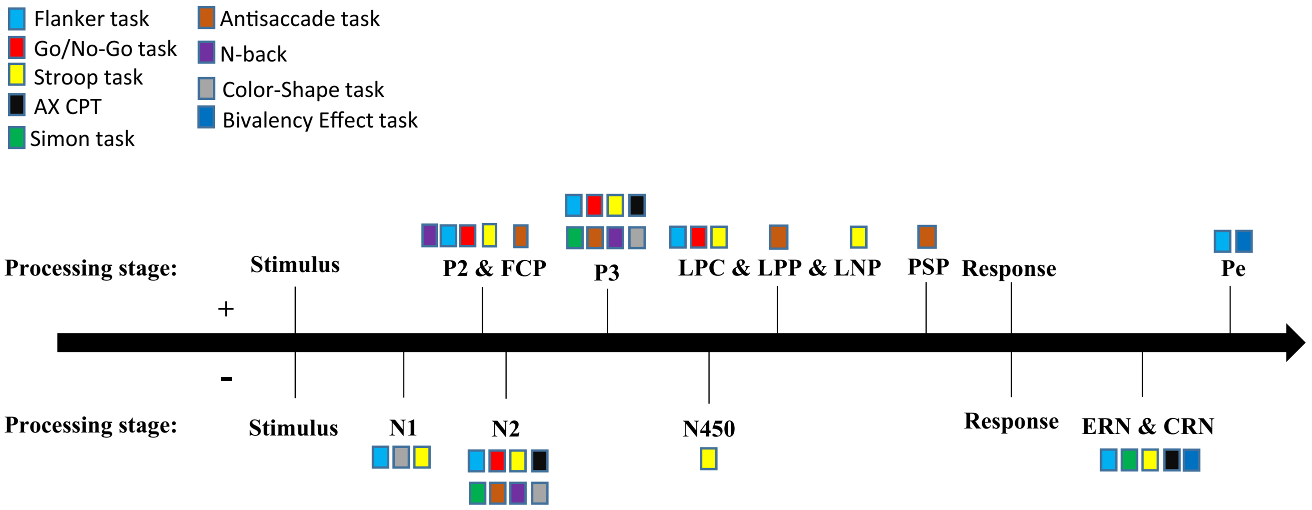

Fig. 7 Timeline of events and of the ERP components investigated in the bilingualism literature on executive control using different paradigms. FCP = frontal cue-locked positivity, LPC = late positive component or late positive complex, LPP = late parietal positivity, LNP = late sustained negative-going potential, PSP = presaccadic positivity, ERN = error-related negativity, CRN = correct-related negativity, Pe = error positivity

et al., 2018; Karayanidis & Jamadar, 2014). However, Clayson and Larson (2011a, 2011b, Larson et al., 2016) suggest that the centro-parietal Flanker P3 is larger for incongruent compared to congruent trials; and may reflect response inhibition and/or the recruitment of attentional control. Thus, for the latter set of EC tasks, it is not clear whether modulations of parietal positivities reflect the same component (e.g., P3b) or whether difficulty increases or reduces the P3 (see also Gratton et al., 2018; Luck, 2014).[6] P3-like ERPs are also elicited at cue onset, such as in cued switching tasks where a cue indicates if the upcoming target requires a switch or repeat. These are typically linked to proactive control processes (Braver, 2012; Gratton et al., 2018; Karayanidis & Jamadar, 2014; see next section).

In the next section, I summarize different EC theories and accounts on the specific locus of the bilingual EC effect. Moreover, based on these EC theories and accounts, and on considerations of neural efficiency, I articulate predictions on how a positive bilingual effect on EC may manifest in ERPs. I then review the methods and results of studies that used ERPs. In this review, I consider not only ERPs that are established EC markers (e.g., N2, P3; e.g., Downes et al., 2017) but also other ERPs from 100 ms post-stimulus (e.g., N1) until later stages of stimulus (500–800 ms; e.g., late sustained negative-going potential) and response processing (e.g., Pe). These ERPs have been also linked in some cases to attention or EC (e.g., Pires et al., 2014; Luck, 2014; Luck & Kappenman, 2012; van Veen & Carter, 2006; see Table 1). Also, early processing possibly has consequences for later EC (e.g., Grundy, Anderson, & Bialystok, 2017a; Luck & Kappenman, 2012; Pires et al., 2014; Roche et al., 2005).

Executive control theories and accounts of the positive bilingual effect on neuro-cognitive executive control

There are different accounts about which EC aspects are impacted by bilingualism. Some accounts were proposed to explain bilingual effects on the brain (e.g., Grundy, Anderson, & Bialystok, 2017a) or draw on EC theories with a clear neuroscientific understanding (e.g., Hilchey & Klein, 2011). For these, I articulate, where appropriate, predictions about which specific ERPs should be influenced. Other accounts, however, are based on behavioral EC models (e.g., Miyake et al., 2000) and/or were proposed to explain the behavioral effects of bilingualism in EC tasks (e.g., Houtzager et al., 2017; Prior & MacWhinney, 2010; Santillán & Khurana, 2018). These accounts as such offer no clear theoretical basis on which to predict the specific neural effects of bilingualism. Thus, for these accounts, I form no predictions about which specific ERPs should be affected. In this case, however, bilingual effects on a given process are expected to appear in ERPs across tasks and conditions tapping into that same function. A bilingual inhibition effect, for example, should appear in ERPs for incongruent trials across inhibition tasks. Table S2 in the Online Supplementary Material

---

[6] This review refers to the larger P3 when withholding a response (e.g., No-Go trials) as “P3a” and to other positivities (250–500 ms) as “P3”. It distinguishes the latter based on task and event (target or cue).

(OSM) presents predictions on how a positive bilingual effect may appear in ERPs, with reference to specific components, ERP measures, tasks, task events, and trial types, as appropriate. These predictions are based on considerations of neural efficiency, different EC theories, and accounts regarding the specific locus of the bilingual EC effect. Table S2 also states if predicted bilingual ERP effects reflect a benefit or more efficiency (= positive effect) depending on the presence of a bilingual behavioral effect or not; and, for amplitude, based on the direction of the bilingual effect.

Grundy, Anderson, and Bialystok (2017a) proposed the Bilingual Anterior to Posterior and Subcortical Shift (BAPSS) model. This maintains that bilinguals show greater neural efficiency; specifically, they exhibit more automatic processing and devote more resources earlier in EC tasks. This reduces the need for later effortful processing. For ERPs, this manifests as larger and earlier stimulus-locked ERPs such as N2 and P3; and smaller later ERPs such as stimulus-locked N450, late negative-going potential (LNP), and response-locked ERN.

A second account draws on conflict-monitoring theory (Botvinick et al., 2001; Botvinick et al., 2004) and suggests that the positive bilingual effect lies in monitoring (Costa et al., 2009; Hilchey & Klein, 2011). Monitoring is an evaluative EC system, supported by ACC areas, that is responsible for assessing and signalling the need for regulative top-down EC through the detection of conflict or effort in general. Conflict detection, in turn, subsequently leads to regulative EC use (e.g., inhibition or attentional focus) to resolve conflict, implemented by brain areas in the lateral prefrontal cortex (Botvinick et al., 2004; Carter & Van Veen, 2007; Larson et al., 2014). Conflict is typically present in EC paradigms such as the Stroop and Flanker tasks but can also occur in other EC tests, such as in Go/No-Go (between Go and No-Go responses; e.g., Botvinick et al., 2001; Nieuwenhuis et al., 2003; Pires et al., 2014), working memory (e.g., non-target lure trials in the N-back task, where the current stimulus appeared recently but not in the correct sequence position; Braver et al., 2007), and switching tasks (e.g., between the relevant and irrelevant task set for switch trials; Gajewski et al., 2018; Jamadar et al., 2015; Karayanidis & Jamadar, 2014). Monitoring is thought to be bolstered in bilinguals because of the constant need to monitor the appropriate language to use with speakers of different languages (Costa et al., 2009), or due to regular experience of monitoring conflict between translation-equivalent words in two languages, which are simultaneously active when speaking in one language (Hilchey & Klein, 2011).

Within the bilingualism literature, the monitoring account was originally offered to explain behavioral results from inhibition and switching tests showing a bilingual RT benefit across trials within mixed, EC-demanding blocks, rather than specifically for incongruent or switch trials (e.g., Costa et al., 2009; Hilchey & Klein, 2011; Lehtonen et al., 2018). Thus, a first prediction from this account is that target-locked ERP latency or amplitude should be affected by bilingualism for all trials in mixed blocks, such as for both incongruent and congruent trials in inhibition tasks (e.g., Cespón & Carreiras, 2020). This positive bilingual effect may appear as shorter latencies linked to bilingualism if there is evidence for better bilingual or no effect on behavior. Also, the direction of a positive bilingual effect on amplitude (larger or smaller) depends on component – whether amplitude increases or reduces with difficulty – and the presence of behavioral differences or not (see Table S2 (OSM)).

Additional predictions arise from conflict-monitoring theory as a general neuro-cognitive framework of EC (Botvinick et al., 2001; Botvinick et al., 2004). First, the theory has offered accounts for the N2, N450, and ERN; specifically, it assumes that the N2 and N450 are functionally equivalent (Botvinick et al., 2001; Carter & van Veen, 2007). Moreover, it proposes an integrative explanation of the N2/N450 and ERN (van Veen & Carter, 2002; Yeung et al., 2004): Both indicate ACC-based conflict detection, with more negative amplitudes showing greater ACC activity. The N2/N450 reflects conflict during stimulus processing, before a correct response; and is determined by attention to irrelevant information. For example, the N2 becomes larger with more attention to flankers in the Flanker task. In contrast, the response-locked ERN corresponds to conflict after an error: Continuous processing of relevant target information (e.g., central arrow in Flanker task), after an error, leads to the post-error activation of the correct response and, hence, to a transient period of conflict between the correct and error responses. Thus, any effects on monitoring should be equally evident on both N2/N450 and ERN; and, for amplitude, in the same direction. For instance, less negative N2/N450 for demanding trials and less negative ERN show greater monitoring efficiency if there is evidence for better bilingual or no difference in behavior (see Table S2 (OSM), for more predictions).[7]

Furthermore, monitoring theory proposes that conflict detection leads to high regulative EC that reduces conflict for upcoming trials. These conflict-driven adjustments in

---

[7] Predictions concerning how a positive bilingual effect may manifest in the *direction* of amplitude size (smaller or larger) of a single or a combination of ERPs (e.g., for this prediction, less negative N2 and ERN) should hold (1) in within-subjects designs (e.g., for this prediction, both less negative N2 and ERN are expected if the two ERPs are examined within subjects in the same task); and (2) across independent samples, assuming no bilingual behavioral effect. The direction of the bilingual effect on amplitude may differ across independent samples if the bilingual behavioral effect differs or if all different samples report better bilingual behavior (e.g., in the latter case, less and more negative N2 in different samples may both be interpreted positively, as more efficient and better EC, respectively).

Table 2 List of the reviewed studies and summary and interpretation of their ERP results based on considerations of neural efficiency, different executive control theories, and different accounts regarding the locus of the bilingual executive control effect (apart from the Bilingual Anterior to Posterior and Subcortical Shift model)

		Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation[a]
1	Kousaie & Phillips, 2012	Stroop	N2	Mean amplitude	No	Less negative N2 across neutral, congruent, and incongruent trials	No	More efficient evaluative/reactive control (monitoring)
			P3	Mean amplitude	I<C=N	No		No positive effect
				Peak latency	I=C<N	Earlier P3 latency across neutral, congruent, and incongruent trials		More efficient evaluative/reactive control (monitoring)
			ERN	Mean amplitude	No	Less negative ERN across congruent and incongruent trials		More efficient evaluative/reactive control (monitoring) More efficient attention disengagement
		Simon	N2	Mean amplitude	I=C<N	No		No positive effect
			P3	Mean amplitude	No	Smaller P3 across neutral, congruent, and incongruent trials		Ambiguous
				Peak latency	I>C=N	No		No positive effect
			ERN	Mean amplitude	No	No		No positive effect
		Flanker	N2	Mean amplitude	I>C=N	No		No positive effect
			P3	Mean amplitude	I=C<N	No		No positive effect
				Peak latency	I>C=N	Smaller delay in P3 latency for incongruent trials relative to neutral and congruent trials		More efficient regulative/reactive control (inhibition)
			ERN	Mean amplitude	No	More negative amplitude for neutral and incongruent trials (0–40 ms post-error) and less negative ERN for congruent trials (80–100 ms post-error) Wider ERN difference between incongruent and congruent/neutral trials (only between 40–100 ms post-error and only at sites FCz and Cz)		Ambiguous[b]

Table 2 (continued)

		Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation[a]
2	Festman & Münte, 2012	Flanker	ERN	Mean amplitude	Incor.>Cor.	Less negative ERN (for non-switchers)	Yes. Bilingual non-switchers had better performance in the Wisconsin Card Sort and Flanker tasks; particularly, in the Flanker task (for which ERPs were recorded), they showed smaller RT interference effect and corrected their errors more often	More efficient evaluative/reactive control (monitoring) for bilingual non-switchers More efficient attention disengagement for bilingual non-switchers
3	Fernandez, Tartar, Padron, & Acosta, 2013	Auditory Go/No-Go	N2	Mean amplitude	No	More negative N2 for No-Go trials	No	Less efficient reactive control (monitoring or inhibition)
			P3a	Difference between mean N2 amplitude and the mean amplitude of the 300–440 ms time window	No-Go>Go	No		No positive effect
4	Fernandez, Acosta, Douglass, Doshi, & Tartar, 2014	Auditory Go/No-Go	N2	Difference between mean N2 peak amplitude in the 236–316 ms time window and the P3 mean peak amplitude in the 180–260 ms time window.	No	More negative N2 for No-Go trials	No	Less efficient reactive control (monitoring or inhibition)
		Visual Go/No-Go	N2	Difference between mean N2 peak amplitude in the 236–316 ms time window and the P3 mean peak amplitude in the 180–260 ms time window.	No-Go>Go	No		No positive effect
5	Moreno, Wodniecka, Tays, Alain, & Bialystok, 2014	Visual Go/No-Go	P2	Peak amplitude	No	No	No	No positive effect
				Peak latency	No	No		No positive effect
			N2	Mean amplitude	No-Go>Go	More negative N2 for No-Go trials		Less efficient reactive control (monitoring or inhibition)
			P3a	Mean amplitude	No-Go>Go	No		No positive effect
			LPC	Mean amplitude	No	No		No positive effect[c]

Table 2 (continued)

		Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation[a]
6	Kousaie & Phillips, 2017	Stroop	N2	Mean integrated amplitude	No	No	Yes. Higher accuracy and faster RTs for incongruent trials in the Stroop task + Higher accuracy (across congruent and incongruent trials) in the Flanker task	No positive effect
				Mean peak amplitude	No	No		No positive effect
				Peak latency	No	Earlier N2 latency across congruent and incongruent trials		Advantage in evaluative/reactive control (monitoring)
			P3	Mean integrated amplitude	I<C	Larger P3 amplitude for incongruent trials (indicating a narrower difference between incongruent and congruent trials)		Advantage in regulative/reactive control (inhibition)
				Mean peak amplitude	I<C	No		No positive effect
				Peak latency	No	No		No positive effect
		Simon	N2	Mean integrated amplitude	No	Less negative N2 across congruent and incongruent trials		More efficient evaluative/reactive control (monitoring)
				Mean peak amplitude	I>C	No		No positive effect
				Peak latency	I>C	No		No positive effect
			P3	Mean integrated amplitude	No	Larger P3 amplitude across congruent and incongruent trials		Advantage in evaluative/reactive control (monitoring)
				Mean peak amplitude	No	No		No positive effect
				Peak latency	I>C	Earlier P3 latency across congruent and incongruent trials		Advantage in evaluative/reactive control (monitoring)
		Flanker	N2	Mean integrated amplitude	I>C	No		No positive effect
				Mean peak amplitude	I>C	No		No positive effect
				Peak latency	No	Earlier N2 latency across congruent and incongruent trials		Advantage in evaluative/reactive control (monitoring)
			P3	Mean integrated amplitude	No	No		No positive effect
				Mean peak amplitude	No	No		No positive effect
				Peak latency	I>C	Earlier P3 latency across congruent and incongruent trials		Advantage in evaluative/reactive control (monitoring)

