Tracking components of bilingual language control in speech production: an fMRI study using functional localizers.

Agata Wolna[1], Jakub Szewczyk[2], Michele Diaz[3], Aleksandra Domagalik[4], Marcin Szwed[1], Zofia Wodniecka[1]

[1] Institute of Psychology, Jagiellonian University, Kraków, Poland

[2] Donders Institute for Brain Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

[3] Social, Life, and Engineering Sciences Imaging Center, the Pennsylvania State University, Pennsylvania, United States

[4] Centre for Brain Research, Jagiellonian University, Kraków, Poland

Corresponding authors:

zofia.wodniecka@uj.edu.pl

agata.wolna@doctoral.uj.edu.pl

Language and Bilingualism Lab

Instytut Psychologii UJ

ul. Ingardena 6

30-060 Kraków

Abstract

When bilingual speakers switch back to speaking in their native language (L1) after having used their second language (L2), they often experience difficulty in retrieving words in their L1: this phenomenon is referred to as the *L2 after-effect.* We used the L2 after-effect as a lens to explore the neural bases of bilingual language control mechanisms. Our goal was twofold: first, to explore whether bilingual language control draws on domain-general or language-specific mechanisms; second, to investigate the precise mechanism(s) that drive the L2 after-effect. We used a precision fMRI approach based on *functional localizers* to measure the extent to which the brain activity that reflects the L2 after-effect overlaps with the language network (Fedorenko et al., 2010) and the domain-general Multiple Demand network (Duncan et al., 2010), as well as three task-specific networks that tap into interference resolution, lexical retrieval, and articulation. Forty-two Polish-English bilinguals participated in the study. Our results show that the L2 after-effect reflects increased engagement of domain-general but not language-specific resources. Furthermore, contrary to previously proposed interpretations, we did not find evidence that the effect reflects increased difficulty related to lexical access, articulation, and the resolution of lexical interference. We propose that difficulty in speech production – manifested as the L2 after-effect – reflects interference at a non-linguistic level of task schemas or a general increase of engagement of cognitive control in speech production in L1 after L2.

Keywords: fMRI, language production, bilingualism, language control, cognitive control

Highlights

- We use fMRI to explore brain correlates of the L2 after-effect
- The L2 after-effect is a hallmark of proactive bilingual control
- The effect overlaps with the Multiple Demand network but not the language network
- It likely reflects domain-general but not language-specific difficulty
- It likely reflects interference between higher-level task schemas

Introduction

In their daily lives, bilinguals have to face a constant challenge related to the non-selectivity or constant co-activation of their two languages (Costa et al., 1999; Kroll et al., 2006). To cope with this problem, bilinguals must either have an efficient system of control that resolves interferences between the competing languages (Green, 1998), or a selection mechanism (Blanco-Elorrieta & Caramazza, 2021) that allows them to efficiently use the intended language. In this study we aimed to characterize the neurocognitive mechanisms of control that are engaged in bilingual speech production by capitalizing on the advantages of the functional localizer approach.

Bilingual language control: domain-general or domain-specific?

One of the most fundamental questions regarding the complex system that bilinguals use to control their two languages is whether it is based on a set of mechanisms specifically tailored to handle language (domain-specific), or whether bilingual language control goals are achieved by domain-general cognitive control mechanisms (for a review see: Calabria et al., 2018).

In principle, bilingual language control could rely on the neural network involved in any language processing task (Blanco-Elorrieta & Caramazza, 2021). Language-specific[1] computations, which are putatively supported by the language system, could carry out (at least some of) the tasks related to controlling the two languages of a bilingual. Based on previous findings, it can be argued that the language system supports a set of core language-specific computations that make it possible to monitor speech production (Indefrey, 2011; Indefrey & Levelt, 2004), detect and correct errors (Nozari & Novick, 2017), resolve interference between competing representations (Ivanova & Hernandez, 2021), and select and retrieve lexico-semantic representations (Hu et al., 2022; Nozari & Novick, 2017). Neuroimaging shows that, at the brain level, all these processes recruit a common set of frontotemporal areas located along the lateral sulcus, which together have been dubbed ‘the core language network’ (Fedorenko et al., 2010). This network has a robust specificity for language: it is engaged in reading (Fedorenko et al., 2010), listening (e.g., Pallier et al., 2003; Scott et al., 2017), and producing language (Hu et al., 2022; Liljeström et al., 2015), but it shows little engagement in non-linguistic tasks (Benn et al., 2021; Chen et al., 2021; Fedorenko et al., 2011; Ivanova et al., 2021) or social reasoning (Shain et al., 2022). Importantly, it has been demonstrated that linguistic stimuli activate the language network proportionally to the degree of difficulty of their processing (Wehbe et al., 2021), but at the same time this processing difficulty does not affect the domain-general Multiple

[1] Throughout the text, when using the term 'language-specific' we refer to processes dedicated to language processing, (as opposed to 'domain-general processing’) and not to a specific language, e.g., Polish or English etc.

Demand network (Duncan et al., 2010; see below). This indicates that language processing, even if cognitively demanding, is mainly subserved by the language network (Wehbe et al., 2021). If a bilingual's use of the intended language boils down to regulating the activation of language representations at different levels of processing (i.e., semantic, lexical, morphosyntactic, phonological and syntactic representations) and adjusting the activation thresholds of lexical selection to meet the current contextual requirements (Blanco-Elorrieta & Caramazza, 2021; Branzi et al., 2014), then the language network should be sufficient to do the job.

Alternatively, bilingual language control could be implemented by a broader, domain-general mechanism (e.g. Calabria et al., 2018; Green & Abutalebi, 2013; Wu et al., 2019). Domain-general cognitive control comprises a wide range of general cognitive functions such as attention, problem-solving, working memory, and inhibition that jointly monitor and control goal-driven behavior (Mackie, Can Dam & Fan, 2013), task-set maintenance and general flexibility in the allocation of mental resources (Dosenbach et al., 2006)[2]. This wide range of cognitive functions has been shown to consistently activate a broad frontoparietal network in the brain (Badre et al., 2005; Dosenbach et al., 2006, 2007; Kim et al., 2012; Smith et al., 2021), also referred to as the *cognitive control network* (Cole & Schneider, 2007) or the *Multiple Demand* network (Duncan, 2010). This network includes the anterior (Niendam et al., 2012) and the dorsolateral prefrontal cortex (Chein & Schneider, 2005; Cole & Schneider, 2007; Dosenbach et al., 2006; Niendam et al., 2012), the anterior cingulate cortex / pre-supplementary motor area (Cole & Schneider, 2007; Dosenbach et al., 2006; Niendam et al., 2012), the superior and inferior parietal cortices (Chein & Schneider, 2005; Cole & Schneider, 2007; Niendam et al., 2012) as well as subcortical structures such as the caudate, putamen, thalamus and cerebellum (Niendam et al., 2012). The cognitive control network has been shown to be sensitive primarily to changes in task demands (e.g., Fedorenko et al., 2011). Activity within this network has been shown to decrease with the increasing automatization of task performance (Chein & Schneider, 2005) and increase as a function of task complexity (Crittenden & Duncan, 2014; Fedorenko et al., 2013; Shashidhara et al., 2019), time pressure and reward (Shashidhara et al., 2019). Importantly, the frontoparietal Multiple Demand network shows little or no domain-specificity (Assem, Glasser, et al., 2020): it is robustly activated in a range of different linguistic and non-linguistic cognitive tasks (Fedorenko et al., 2013; Dosenbach et al., 2006; Shashidhara et al., 2019), including low-level perceptual tasks (Cole & Schneider, 2007).