Table 2 (continued)

		Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation[a]
7	Jiao, Zhang, Plummer, Liu, & Chen, 2019	Spatial Stroop	N1	Mean amplitude	No	No	No	No positive effect
			P2	Mean amplitude	No	No		No positive effect
			N2	Mean amplitude	I>C	Larger N2 across congruent and incongruent trials with more L2 proficiency		Less efficient evaluative/reactive control (monitoring) linked to higher L2 proficiency
			P3	Mean amplitude	No	More positive P3 across congruent and incongruent trials with more L2 proficiency		Ambiguous
8	Barac, Moreno, & Bialystok, 2016	Visual Go/No-Go	N2	Mean amplitude	Not explicitly reported	No	Yes. Higher accuracy for congruent and incongruent trials in Flanker task + Higher accuracy (across Go and No-Go trials), faster RTs (for Go trials), and a higher *d* prime discriminability index (indicating better discrimination of Go from No-Go stimuli) in the Go/No-Go task	No positive effect
				Peak latency	Not explicitly reported	Shorter N2 latency across Go and No-Go trials		Advantage in evaluative/reactive control (monitoring)
			P3a	Mean amplitude	Not explicitly reported	Larger P3 across Go and No-Go trials		Advantage in evaluative/reactive control (monitoring)
				Peak latency	Not explicitly reported	Shorter P3 latency for Go trials or across Go and No-Go trials depending on electrode sites		Advantage in evaluative/reactive control (monitoring)
9	Coderre & van Heuven, 2014	Stroop with two SOAs. -400 ms SOA: word (irrelevant stimulus) appeared 400 ms before the color (relevant stimulus). 0 ms SOA: word and color appeared simultaneously.	N450	Mean amplitude difference (I-C) and mean amplitude (for neutral trials).	Mono: I>C (Cz, Pz, both SOAs) I>N (Cz, Pz, both SOAs) Bil. L1: I>C (Pz, both SOAs) I>N (Cz, both SOAs) Bil. L2: I>C (Cz, Pz, both SOAs) I>N (Pz for 0 ms SOA and Pz, Cz for -400 ms SOA)	Less positive waveform (for L1 and L2) for neutral trials within 200-450 ms post-stimulus (0 ms SOA condition); and within 200–400 ms post-stimulus (N450 window, -400 ms SOA condition)	Yes. Faster for neutral trials and higher accuracy across neutral, congruent, and incongruent trials in both testing languages[d]	No positive effect[e]
				Peak amplitude difference (I-C)	Not explicitly reported	No		No positive effect
				Peak latency of difference waves (I-C & I-N)	Not explicitly reported	No		No positive effect

Table 2 (continued)

		Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation[a]
10	Hannaway, Opitz, & Sauseng, 2019	Bilingual Stroop with a language condition (English or German) and a similarity condition (color words were either similar or dissimilar between bilinguals' two languages).	N450	Mean amplitude	I>N C>N	More negative N450 across neutral, congruent, and incongruent trials for bilinguals who had a higher L2 proficiency and an earlier mean age of L2 acquisition (but were not immersed in the L2)	Yes. Smaller RT interference effect across conditions and, particularly, for the most difficult German language condition for bilinguals with a higher L2 proficiency and an earlier L2 acquisition relative to bilinguals who were immersed in their L2	Advantage in evaluative/reactive control (monitoring) for bilinguals who had a higher L2 proficiency and an earlier mean age of L2 acquisition (but were not immersed in the L2)
			LPC	Mean amplitude	I>N	No		No positive effect
11	Heidlmayr, Hemforth, Moutier, & Isel, 2015	Stroop with negative-priming condition.	N2	Mean amplitude	NP>C at Cz (midline electrodes)[f]	No[g]	No	No positive effect
			N450	Mean amplitude	I>C (lateral electrodes) I>C at Pz (midline electrodes) NP>C at posterior electrodes (lateral electrodes) NP>C (midline electrodes)	Narrower N450 effect (=more negative N450 for incongruent than congruent trials)		More efficient reactive control (monitoring or inhibition)
			LNP	Mean amplitude	I>C at Cz NP>C at anterior electrodes but NP<C at posterior electrodes (lateral electrodes) NP>C (midline electrodes)	Narrower LNP effect (=more negative incongruent than congruent trials)		More efficient regulative/reactive control (inhibition)
12	Heidlmayr, Dore-Mazars, Aparicio, & Isel, 2016	Antisaccade	Cue-locked positivity (150–250 ms)	Mean amplitude	AS<PS (lateral & midline electrodes)[f]	Narrower difference in positivity between anti- and prosaccade trials	No	More efficient proactive control
			N2 (target-locked)	Mean amplitude	AS>PS (lateral electrodes)	Wider N2 difference between anti- and prosaccade trials[h]		Less efficient reactive control (monitoring or inhibition)
			P3 (target-locked)	Mean amplitude	AS<PS (lateral electrodes)	Narrower difference in P3 between anti- and prosaccade trials		More efficient regulative/reactive control (inhibition)
			LPP (target- locked)	Mean amplitude	SAS>RAS (lateral electrodes) SPS>RPS (lateral electrodes)	No		No positive effect

Table 2 (continued)

		Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation^{a}
			PSP (saccade-locked)	Mean amplitude	AS<PS (lateral electrodes)	Narrower difference in PSP between anti- and prosaccade trials		More efficient regulative/reactive control (inhibition)
13	Morales, Yudes, Gomez-Ariza, & Bajo, 2015	AX-CPT	P3 (cue-locked)	Mean amplitude	B cues>A cues	No^{i}	Yes. Higher accuracy for AX, BX, and AY trials	No positive effect
			N2 (probe-locked)	Mean amplitude	No	More negative N2 for AY trials		Advantage in reactive control (monitoring or inhibition)
			P3a (probe-locked)	Mean amplitude	AY trials>BX=BY	Larger P3 difference between AY and BY trials		Advantage in regulative/reactive control (inhibition)
			ERN	Mean amplitude	Incor.> Cor.	Narrower ERN difference between incorrect and correct responses across all trials		More efficient evaluative/reactive control (monitoring) More efficient attention disengagement
14	Wu, Zhang, & Guo, 2016	Flanker	P2	Mean amplitude	I>N C>N	Smaller P2 amplitude for neutral trials^{j}	No	No positive effect
				Peak latency	I<N C<N	Shorter P2 latency across congruent and incongruent trials		More efficient evaluative/reactive control (monitoring)
			N2	Mean amplitude	I>N I>C	No		No positive effect
			LPC	Mean amplitude	No	Larger LPC across trials		Ambiguous
15	Kałamała, Drożdżowicz, Szewczyk et al., 2018	Lateralized Flanker (arrows pointing upwards or downwards) with Visual Field Condition (arrows appearing either at the left or at the right side of a fixation cross).	P3^{k}	Mean amplitude	I<C	No	Yes. Higher accuracy across congruent and incongruent trials and, particularly, in the incongruent condition; and the most difficult Right Visual Field trials of the incongruent condition	No positive effect
			ERN/CRN	Mean amplitude	Incor.>Cor.	More negative ERN and CRN across trials		Advantage in evaluative/reactive control (monitoring) Less efficient attention disengagement
			Pe	Mean amplitude	Incor.>Cor.	No		No positive effect
16	Timmer, Grundy, & Bialystok, 2017	Color-Shape	N2 (switch effect)	Mean amplitude	S<R	Wider N2 effect (repeat-switch trials)	Yes. Higher accuracy across repeat and switch trials in mixed blocks	Advantage in reactive control (monitoring or switching) Less efficient attention disengagement
			P3/LPC (switch effect)	Mean amplitude	S>R	No		No positive effect
			N1 (mixing cost)	Mean amplitude	R<P	Condition effect on N1 (pure-repeat trials) was observed at more scalp sites		Ambiguous^{l}

Table 2 (continued)

	Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation^{a}
		N2 (mixing cost)	Mean amplitude	R<P at frontal regions for both groups, R<P at the occipital region only for bilinguals	Condition effect on N2 (pure-repeat trials) was observed at more scalp sites		Ambiguous^{l}
		P3/LPC (mixing cost)	Mean amplitude	R<P	Condition effect on P3/LPC (pure-repeat trials) was observed at more scalp sites		Ambiguous^{l}
17 Grundy, Chung-Fat-Yim, Friesen, Mak, & Bialystok, 2017b	Flanker	N2	Mean amplitude	I>C iI>iC	Narrower N2 difference between incongruent and congruent trials, after incongruent trials^{m}	No	More efficient attention disengagement
		P3	Mean amplitude	I>C cI>cC	Smaller P3 difference between incongruent and congruent trials, after congruent trials^{m}		More efficient attention disengagement
18 Grundy & Bialystok, 2018	Bivalency effect	Negativity (110–130 ms)	Mean amplitude	UC>UP	No	No	No positive effect
		Negativity (300–400 ms)	Mean amplitude	UC<UP	No		No positive effect
		Negativity (400–800 ms)	Mean amplitude	No	Narrower negativity difference between univalent trials in conflict blocks and univalent trials in pure blocks		More efficient attention disengagement
		ERN/CRN	Mean amplitude	Incor.>Cor.	Less negative ERN across trials in conflict blocks		More efficient evaluative/reactive control (monitoring) More efficient attention disengagement
		Pe/Pc	Mean amplitude	Incor.>Cor.	Less positive Pe across trials in conflict blocks		More efficient attention disengagement
19 Morrison, Kamal, & Taler, 2019	N-back	P2	Peak amplitude	2>1=0	No	No	No positive effect
			Peak latency	2=1>0	No		No positive effect
		N2	Peak amplitude	2<1<0	No		No positive effect
			Peak latency	No	No		No positive effect
		P3	Mean amplitude	2=1>0	Larger P3 across trials		Less efficient reactive control (working memory/updating)

Table 2 (continued)

		Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation[a]
20	Barker & Bialystok, 2019	Emotion N-back (target letters superimposed over emotional stimuli).	P3	Mean amplitude	2<1 T>NT Ang.>Neut.	Smaller P3 for target 2-back trials (wider difference between 2-back and 1-back target trials)	Yes. Slower for 2-back trials + More accurate for angry trials	Disadvantage in reactive control (working memory/updating)
				Peak latency	T>NT	Shorter P3 latency for emotion (happy and angry) trials		No positive effect (on executive control)
		Control N-back	P3	Mean amplitude	2<1 T>NT	No	No	No positive effect
				Peak latency	T>NT	No	No	No positive effect
21	Dong & Zhong, 2017 (Experiment 1)	Flanker	N1	Mean amplitude	No	More negative N1 across congruent and incongruent trials for the more-IE group	Yes. More-IE group had a smaller RT interference effect	More interpreting experience linked to an advantage in evaluative/reactive control (monitoring)
			N2	Mean amplitude	I>C	More negative N2 across congruent and incongruent trials for the more-IE group		More interpreting experience linked to an advantage in evaluative/reactive control (monitoring)
			P3	Mean amplitude	I<C	Smaller P3 across congruent and incongruent trials for the more-IE group (320-520 ms time window, electrodes F5, F3, F1, F2, FC3, FC1, FCz, FC2, FC4, FC6)		For the whole P3 time window (320-520 ms): More interpreting experience linked to an advantage in evaluative/reactive control (monitoring) for most electrodes but linked to an advantage in regulative/reactive control (inhibition) for one electrode For the early P3 time window (320-440 ms): More interpreting experience linked to an advantage in evaluative/reactive control (monitoring) For the late P3 time window (440-520 ms): More interpreting experience has a positive effect on regulative/reactive control (inhibition) for most electrodes but on evaluative/reactive control (monitoring) for three electrodes

Table 2 (continued)

		Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation[a]
					I<C	Smaller P3 for incongruent trials (320-520 ms time window, electrode F4), indicating a wider P3 difference between incongruent and congruent trials for the more-IE group		
					I<C	Smaller P3 across congruent and incongruent trials for the more-IE group (320-440 ms time window, electrodes AF4, F5, F3, F1, F2, Fz, FC3, FC1, FCz, FC2, FC4, T8)		
					I>C	Smaller P3 for incongruent trials for the more-IE group (440-520 ms time window, electrodes AF4, Fz, F4, F6, FCz, FC2, FC4, C4, C6, T8)		
					No	Smaller P3 across congruent and incongruent trials for the more-IE group (440-520 ms time window, electrodes F5, F3, FC3)		
21	Dong & Zhong, 2017 (Experiment 2)	Flanker	N1	Mean amplitude	No	More negative N1 across congruent and incongruent trials for the more-IE group	No	Ambiguous
			N2	Mean amplitude	I>C	More negative N2 across congruent and incongruent trials for the more-IE group		More interpreting experience linked to less efficient evaluative/reactive control (monitoring)

Table 2 (continued)

	Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation[a]
		P3	Mean amplitude	I<C	Smaller P3 for incongruent trials (320-520 ms time window), indicating a wider P3 difference between incongruent and congruent trials for the more-IE group		For the whole P3 time window (320-520 ms): More interpreting experience linked to less efficient regulative/reactive control (inhibition) For the early P3 time window (320-440 ms): Ambiguous For the late P3 time window (440-520 ms): Ambiguous
				Not explicitly reported	Smaller P3 across congruent and incongruent trials for the more-IE group (320-440 ms time window, FT8, T8)		
				Not explicitly reported	Smaller P3 for incongruent trials for the more-IE group (320-440 ms time window, electrodes F4, F6, CP4)		
				Not explicitly reported	Smaller P3 for incongruent trials for the more-IE group (440-520 ms time window, electrodes FP1, FP2, AF4, Fz, F2, F4, F8, FC4, FT8, T8)		
22 Sullivan, Janus, Moreno, Astheimer, & Bialystok, 2014	Visual Go/No-Go	N2	Peak amplitude	No-Go>Go	No	No	No positive effect
			Peak latency	No	N2 latency reduced across Go and No-Go trials from pre-test to post-test, after intensive six-month introductory course in L2 (no such effect in control group)		L2 training linked to more efficient evaluative/reactive control (monitoring)
		P3a	Peak amplitude	No-Go>Go	P3 amplitude increased across Go and No-Go trials from pre-test to post-test, after intensive six-month introductory course in L2 (no such effect in control group)		L2 training linked to less efficient evaluative/reactive control (monitoring)

Table 2 (continued)

	Task	ERP component	Measure	Trial effect on ERP measure	Bilingual neural effect	Bilingual behavioral effect	Interpretation[a]
			Peak latency	No	P3 latency reduced for No-Go trials from pre-test to post-test, after intensive six-month introductory course in L2 (no such effect in control group)		L2 training linked to more efficient regulative/reactive control (inhibition)
23 Zhang, Kang, Wu, Ma, & Guo, 2015	AX-CPT	N2 (cue-locked)	Mean amplitude	Not explicitly reported	N2 amplitude increased from pre-test to post-test across trials, after 10 days of training sessions in language switching (no such effect in control group)	Yes. Significant increase in the (i.e., more positive) Behavioral Shift Index (showing preference for proactive control), after 10 days of training sessions in language switching (no such effect in control group)	Language-switching training linked to an advantage in proactive control
		N2 (probe-locked)	Mean amplitude	Not explicitly reported	No		No positive effect
24 Chen, Bobb, Hoshino, & Marian, 2017	Flanker with No-Go trials.	N1	Mean amplitude	No	No	No	No positive effect
		P2	Mean amplitude	Difference between No-Go and neutral trials	Wider P2 difference between congruent and neutral trials; and between No-Go and neutral trials		Less efficient regulative/reactive control (inhibition)
		P3	Mean amplitude	I>C I>N	Shorter-lasting P3 effect (=larger P3 for incongruent than congruent and/or neutral trials)		More efficient regulative/reactive control (inhibition)

---

[a] All neural effects are interpreted in the direction of a bilingual effect unless otherwise stated. Moreover, interpretation of the neural effects may not coincide with interpretation in the original studies. Finally, each neural effect is interpreted based on the following assumptions. First, that no other ERP component was tested in the same participants (see section *A positive bilingual effect on executive control in ERPs? Identifying and explaining the pattern* for the interpretation of patterns of effects on multiple components within subjects). Second, that larger N2, N450, ERN, and P3a amplitudes reflect greater executive control demands and effort, if no condition effect was reported. Third, I interpret neural effects as indicating a bilingual advantage in the presence of any bilingual behavioral benefit in executive control, on the assumption that there may not exist a single neural mechanism likely responsible for a bilingual behavioral benefit or a one-to-one correspondence between brain and behavioral effects linked to bilingualism (see Clayson & Larson, 2012: p. 635; Gajewski et al., 2018: p. 271; Kappenman & Luck, 2012: p. 27; Yeung & Cohen, 2006) No positive effect: no bilingual neural effect or no bilingual neural effect in line with the predictions of the accounts or with the neural efficiency considerations described in section *Executive control theories and accounts of the positive bilingual effect on neuro-cognitive executive control* and Table S2 (OSM). Ambiguous: bilingual neural effect cannot be interpreted as clearly indicating a positive or negative bilingual effect

[b] I consider the results ambiguous because they suggest contrasting interpretations. The less negative bilingual ERN for congruent trials indicates a positive bilingual effect on monitoring and/or attention disengagement. However, the more negative bilingual ERN for incongruent and neutral trials, and the wider difference between incongruent and congruent/neutral trials for bilinguals suggest a negative bilingual effect

[c] Moreno et al. (2014) report that there was a marginally non-significant interaction between Group and Condition (p=.09). This suggests that bilinguals had a larger LPC than monolinguals for No-Go trials. I take this result as indicating no positive effect because it was only a non-significant trend

[d] Other reported group differences were not evident for both of bilinguals' languages (across stimulus languages). Thus, they cannot unambiguously be considered as differences in domain-general executive control

Table 2 (continued)

[e] I take the ERP results in this study as indicating no positive bilingual effect because bilingual-monolingual differences were restricted to (and driven by) neutral trials

[f] Heidlmayr et al. (2015) and Heidlmayr et al. (2016) conducted separate analyses for lateral and midline electrodes

[g] There was also a more negative N2 effect (=more negative N2 for negative-priming trials compared to congruent trials) in bilinguals relative to monolinguals, which Heidlmayr et al. (2015) do not interpret as indicating group differences in the N2. This effect was due to an effect inversion at electrode Fz (Heidlmayr et al., 2015: p. 9): a larger N2 for negative-priming relative to congruent trials in bilinguals and a less negative N2 for negative-priming compared to congruent trials in monolinguals

[h] There was also a narrower N2 effect for bilinguals at site Pz. However, Heidlmayr et al. (2016: p. 21) argue that since the control-related N2 is typically described as having a frontocentral distribution, the group difference in N2 at site Pz does not reflect an executive control effect

[i] Morales et al. (2015) report that bilinguals tended to have a larger P3 difference between B and A cues only in a second analysis where bilinguals and monolinguals were matched for AY behavioral performance (but bilinguals still had better BX behavioral performance). I do not report this effect here because it was only a non-significant trend ($p = .06$)

[j] This suggests a larger P2 difference between incongruent/congruent and neutral trials. I interpret this result as indicating no effect of bilingualism on executive control because the effect is driven by the non-demanding neutral trials

[k] Kałamała, Drożdżowicz, et al. (2018a) also intended to look at the N2. However, in their analyses, they did not identify a component with the characteristics of the N2 and did not report any group comparisons on this component

[l] I take the results as ambiguous because the group differences were not in overall component amplitude. Rather, the condition effects on N1, N2, and P3/LPC were observed at more scalp sites for bilinguals compared to monolinguals

[m] Grundy, Chung-Fat-Yim, et al. (2017b) did not intend to examine the conflict-monitoring account of the SCE and, thus, did not take any steps to control for low-level confounds in their sequential congruency measures. For this reason, their results cannot be interpreted as indicating a bilingual effect on regulative/reactive control

$>$ more negative for negative components, more positive for positive components, $<$ less negative for negative components, less positive for positive components, *n.a.* not applicable, *I* incongruent, *C* congruent, *N* neutral, *NP* negative priming, *SOA* stimulus onset asynchrony, *Mono* monolinguals, *Bil. L1* bilinguals tested in their first language, *Bil. L2* bilinguals tested in their second language, *ERN* error-related negativity, *RT* reaction time, *CRN* correct-related negativity, *LNP* late sustained negative-going potential, *AS* antisaccade, *PS* prosaccade, *LPP* late parietal positivity, *PSP* presaccadic positivity, *LPC* late positive component, *SAS* switch antisaccade, *RAS* repeat antisaccade, *SPS* switch prosaccade, *RPS* repeat prosaccade, *S* switch, *R* repeat, *P* pure, *iI* current incongruent trial preceded by an incongruent trial, *iC* current congruent trial preceded by an incongruent trial, *cI* current incongruent trial preceded by a congruent trial, *cC* current congruent trial preceded by a congruent trial, *UC* univalent conflict, *UP* univalent pure, *Incor.* incorrect, *Cor.* = correct, *Ang.* = angry emotion trials, *Neut.* = neutral emotion trials, *T* = target, *NT* = non-target, *L2* = second language, *more-IE* = more interpreting experience regulative EC are evident in phenomena such as the sequential congruency effect (SCE; or conflict adaptation or Gratton effect; e.g., Gratton et al., 1992). The interference effect is larger after congruent than after incongruent trials. This SCE – the difference between the two interference effects – has been explained as reflecting conflict-driven, trial-by-trial adjustments in EC: The detection of conflict during trial *n* leads to high regulative EC. This high EC persists and results in better conflict resolution for the following incongruent trial(s) (e.g., $n+1$) compared to when trial *n* does not include conflict. Thus, the theory predicts that greater conflict detection for trial *n* (more negative N2/N450 or ERN) leads to high regulative EC (i.e., less conflict) for the following incongruent trial(s). Neurally, this appears as less N2/N450 for conflict targets after incongruent or error trials than after congruent or correct trials, respectively. Also, logically, post-conflict EC adjustments should be further evident in post-N2/N450 ERPs linked to EC (e.g., P3 and LNP or conflict slow potential; Larson et al., 2011; Larson et al., 2014; Larson et al., 2016; Yeung & Cohen, 2006), even though the theory makes no explicit predictions about these ERPs.

There are, however, alternative accounts of the SCE, which explain it as resulting from low-level associative processes. For example, when the response and/or stimulus feature(s) from an immediately past trial repeat and, hence, may facilitate or hinder current-trial performance (e.g., Braem et al., 2019). Research on the SCE has tried to control for such low-level confounds by using tasks with large (> 3) stimulus and response sets, which allow for the removal of all partial or complete stimulus feature and/or response first-order repetitions. This work has often still observed the SCE (e.g., Egner, 2007, 2014; Larson et al., 2014; but see Cespón et al., 2020, for the Simon task), suggesting that, at least partly, the SCE reflects adaptive control-based mechanisms. However, this strategy may not account for all low-level confounds (e.g., Braem et al., 2019). For purer measures of adaptive EC, Braem et al. (2019) recommend using separate stimulus and associated response sets for inducer and diagnostic items, which trigger and reflect EC adaptation, respectively.

Crucially, less negative conflict ERPs and smaller behavioral and narrower neural SCEs are often found in neurologic and psychiatric populations with cognitive deficits. Such differences have often been interpreted as showing reduced conflict detection and poor regulative EC adaptation, respectively (e.g., Botvinick et al., 2001; Carter & van Veen, 2007; Clayson & Larson, 2012; Larson et al., 2014). Thus, more negative conflict ERPs exhibit better monitoring; and wider SCEs reflect more adaptive regulative EC (for the SCE, see, e.g., Clayson & Larson, 2012; Goldsmith & Morton, 2018). However, less negative conflict ERPs and narrower neural SCEs may be interpreted positively – as showing greater efficiency – in the absence of (known) behavioral effects or in the presence of behavioral advantages.

Finally, the theory allows for dissociations between the N2/N450 and ERN. These have been generally attributed to processes other than or to functions that act in combination with monitoring (e.g., Yeung et al., 2007; Yeung & Cohen, 2006). Deficits in (conflict-driven adjustments of) regulative EC, for example, result in (a) deficient attentional focus on the relevant target feature for (subsequent) incongruent trials (e.g., central arrow in Flanker task) and, hence, to more negative N2 because of high influence from the irrelevant stimulus dimension (e.g., flankers); and (b) less negative ERN because of reduced conflict with the less-activated – due to deficient focus – correct response after an error (Yeung et al., 2007; Yeung & Cohen, 2006). Thus, given better behavior, less negative N2/N450 for conflict trials but more negative ERN shows a positive bilingual effect on regulative EC. Also, the opposite ERP pattern suggests more efficient regulative EC if there is no difference in behavior: More negative N2/N450 shows more conflict detected, but this high conflict is due to less regulative EC use (e.g., less focus on center arrow and, thus, more flanker influence). Post error, this further results in less conflict between the error and correct response (i.e., less negative ERN). Thus, in this case, more efficient regulative EC is linked to less ERN but at the expense of more N2/N450 conflict. Finally, greater regulative EC efficiency may be evident in another pattern, with less negative bilingual ERN linked to no effect on N2/N450. This shows efficient regulative EC but also an effect on other processes, possibly including monitoring. One possibility, for example, is that it reflects less conflict detection (i.e., more efficient monitoring) leading to less subsequent EC use (i.e., smaller ERN). Less EC, however, does not, in turn, result in a proportional increase of conflict, given the system's efficiency to detect less conflict than present or than others experience.

A third account proposes that the positive bilingual effect on EC may be (partly) attributed to bilinguals' faster or more efficient ability to disengage attention (Grundy & Bialystok, 2018; Grundy, Chung-Fat-Yim, et al., 2017b). This benefit possibly stems from bilinguals' constant experience in paying attention to multiple sources of information in order to use the right language. This, in turn, requires rapid disengagement from one language context (e.g., language, speaker, situation) in order to switch and engage attention to another; and use the appropriate language.

According to Grundy and colleagues (Grundy & Bialystok, 2018; Grundy, Chung-Fat-Yim, et al., 2017b), disengagement can be examined by looking at error-linked neural activity, and at phenomena such as the SCE and the post-conflict slowing effect. Equally, a bilingual effect on disengagement may manifest on switch costs in switching tasks (Grundy, Chung-Fat-Yim, et al., 2017b). This is because, presumably, rapid or efficient disengagement from trial *n*-1 information facilitates performance when *n* is a switch trial, but reduces facilitation (i.e., hinders performance) when *n* is a repeat trial. This results in smaller switch costs. The post-conflict slowing effect reflects conflict-related processes but is distinct from and has a different neural signature to the SCE; specifically, it is indexed by two negative components in the Bivalency Effect switching test (Fig. 5), where it is typically observed. First, an early (100 ms post-stimulus) frontal negativity that is more negative for univalent trials in conflict, more-demanding blocks than in pure, less-demanding blocks. This reflects extra visual processing in complex bivalent blocks. Second, a later frontal negativity with a likely source in the ACC that appears about 300 ms post-stimulus and is sustained for a few hundred milliseconds. This is less negative for conflict univalent (demanding) than pure univalent (less-demanding) trials, and indexes EC (Grundy et al., 2013; Grundy & Shedden, 2014). If disengagement is faster or more efficient in bilinguals, they should show smaller error ERPs (ERN and Pe) and/or a smaller behavioral, and/or narrower neural SCE or post-conflict slowing effect (e.g., Grundy & Bialystok, 2018; Grundy, Chung-Fat-Yim, et al., 2017b). Also, the neural SCE and post-conflict slowing effect may be shorter-lasting. For example, they may be present in earlier ERPs for both bilinguals and monolinguals but not in later ERPs for bilinguals. Finally, according to the same logic, bilinguals may exhibit narrower or shorter-lasting neural switch costs. Each of these bilingual effects shows faster or more efficient attention disengagement from past-trial errors and information, or less effort during later processing due to more efficient disengagement.

A fourth proposal, based on the Dual Mechanisms of Control (DMC) theory (e.g., Braver, 2012), is that the bilingual effect is located in proactive control (e.g., Dash et al., 2021). The DMC postulates that EC can be exerted via a proactive and/or reactive mode, which are potentially (semi-)independent (e.g., Braver, 2012; Braver et al., 2007; Chiew & Braver, 2017). Proactive control refers to the sustained maintenance of goal-relevant information within the lateral prefrontal cortex (PFC), which biases attention, perception, and action systems, during effortful tasks. Thus, proactive control operates in an anticipatory fashion, in that it reduces the influence of demanding events before they occur. For example, it resolves the conflict of a Flanker incongruent trial through proactive focus on the location of the center arrow. In contrast, reactive control reflects the transient, event-triggered reactivation of goal-relevant information that recruits the lateral PFC and other brain regions. It operates upon the occurrence of a demanding event. For example, it resolves conflict only after its onset.[8] Overall, both modes have benefits and costs; and, optimally, the same individual should use both to different degrees, and flexibly shift from one to the other mode based on changing task demands and contexts. However, proactive control is more resource demanding and positively correlates with working memory capacity and fluid intelligence. Thus, individual differences in cognitive resources affect the tendency and ability for proactive control. Moreover, in contexts where a proactive mode is possible (e.g., predictive information is available and reliable) and confers benefits, and a reactive strategy is not as effective and cannot compensate for low proactive control use, reduced cognitive capacity may lead to poorer behavioral performance. Finally, proactive control underlies performance variability in various EC tasks (e.g., switching, working memory, inhibition) and proactive deficits have been argued to underpin cognitive difficulties in healthy older adults and schizophrenic and dementia patients (e.g., Braver, 2012; Braver et al., 2007; Chiew & Braver, 2017).

A popular test for examining both proactive and reactive control is the AX-Continuous Performance Task (AX-CPT; e.g., Braver, 2012). For each trial, subjects see two consecutive stimuli (e.g., Fig. 4). They have to respond “Yes” to every X probe preceded by an A cue (AX trials) and “No” to other trials (e.g., BX, AY, BY). AX trials are very frequent (e.g., 70%). Thus, if guided by proactive control, subjects prepare to respond “Yes” when they see an A cue. However, they need to use reactive control to resolve conflict with the prepared response when the Y instead of the X probe follows. The task is normally performed proactively by healthy young adults, shown by better BX than AY behavioral performance. Higher BX performance results from high proactive use of the B cue, which reliably shows a “No” response. This proactively reduces the bias to respond “Yes” to X, which is habitually linked to a “Yes” response. For AY trials, however, high proactive use of the A cue causes high interference from the error “Yes” response and worse performance. Proactive deficits lead to a (more) reactive (and less proactive) mode, evident in the opposite pattern, or in worse BX and/or better AY results than controls (Braver, 2012; Braver et al., 2007).