[2] Other more constrained definitions of cognitive control point to its role in detecting and resolving conflicts during information processing. Framed this way, cognitive control is a sub-system of the executive functions: a family of cognitive mechanisms that manage information processing in the brain and support the flexible guidance of goal-directed behavior (Salig et al., 2021). In this paper we will refer to *cognitive control* in a wider sense that is synonymous to *executive functions*.

So far, most evidence for the engagement of domain-general control mechanisms in bilingual language production comes from behavioral studies in which bilinguals were required to switch between their languages. The overall conclusion from these experiments is that switching between two languages incurs additional processing costs (Costa & Santesteban, 2004; Gollan & Ferreira, 2009; Meuter & Allport, 1999; Verhoef et al., 2009; for review see: Bobb & Wodniecka, 2013; Declerck et al., 2020; Declerck & Koch, 2022; Declerck & Philipp, 2015). This additional cognitive effort has also been observed as increased neural activation in structures associated with broadly defined cognitive control mechanisms (e.g., Abutalebi & Green, 2007; Coderre et al., 2016; De Baene et al., 2015; Wu et al., 2019; for reviews see: Abutalebi & Green, 2016; Luk et al., 2012; Sulpizio et al., 2020). It has been argued that bilingual language control is supported by domain-general mechanisms because task switching (which does recruit domain-general control) and language switching share a number of computational processes, e.g., context monitoring, reconfiguring task-sets, engaging and disengaging from a task, suppressing the interference between competing representations and selective response inhibition (Green, 1998; Green & Abutalebi, 2013; Wu et al., 2019). This is also supported by evidence for the overlap between brain networks that support language switching and task switching (Blanco-Elorrieta & Pylkkänen, 2016; Coderre et al., 2016; De Baene et al., 2015; Kim et al., 2012; Luk et al., 2012; Mendoza et al., 2021; Nee et al., 2007; Niendam et al., 2012; Weissberger et al., 2015; Wu et al., 2019).

On the other hand, the overlap between brain networks activated by language- and task-switching may not be the most adequate evidence for the domain-general nature of bilingual language control. First, some evidence shows that language switching also engages brain structures that are not implicated in non-linguistic switching tasks (De Baene et al., 2015; Mendoza et al., 2021; Weissberger et al., 2015). Second, the overlap between the brain networks recruited by language- and task-switching may result from idiosyncratic properties of the switching task itself rather than bilingual language control in general. More specifically, it can be argued that language switching is an “artificial” task in the sense that it requires changing the language in use at a very frequent rate and depending on the external cues, while switching in natural contexts is dictated more by the speaker (i.e., internal cues; for a discussion on how cues affect switch costs, see Blanco-Elorrieta & Pylkkänen, 2017 for natural vs. artificial cues and Gollan et al., 2014 and Gollan & Ferreira, 2009 for voluntary vs. forced cues). As such, a language switching task may require the use of a very specific task set and keeping both languages co-activated. Such extra demands may recruit mechanisms engaged in performing tasks that are untrained, difficult, or not automated (Chein & Schneider, 2005; Crittenden & Duncan, 2014; Shashidhara et al., 2019), and these types of tasks are known to reliably activate the domain-general Multiple Demand network. What is more, bilinguals rarely switch between their languages on

a word-to-word basis (Wodniecka et al., 2020; see also: Bobb & Wodniecka, 2013); forced switching is related to greater engagement of cognitive control than voluntary switching, in which speakers can choose the language of utterance at will (*for behavioral evidence see:* Gollan et al., 2014; Gollan & Ferreira, 2009; *for fMRI evidence see:* Reverberi et al., 2018; Zhang et al., 2015).

In summary, it remains an open question which brain networks underlie bilingual language control, especially if the control is tested in paradigms that do not require frequent switching and, as such, do not confound mechanisms that subserve task-switching with controlling the two languages of a bilingual.

Digging deeper – specific processes involved in bilingual language control

Besides the more general question of whether bilinguals rely on domain-general or language-specific networks when using their two languages, another unresolved issue is what specific processes are involved in bilingual language control. Addressing this question requires sketching a more detailed model of bilingual language production to identify specific processes that may be involved in controlling the two languages.

Models of bilingual language production assume that each language of a bilingual speaker is subserved by a set of different lexical, phonological, and articulatory representations (e.g., Blanco-Elorrieta & Caramazza, 2021; Costa, 2005; La Heij, 2005). Successful production requires using representations that belong to the target language but not to the non-target language. In the case of bilingual speakers (especially L1-dominant speakers), it is typically assumed that using their native language (L1) does not require additional control over representations of their second language (L2) because L1 representations are readily available and receive no interference from L2 representations. The situation becomes more complicated when an L1-dominant bilingual wants to use their weaker L2. The representations of L2 are less active, so accessing them may require help from control processes. In addition, representations of L2 receive interference from strongly activated L1 representations. The most prominent theory of bilingual language control proposes that resolution of between-language interference is implemented by active inhibition of L1 representations (Green, 1998); however, it does not specify at which level of word processing (e.g., lexical, phonological, articulatory) the inhibition is implemented. On the other hand, other theoretical models argue that inhibition is not necessary for the successful selection of the target language (Blanco-Elorrieta & Caramazza, 2021; Costa, 2005; Koch et al., 2010). For example, the persisting activation account assumes that when a bilingual wants to use the weaker L2, its representations receive additional activation, which allows them to surpass the level of activation of the stronger L1 (for a review see: Koch et al., 2010). A recent extension of the activation-based account proposed that activation of

target representations does not operate on a whole-language level (i.e., it does not apply to all items of L1 or L2) but it is rather a sum of activations driven by different factors, such as baseline word frequency, recency of use and communicative context (Blanco-Elorrieta & Caramazza, 2021). As such, both bilingual and monolingual speech production can be explained by the same selection mechanisms that select the most active representation at a given moment. An important assumption of this proposal is that this selection applies to all linguistic levels of processing, i.e., semantic, lexical, morphosyntactic, phonological, and syntactic representations.

The available theoretical models of bilingual word production do not provide a unanimous response to the question of the nature of bilingual language control (domain-general vs. language specific), its mechanisms (inhibition vs. activation) nor its locus (lexico-semantic, phonological, or articulatory representations).

The L2 after-effect as a window to study neurocognitive mechanisms of bilingual language control

As has been already indicated, the lion's share of evidence on the involvement of control in bilingual language production comes from studies that involved the language-switching paradigm (for a recent review see Sulpizio et al., 2020) which – as we argued above - has several important limitations. Some of these limitations are circumvented in a blocked language picture naming task that does not require as frequent switches between languages as the language switching task and yet allows control mechanisms involved in bilinguals' language production to be tested. Several studies have observed that completing a task in L2 results in a subsequent slow-down of naming or an increase of errors committed in L1 (Branzi et al., 2014; Degani et al., 2019; Gollan & Ferreira, 2009; Van Assche et al., 2013; Wodniecka, Szewczyk, et al., 2020). This difficulty in using L1 after a short exposure to L2 has been referred to as the L2 after-effect. Any task involving language production in L2 can lead to the L2 after-effect: simply reading a list of words aloud (Degani et al., 2019), naming a set of pictures (Branzi et al., 2014, 2016; Gollan & Ferreira, 2009; Guo et al., 2011; Wodniecka et al., 2020) or coming up with words belonging to a given semantic category (Van Assche et al., 2013).