The DMC suggests various predictions on how a positive bilingual effect on proactive control may manifest neurally. First, given the sustained and anticipatory nature of proactive control, a bilingual effect should appear before the onset and regardless of EC demands of targets within a demanding block. For example, at a cue preceding both switch and repeat targets in a mixed switching block. However, this effect may be more prominent for cues indicating high EC demands for upcoming targets (e.g., Jamadar et al., 2015) because this information may further boost the already high level of sustained proactive control in bilinguals. Second, bilingual effects may also be evident on EC-demanding targets (e.g., switch) but in the opposite direction relative to pre-target activity in the same block. This is because proactive control reduces anticipatorily the reactive demands of the upcoming target. Third, the bilingual neural effect may also appear on targets that require less EC in

---

[8] Switching, updating, inhibition, and monitoring are generally considered reactive control processes.

a demanding block; specifically, proactive control is expected to decrease the demands of repeat targets in switching blocks with predictable switch and repeat trials (Braver et al., 2003). It may also affect the demands of congruent trials in mixed inhibition blocks. In a Flanker task, for instance, anticipatory focus on the position of the center arrow reduces facilitation from congruent flankers (de Pisapia & Braver, 2006).[9]

The following example predictions assume (known) evidence for better bilingual behavioral performance, unless otherwise stated. Also, they concern tasks with pre-target cues, and ERPs whose amplitude increases with difficulty. Further predictions are given in Table S2 (OSM). In this scenario, higher bilingual than monolingual pre-target (e.g., cue) amplitude shows better bilingual proactive control (e.g., Braver, 2012; Braver et al., 2007; Chiew & Braver, 2017). This effect may appear coupled with lower bilingual amplitude for demanding targets in the same block (e.g., switch and repeat) indicative of less reactive control use. In contrast, higher bilingual activity for demanding targets but not pre-target, suggests superior bilingual reactive control. Also, specifically for the AX-CPT, interpretation may depend on the specific bilingual-monolingual behavioral difference. A first possibility is that bilinguals show better BX and lower AY behavioral performance (= more proactive strategy). In this context, higher cue neural activity in bilinguals (for all or for B cues; e.g., Paxton et al., 2008), coupled with lower amplitude for BX and increased amplitude for AY targets suggests more bilingual proactive reliance at the neural level. A second possibility is that the same ERP effects (for cue, BX, and AY probes) are evident in the presence of better bilingual behavior for both BX and AY or only BX probes. This suggests more effective neural use of both modes. Finally, higher bilingual amplitude for only BX or only AY or both BX and AY probes, indicates, in the absence of a bilingual cue effect, more bilingual reactive use at the neural level. In the latter case, the AY prediction assumes that the AX-CPT is normally solved with a mainly proactive strategy. On this assumption, AY probes have more reactive demands than BX probes. This is because, under a proactive mode, increased use of the A cue leads to higher activation of the error "Yes" response for AY trials, while reliance on the B cue results in lower activation of the error "Yes" response for BX trials.

Also, in mixed blocks of cued tasks, if there is evidence for better bilingual or no effect on behavior, lower bilingual than monolingual pre-target amplitude alone or lower bilingual activity for demanding targets alone shows more efficient proactive or reactive control, respectively. If the two effects appear together, this suggests higher efficiency in both modes. Finally, in the absence of (known) bilingual behavioral effects, higher bilingual pre-target activity coupled with lower bilingual amplitude for demanding targets – or the opposite pattern for the two effects – shows a strategy difference but no positive bilingual effect: Bilinguals rely more on one but monolinguals use more of the other mode.

A final broad view explains the bilingual behavioral benefit in EC tasks by relying on the behavioral EC model of Miyake et al. (2000): It suggests that the bilingual benefit is found in separate EC functions, such as in switching, working memory, and/or inhibition (e.g., Houtzager et al., 2017; Prior & MacWhinney, 2010; Santillán & Khurana, 2018). Bilinguals may enjoy an inhibition benefit, for instance, because, based on Green's (1998) model, the effective use of one language requires the inhibition of the non-relevant language, which is always active (e.g., Santillán & Khurana, 2018). Thus, bilinguals gain extensive practice in inhibition during daily life. For ERPs, this view predicts bilingual effects on target-locked ERPs in conditions that place demands on specific functions: switch trials for switching, critical events (e.g., targets in N-back task) for working memory, and incongruent and No-Go trials for inhibition (Figs. 1, 2, 3, 4, 5 and 6). These positive bilingual effects may appear as shorter ERP latencies, or as larger or smaller amplitudes depending on behavioral results (e.g., bilingual benefit or no effect) and the direction of the EC condition effect (e.g., incongruent vs. congruent) on amplitude (see section *Event-related potentials, bilingualism, and executive control: The present review* and Table S2).[10]

---

[9] This and the monitoring explanation of the bilingual behavioral advantage across trials in EC tasks are two reasons why repeat and congruent trials in mixed blocks may be inappropriate to use as baseline performance measures (e.g., in the calculation of difference scores to isolate EC processing).

[10] This, of course, cannot be an exhaustive list of all accounts on the locus of the positive bilingual effect on EC, or of all (patterns of) neural effects that may suggest a positive bilingual effect in the (types of) tasks mentioned, or of all tasks and phenomena relevant to the accounts described. This is a limitation of the present review. Other accounts, for example, include the bilingual expertise hypothesis (Incera & McLennan, 2016) and proposals that bilingualism positively impacts executive attention (Bialystok, 2017) or the coordination of EC functions (e.g., Morales et al., 2015). These accounts are not discussed because the author did not find them clear enough (at least when writing this review) or it was not possible to form (more) detailed ERP predictions to justify their inclusion as separate accounts. For instance, coordination may be interpreted as suggesting that the bilingual effect appears jointly in more than one process in EC tasks, or that it appears in conditions that require joint use of multiple processes. The BAPSS model can also be considered as a possible instantiation of the coordination account. I provide several predictions for the BAPSS model. Also, for how a bilingual effect may appear jointly in monitoring and regulative EC or in proactive and reactive control, and for how it may appear on the cascade of these functions in the same condition, based on monitoring and DMC theory. However, these are only some possible versions of this account. Also, the processes considered, based on monitoring and DMC theory, are just a few of many other executive functions. Second, in terms of (patterns of) neural effects, not all possible predictions are given because not all have been tested, and to avoid a lengthy discussion. Finally, regarding other phenomena, monitoring, for example, could be tested with the list-wide proportion congruency or post-error slowing effects (e.g., Botvinick et al., 2001). Again, these are not described because no bilingualism ERP study has examined them.

Literature search

A literature search was performed in the Web of Science (Core Collection), PsycINFO (EBSCO host) and PubMed databases. The search included terms referring to bilingualism, EC in general, specific EC processes, and the ERP method. The exact search string was as follows[11]: *(biling* OR multiling* OR bidialect* OR "second language" OR "dual language") AND (ERP OR "Event Related Potential*" OR EEG) AND ("executive control" OR "cognitive control" OR monitoring OR "executive function*" OR inhibit* OR suppression OR interference OR conflict OR "working memory" OR updating OR switching OR shifting OR attention*).*

The search targeted published peer-reviewed articles (in English) through 31 August 2019. For a study to be included in the review, it had to meet all three of the following criteria. (1) Examine the lasting effect (after years/months of using an additional language or after a few days of training) of a bilingual experience on EC, either by comparing groups or through training, or on a continuous scale. (2) Use a non-verbal EC task or a common EC task with verbal stimuli (Stroop, N-back, Go/No-Go, and task-switching). (3) Report a measure (latency and/or amplitude) of an ERP component with reference to at least one of the EC tasks employed.

After screening the titles and abstracts of the detected studies, 23 articles were identified for inclusion in this review (see Table 2). The study of Fernandez et al. (2014) was identified while reviewing these 23 articles.

Bilingualism effects on event-related potential (ERP) components linked to executive control

Tables S1–S4 (OSM) and Table 2 summarize the studies in this review. Table 2 presents the ERP results (focusing on amplitude and latency measures) and their interpretation based on considerations of neural efficiency, different EC theories and accounts of the bilingual EC effect. The other tables provide information about each study's sample size, general participant characteristics (age, gender), possible confounds (Table S1), and details on the participants' dual-language experience (Table S3). Finally, they give information on the tasks used, the ERP components examined, and how they were measured (Table S4). In interpreting the ERP results in Table 2, it was assumed that, when no EC condition neural effect was reported, larger N2, N450, ERN, and P3a amplitudes reflect greater EC demands and effort. Hence, smaller amplitudes reflect greater efficiency in the presence of behavioral benefits or in absence of behavioral effects. This is because these are fairly established findings (see e.g., Table 1). For other ERPs, the literature is less clear about how effort or difficulty manifests in amplitude. For these ERPs, in the absence of a bilingual behavioral effect, results were interpreted only if an EC neural condition effect was reported and based on the direction of this effect (i.e., whether EC was linked to larger or smaller amplitude). In any case, where appropriate, I also discuss the results for a bilingual EC effect when excluding ERP measures for which an EC condition neural effect was not found or reported.

Overall, 18 studies compared bilinguals and monolinguals (Barac et al., 2016; Barker & Bialystok, 2019; Chen et al., 2017; Coderre & van Heuven, 2014; Fernandez et al., 2013, 2014; Grundy & Bialystok, 2018; Grundy, Chung-Fat-Yim, et al., 2017b; Heidlmayr et al., 2015, 2016; Kałamała, Drożdżowicz, et al., 2018a; Kousaie & Phillips, 2012, 2017; Morales et al., 2015; Moreno et al., 2014; Morrison et al., 2019; Timmer et al., 2017; Wu et al., 2016). Three other studies compared different bilingual groups: participants with more interpreting experience versus less interpreting experience (two experiments in Dong & Zhong, 2017); bilinguals who committed very few switching errors ("non-switchers") versus bilinguals who frequently switched to the non-target language ("switchers"; hence, had lower language control skills) in a language-switching task (Festman & Münte, 2012); and bilinguals who were immersed in the L2 versus bilinguals who had a higher L2 proficiency and an earlier age of L2 acquisition (Hannaway et al., 2019). Another study examined how EC is affected by L2 proficiency, frequency of language switching, and SES on a continuous scale in bilinguals (Jiao et al., 2019). Finally, two studies employed training experiments. Sullivan et al. (2014) examined the impact of a 6-month, intensive L2 course. Also, Zhang et al. (2015) investigated the effect of 10 days of language-switching training on EC in experimental and control groups who received training or not, respectively.

In the next sections, I first summarize and discuss the findings for each ERP component separately. I organize the ERP data based on the temporal course of processing, starting with early stimulus- and moving on to response-locked ERPs. For each ERP component, I interpret the results assuming that no other component was tested in the same participants. For predictions concerning more than one component (e.g., predictions 8–9, 12–14 from conflict-monitoring and predictions 17–26 from DMC theory in Table S2 (OSM)), this allows us to examine whether the

---

[11] I decided to look for and include research with bidialectals (i.e., speakers of two linguistically similar and genetically related dialects of the same language), for several reasons. First, because there is some evidence (though controversial) that bidialectals exhibit similar to bilinguals EC advantages (e.g., Antoniou et al., 2016; Antoniou & Spanoudis, 2020). Relatedly, superior EC has been reported for bilingual speakers of very similar languages, such as Spanish-Catalan (e.g., Costa et al., 2009). Second, meta-analyses indicate that language similarity (small or large) between bilinguals' languages does not affect the emergence or absence of a bilingual EC effect (Adesope et al., 2010; Lehtonen et al., 2018).

predicted pattern of ERP effects holds *across independent samples*. However, I also consider the evidence from studies that examined the relevant ERPs within subjects (section *A positive bilingual effect on executive control in ERPs? Identifying and explaining the pattern*). I then move on to examine the evidence on whether specific bilingual experiences are related to ERP indicators of EC. Next, I discuss the overall pattern of results with reference to different EC theories and accounts regarding the locus of the bilingual EC benefit. I close this review by examining methodological concerns from this literature.

Processing in the first 200 ms post-stimulus

N1 and other early negativities

Five studies (six measures) looked at the N1 and other negativities in the Bivalency Effect (Grundy & Bialystok, 2018), Color-Shape (Timmer et al., 2017), Flanker (Dong & Zhong, 2017), Flanker with No-Go trials (Flanker/No-Go; Chen et al., 2017), and spatial Stroop (Jiao et al., 2019) tasks. Dong and Zhong (2017) found more negative N1 across trials (two measures) for the more-interpreting-experience group. However, this suggests a positive effect (on monitoring) only in their Experiment 1.

P2 and other early positivities

Five studies (eight measures) looked at the target-locked P2 during the Flanker (Wu et al., 2016), Flanker/No-Go (Chen et al., 2017), Go/No-Go (Moreno et al., 2014), spatial Stroop (Jiao et al., 2019), and N-back (Morrison et al., 2019) tasks. Another study examined an early cue-locked positivity in the Antisaccade task (Heidlmayr et al., 2016). Chen et al. (2017), Heidlmayr et al. (2016), and Wu et al. (2016) reported bilingual-monolingual differences. In Chen et al. (2017), bilinguals showed a larger P2 amplitude effect (P2 congruent/No-Go vs. P2 neutral), suggesting less efficient inhibition. Furthermore, Wu et al. (2016) reported that bidialectals exhibited smaller P2 than monolinguals for neutral trials and shorter P2 latency for incongruent and congruent trials. However, only the latter result is consistent with a positive bilingual effect on EC (monitoring) because the former effect was restricted to the non-demanding neutral trials. Finally, Heidlmayr et al. (2016) found a narrower cue-locked positivity effect for bilinguals, indicating more efficient proactive control.

Processing in the first 200 ms post-stimulus: Conclusion

Nine studies (15 measures) examined processing within the first 200 ms post-stimulus. Two measures from studies comparing bilinguals and monolinguals are consistent with a positive bilingual effect (Heidlmayr et al., 2016; Wu et al., 2016), but another suggests a negative bilingual effect (Chen et al., 2017). A further measure from Dong and Zhong (2017) indicates a positive effect of more interpreting experience. The bilingual effect (positive or negative) is evident as a larger target-locked P2 effect (Chen et al., 2017), earlier target-locked P2 (Wu et al., 2016) for bilinguals compared to monolinguals, and as larger target-locked N1 (two measures) with more interpreting experience (Dong & Zhong, 2017). For cue-locked ERPs, a narrower positivity effect has been linked to bilingualism (Heidlmayr et al., 2016). Overall, there is little support for a positive bilingual effect in this time window. Also, there is some weak evidence that a bilingual effect (positive or negative) appears as a larger amplitude, wider amplitude effect or earlier latency, at least for target-locked ERPs.