Available theories of bilingual speech production provide different explanations of the L2 after-effect. According to inhibition-based accounts, the retrieval difficulty that speakers experience in L1 after using L2 is a lingering consequence of the strong inhibition that had to be applied to L1 representations during L2 use (Green, 1998; Guo et al., 2011; Misra et al., 2012). Under the inhibitory accounts, the control mechanism responsible for the resolution of the interference that drives the L2 after-effect relies on a set of domain-general mechanisms (e.g., Abutalebi & Green, 2016; Green, 1998; Green & Abutalebi, 2013). An alternative proposal assumes that the L2 after-effect is driven by

increased interference between the representations of L1 and L2 due to the carry-over effects of increased L2 activation (Branzi et al., 2014, 2016; Koch et al., 2010). The common selection mechanism account (Blanco-Elorrieta & Caramazza, 2021) offers a similar explanation: within this framework, the L2 after-effect can also be explained as a consequence of increased activation of L2 representations (due to their recent use), thus leading to increased interference between L2 and L1. Under this account, the selection mechanism is common across the languages known to a bilingual; as such, it can be assumed to be a part of a domain-specific language system (Blanco-Elorrieta & Caramazza, 2021).

To date, the origin of the L2 after-effect has been explored in three neuroimaging studies using fMRI. They revealed that the difficulty in word retrieval in L1 after using L2 results in increased engagement of brain regions often linked to domain-general control mechanisms (Branzi et al., 2016; Guo et al., 2011; Rossi et al., 2018). An experiment by Guo et al (2011) found that naming pictures in L1 after L2 as compared to an L1 baseline led to increased brain activation in brain regions that have been linked to control mechanisms, such as interference suppression (left middle frontal gyrus, left inferior parietal gyrus), and visual attention shifts (left parietal cortex). Additionally, this study found increased brain activity corresponding to the L2 after-effect in brain regions linked to articulation (right postcentral gyrus). Results that partially overlap with those of Guo et al. (2011) have been reported by an fMRI study by Branzi et al. (2016), which showed that the L2 after-effect is linked to increased activity in structures that neurocognitive models of bilingual language control (Abutalebi & Green, 2007; Green & Abutalebi, 2013; Luk et al., 2012) have linked to response inhibition (right inferior frontal cortex), response selection (left prefrontal cortex), and attention (bilateral inferior parietal lobules). Finally, in a study by Rossi et al. (2018), increased brain activity related to speech production in L1 after L2 in bilingual speakers compared to a group of monolingual speakers was found in a network of structures that have been previously linked (among many other functions) to inhibitory control (right frontal pole, right superior parietal cortex) and conflict monitoring and resolution (bilateral anterior cingulate cortex). To sum up, previous studies looking at the brain mechanisms behind the L2 after-effect have identified a broad network of brain regions. While most of these structures have been linked to several different domain-general control mechanisms, some structures are thought to support language-specific processing as well as articulation. As such, the results of previous studies do not unanimously identify the mechanisms that drive the L2 after-effect.

Capturing the neural basis of the L2 after-effect: methodological challenges

Importantly, three studies that explored the L2 after-effect were not free of methodological challenges. In each, certain methodological choices may have limited the scope of possible conclusions. Two out of three studies report neural contrasts that involved different samples of participants (Guo

et al., 2011; Rossi et al., 2018). This could be problematic as the engagement of language control mechanisms may be dependent on some aspects of language experience (see Casado et al., 2022 for evidence of the influence of language balance on the magnitude of engagement of global language control). What is more, a study by Rossi et al. (2018) compared a sample of bilinguals to monolinguals, but this does not make it possible to disentangle the effects of previous use of L2 from more general differences between bilingual and monolingual language processing. Branzi and colleagues (2016) addressed these limitations by comparing preceding context effects within the same bilingual participants; however, in this study, the baseline L1 block, which was compared to L1 after L2 in order to estimate the L2 after-effect, was in some participants preceded by a task in L2. Since the L2 after-effect is relatively long-lasting (at least 5 minutes, see Branzi et al., 2014; Casado et al., 2021; Wodniecka, Szewczyk, et al., 2020; for review see: Wodniecka, et al., 2020), a baseline for comparison (i.e., L1 naming) should not be preceded by a task in L2. It is therefore possible that the observed L2 after-effect was confounded by factors such as training, fatigue, or accumulating semantic interference (Runnqvist et al., 2012; for a discussion of how trial-related factors can affect the measurement of L2 after-effect see: Casado et al., 2022).

Moreover, previous studies in the area of bilingualism have relied on group-level fMRI analysis. This approach focuses on the commonalities in activations between participants, but this makes this approach less sensitive to individual variability in brain organization. At the same time, the language experience of bilingual speakers, which is subject to large inter-individual variability (related to a number of factors, such as immersion, age of acquisition, proficiency, daily use of language), translates into changes in the functional and structural organization of the brain (e.g., DeLuca et al., 2019, 2020; Fedeli et al., 2021). Therefore, we may expect much more variability in the functional organization of a bilingual brain than we usually see in monolingual populations. An analytical approach based on the group-level averaging of brain activations may not be sensitive enough to account for the individual variability between subjects, therefore it may miss some of the important characteristics of the neural basis of the tested mechanisms. Another limitation of previous studies is related to the interpretation of the observed activations. While previous studies undoubtedly identified brain structures engaged in the L2 after-effect, the interpretations regarding the precise mechanisms represented by the observed brain response were not directly tested and were proposed based on the function assigned to a given anatomical structure (the so-called *reverse inference* problem, Poldrack, 2006; for a discussion of issues related to this approach in studying language see Fedorenko, 2021). It has been shown that macroanatomical cortical areas are characterized by vast inter-individual variability, which means that the link between anatomical structure and functional organization is not straightforward, especially for high-level cognitive processes such as language (Frost & Goebel, 2012; Tahmasebi et al.,

2012; for a discussion see Fedorenko, 2021; Michon et al., 2022). In the context of language processing, a brain area that has been shown to be functionally heterogeneous is, for example, the left inferior frontal gyrus (*Broca's area*), which includes both non-overlapping language-specific and domain-general subregions (Fedorenko et al., 2012). What is more, the inferior frontal gyrus shows robust inter-individual variability in its anatomical location and size (Juch et al., 2005; Tomaiuolo et al., 1999). As such, a clear dissociation between the functional subsections of Broca's area can be lost when its interpretation is based on anatomically based parcellations. In the current study, we proposed a different approach to interpreting the brain responses related to the L2 after-effect: instead of focusing on group-level whole-brain analysis, we used a precision fMRI approach based on a set of functional localizers (Fedorenko, 2021; Fedorenko et al., 2010).

Current study

In the current study, we explored the nature of bilingual language control by looking at the brain basis of the L2 after-effect. We had two specific goals: first, to determine to what extent the control mechanisms engaged in bilingual language control are domain-general vs. language-specific; second, to zoom in to the specific process that may underlie difficulties in L1 production following L2 use, namely lexical retrieval difficulty (Casado et al., 2022), articulatory difficulty (Guo et al., 2011), and increased interference between languages (Branzi et al., 2014). We addressed these questions by using fMRI and a set of functional localizers, which - to the best of our knowledge - is an approach that has not been used before in studies exploring the mechanisms of language control in bilinguals. A crucial advantage of functional localizers is that they allow the modeling of brain responses at a subject-specific level; this increases the sensitivity of the statistical analyses and models inter-individual variability better than can be done with standard statistics that rely on averaging the brain response at the group level.