Processing between 200 and 400 ms post-stimulus

N2 and other negativities

Eighteen studies (35 measures) examined the impact of bilingual experiences on N2 and other negativities in this time window. Twenty-seven measures came from inhibition tasks (Antisaccade, Flanker, Go/No-Go, Stroop, and Simon), three from the AX-CPT, two from the Color-Shape, two from the N-back, and one from the Bivalency Effect task. Focusing on bilingual-monolingual comparisons (28 measures), eight measures suggest a positive bilingual effect (Barac et al., 2016; Grundy, Chung-Fat-Yim, et al., 2017b; Kousaie & Phillips, 2012, 2017; Morales et al., 2015; Timmer et al., 2017), four a negative effect (Fernandez et al., 2013, 2014;Heidlmayr et al., 2016 ; Moreno et al., 2014), the other showing ambiguous or no effects. The N2 result in Timmer et al. (2017) could be interpreted as a positive effect on monitoring or switching, or as a negative effect on attention disengagement by different accounts. However, here and in subsequent analyses, it is counted as positive evidence due to the behavioral EC benefit for bilinguals. Some studies had multiple N2 measures, so it is appropriate to also consider the studies independently. This reveals that six (of 14) studies with bilingual-monolingual comparisons suggest only (one or more) positive results, with no or ambiguous effects on other N2 measures, if any (Barac et al., 2016; Grundy, Chung-Fat-Yim, et al., 2017b; Kousaie & Phillips, 2012, 2017; Morales et al., 2015). In contrast, four studies show only negative effects (Fernandez et al., 2013, 2014;Heidlmayr et al., 2016 ; Moreno et al., 2014). Other studies indicate a positive effect of language-switching and L2 training on one N2 measure each (of four; Sullivan et al., 2014; Zhang et al., 2015). Another study shows a negative effect of L2 proficiency (Jiao et al., 2019). Finally, there is one positive and one a negative effect of more interpreting experience (Dong & Zhong, 2017).

Four secondary analyses further examined the robustness of the above N2 results (including the negativity in Grundy & Bialystok, 2018). First, I excluded measures for which no condition neural effect or, for the N2, no larger amplitude or later latency linked to a more EC-demanding condition was reported. This analysis (15 measures) shows one difference favoring bilinguals over monolinguals (Grundy, Chung-Fat-Yim, et al., 2017b) and a positive effect of more interpreting experience (Dong & Zhong, 2017). However, two other measures from monolingual-bilingual comparisons suggest a negative bilingual effect (Heidlmayr et al., 2016; Moreno et al., 2014) and another two show a negative effect of interpreting experience (Dong & Zhong, 2017) and L2 proficiency (Jiao et al., 2019). Second, focusing on measures for which any EC condition effect was reported (e.g., larger or smaller N2 for the EC condition, 20 measures) shows two positive (Grundy, Chung-Fat-Yim, et al., 2017b; Timmer et al., 2017) and two negative effects (Heidlmayr et al., 2016; Moreno et al., 2014) from bilingual-monolingual comparisons. Other measures show one positive and one negative effect of interpreting experience (Dong & Zhong, 2017), and one negative effect of L2 proficiency (Jiao et al., 2019). A third analysis focused on measures for which either any condition effect or a bilingual behavioral benefit was reported. This shows that seven (of 24) bilingual-monolingual comparisons suggest a positive (Barac et al., 2016; Grundy, Chung-Fat-Yim, et al., 2017b; Kousaie & Phillips, 2012; Morales et al., 2015; Timmer et al., 2017) and two a negative effect (Heidlmayr et al., 2016; Moreno et al., 2014). Other measures show a positive effect of language-switching training (Zhang et al., 2015) and more interpreting experience (Dong & Zhong, 2017), and two show negative effects of L2 proficiency (Jiao et al., 2019) and more interpreting experience (Dong & Zhong, 2017).

Finally, because not all N2s are possibly created equally (Larson et al., 2014: p. 290), I considered the N2 data for the Go/No-Go (including the AX-CPT) and Flanker tasks separately. The N2 in the reviewed studies has been most often examined with Go/No-Go tasks, while the Flanker N2 may be thought of as a more unequivocal index of monitoring (Larson et al., 2014). For Go/No-Go tasks, two monolingual-bilingual comparisons (of seven) suggest a positive (Barac et al., 2016; Morales et al., 2015) and three a negative bilingual effect (Fernandez et al., 2013, 2014; Moreno et al., 2014). Other measures show one positive effect of L2 (Sullivan et al., 2014) and one of language-switching (Zhang et al., 2015) training. For the Flanker, two bilingual-monolingual comparisons (of six) suggest a positive bilingual effect (on monitoring and attention disengagement; Grundy, Chung-Fat-Yim, et al., 2017b; Kousaie & Phillips, 2017), and two measures show a positive and a negative effect of interpreting experience (Dong & Zhong, 2017). Thus, results from the third secondary and the Flanker analysis are somewhat improved in favor of a positive bilingual effect. However, they need to be taken with caution because a positive effect did not appear in three other analyses with more data. Also, for the third analysis, many measures suggest a null bilingual effect, besides the negative results. Finally, for the Flanker, there were only very few data, while one positive effect could not be attributed to monitoring, the process assumed to be primarily indexed by the Flanker N2.

Looking at the bilingual effect regardless of interpretation reveals that six measures indicate a larger N2 or wider N2 effect (in inhibition, AX-CPT, and Color-Shape tasks) for bilinguals than monolinguals (Fernandez et al., 2013, 2014; Heidlmayr et al., 2016; Morales et al., 2015; Moreno et al., 2014; Timmer et al., 2017). For three other measures from inhibition tasks, the bilingual effect was in the opposite direction (Grundy, Chung-Fat-Yim, et al., 2017b; Kousaie & Phillips, 2012, 2017). Another four measures show larger N2 after language-switching training (Zhang et al., 2015), with higher L2 proficiency (Jiao et al., 2019) and with more interpreting experience (Dong & Zhong, 2017). Finally, shorter N2 latencies were found for bilinguals compared to monolinguals (three measures; Barac et al., 2016; Kousaie & Phillips, 2017) and after L2 training (one measure; Sullivan et al., 2014).

P3

Thirteen studies (32 measures) compared bilinguals and monolinguals on the P3. Twenty-three measures are from inhibition tasks, two from the AX-CPT, two from the Color-Shape, and five from the N-back paradigm. Twelve target-locked measures from inhibition tasks (including the probe-locked P3a in the AX-CPT; Barac et al., 2016; Chen et al., 2017; Grundy, Chung-Fat-Yim, et al., 2017b; Heidlmayr et al., 2016; Kousaie & Phillips, 2012, 2017; Morales et al., 2015) support a positive bilingual effect. Also, two target-locked measures from the N-back test show a negative effect (Barker & Bialystok, 2019; Morrison et al., 2019). At the individual study level, seven studies indicate only positive evidence for a bilingual effect (all in inhibition tasks) and two only negative evidence in working memory tasks (Barker & Bialystok, 2019; Morrison et al., 2019). Thus, overall, these findings suggest a positive bilingual effect on the target P3 in inhibition tasks. Results of other studies are mixed. In Sullivan et al. (2014), one latency measure shows a positive and one amplitude measure suggests a negative effect of L2 training. Similarly, the Dong and Zhong (2017) results suggest a positive and a negative effect of more interpreting experience.[12]

---

[12] Dong and Zhong (2017) reported an analysis for the whole P3 time window and two further analyses for an early and late P3 window. Here and later on, I consider the results only from the former analysis.

Given that the bilingual effect is more consistent at the target-locked P3 – at least for bilingual-monolingual comparisons in inhibition tasks – it is worth pursuing what this effect means in terms of the specific processes impacted. Of the 14 positive bilingual effects (including Dong & Zhong, 2017, and Sullivan et al., 2014), seven suggest a monitoring, six an inhibition, and one an attention-disengagement effect. Negative results show two negative effects on working memory (Barker & Bialystok, 2019; Morrison et al., 2019), one on monitoring (Sullivan et al., 2014), and one on inhibition (Dong & Zhong, 2017). Thus, the P3 results are more consistent with a bilingual monitoring effect, even though a bilingual inhibition effect is also supported.

Finally, the pattern of results is roughly the same when excluding (a) measures for which no EC condition effect was found or reported and (b) only measures for which neither a condition effect nor a bilingual behavioral advantage was found or reported. The only difference is that inhibition receives equal or more support (five positive effects in both secondary analyses) than monitoring (two positive effects in the first secondary analysis) when considering bilingual-monolingual comparisons.

Finally, regardless of interpretation, the bilingual P3 effect most often (five measures) manifests as a larger P3 or wider P3 effect when considering bilingual-monolingual comparisons. Moreover, five other measures show an earlier P3 latency for bilinguals. These findings are reinforced by Jiao et al. (2019) and Sullivan et al. (2014). The former study found a larger P3 with higher L2 proficiency, while the latter reported shorter P3 latency and increased P3 amplitude after L2 training. Five other measures suggest a smaller P3, narrower P3 effect (Grundy & Bialystok, 2018; Heidlmayr et al., 2016; Kousaie & Phillips, 2012, 2017), and a shorter-lasting P3 effect for bilinguals (Chen et al., 2017). Finally, in Dong and Zhong (2017) results are conflicting, indicating a smaller P3 (Experiment 1) and a wider P3 effect (Experiment 2) with more interpreting experience.

Processing between 200 and 400 ms post-stimulus: Conclusion

Processing within this time window has been the focus of most studies. Overall, there is no clear evidence for a positive bilingual effect at the N2 and other negativities. However, a positive bilingual effect is more consistently evident at the target-locked P3 in inhibition tasks. This evidence more often indicates a monitoring locus of the bilingual effect, with inhibition also receiving some support. Also, some evidence suggests that, when a bilingual effect appears, it often manifests as larger N2 and P3, wider N2 and P3 amplitude effect or shorter N2 and P3 latency (27 of 71 measures). However, for amplitude, opposite results are also reported (three for N2, five for P3).

Processing after 400 ms post-stimulus

N450

Three studies (three measures) looked at the N450 during Stroop tests (Coderre & van Heuven, 2014; Hannaway et al., 2019; Heidlmayr et al., 2015). Heidlmayr et al. (2015) reported a narrower N450 effect (incongruent-congruent trials) in bilinguals compared to monolinguals, suggesting more efficient monitoring or inhibition. Coderre and van Heuven (2014) also reported a bilingual-monolingual difference, but the effect was found for neutral trials. Thus, the group effect does not reflect EC differences. Hannaway et al. (2019) used a bilingual Stroop task and found a larger N450 across trials – suggesting better monitoring – for bilinguals with higher L2 proficiency and earlier L2 acquisition relative to bilinguals immersed in their L2. Crucially, a larger N450 was found for bilinguals, who exhibited a smaller RT interference effect overall and, particularly, in the difficult condition where the stimulus language was German. Thus, one study suggests no positive bilingual effect, one study suggests more efficient monitoring or inhibition in bilinguals compared to monolinguals, and another better monitoring for bilinguals with higher L2 proficiency and earlier age of L2 acquisition relative to bilinguals immersed in their L2.

Other late negativities

Two studies compared bilinguals and monolinguals on other late negativities using the Stroop (Heidlmayr et al., 2015) and Bivalency Effect task (Grundy & Bialystok, 2018). Heidlmayr et al. (2015) found a narrower LNP effect for bilinguals. Furthermore, Grundy and Bialystok (2018) reported a significant post-conflict negativity effect (400–800 ms) only for monolinguals. These effects suggest more efficient neural inhibition and attention disengagement in bilinguals, respectively.

Late positivities

Four studies examined late positivities (late positive component and late parietal positivity) in Antisaccade (Heidlmayr et al., 2016), Flanker (Wu et al., 2016), Go/No-Go (Moreno et al., 2014), and Stroop (Hannaway et al., 2019) tasks. Wu et al. (2016) found a larger late positive component (LPC) across trials for bidialectals than monolinguals. However, this result cannot unambiguously be interpreted.

Processing after 400 ms post-stimulus: Conclusion

Overall, the evidence for late target-locked components is scant. Two studies (three measures) with bilingual-monolingual comparisons suggest a positive bilingual effect (Grundy

& Bialystok, 2018; Heidlmayr et al., 2015), one shows an ambiguous effect (Wu et al., 2016), and three indicate no effect (Coderre & van Heuven, 2014; Heidlmayr et al., 2016; Moreno et al., 2014). Moreover, the bilingual effect on late negative components appears as a narrower negativity, N450, and LNP effect (Grundy & Bialystok, 2018; Heidlmayr et al., 2015).[13] However, this evidence contrasts with Hannaway et al. (2019), who compared bilingual groups and reported overall larger N450 for bilinguals with higher L2 proficiency, earlier L2 acquisition, and who had better behavioral results. Moreover, it contrasts with the results for late positivities, which show a larger LPC across trials for bidialectals (Wu et al., 2016) or no bilingual effect (Hannaway et al., 2019; Heidlmayr et al., 2016; Moreno et al., 2014).

Response-locked processing

Presaccadic positivity (PSP)

Heidlmayr et al. (2016) looked at the saccade-locked PSP in the Antisaccade task. They found a narrower PSP effect for bilinguals compared to monolinguals, suggesting a positive bilingual effect on inhibition.

Error-related negativity (ERN) and correct-related negativity (CRN)

Five studies (seven measures) examined the ERN using the AX-CPT (Morales et al., 2015), Bivalency Effect (Grundy & Bialystok, 2018), Flanker (Festman & Münte, 2012; Kałamała, Drożdżowicz, et al., 2018a; Kousaie & Phillips, 2012), Simon, and Stroop tasks (Kousaie & Phillips, 2012). Four studies (four measures) with bilingual-monolingual comparisons (of four total studies with six measures) suggest a positive bilingual effect on monitoring or attention disengagement (Grundy & Bialystok, 2018; Kałamała, Drożdżowicz, et al., 2018a; Kousaie & Phillips, 2012; Morales et al., 2015). In Kałamała, Drożdżowicz, et al. (2018a), the larger ERN/CRN for bilinguals could indicate both a positive (on monitoring) and a negative effect (disengagement). However, the bilingual behavioral benefit suggests the former interpretation. Thus, this measure was counted as positive evidence. Also, in Kousaie and Phillips (2012), no EC condition effect was found for the two measures indicating no bilingual effect. This reinforces the results of a positive bilingual effect at the ERN. The bilingual effect (positive or negative) materialized as smaller ERN (Grundy & Bialystok, 2018; Kousaie & Phillips, 2012) or narrower ERN effect (Morales et al., 2015) for three measures but as larger ERN for one measure (Kałamała, Drożdżowicz, et al., 2018a). Finally, Festman and Münte (2012) compared bilingual groups and found smaller ERN for bilingual non-switchers, who had better language control and EC skills.

Error positivity (Pe) and correct positivity (Pc)

Two studies compared bilinguals and monolinguals on the Pe/Pc using the Bivalency Effect (Grundy & Bialystok, 2018) and a lateralized Flanker task (Kałamała, Drożdżowicz, et al., 2018a). The study by Grundy and Bialystok (2018) suggests a positive bilingual effect on disengagement, evident as a smaller Pe for bilinguals.

Response-locked processing: Conclusion

The five studies that compared bilinguals and monolinguals suggest a positive bilingual effect on six (of nine) measures. Considering each study independently, five studies provide only positive evidence (Grundy & Bialystok, 2018; Heidlmayr et al., 2016; Kałamała, Drożdżowicz, et al., 2018a; Kousaie & Phillips, 2012; Morales et al., 2015). Thus, there is evidence for a positive bilingual effect on response-locked ERPs. These positive effects suggest monitoring and attention disengagement as the locus of the bilingual effect. Moreover, for five measures, bilingual effects manifest as smaller ERPs or smaller or narrower amplitude effects (smaller PSP effect, smaller ERN, narrower ERN effect, and smaller Pe). These results are in line with the less negative ERN for bilingual non-switchers in Festman and Münte (2012). However, one measure indicates a larger ERN/CRN for bilinguals (Kałamała, Drożdżowicz, et al., 2018a).

Do specific bilingual experiences affect executive control in ERPs?