To explore the domain-specificity or domain-generality of bilingual language control mechanisms engaged in the L2 after-effect, we used two well-established functional localizers which make it possible to track down subject-specific networks that correspond to the language system (Malik-Moraleda, Ayyash et al., 2022; Fedorenko et al., 2010) and the Multiple Demand network (Duncan, 2010; Fedorenko et al., 2013). The Language localizer is a well-established paradigm that allows the identification of a robust network in the brain which spans the frontal, temporal, and parietal regions, which support the processing of language (e.g., Malik-Moraleda, Ayyash et al., 2022; Fedorenko, 2021; Fedorenko et al., 2010, 2011). The multiple-demand localizer, on the other hand, allows the identification of another robust network in the brain which, at a most general level, responds to increased demands related to performing different types of tasks (Assem, Blank, et al., 2020; Braver et al., 2003; Cole & Schneider, 2007; Dosenbach et al., 2006; Duncan, 2010; Fedorenko

et al., 2011, 2013; Miller & Cohen, 2001). While both the Language and the Multiple Demand networks show vast cross-subject variability (Fedorenko et al., 2010; Smith et al., 2021), they show a clear functional and topographic dissociation on a single-subject level (e.g., Blank et al., 2014; Diachek et al., 2020). As such, if brain activity corresponding to the L2 after-effect is found within one of these two networks, it will provide straightforward evidence for the engagement of domain-general or language-specific mechanisms.

To investigate the specific mechanisms that drive the L2 after-effect, we tested whether brain activity associated with naming in L1 after L2 compared to naming in L1 after L1 overlaps with functional networks recruited by a Stroop task, a Verbal Fluency task, and a simple articulation task. Each of them tapped into one of the previously described mechanisms that possibly underlies the L2 after-effect. The Stroop task (Stroop, 1935) was used to capture brain structures engaged in the resolution of interference and inhibition of non-target responses in speech production (Diamond, 2013). An overlap between these structures and those involved in the L2 after-effect would provide an argument that the L2 after-effect is driven by increased interference between the two languages. We used a classic within-language version of a Stroop-task. Although the interference driving the L2 after-effect may be driven by between-language effects, electrophysiological research has shown that both the within-language (i.e., response language matched with the written word language) and between-language (i.e., response language different from the written word language) congruency effects in the Stroop task elicit similar brain responses in the N450 time window (Naylor et al., 2012). This result suggests that the mechanisms of response interference detection and resolution are the same regardless of whether the stimulus (word) and response (font name) belong to the response set or not[3].

The verbal fluency task was used to identify brain regions that are sensitive to the difficulty of lexical access. As previously discussed, difficulties in lexical access seem to be one of the most parsimonious explanations of the L2 after-effect (Wodniecka, Szewczyk, et al., 2020; Casado et al., 2022). The verbal fluency task is thought to engage cognitive mechanisms that support word retrieval (Birn et al., 2010; Friesen et al., 2021; Shao et al., 2014). Specifically, the verbal fluency task was shown to tap into the efficiency of the lexical search and selection processes (Shao et al., 2014) as well as executive function mechanisms engaged in word retrieval (Friesen et al., 2021; Martin et al., 1994). Therefore, if L1 retrieval difficulty after L2 is due to L1 lexical access being hampered, then the brain

[3] Differences between the response language (L1 vs. L2) have also been shown to affect ERPs; however, the effect of language was observed in an earlier time-window (N2) and showed no interaction with color-word congruence (Naylor et al., 2012). This suggests that language-related effects are separate from the interference-resolution mechanisms that we wanted to target with the Stroop localizer in the current study.

activity corresponding to the L2 after-effect should overlap with the network identified using a verbal fluency localizer. Finally, verbal fluency performance has been shown to depend either on individual differences between subjects that are related to vocabulary size and working memory capacity, or on the mere speed of information processing (Unsworth et al., 2011). Therefore, a subject-specific localizer based on a verbal fluency task allows us to better capture the inter-individual variability in lexical retrieval between subjects.

Finally, to test the extent to which the L2 after-effect is driven by the articulatory difficulty of word production in L1, we used a localizer that allows us to identify brain regions specifically linked to articulation (Basilakos et al., 2018). A fMRI experiment which looked into the brain basis of the L2 after-effect suggested that speaking in L1 after being exposed to L2 entails the engagement of additional articulatory resources (Guo et al., 2011). Moreover, more-recent behavioral evidence (Broos et al., 2018) suggests that greater difficulty in speaking L2 compared to L1 only affects the late articulatory stage; if so, it is also plausible that exposure to L2 results in articulatory difficulties for subsequent L1 use.

To address the limitations of experimental designs previously used to measure the L2 after-effect, we used a within-participant design that carefully controlled for the effects of the preceding language context as well as for possible effects of task order and fatigue. The exact paradigm used in the current study was also tested in a series of behavioral studies (for a discussion see: Casado et al., 2022).

Method

Participants

Forty-two Polish-English late bilinguals took part in the study. One participant was excluded due to excessive head motion (>2mm) in the main fMRI task, resulting in a final sample of forty-one participants (31 females, 10 males, mean age = 23.29, SD = 3.24, range: 19-32). All participants were Polish native speakers who learned English as their second language. To participate in the experiment they had to declare sufficient proficiency in English (B2 or higher) and obtain at least 20/25 points on the General English Test by Cambridge Assessment[4] (mean score = 21.32, SD = 1.94). In the behavioral session, the participants' proficiency in English was assessed again using LexTale (mean score = 71.58%, SD = 9.91%; Lemhöfer & Broersma, 2012). In addition, participants completed a questionnaire in which they rated their proficiency in all known languages with regards to reading, writing, listening, speaking, and accent; participants also gave information on their daily use of all languages and the age at which they started to learn each of them. Detailed information on the participants' proficiency, age of acquisition, and daily use of languages can be found in Table 1. All participants gave written consent for participation in the study. The study was approved by the Ethics Committee of Jagiellonian University Institute of Psychology concerning experimental studies with human subjects.

[4] https://www.cambridgeenglish.org/test-your-english/general-english/

Table 1. Language experience of participants. Information on self-rated proficiency, language learning and daily use of languages is given for all languages that participants declared to know. L1 always refers to Polish and L2 to English. L3 and L4 were a variety of different languages (incl. German, French, Italian, Spanish, Russian, Czech, Japanese, Korean, Norwegian, Latin and Esperanto). Not all of the participants reported knowing an L3 or L4.

Language	**L1** (Polish)	**L2** (English)	**L3** (various)	**L4** (various)
n	41	41	36	17
Self-rated proficiency (1-10)				
reading	9.72	4.34	2.76	1.79
listening	9.62	4.10	2.45	1.84
writing	9.26	3.81	2.21	1.48
speaking	9.49	3.89	2.48	1.68
accent	9.56	3.45	2.45	1.65
Language learning				
Age of acquisition *onset of learning*	0.00	6.44	14.34	15.24
Age of acquisition: *onset of using*	3.17	9.15	15.41	16.88
Years of formal education	14.72	11.56	4.03	4.18
Daily use of languages				
passive	63.26%	31.79%	4.03%	3.25%
active	81.31%	16.31%	2.06%	1.31%
Verbal Fluency (mean number of words produced)				
category	20.04	13.36	-	-
letter	12.64	10.62	-	-

Study Design

The experiment consisted of 3 testing sessions: one behavioral and two MRI scanning sessions. The design of the behavioral and MRI sessions is presented in detail in Figure 1, and the details of each task are discussed in the Design section.