Some studies examined the effect of specific bilingual experiences on ERPs, either on a continuous scale (Fernandez et al., 2013, 2014; Heidlmayr et al., 2015, 2016; Jiao et al., 2019; Sullivan et al., 2014) or by comparing bilingual groups (Dong & Zhong, 2017; Festman & Münte, 2012; Hannaway et al., 2019), or through training experiments (Zhang et al., 2015). Some of these results have been discussed in previous sections. Here, I integrate them with the findings of studies that compared bilinguals and monolinguals, but further examined correlations between ERPs and bilingual experiences on a continuous scale.

For the N2, Fernandez et al. (2013, 2014) and Jiao et al. (2019) showed that higher L2 proficiency is linked to larger N2, even though this correlation did not reach statistical significance in Fernandez et al. (2014). However, it is

---

[13] This is also in line with Chen et al.'s (2017) finding of a shorter-lasting (up to 550–600 ms) bilingual P3 compared to monolinguals (discussed in the 200- to 400-ms time window).

noteworthy that, in the absence of behavioral effects, these correlations suggest a negative L2 proficiency effect on monitoring (one measure) and monitoring or inhibition (two measures). Moreover, two other studies found no effect of L2 proficiency (Heidlmayr et al., 2015; Heidlmayr et al., 2016). Beyond L2 proficiency, Dong and Zhong (2017) reported larger N2 for subjects with more compared to subjects with less interpreting experience, who otherwise had equal L2 proficiency and use. However, this effect suggests a positive effect in their Experiment 1 and a negative effect in their second. Finally, Zhang et al. (2015) showed that a brief period of language-switching training enhanced the N2, suggesting a positive effect on proactive control. However, again, two other studies did not identify frequency of language switching (Jiao et al., 2019) or language-switching experience (Heidlmayr et al., 2016) as factors affecting the N2.

For the P3, Jiao et al. (2019) found larger P3 with higher L2 proficiency. Consistent with this finding, the training study by Sullivan et al. (2014) reported that a wider P3 difference between post- and pre-test in the training group positively correlated with self-reported expected grade in the L2 course. However, in terms of interpretation, the results of Jiao et al. (2019) are ambiguous and the correlation in Sullivan et al. (2014) suggests a negative effect of L2 grade on monitoring. Dong and Zhong (2017) also provided mixed findings for interpreting experience, with a smaller P3 for the more-interpreting-experience group in Experiment 1 indicating a positive effect; and a wider P3 effect in Experiment 2 suggesting a negative effect. Furthermore, Heidlmayr et al. (2016) found no effect of L2 proficiency on the P3.

For late components (stimulus- or response-locked), smaller amplitudes or amplitude effects have been associated with various bilingual experiences: a narrower N450 effect with more frequent L2 use (but no effect of L2 proficiency; Heidlmayr et al., 2015), a narrower PSP effect with more L2 immersion experience and higher L2 proficiency (Heidlmayr et al., 2016); and a smaller ERN for non-switcher bilinguals who had better language control skills, and higher L2 use (in some domains) and L2 proficiency (Festman & Münte, 2012). All these correlations suggest positive effects of various bilingual experiences on neural EC; specifically, on monitoring or inhibition (one measure), monitoring or attention disengagement (one), and inhibition (one). In addition, Hannaway et al. (2019) reported larger N450 for bilinguals who had a higher L2 proficiency and earlier age of L2 acquisition compared to bilinguals who were L2 immersed. This suggests better monitoring in the former group. Finally, no effect of any bilingual experience (frequency of L2 and L3 use, L2 proficiency, duration and age of immersion in L2) was found on the LNP (Heidlmayr et al., 2015).

To sum, there is some evidence that increased bilingual experiences (e.g., language-switching, interpreting experience), and particularly L2 proficiency, correlate with larger amplitudes or wider amplitude effects at the N2 and P3. However, there is little evidence that these correlations reflect a positive effect on neural EC. Moreover, there is some consistency in the findings of a smaller ERN and narrower amplitude effects for some late – stimulus- or response-locked – components (N450, PSP) with increased bilingual experiences. Again, higher L2 proficiency seems to be a common, perhaps, important bilingual variable (if the Hannaway et al., 2019, results are also considered) in these studies. Correlations with late ERPs suggest positive effects on neural processing (monitoring and inhibition).

In general, however, any evidence in this section is only suggestive at best. First, only a few studies have examined the effect of specific bilingual experiences. Second, L2 proficiency appears as a relevant variable for early and late ERPs, but this evidence indicates positive effects only for late components. Also, for both early and late ERPs, other bilingual experiences that co-varied (or are known to co-vary) with L2 proficiency were often not controlled for. Thus, it is unclear whether effects can be attributed to L2 proficiency or other bilingual variables. Third, to foreground the discussion in the section *Small sample sizes*, results in this section are limited by low statistical power. This is especially true for studies in which correlations between ERPs and bilingual experiences were examined in a sub-sample and, thus, on a small number of subjects (Fernandez et al., 2013, 2014; Heidlmayr et al., 2015, 2016; Sullivan et al., 2014).

A positive bilingual effect on executive control in ERPs? Identifying and explaining the pattern

Table 2 presents the interpretation of 108 measures (13 ERP components) from the 24 reviewed articles. In the 18 studies (89 measures) that reported bilingual-monolingual comparisons, 31 measures can be interpreted as a positive bilingual effect (including Kałamała, Drożdżowicz, et al., 2018a, and Timmer et al., 2017), 45 showed no bilingual effect, seven revealed a negative effect, and six could not be interpreted. For most positive evidence, the bilingual effect was in reactive control: monitoring (13 measures) or inhibition (seven). Also, there is a positive effect specific to attention disengagement and proactive control for four and one measure, respectively. Finally, six effects are consistent with a positive effect on more than one aspect: monitoring or attention disengagement (three), monitoring or inhibition (two), and monitoring or switching (one). At the individual study level, of the 18 articles that included bilingual-monolingual comparisons, ten suggest only (one or more) positive bilingual effects (with no negative evidence), five indicate only (one or more) negative results (with no positive evidence), and three studies revealed no effect (Coderre & van Heuven, 2014) or mixed findings (Chen et al., 2017; Heidlmayr et al., 2016).

Processing within 200 ms post-stimulus may not (clearly or strongly) reflect EC. Thus, it is important to examine whether results change when excluding the 11 measures from this early time window. This reveals that 29 measures suggest a positive effect (with most evidence supporting a monitoring account), six measures a negative effect, the rest showing no bilingual effect or ambiguous results. Also, ten independent studies show only positive and five only negative evidence. Thus, the pattern is unchanged, when considering components which more strongly reflect EC.

Moreover, both studies that did not include young adults, but tested children (Barac et al., 2016) or older adults (Kousaie & Phillips, 2017), showed a positive bilingual effect, with most measures suggesting bilingual-monolingual differences in monitoring. This is in line with claims that the bilingual benefit may be found only in populations who are not at the peak of cognition (e.g., Bialystok & Craik, 2010). However, more studies are needed for firm conclusions to be drawn on this issue.

Focusing on measures for which sensitivity to EC was reported[14] (52 measures), 17 measures (including Kałamała, Drożdżowicz, et al., 2018a) suggest a positive bilingual effect and five indicate a negative effect. From the positive results, seven suggest an effect on inhibition, three on monitoring, three on attention disengagement, two on monitoring or attention disengagement, one on monitoring or inhibition, and one on proactive control. For the negative results, two show a negative effect on monitoring or inhibition, two on working memory, and one on inhibition. Thus, this analysis provides some evidence for a positive bilingual effect too. However, the evidence is less clear in terms of the locus of this effect.

To sum, the general picture indicates that there is some (though not strong) evidence in support of a positive bilingual effect on EC in ERPs. This positive effect has been inconsistently located in various EC aspects, with most analyses and evidence supporting a monitoring account. Inhibition is the second most supported account in the global analysis in this section. Also, the more focused analysis – excluding measures with no EC sensitivity reported – provides slightly more support to inhibition than monitoring. Other accounts also receive some support, even though the evidence base for these is relatively small; specifically, in the global analysis, seven measures (in four studies) are consistent with a positive bilingual effect on attention disengagement. However, three of these measures are also in line with a bilingual impact on other EC aspects. Moreover, two other measures indicate, somewhat ambiguously, less efficient functioning in this aspect (Kałamała, Drożdżowicz, et al., 2018a; Timmer et al., 2017). In addition, a positive bilingual effect on proactive control was evidenced for one of two measures. A proactive control locus is also supported by the training study of Zhang et al. (2015). Also, three results are relevant to switching (LPP switch effect in Heidlmayr et al., 2016, N2 and P3 switch effects in Timmer et al., 2017). One of these shows a bilingual benefit but is also in line with a monitoring locus (Timmer et al., 2017). For working memory, two studies suggest negative bilingual effects (Barker & Bialystok, 2019; Morrison et al., 2019).

Moreover, results by time window and component show that positive bilingual effects appear more consistently at the target P3 and response-locked ERPs. This evidence more often supports a monitoring (P3 and response-locked ERPs), secondarily an inhibition (P3), and, to some extent, an attention disengagement account (response-locked ERPs). Finally, training experiments, studies that compared bilingual groups or examined continuously specific bilingual experiences provide some suggestive evidence that higher L2 proficiency is relevant to EC in ERPs: L2 proficiency has been linked to larger ERPs or wider effects at the N2 and P3, and to smaller ERPs or narrower effects for later stimulus- and response-locked ERPs. These results show a positive bilingual impact only for later ERPs, with a likely monitoring or inhibition locus of effects. In general, however, the evidence regarding which specific bilingual experiences affect EC in ERPs is scant, unclear, and only suggestive at best.

Evidence for monitoring, proactive, and reactive control, when considering the pattern of results from multiple components

Overall, this review finds some evidence for a positive (likely small) bilingual effect. Moreover, when considering each ERP component independently (assuming no evidence for other ERPs) positive effects are more often located in monitoring. Also, the positive bilingual effect is more consistently observed for the P3 and for late response-locked components. Regarding proactive and reactive control specifically, most studies focused on reactive and only three studies provided measures relevant to proactive control. The latter studies provide some evidence for a positive effect of bilingualism-related experiences. In the following, I consider the evidence for predictions which concern *patterns of effects on multiple components*. I first discuss predictions from monitoring and then from DMC theory.

---

[14] That is, I excluded all measures for which no condition effect was found or for which EC sensitivity was not reported, as follows: larger N2, N450, and P3a linked to a more EC-demanding versus less-demanding condition or larger ERN and Pe linked to error versus correct trials, for the N2, N450, P3a, ERN, Pe, respectively; or, for latency measures, later latency linked to a more- versus less-demanding condition. For other ERPs, amplitude measures were kept as long as any condition effect was found.

Looking at the evidence across independent studies for the N2[15] and ERN combined indicates that a positive bilingual effect is more consistently found as a less negative ERN but does not appear at the N2. This pattern is in line with a likely monitoring locus of the bilingual effect. However, the N2/ERN dissociation suggests that other, post-monitoring processes are possibly also affected (monitoring predictions 9 and 14 in Table S2 (OSM)). Studies that examined both components within subjects are also relevant here, but the findings are mixed; specifically, Kousaie and Phillips (2012) reported smaller bilingual N2 and ERN for the Stroop task, in absence of bilingual-monolingual behavioral differences in three EC tasks. This shows a positive effect on monitoring. However, they also reported a complex N2/ERN pattern in the Flanker task suggesting N2/ERN dissociation. A dissociative pattern was also reported in Morales et al. (2015), with a larger N2 but smaller ERN in bilinguals, suggesting better post-conflict regulative control rather than monitoring. A final consideration also suggests that monitoring may not be the (only) locus of a bilingual effect; specifically, the finding that a positive bilingual effect is more consistently observed for the P3 strongly suggests that, since the P3 occurs after the N2, the effect is not found (only) in monitoring but in some later, post-conflict mechanism(s). This may or may not involve inhibition, given that inhibition received some support but was not the process identified as primarily impacted at the P3 level.

Regarding DMC theory, based on the results across studies, it is possible that a positive bilingual effect is found on both proactive and reactive control. This is evident in the finding of a positive bilingual effect at the target-locked P3 combined with the few results showing a positive bilingual effect on proactive control. However, the three studies that examined both proactive and reactive control measures within subjects provide little support to this proposal; specifically, in Heidlmayr et al. (2016), the only proactive control (cue-locked) measure suggested a positive bilingual effect, and this was combined with one measure indicating a negative bilingual effect (wider target N2) and two measures (narrower target P3 and PSP) suggesting a positive effect on reactive control. Of note is that, in absence of behavioral differences in this study, the pattern of narrower cue amplitude and wider early target N2 for bilinguals is compatible with a simple strategy difference (DMC prediction 11 in Table S2 (OSM)) rather than a positive or negative bilingual effect, respectively. However, this pattern is further complicated by the subsequent smaller target-locked effects (P3 and PSP) for bilinguals. Also, in Morales et al. (2015) there was no bilingual effect on proactive but positive effects on reactive control. Finally, Zhang et al. (2015) found a positive effect of language-switching training on proactive but not on a reactive control measure. Nevertheless, generally, strong conclusions for bilingual effects on proactive control or on both control modes are not warranted due to the limited relevant evidence. Next, I discuss the evidence for the BAPSS model.

Evidence for the BAPSS model

The general pattern that emerges across independent studies regarding the bilingual effect on amplitude and latency at different processing stages is the following. During very early stages (within 200 ms), there is weak evidence for larger component amplitudes, larger amplitude effects, and earlier latencies linked to bilingualism. This evidence concerns target-locked measures, with one study reporting a narrower cue-locked amplitude effect for bilinguals. Similarly, at 200–400 ms post-stimulus, the bilingual effect, if found, often manifests as a larger N2 and P3, wider N2 or P3 effects, and shorter N2 and P3 latencies. At late stages (after 400 ms post-stimulus), bilingualism is linked to narrower negativity effects (e.g., N450, LNP). Finally, at the response and error processing stage, bilingualism is associated to smaller amplitudes or narrower amplitude effects (PSP, ERN/CRN, and Pe/Pc). These results are further reinforced by findings indicating larger N2, P3, wider N2 and P3 effects, and smaller late ERP amplitudes or amplitude effects with increased bilingual experiences. Overall, this pattern of findings largely fits the predictions of the BAPSS model. However, this is not without challenges. First, at all processing stages, there are exceptions to the overall trend with many studies not reporting the predicted effects or even some studies reporting effects in the opposite direction. Second, focusing on the eight studies (11 measures) that examined at least one relatively earlier (P3 and earlier) and at least one later ERP (after the P3) within the same participants and task (Grundy & Bialystok, 2018; Heidlmayr et al., 2015, 2016; Kałamała, Drożdżowicz, et al., 2018a; Kousaie & Phillips, 2012; Morales et al., 2015; Moreno et al., 2014; Wu et al., 2016), only three studies show a pattern that may – at large – be considered consistent with the BAPSS model; specifically, Kousaie and Phillips (2012) reported earlier P3 and less negative ERN (but also less negative N2) in the Stroop task; Morales et al. (2015) found more negative N2, larger P3, and narrower ERN effect; and Heidlmayr et al. (2015) reported a wider target-locked N2 (but narrower target-locked P3) and narrower PSP effect. Third, as discussed previously and notwithstanding the cautionary note on low statistical power, it is unclear which specific bilingual experiences affect neural EC in ERPs. Fourth, as discussed in the *Introduction*, larger P3 amplitudes may not always unambiguously reflect the devotion of more early neural resources, as suggested in the BAPSS model (Grundy, Anderson, & Bialystok, 2017a). Finally, it is puzzling that,

---

[15] I do not consider the N450 because it has hardly been examined (three measures) in this literature.

after 400 ms post-stimulus, a bilingualism effect is identified for negative but is not found for positive ERPs linked to EC; or, if found, it appears as larger ERP amplitude (Wu et al., 2016).