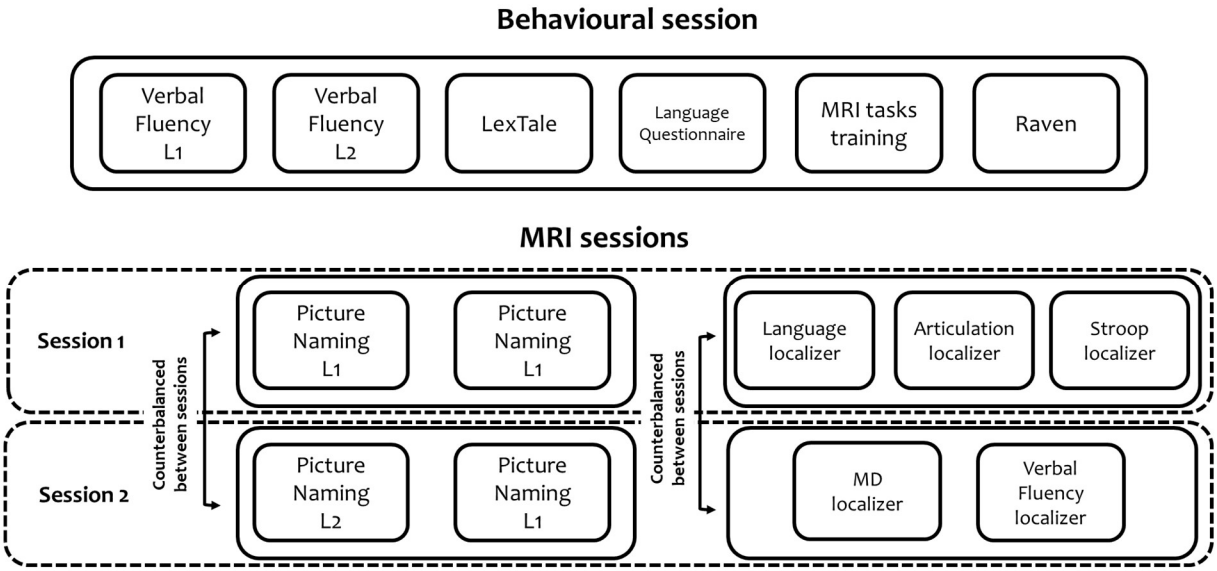

Figure 1 . Summary of the design of the behavioral and MRI sessions. In the behavioral sessions, all participants completed the tasks in a fixed order. In the MRI sessions, the order of conditions completed in the first and the second session (experimental manipulation with L1 or L2) was independently counterbalanced between participants; the set of localizer tasks assigned to each session was also counterbalanced between participants.

In the behavioral session, participants completed a series of tasks, including semantic and phonetic verbal fluency in L1 and L2 (see details below), LexTale in English (Lemhöfer & Broersma, 2012), a language questionnaire[5], and a short version of Raven's Progressive Matrices test (only uneven items) as a measure of a non-verbal intelligence. The behavioral session also included a short training session on the tasks to be performed in the MRI sessions.

During the MRI sessions, participants completed the main task, namely picture naming, and a series of localizer tasks. The main task was split between two sessions, with one condition (L1 or L2 in the first block) completed in the first session and the other condition completed in the second session, with the order counterbalanced across sessions. Localizer tasks included an auditory language localizer task (Malik-Moraleda, Ayyash et al., 2022; Scott et al., 2017), an articulation localizer, a Stroop task, a visual working memory task corresponding to the Multiple Demand localizer (Fedorenko et al., 2013), and a verbal fluency task (order counterbalanced, see Figure 1 for details). In tasks requiring overt spoken responses (i.e., picture naming, Stroop task, articulation task, and verbal fluency) participants' responses were recorded with FOMRI III, which is an MR-compatible, fiber-optic microphone system

[5] The questionnaire in its original Polish version is available under the following link: https://psychocracow.fra1.qualtrics.com/jfe/form/SV_eb18xIkbaFOZeM5

(OptoacousticsLtd., Or-Yehuda, Israel). Participants were instructed to minimize jaw movements while giving the spoken responses, which they had practiced during the behavioral session.

Materials and Procedure

Behavioral session

Verbal Fluency task. In this task, participants were asked to produce as many words belonging to a given category (fruits or animals) or beginning with a given letter (B or M) as they could within a 1-minute time limit. Cues were counterbalanced between languages. Each participant first completed the task in Polish (with one category and one letter) and then in English (also with one category and one letter). Each cue (letter or category) was displayed on the screen for 1 minute, during which participants provided verbal responses that were recorded with the microphone. This task was used as a complementary measure of relative proficiency in L1 and L2; the mean number of words produced for letter and category tasks in each language is reported in Table 1.

LexTale. In the LexTale task, participants were instructed to say whether the string of letters presented on a screen corresponded to an existing word in English. The response (yes/no) was provided via a keyboard. Each word or nonword appeared in the center of the screen and remained there until the participant responded. The task used the same words and nonwords as the original LexTale task (Lemhöfer & Broersma, 2012).

Language questionnaire. In the language questionnaire, participants were asked to respond to four categories of questions. The first category concerned proficiency in the native and foreign languages that the participants knew. It included a self-rating scale on the five language skills (reading, writing, listening, speaking, and accent) and questions concerning the age of acquisition. The second category included questions on patterns of language use: the participants were asked to estimate how much time during a typical day they actively (i.e., speaking, writing) and passively (i.e., watching movies, reading, listening to music/radio/podcasts) use the languages they know. In the third category, participants were asked to provide information about language use contexts: home, work, school, and free time. Specifically, they were asked to estimate the number of hours that they use each language they know in each of these contexts. Finally, the fourth category of questions referred to demographic information, such as age, gender, and education. Filling out the entire questionnaire took approximately 15-20 minutes.

MRI task training. In this part of the behavioral session, participants completed short versions of the tasks that they were going to perform during the MRI session (for details of the tasks see the MRI session design). Each task lasted approximately 1-2 minutes and was repeated if a participant asked

for more practice. In tasks requiring overt responses, participants were asked to practice minimizing jaw movements while speaking.

Raven test. In this task, participants completed a shortened version of Raven's Advanced Progressive Matrices test (only odd-numbered items). They had 20 minutes to complete as many matrices as they could. The test was used in a paper-pencil format and was aimed to provide a measure of fluid intelligence.

MRI sessions – the main task

In the main task, participants were asked to overtly name pictures of objects in four task blocks: two blocks in each session. One task block corresponded to one functional run. The first block of each session corresponded to one of the experimental manipulations: exposure to L1 (which required naming pictures in L1) or exposure to L2 (which required naming pictures in L2). The order of experimental manipulation blocks was counterbalanced between participants, i.e., some of them started with the L2 exposure conditions in the first session and completed the L1 exposure condition in the second session; some of them started with the L1 exposure condition in the first session and completed the L2 exposure condition in the second session. In the second block of each session, participants were asked to name pictures in L1. Each block contained 55 colored pictures representing objects (from the CLT database: Haman et al., 2015; Wolna et al., 2022). Pictures were not repeated between blocks. Each block started with a fixation cross ("+") displayed on a screen for 7000ms. Subsequently, 55 pictures were presented following a rapid event-related design. Each picture was displayed on a white background for 2000ms and was followed by a fixation cross, jittered at an interstimulus interval ranging from 607ms to 10144ms (M = 3091ms, SD = 2225ms). The order of picture presentation and interstimulus intervals was optimized using optseq2 (Dale, 1999). Finally, at the end of each block, a fixation cross was presented for 8000ms. Overall, each block of the Picture Naming task lasted 295s.

MRI sessions – localizer tasks

After completing the main task, in each session participants completed two or three localizer tasks. Each localizer consisted of two functional runs.

Language localizer. Participants listened to fragments of *Alice in Wonderland* in L1 and distorted speech recordings in which it was impossible to understand what the speaker was saying. The task followed the design introduced by Scott et al. (2017) and Malik-Moraleda, Ayyash et al. (2022), in which the localizer contrast was based on the intact > distorted speech condition contrast. In each functional run, participants listened to 6 short passages of Alice in Wonderland in their native language (18s each)

and 6 passages of distorted speech (18s each). Additionally, 4 fixation blocks (12s each) were included in each run. The total duration of each run was 264s. Each participant completed 2 runs of the task.

Articulation localizer. Participants performed two tasks: an articulation task in which they were asked to repeat out loud a string of previously memorized syllables ("bła" [bwa], "sza" [ʂa], "tra" [tra] – all frequently occurring in Polish as word constituents, but meaningless in isolation) and a motor task in which they were asked to perform a simple motor sequence with the fingers of both hands (touching the thumb with their index, ring, and little finger, respectively). The localizer was based on the articulation > motor sequence contrast. In each functional run, participants completed 6 articulation and 6 finger tapping blocks (18 s each). Additionally, 4 fixation blocks (12 s each) were included in each run. Each participant completed 2 runs of the task.

Stroop localizer. Participants were asked to overtly name the color of the presented word's font. The design of the task followed the one used by Fedorenko et al. (2011), in which participants completed blocks of the easy and hard versions of the task. In the easy condition, the words corresponded to non-color adjectives ("daleki" (eng. *far*), "wysoki" (eng. *tall*), "głęboki" (eng. *deep*)); in the hard condition, the words corresponded to color adjectives ("niebieski" (eng. *blue*), "zielony" (eng. *green*), "czerwony" (eng. *red*)). In the hard condition, half of the trials were conflict trials in which the word's font color did not match its meaning; the other half were congruent trials. The order of trials within the hard condition was randomized. Each run started with a fixation cross ("+") displayed on a gray background for 8000 ms. Following that, participants completed 16 task blocks (8 per condition, 18s each). Task blocks contained 12 trials which started with a fixation cross displayed for 500 ms followed by a word displayed for 1000 ms. Additionally, five fixation blocks (18s) were included in each run. The order of experimental and fixation blocks was counterbalanced between participants and runs (four counterbalance lists were created and each participant completed two of them). The localizer was based on the hard > easy condition contrast. The total duration of one run was 394s. Each participant completed 2 runs of the task.

Multiple Demand localizer. Participants were asked to perform a spatial working memory task (e.g., Fedorenko et al., 2011, 2013). The design of this task followed the one used by Fedorenko et al. (2011), in which the localizer contrast was based on the hard > easy condition contrast. In each trial, participants saw four 3 x 4 grids and were asked to memorize the locations of the fields that were marked as black on each of the grids. In the easy condition, they had to memorize one location per grid (four in total); in the hard condition, they had to keep track of two locations per grid (eight in total). After memorizing the grids, participants had to choose a grid that contained all the memorized locations from among two grids presented on the screen. Each trial started with a fixation cross

displayed on the screen for 500 ms; this was followed by four grids, each displayed for 1000 ms. After that, the choice task appeared on the screen until a response was given (for a max of 3750 ms); this was followed by feedback (right or wrong) displayed for 250 ms and a fixation cross, which was displayed for 3750 ms minus the reaction time in the choice task. In each of the conditions (easy and hard) participants completed 4 trials per block (34s in total). Each run contained five easy condition blocks and five hard condition blocks. Additionally, six fixation blocks (16s) were included in each run. The order of experimental and fixation blocks was counterbalanced between participants and runs (four counterbalance lists were created and each participant completed two of them). The total duration of one run was 438s.

Verbal fluency localizer. Within 10 s blocks, participants were asked to produce as many words as they could that either pertained to a given category (semantic fluency) or started with a given letter (phonemic fluency). The entire task was completed in L1. In each run, participants completed six semantic fluency and six phonemic fluency blocks, each focusing on a separate letter or semantic category. Each letter and category was presented only once and they did not repeat between blocks and runs. A full list of stimuli used in this task is available in the Supplementary Materials. In addition, each run included five baseline blocks, in which participants were asked to enumerate the months starting from January (Birn et al., 2010). The localizer contrast was based on task (semantic or phonemic fluency) > baseline condition contrast[6]. Following Birn et al. (2010), each block was followed by a fixation cross displayed for 7s–13 s. Each run lasted 348s. Each participant completed 2 runs of the task.

MRI Data Acquisition

MRI data were acquired using a 3T scanner (Magnetom Skyra, Siemens) with a 64-channel head coil. High-resolution, whole-brain anatomical images were acquired using a T1-weighted MPRAGE sequence (208 sagittal slices; voxel size 0.9×0.9×0.9 mm^3; TR = 1800ms, TE = 2.32ms, flip angle = 8°). A gradient field map (magnitude and phase images) was acquired with a dual-echo gradient-echo sequence, matched spatially with fMRI scans (TE1 = 4.92ms, TE2 = 7.3ms, TR = 503ms, flip angle = 60°). Functional T2*-weighted images were acquired using a whole-brain echo-planar (EPI) pulse sequence (50 axial slices, 3x3x3 mm isotropic voxels; TR = 1400ms; TE = 27ms; flip angle = 70°;

[6] Even though semantic and phonemic fluency tasks rely on different strategies and have been linked to differences in cognitive control engagement and their corresponding neural responses (Birn et al., 2010), they have also been shown to largely measure similar processes rather than rely on fundamentally different mechanisms (Unsworth et al., 2011). As these similarities most likely reflect the mechanism of word retrieval that we are interested in identifying using the Verbal Fluency localizer, we decided to use the broader contrast instead of going for separate contrasts for semantic and phonemic fluency as we believe this better fits our research questions.

FOV 192 mm; MB acceleration factor 2; in-plane GRAPPA acceleration factor 2 and phase encoding A>>P) using 2D multiband echo-planar imaging (EPI) sequences from the Center for Magnetic Resonance Research (CMRR), University of Minnesota (Xu et al, 2013). In order to account for magnetic saturation effects, the first four volumes of each sequence were not acquired.

MRI Data preprocessing

All data were visually inspected for artifacts. The non-brain tissue was removed using the FSL Brain Extraction Tool (Smith, 2002). For the preprocessing of functional data, we used FSL FEAT (version 6.0.0; Woolrich et al., 2004). The preprocessing steps included high-pass temporal filtering (100s), spatial smoothing (FWHM = 5 mm), distortion correction of functional images using fieldmap with FUGUE, co-registration and normalization using FLIRT (Jenkinson et al., 2002; Jenkinson & Smith, 2001) and motion correction using MCFLIRT (Jenkinson et al., 2002) with 6 rigid-body transformations. Functional images were co-registered with their anatomical images; subsequently, they were registered to an MNI standard space (FSL's 2mm standard brain template).

MRI Data analysis

In the 1st-level statistical analysis, a double-gamma hemodynamic response function was used to model the BOLD signal corresponding to each event (trial in the main task and blocks in the localizer tasks). The estimates of motion obtained with MCFLIRT (Jenkinson et al., 2002) were included in the 1st-level GLM as nuisance covariates. For the localizer task, the two functional runs of each localizer task were combined in a second-level analysis for each participant using an FSL fixed-effect model. Following that, group-constrained subject-specific ROIs were defined for each participant following the approach proposed by Fedorenko et al. (2010). In the first step, the individual subject activation maps (thresholded at 5% of most active voxels[7]) were binarized and overlaid on top of each other. The probabilistic map obtained this way was then thresholded so that only voxels active in more than 5 participants (~20%) would be included. In the second step, the group-level partitions were created using a watershed algorithm for segmentation. This procedure yielded nineteen group-level parcels for the Stroop task and fifteen group-level parcels for the Verbal Fluency task. Before further analyses, we excluded very small parcels (mean size across subjects smaller than 10 voxels). This yielded a final set of thirteen parcels within the Stroop network and thirteen parcels within the verbal fluency network. To ensure better comparability of our findings with previous studies, for language, Multiple

[7] In most cases, this threshold was slightly more conservative than z > 3.1, uncorrected at whole-brain level. We chose to use the 5% of most active voxels instead of a standard threshold because it yielded maps with a comparable number of voxels between subjects that consequently produced more coherent group-level parcels.