Regarding the latter finding, it is fair to stress that BAPSS (at least in Grundy, Anderson, & Bialystok, 2017a) does not explicitly consider these late positivities (or ERPs within 200 ms), even though it does claim less effort for bilinguals at later processing stages. Moreover, one way to reconcile this result with the BAPSS model would be to suggest that what this review called (or are often called) "late positivities" are in fact (late) P3s or extensions of the P3, as has been suggested for the LPC (see Table 1). This is possible if one considers that measurement time windows for the P3 and late positivities in the reviewed studies overlap; specifically, in some studies, the P3 time window goes beyond or starts at 500 ms and extends until around 800 ms post-stimulus (e.g., Barac et al., 2016; Chen et al., 2017; Kousaie & Phillips, 2017; Morales et al., 2015), while late positivities were often measured within 400–650 ms (Heidlmayr et al., 2016; Moreno et al., 2014) or even earlier (Wu et al., 2016). From this perspective, the larger LPC for bidialectals in Wu et al. (2016) may in fact be consistent with the overall pattern for the bilingual effect on the P3 (hence, with BAPSS). However, the two studies reporting no bilingual-monolingual differences in late positivities would add to the null results for a bilingual P3 effect. Moreover, this may pose additional challenges; specifically, if stimulus processing, as reflected in the P3, may extend up to 800 ms, then, no room is left for further, less effortful later processing linked to bilingualism as argued by BAPSS. Of course, this is unless it is assumed that, in these cases, less effortful later processing manifests in response-locked ERPs, which would be in line with BAPSS. Next, I discuss two factors that suggest further caution in the conclusions we can draw from this literature.

Methodological concerns

Small sample sizes

Small sample sizes are particularly problematic because they reduce the power of an experiment. In turn, underpowered studies can lead to erroneous findings in three ways: they might not detect a true effect, they have increased chances of detecting an effect that does not exist; and they might detect a real but inflated in size effect (Bakker et al., 2012; Brysbaert, 2019; Button et al., 2013). A recent assessment of statistical power in research using ERPs reported an average sample of 21 participants per group, which suggests low power (= .72, below the recommended .80 power level) to detect even a large effect size of Cohen's d =.8 (Clayson et al., 2019). This is on a par with power reports for the fields of psychology and cognitive neuroscience in general (e.g., Szucs & Ioannidis, 2017). A similar picture arises in the literature reviewed here, with a mean sample size of 21.76 across all studies and a mean sample of 22.6 for studies that compared bilinguals and monolinguals. Using G*Power (Faul et al., 2007) and specifying a desired power of .80 and a two-tailed alpha of .05, a study with two independent groups of 23 subjects each has enough power to detect a large effect of d = .84. However, it is unlikely that this reflects the true size of the bilingual effect. First, the most recent meta-analyses on the bilingual behavioral effect on EC suggest that, if there is a cognitive benefit, it is at most of small-to-moderate size (Grundy & Timmer, 2017, for working memory; Gunnerud et al., 2020, for switching in children). Second, it has been recently suggested that a good first estimate of the typical effect size in psychological research is d = .4 (Brysbaert, 2019). This is the average effect size reported in recent large-scale replication studies of published research in psychology and is also the value often reported in various meta-analyses of psychological studies (Brysbaert, 2019). Similarly, for both psychology and cognitive neuroscience, large-sample-size studies rarely report large effects (Szucs & Ioannidis, 2017; Szucs & Ioannidis, 2020: p. 8; see also Poldrack et al., 2017: p. 119; Yarkoni, 2009). Moreover, expecting relatively large medium-sized effects has been suggested to be overly optimistic, at least for fMRI neuroimaging (Szucs & Ioannidis, 2020: p. 8). Thus, overall, low power may be one factor contributing to the inconsistent findings in this body of work.[16]

Confounding variables

Table S1 (OSM) includes information about whether the reviewed studies took into account various potential confounds; specifically, culture, SES, and immigration status. Of the 21 studies that compared groups, 11 did not control for, while an additional three studies did not consider – or did not provide enough information for a judgment to be made on whether they controlled for – at least one of these factors. Moreover, non-verbal general intelligence was a potential confound in the training study by Zhang et al. (2015), given that the training group had a higher general intelligence at pre-test.

Moving forward

The controversy on the neuro-cognitive benefits of bilingualism has led to extensive discussions on how to move this field forward (e.g., Bialystok, 2017; Cespón, 2021; de Bruin

---

[16] Other methodological factors may also play a role because of the many methodological choices available when collecting and analyzing ERP data (e.g., choosing specific time window and electrodes for ERP measurement). I do not discuss these issues here, but information regarding such decisions in the reviewed studies is presented in Table S4 (OSM; see also Clayson et al., 2019, for relevant discussion).

et al., 2021; García-Pentón et al., 2016, and commentaries; Navarro-Torres et al., 2021; Paap et al., 2015, and commentaries; i.a.). Here, I comment on four factors that may help extract reliable signals from ERP research on bilingualism.

First, there is the obvious need for research with higher statistical power. A straightforward solution is to recruit larger participant numbers. This is easier to achieve through collaborative work that includes data collection from multiple sites (Clayson et al., 2019). That said, another, often neglected, factor that is known to improve power is measurement reliability, which, in turn, is related to the number of trials in a test (e.g., Clayson et al., 2019; Cohen, 1988; Goulet & Cousineau, 2019). Brysbaert (2019), for instance, suggests that, for within-group designs, a given effect can be assumed to be 1.5 times larger if the dependent measure has a high (intraclass) reliability of $r = .8$. Similarly, Goulet and Cousineau (2019) show that the number of trials and their correlation within a test condition, which are linked to measurement reliability, may reduce the number of participants for adequate power. In this regard, they present formulas to calculate power and the potential gain in participant numbers by taking such information into account. Finally, for ERP studies, increased reliability has been found to improve power in both between-group (Clayson et al., 2021; Hajcak et al., 2017) and within-group studies (Clayson & Miller, 2017; see also in Clayson et al., 2019). Thus, researchers should carefully consider this design feature in power calculations because it can often substantially reduce the sample size required to achieve sufficient power (see also Baker et al., 2021; Luck et al., 2021).

Second, future work should more carefully exclude the possibility that bilingual (or lack of bilingual) effects can be explained by various confounds. This review focused on three main potential confounds in bilingualism studies, but this list is not exhaustive. Other possible confounds include, for example, videogame play, music performance and training, and experience or ability in sports (Paap, 2019). Relatedly, experimental designs, such as longitudinal and/or training studies, in which groups, if used, do not differ at pre-test may permit stronger causal conclusions on a bilingual EC effect (Bialystok, 2017; Cespón & Carreiras, 2020).

Third, it is important to move away from simple binary classifications of bilinguals and monolinguals, and address the inherent heterogeneity that characterizes the bilingual experience (see Bialystok, 2017; de Bruin et al., 2021; García-Pentón et al., 2016; Luk & Bialystok, 2013; Navarro-Torres et al., 2021; i.a., for previous, more extensive discussions). Bilingualism is a multidimensional construct. Dimensions of bilingualism include, for example, similarity of, age of onset of acquisition of, exposure to, proficiency in, and patterns of use of two or more languages. Thus, it is possible that different (degrees of) bilingual characteristics have quantitatively and qualitatively varying effects on neuro-cognition. Most of the reviewed work, however, has mainly focused on bilingual-monolingual comparisons; and diversity in the characteristics of bilingual and monolingual samples in the different studies (see Table S3) possibly contributes to the inconsistent results. Moreover, studies in this review that directly examined the impact of individual bilingual experiences indicate some trends – particularly, for an L2 proficiency effect – but, generally, no conclusive evidence on which specific bilingual experiences affect neural EC, and on whether these associations reflect positive effects. The inconclusive evidence in these studies is possibly due to this question often being secondary to the main bilingual-monolingual comparison; and, perhaps relatedly, due to low statistical power. Thus, there is need for well-powered studies that will examine the independent effects of different bilingual characteristics either on a continuous scale (e.g., from low to high L2 proficiency) or by forming "extreme" groups that differ only on the specific bilingual experience of interest (e.g., early balanced bilinguals in single-language vs. early balanced bilinguals in dual-language context). To better achieve this, however, it is necessary to have clear theories and a priori hypotheses, as, for example, in the adaptive control hypothesis (Green & Abutalebi, 2013), about which bilingual experiences are relevant to which EC aspects.

Fourth, it would be beneficial to design studies that enable the examination of multiple components – selected based on particular theories and specific hypotheses – in EC tasks. This is because the pattern of results for two or more ERPs is often theoretically more informative in terms of interpretation and the locus of a potential bilingual effect. For instance, examining a cue- (e.g., P3), target- (e.g., N2), and response-locked ERP (e.g., ERN) can provide information on whether a bilingual effect lies in proactive, reactive control, and on whether possible reactive control differences are located, for instance, in monitoring or post-conflict regulative control. Also, this would allow to test whether a bilingual effect is found in more than one process, and to examine accounts such as the BAPSS, which posits different bilingual effects on response-locked (and later stimulus-locked) compared to relatively earlier stimulus-locked ERPs. Moreover, examination of cue- and target-locked ERPs would be useful in interpreting effects as a simple strategy difference or as showing greater efficiency or benefits (e.g., DMC prediction 11 in Table S2 (OSM)). Finally, future studies should scrutinize other accounts on the locus of the bilingual neurocognitive effect, such as the executive attention (Bialystok, 2017) and bilingual expertise (Incera & McLennan, 2016) proposals, which were not directly considered in this review.

Conclusion

This review examined 24 published studies that investigated the effect of bilingualism-related experiences on executive control using the ERP method. The ERP technique records brain activity directly and with high temporal accuracy. Thus, it can provide critical evidence for predictions casted in terms of timing, such as that bilingualism positively impacts the speed and efficiency of EC. This review evaluated the evidence based on considerations of neural efficiency, different EC theories, and different accounts regarding the locus of the bilingual neuro-cognitive effect. Most studies focused on the N2 and P3. Other components have been also examined (e.g., N1, P2, N450, ERN/CRN), but to a lesser extent. Most studies included young adults and used inhibition tasks. Overall, this review finds some evidence for a positive (likely small) bilingual effect. This effect is more consistent for the P3 and response-locked ERP components (including the ERN). Moreover, the bilingual effect is inconsistently found in various EC domains. When considering each ERP component independently, most positive evidence supports primarily a monitoring and, secondarily, an inhibition account. Moreover, an N2/ERN dissociation (no bilingual effect on N2 but positive effect at the ERN, evident as smaller bilingual ERN), coupled with the P3 results, suggest that monitoring may not be the (only) locus of a positive bilingual effect but (an)other post-monitoring, later control mechanism(s). In addition, studies that examined attention disengagement and proactive control generally suggest positive evidence, even though only a few studies have, to date, provided data relevant to these processes; especially, for proactive control. Working memory and switching have been hardly examined, with two studies suggesting a negative bilingual effect on the former and one study indicating a positive effect on the latter process. Finally, the pattern of results at different processing stages is largely consistent with the BAPSS model; specifically, when bilingual effects are found, they often manifest as shorter latencies, larger components or wider amplitude effects during earlier stages of processing (within 200 ms and for the N2 and P3) but as smaller components or narrower amplitude effects at later stages of stimulus (e.g., N450, LNP) and response processing (presaccadic positivity, ERN, and error positivity). However, various findings and methodological issues suggest that the evidence from this literature is inconclusive. First, many studies comparing bilinguals and monolinguals suggest null or some even suggest negative or opposite to prediction bilingual effects. Second, the scant evidence on the bilingual characteristics that affect ERPs is, in general, unclear in terms of which specific bilingual experiences positively impact neural processing, while some correlations suggest negative effects. Third, BAPSS is often not supported by studies that examined multiple components within subjects. Finally, this literature is further complicated by methodological challenges, such as small sample sizes and the presence of confounds. Overall, I hope that this review has detected patterns in the data, identified methodological issues, and provided theoretical tools and methodological recommendations that will help advance the neuro-cognitive study of EC and bilingualism using the ERP technique.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-023-02245-x.

Acknowledgements This work was supported from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 800305; and from a Cyprus University of Technology postdoctoral fellowship. I would also like to thank George Spanoudis for reading and providing feedback on an earlier version of the manuscript. I am also grateful to Iva Ivanova and two anonymous reviewers for their helpful and constructive comments on previous versions of the manuscript.

References

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*, 207–245.

Antoniou, K., & Spanoudis, G. (2020). An investigation of the multilingual and bi-dialectal advantage in executive control. In *Proceedings of the 42nd Annual Virtual Meeting of the Cognitive Science Society* (pp. 2050–2056). Curran Associates, Inc..

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition, 149*, 18–30.

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods, 26*, 295.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*, 543–554.

Barac, R., Moreno, S., & Bialystok, E. (2016). Behavioural and electrophysiological differences in executive control between monolingual and bilingual children. *Child Development, 87*, 1277–1290.

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.

Barulli, D., & Stern, Y. (2013). Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. *Trends in Cognitive Sciences, 17*, 502–509.

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

Bialystok, E., & Craik, F. I. (2010). Cognitive and linguistic processing in the bilingual mind. *Current directions in Psychological Science, 19*, 19–23.

Bialystok, E., Craik, F. I., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest, 10*, 89–129.

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.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8*, 539–546.

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J., Weissman, D. H., Notebaert, W., & Egner, T. (2019). Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences, 23*, 769–783.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences, 16*, 106–113.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron, 39*, 713–726.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. *Variation in Working Memory, 75*, 106.

Brysbaert, M. (2019). How many participants do I have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition, 2*, 1–38.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365–376.

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

Cespón, J. (2021). Neural Processing Underlying Executive Functions in Bilinguals: "Heads I Win, Tails You Lose". *Frontiers in Human Neuroscience, 15*, 710905.

Cespón, J., & Carreiras, M. (2020). Is there electrophysiological evidence for a bilingual advantage in neural processes related to executive functions? *Neuroscience & Biobehavioral Reviews, 118*, 315–330.

Cespón, J., Hommel, B., Korsch, M., & Galashan, D. (2020). The neurocognitive underpinnings of the Simon effect: An integrative review of current research. *Cognitive, Affective, & Behavioral Neuroscience, 20*, 1133–1172.

Chen, P., Bobb, S. C., Hoshino, N., & Marian, V. (2017). Neural signatures of language co-activation and control in bilingual spoken word comprehension. *Brain Research, 1665*, 50–64.

Chiew, K. S., & Braver, T. S. (2017). Context processing and cognitive control: From gating models to dual mechanisms. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 143–166). Wiley Blackwell.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology, 69*, 817–828.

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition, 40*, 693–706.

Clawson, A., Clayson, P. E., Keith, C. M., Catron, C., & Larson, M. J. (2017). Conflict and performance monitoring throughout the lifespan: An event-related potential (ERP) and temporospatial component analysis. *Biological Psychology, 124*, 87–99.

Clayson, P. E., & Larson, M. J. (2011a). Conflict adaptation and sequential trial effects: Support for the conflict monitoring theory. *Neuropsychologia, 49*, 1953–1961.

Clayson, P. E., & Larson, M. J. (2011b). Effects of repetition priming on electrophysiological and behavioral indices of conflict adaptation and cognitive control. *Psychophysiology, 48*, 1621–1630.