553 Demand, and articulation localizers we used sets of functional parcels established and validated by
554 previous studies instead of creating group-level parcels based on our data (twelve language parcels:
555 Fedorenko et al., 2010; twenty multiple demand parcels: Fedorenko et al., 2013; eleven articulation
556 parcels: Basilakos et al., 2018)[8]. In the last step, group-level partitions were intersected with subject-
557 specific data, thus yielding a set of subject-specific group-constrained ROIs. Finally, we selected the
558 top 10% most responsive voxels within each parcel based on the *z*-value map for a given localizer
559 contrast (Language: speech > degraded speech, MD: hard > easy visual working memory task,
560 Articulation: syllables > motor sequence, Stroop: color words > neutral words, and Verbal Fluency:
561 fluency > automated speech) and we binarized the obtained ROIs. We then used the ROIs
562 corresponding to all localizer tasks to extract parameter estimates corresponding to % signal change
563 in the two critical conditions of the main task: naming in L1 after L1, and naming in L1 after L2.
564 Parameter estimates were extracted from the 1st-level analyses for each participant separately. This
565 was done using FSL's FEATquery tool (http://www.FMRIb.ox.ac.uk/fsl/feat5/featquery.html). Further
566 analyses were performed on the extracted parameter estimates in R (version: 4.0.2; R Core Team,
567 2020). Following an approach that proposes modeling activity within a functional network on the
568 network level instead of relying on separate analyses for each ROI (Malik-Moraleda et al., 2021), we
569 fitted one linear mixed-effect model for each functional network according to the following formula:

570 % signal change ~ PrecedingLanguage +
571 (1 + PrecedingLanguage | Subject) +
572 (1 + PrecedingLanguage | ROI)

573 In each model, the fixed effect of PrecedingLanguage corresponded to a predictor with two
574 levels: *L1 after L1* and *L1 after L2*. Before the analysis, this categorical predictor was deviation-coded
575 (L1 after L1 = -0.5; L1 after L2 = 0.5). The analysis was performed using the lmer() function from the
576 lme4 package (Bates et al., 2015). For each functional network, we first fitted a maximal model and
577 then identified the best random-effects structure following the recommendations of Bates et al.
578 (2018). The reason for modeling the activity within each network with ROI as a random effect instead
579 of a fixed effect relies on the assumption that the different brain regions engaged within a given
580 functional network exhibit a significant degree of functional integration. To further explore the fine-
581 grained differences between our critical conditions, which might only affect a subset of functional
582 parcels within each network, we also fitted separate models for each ROI in each of the functional
583 networks. The results of the ROI-specific analyses are reported in the *Supplementary Materials*. Raw
584 neuroimaging data used in this study are freely available at

[8] The group-level parcels from previous studies showed a high overlap with parcels for the language, Multiple Demand, and articulation networks defined based on our data.

https://openneuro.org/datasets/ds004456. Data and code necessary to reproduce the ROI and behavioral analyses are freely available at https://osf.io/59za8/?view_only=866d417f0dd24cf683b37ad025baa369.

Behavioral data analysis

Behavioral naming data recorded in the scanner were transcribed for each participant and analyzed for naming latencies and accuracy. One participant's data were excluded because of very poor recording quality that did not allow us to determine the naming latencies in one of the sessions[9]. For the picture naming task, all answers corresponding to the dominant response and its synonyms were considered correct. Trials with no responses, inaudible responses, and names unrelated to a given picture were classified as errors. The naming latencies for each answer were set manually. Subsequently, we fitted a linear mixed-effect model to the naming latency data and a generalized mixed-effect linear model to the accuracy data (using the lmer() and glmer() functions from the lme4 package (Bates et al., 2015), respectively). For both models, we used the same formula:

Naming latencies / Accuracy ~ PrecedingLanguage +
(1 + PrecedingLanguage | Subject) +
(1 + PrecedingLanguage | Item)

Language corresponded to a predictor with two levels: *L1* and *L2*. Before the analysis, the categorical predictor was deviation-coded (L1 = -0.5; L2 = 0.5); for the analysis of naming latencies, the reaction times were log-transformed to reduce the skewness of the data (skewness before and after transformation was equal to 0.80 and 0.26, respectively). For each analysis, we first fitted a maximal model and then identified the best random-effects structure following the recommendations of Bates et al. (2018).

Results

Bilingual language control: domain-general or language-specific

To address the question of whether the bilingual language control mechanisms driving the L2 after-effect rely on domain-general or domain-specific mechanisms, we compared brain activity corresponding to naming pictures in *L1 after L1* and *L1 after L2* within two functional networks: the Language network and the Multiple Demand network. The results on the network level revealed no significant effect of the preceding language within the language network (t = 0.67, p = .509, effect size = 0.21); however, a significant effect of the preceding language was found within the MD network (t =

[9] As this problem was limited to the behavioral data, this participant was still included in the neuroimaging analyses.

2.65, p = .011, effect size = 0.80). The results of the comparison between *L1 after L1* and *L1 after L2* within the language and MD networks are presented in **Figure 2.**