Clayson, P. E., & Larson, M. J. (2012). Cognitive performance and electrophysiological indices of cognitive control: A validation study of conflict adaptation. *Psychophysiology, 49*, 627–637.

Clayson, P. E., & Miller, G. A. (2017). Psychometric considerations in the measurement of event-related brain potentials: Guidelines for measurement and reporting. *International Journal of Psychophysiology, 111*, 57–67.

Clayson, P. E., Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology, 56*, e13437.

Clayson, P. E., Brush, C. J., & Hajcak, G. (2021). Data quality and reliability metrics for event-related potentials (ERPs): The utility of subject-level reliability. *International Journal of Psychophysiology, 165*, 121–136.

Coderre, E. L., & van Heuven, W. J. (2014). Electrophysiological explorations of the bilingual advantage: evidence from a Stroop task. *PLoS One, 9*(7), e103424.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Erlbaum.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience, 15*, 336–345.

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.

Dash, T., Berroir, P., Ghazi-Saidi, L., Adrover-Roig, D., & Ansaldo, A. I. (2021). A new look at the question of the bilingual advantage: Dual mechanisms of cognitive control. *Linguistic Approaches to Bilingualism, 11*, 520–550.

Davidson, G. D., & Pitts, M. A. (2014). Auditory event-related potentials associated with perceptual reversals of bistable pitch motion. *Frontiers in Human Neuroscience, 8*, 572.

De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science, 26*, 99–107.

De Bruin, A., Dick, A. S., & Carreiras, M. (2021). Clear theories are needed to interpret differences: Perspectives on the bilingual advantage debate. *Neurobiology of Language, 2*, 433–451.

De Houwer, A., & Ortega, L. (Eds.). (2018). *The Cambridge Handbook of Bilingualism*. Cambridge University Press.

De Pisapia, N., & Braver, T. S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing, 69*, 1322–1326.

Dong, Y., & Zhong, F. (2017). Interpreting experience enhances early attentional processing, conflict monitoring and interference suppression along the time course of processing. *Neuropsychologia, 95*, 193–203.

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*, 1122–1147.

Downes, M., Bathelt, J., & De Haan, M. (2017). Event-related potential measures of executive functioning from preschool to adolescence. *Developmental Medicine & Child Neurology, 59*, 581–590.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin, 145*, 508.

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 380–390.

Egner, T. (2014). Creatures of habit (and control): A multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology, 5*, 1247.

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

Evdokimidis, I., Liakopoulos, D., Constantinidis, T. S., & Papageorgiou, C. (1996). Cortical potentials with antisaccades. *Electroencephalography and Clinical Neurophysiology, 98*, 377–384.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behaviour Research Methods, 39*, 175–191.

Fernandez, M., Tartar, J. L., Padron, D., & Acosta, J. (2013). Neurophysiological marker of inhibition distinguishes language groups on a non-linguistic executive function test. *Brain and Cognition, 83*, 330–336.

Fernandez, M., Acosta, J., Douglass, K., Doshi, N., & Tartar, J. L. (2014). Speaking two languages enhances an auditory but not a visual neural marker of cognitive inhibition. *AIMS Neuroscience, 1*, 145.

Festman, J., & Münte, T. F. (2012). Cognitive control in Russian–German bilinguals. *Frontiers in Psychology, 3*, 115.

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

Gajewski, P. D., Ferdinand, N. K., Kray, J., & Falkenstein, M. (2018). Understanding sources of adult age differences in task switching: Evidence from behavioral and ERP studies. *Neuroscience & Biobehavioral Reviews, 92*, 255–275.

García-Pentón, L., Fernandez Garcia, Y., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016). The neuroanatomy of bilingualism: how to turn a hazy view into the full picture. *Language, Cognition and Neuroscience, 31*, 303–327.

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN/Ne). In S. J. Luck & E. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 231–291). Oxford University Press.

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex, 10*, 829–839.

Goldsmith, S. F., & Morton, J. B. (2018). Time to disengage from the bilingual advantage hypothesis. *Cognition, 170*, 328–329.

Goulet, M. A., & Cousineau, D. (2019). The power of replicated measures to increase statistical power. *Advances in Methods and Practices in Psychological Science, 2*, 199–213.

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.

Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology, 55*, e13016.

Gray, J. R., Burgess, G. C., Schaefer, A., Yarkoni, T., Larsen, R. J., & Braver, T. S. (2005). Affective personality differences in neural processing efficiency confirmed using fMRI. *Cognitive, Affective, & Behavioral Neuroscience, 5*, 182–190.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67–81.

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

Grosjean, F., & Li, P. (2013). *The Psycholinguistics of Bilingualism*. Wiley-Blackwell.

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., & 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., & Shedden, J. M. (2014). A role for recency of response conflict in producing the bivalency effect. *Psychological Research, 78*, 679–691.

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

Grundy, J. G., Benarroch, M. F., Woodward, T. S., Metzak, P. D., Whitman, J. C., & Shedden, J. M. (2013). The Bivalency effect in task switching: Event-related potentials. *Human Brain Mapping, 34*, 999–1012.

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.

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.

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.

Hajcak, G., Meyer, A., & Kotov, R. (2017). Psychometrics and the neuroscience of individual differences: Internal consistency limits between-subjects effects. *Journal of Abnormal Psychology, 126*, 823.

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

Heidlmayr, K., Hemforth, B., Moutier, S., & Isel, F. (2015). Neurodynamics of executive control processes in bilinguals: Evidence from ERP and source reconstruction analyses. *Frontiers in Psychology, 6*, 821.

Heidlmayr, K., Doré-Mazars, K., Aparicio, X., & Isel, F. (2016). Multiple language use influences oculomotor task performance: Neurophysiological evidence of a shared substrate between language and motor control. *PloS one, 11*.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review, 18*, 625–658.

Houtzager, N., Lowie, W., Sprenger, S., & De Bot, K. E. E. S. (2017). A bilingual advantage in task switching? Age-related differences between German monolinguals and Dutch-Frisian bilinguals. *Bilingualism: Language and Cognition, 20*, 69–79.

Hsieh, S., & Fang, W. (2012). Elderly adults through compensatory responses can be just as capable as young adults in inhibiting the flanker influence. *Biological Psychology, 90*, 113–126.

Incera, S. (2018). Measuring the Timing of the Bilingual Advantage. *Frontiers in Psychology, 9*, 1983.

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

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta psychologica, 127*, 277–288.

Jamadar, S. D., Thienel, R., & Karayanidis, F. (2015). Task switching processes. *Brain Mapping: An Encyclopedic Reference, 3*, 327–335.

Jared, D. (2015). What is the theory? *Cortex, 73*, 361–363.

Jiao, L., Zhang, Y., Plummer, P., Liu, C., & Chen, B. (2019). The influence of bilingual language experience on executive control: An ERPs study. *Journal of Neurolinguistics, 51*, 42–52.

Johnstone, S. J., Barry, R. J., Markovska, V., Dimoska, A., & Clarke, A. R. (2009). Response inhibition and interference control in children with AD/HD: A visual ERP investigation. *International Journal of Psychophysiology, 72*, 145–153.

Kałamała, P., Drożdżowicz, A., Szewczyk, J., Marzecová, A., & Wodniecka, Z. (2018a). Task strategy may contribute to performance differences between monolinguals and bilinguals in cognitive control tasks: ERP evidence. *Journal of Neurolinguistics, 46*, 78–92.

Kałamała, P., Szewczyk, J., Senderecka, M., & Wodniecka, Z. (2018b). Flanker task with equiprobable congruent and incongruent conditions does not elicit the conflict N2. *Psychophysiology, 55*, e12980.

Kappenman, E. S., & Luck, S. J. (2012). ERP Components: The Ups and Downs of Brainwave Recordings. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 1–30). Oxford University Press.

Kappenman, E. S., & Luck, S. J. (2016). Best practices for event-related potential research in clinical populations. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1*, 110–115.

Karayanidis, F., & Jamadar, S. D. (2014). Event-related potentials reveal multiple components of proactive and reactive control in task switching. In J. Grange & G. Houghton (Eds.), *Task switching and cognitive control* (pp. 200–236). Oxford University Press.

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin, 144*, 1147.

Key, A. P. F., Dove, G. O., & Maguire, M. J. (2005). Linking brainwaves to the brain: An ERP primer. *Developmental Neuropsychology, 27*, 183–215.

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

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.

Larson, M. J., Kaufman, D. A., & Perlstein, W. M. (2009). Neural time course of conflict adaptation effects on the Stroop task. *Neuropsychologia, 47*, 663–670.

Larson, M. J., Farrer, T. J., & Clayson, P. E. (2011). Cognitive control in mild traumatic brain injury: Conflict monitoring and conflict adaptation. *International Journal of Psychophysiology, 82*, 69–78.

Larson, M. J., Clayson, P. E., & Clawson, A. (2014). Making sense of all the conflict: A theoretical review and critique of conflict-related ERPs. *International Journal of Psychophysiology, 93*, 283–297.

Larson, M. J., Clayson, P. E., Keith, C. M., Hunt, I. J., Hedges, D. W., Nielsen, B. L., & Call, V. R. (2016). Cognitive control adjustments in healthy older and younger adults: Conflict adaptation, the error-related negativity (ERN), and evidence of generalized decline with age. *Biological Psychology, 115*, 50–63.

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.

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

Lowe, C. J., Cho, I., Goldsmith, S. F., & Morton, J. B. (2021). The bilingual advantage in children's executive functioning is not related to language status: A meta-analytic review. *Psychological Science, 32*, 1115–1146.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.

Luck, S. J., & Kappenman, E. (2012). ERP components and Selective Attention. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 295–327). Oxford University Press.

Luck, S. J., Stewart, A. X., Simmons, A. M., & Rhemtulla, M. (2021). Standardized measurement error: A universal metric of data quality for averaged event-related potentials. *Psychophysiology, 58*(6), e13793.

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

Mahé, G., Doignon-Camus, N., Dufour, A., & Bonnefond, A. (2014). Conflict control processing in adults with developmental dyslexia: an event related potentials study. *Clinical Neurophysiology, 125*, 69–76.

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*, 49–100.

Monnier, C., Boiché, J., Armandon, P., Baudoin, S., & Bellocchi, S. (2022). Is bilingualism associated with better working memory capacity? A meta-analysis. *International Journal of Bilingual Education and Bilingualism, 25*, 2229–2255.

Morales, J., Yudes, C., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. *Neuropsychologia, 66*, 157–169.

Moreno, S., Wodniecka, Z., Tays, W., Alain, C., & Bialystok, E. (2014). Inhibitory control in bilinguals and musicians: event related potential (ERP) evidence for experience-specific effects. *PloS one, 9*.

Morrison, C., Kamal, F., & Taler, V. (2019). The influence of bilingualism on working memory event-related potentials. *Bilingualism: Language and Cognition, 22*, 191–199.

Mueller, S. C., Swainson, R., & Jackson, G. M. (2009). ERP indices of persisting and current inhibitory control: A study of saccadic task switching. *Neuroimage, 45*, 191–197.

Navarro-Torres, C. A., Beatty-Martínez, A. L., Kroll, J. F., & Green, D. W. (2021). Research on bilingualism as discovery science. *Brain and Language, 222*, 105014.

Naylor, L. J., Stanley, E. M., & Wicha, N. Y. (2012). Cognitive and electrophysiological correlates of the bilingual Stroop effect. *Frontiers in Psychology, 3*, 81.

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience, 3*, 17–26.

Paap, K. R. (2019). *The bilingual advantage debate: Quantity and quality of the evidence*. In J. W. Schwieter (Ed.), *The handbook of the neuroscience of multilingualism* (pp. 701–735). Wiley-Blackwell.

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

Paxton, J. L., Barch, D. M., Racine, C. A., & Braver, T. S. (2008). Cognitive control, goal maintenance, and prefrontal function in healthy aging. *Cerebral Cortex, 18*, 1010–1028.

Pires, L., Leitão, J., Guerrini, C., & Simões, M. R. (2014). Event-related brain potentials in the study of inhibition: Cognitive control, source localization and age-related modulations. *Neuropsychology Review, 24*, 461–490.

Pliatsikas, C. (2019). Multilingualism and brain plasticity. In J. W. Schweiter (Ed.), *The Handbook of the Neuroscience of Multilingualism* (pp. 230–251). Wiley-Blackwell.

Pliatsikas, C., & Luk, G. (2016). Executive control in bilinguals: A concise review on fMRI studies. *Bilingualism: Language and Cognition, 19*, 699–705.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., ... & Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience, 18*, 115-126.

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

Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 159–188). Oxford University Press.

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

Rey-Mermet, A., Gade, M., & Steinhauser, M. (2019). Sequential conflict resolution under multiple concurrent conflicts: An ERP study. *NeuroImage, 188*, 411–418.

Roche, R. A., Garavan, H., Foxe, J. J., & O'Mara, S. M. (2005). Individual differences discriminate event-related potentials but not performance during response inhibition. *Experimental Brain Research, 160*, 60–70.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and US preschoolers. *Psychological Science, 17*, 74–81.

Samuel, S., Roehr-Brackin, K., Pak, H., & Kim, H. (2018). Cultural effects rather than a bilingual advantage in cognition: A review and an empirical study. *Cognitive Science, 42*, 2313–2341.

Santillán, J., & Khurana, A. (2018). Developmental associations between bilingual experience and inhibitory control trajectories in Head Start children. *Developmental Science, 21*, e12624.

Steinhauer, K. (2014). Event-related Potentials (ERPs) in Second Language Research: A Brief Introduction to the Technique, a Selected Review, and an Invitation to Reconsider Critical Periods in L2. *Applied Linguistics, 35*, 393–417.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652–654.

Sullivan, M. D., Janus, M., Moreno, S., Astheimer, L., & Bialystok, E. (2014). Early stage second-language learning improves executive control: Evidence from ERP. *Brain and Language, 139*, 84–98.

Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS biology, 15*, e2000797.

Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage, 221*, 117164.

Timmer, K., Grundy, J. G., & Bialystok, E. (2017). Earlier and more distributed neural networks for bilinguals than monolinguals during switching. *Neuropsychologia, 106*, 245–260.

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

van Veen, V., & Carter, C. S. (2006). Error detection, correction, and prevention in the brain: a brief review of data and theories. *Clinical EEG and Neuroscience, 37*, 330–335.

West, R., & Alain, C. (1999). Event-related neural activity associated with the Stroop task. *Cognitive Brain Research, 8*, 157–164.

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

Wu, Y. J., Zhang, H., & Guo, T. (2016). Does speaking two dialects in daily life affect executive functions? An event-related potential study. *PloS one, 11*, e0150492.

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—Commentary on Vul et al. (2009). *Perspectives on Psychological Science, 4*, 294–298.

Yeung, N., & Cohen, J. D. (2006). The impact of cognitive deficits on conflict monitoring: Predictable dissociations between the error-related negativity and N2. *Psychological Science, 17*, 164–171.

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

Yeung, N., Ralph, J., & Nieuwenhuis, S. (2007). Drink alcohol and dim the lights: The impact of cognitive deficits on medial frontal cortex function. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 347–355.

Zhang, J. X., Wu, R., Kong, L., Weng, X., & Du, Y. (2010). Electrophysiological correlates of proactive interference in the 'Recent Probes' verbal working memory task. *Neuropsychologia, 48*, 2167–2173.

Zhang, H., Kang, C., Wu, Y., Ma, F., & Guo, T. (2015). Improving proactive control with training on language switching in bilinguals. *NeuroReport, 26*, 354–359.

Open practices statement Data sharing is not applicable to this article. The review was not preregistered.