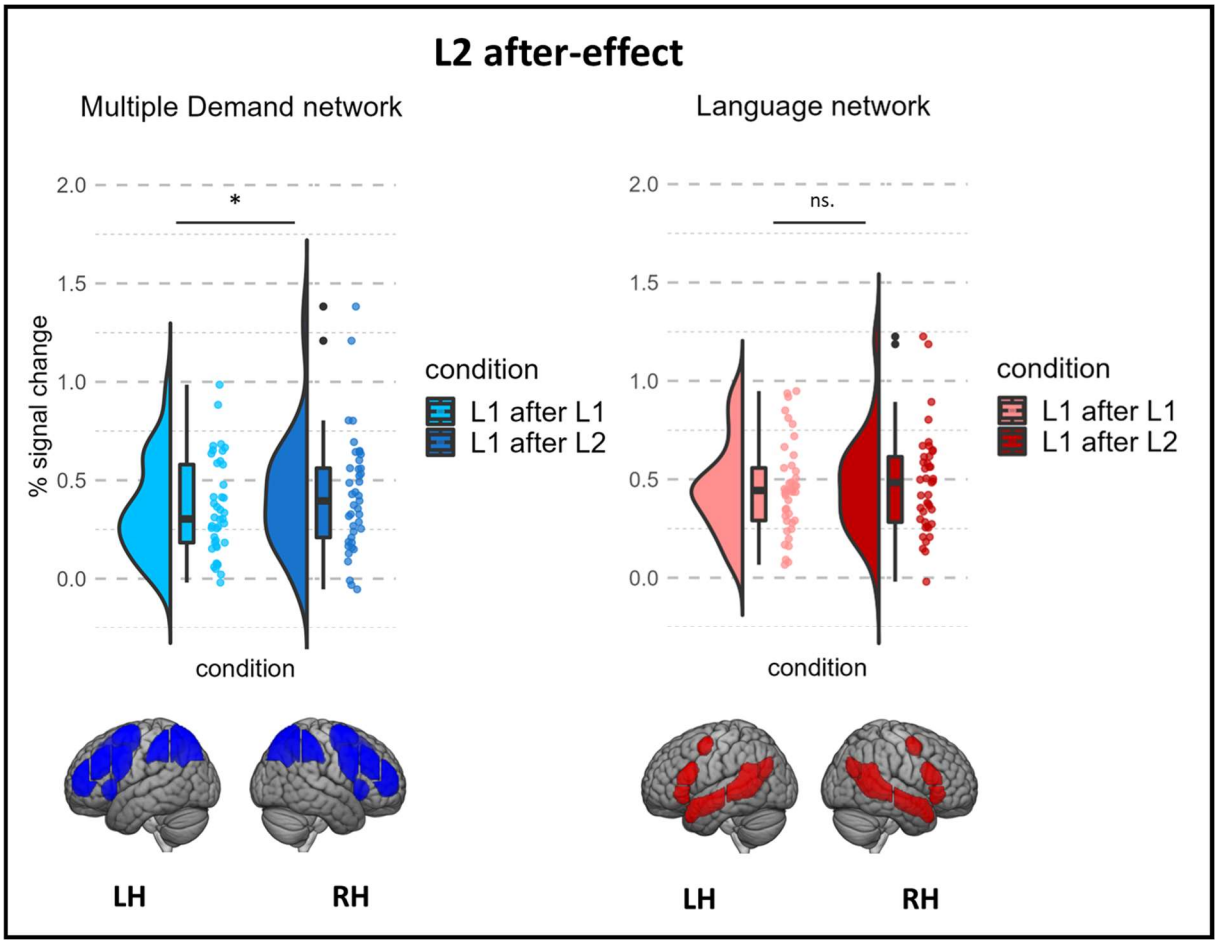

Figure 2. Neural activation corresponding to the L2 after-effect within the Multiple Demand and Language networks. Plots represent % BOLD signal change in response to naming pictures in L1 after L1 (light hues) and L1 after L2 (dark hues). Rendered maps present sets of group-level parcels used in the subject-specific analysis (LH = left hemisphere; RH = right hemisphere).

Separate analyses on each individual ROI showed significant differences between *L1 after L1* and *L1 after L2* in four ROIs in the Multiple Demand network, but no significant results were found in the language network. Full results showing comparisons between conditions for each ROI and each network are presented in **Tables A1 and A2** in the *Supplementary Materials*.

Digging deeper: specific mechanisms of bilingual language control

To explore which specific mechanisms drive the L2 after-effect, we compared the brain activity corresponding to naming pictures in *L1 after L1* and *L1 after L2* within three functional networks identified with the Articulation, Stroop and Verbal Fluency tasks. We did not find a significant effect of the preceding language in any of the three networks (Articulation: t = 0.09, p = .926, effect size = 0.03;

Stroop: t = 1.02, p = .314, effect size = 0.32; Verbal Fluency: t = 0.07, p = .941, effect size = 0.02; see Figure 3). Separate analyses on each individual ROI for each of these networks are presented in **Tables A3-5** in the *Supplementary Materials*.

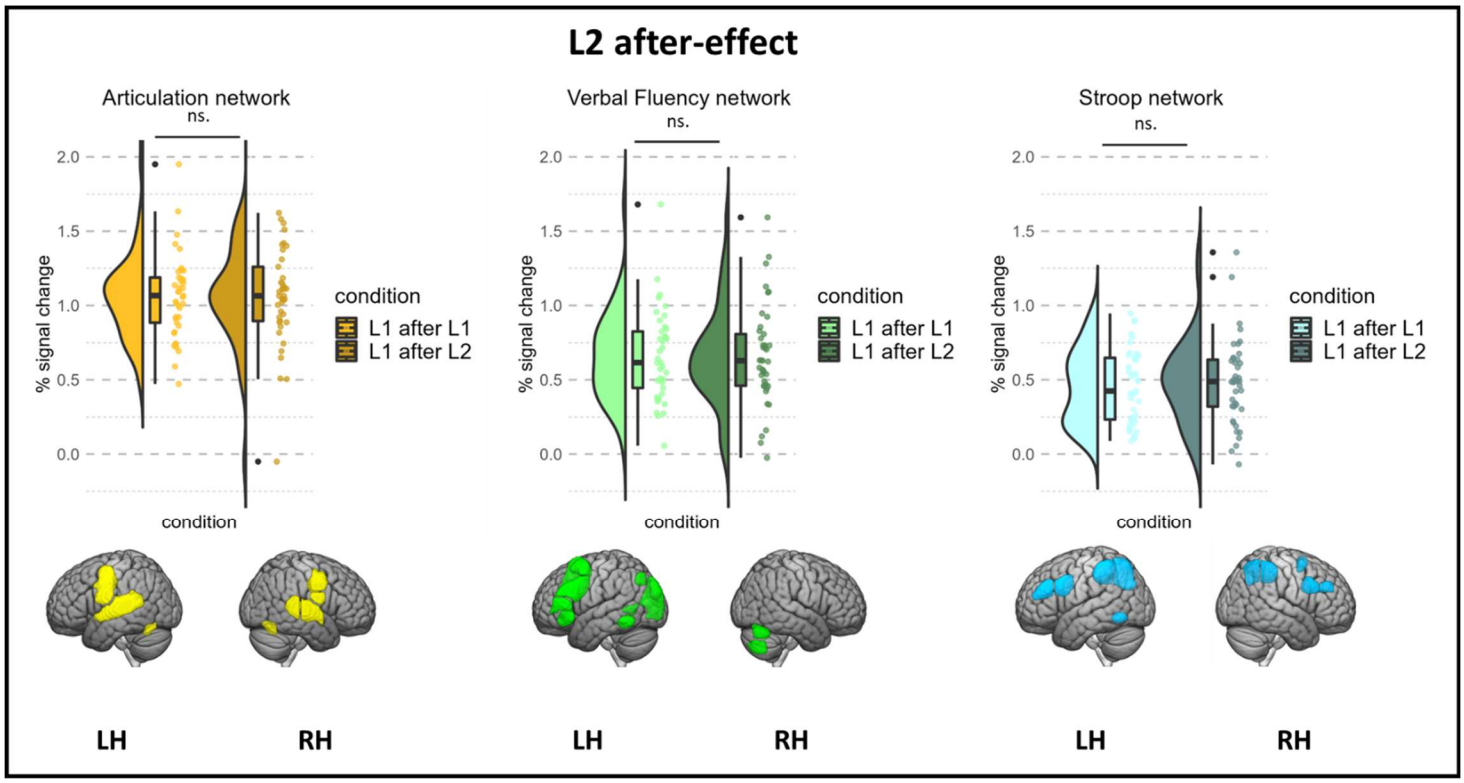

Figure 3. Neural activation corresponding to the L2 after-effect within the Articulation, Verbal Fluency and Stroop networks. Plots represent % BOLD signal change in response to naming pictures in L1 after L1 (bright hues) and L1 after L2 (dark hues). Rendered maps present sets of group-level parcels used in the subject-specific analysis (LH = left hemisphere; RH = right hemisphere).

Behavioral results

The analysis of naming latencies did not reveal significant differences between naming speed in *L1 after L1* and *L1 after L2* (t = 0.292, p = 0.771). Naming latencies predicted by the model corresponded to 1072ms (CI [1042ms–1101ms]) for *L1 after L1* and 1075ms (CI [1045ms–1105ms]) *L1 after L2*. Similarly, no differences in accuracy were found for the two conditions (z = 0.44, p = 0.660). Accuracies predicted by the model correspond to 99.5% (CI[98.59% - 99.83%]) for the *L1 after L1* condition, and 99.65% (CI[98.89% - 99.89%]) for the *L1 after L2* condition. The lack of results may be explained by the generally noisy and challenging environment because data were collected during scanning. It is also important to note that even though participants were instructed to name pictures as quickly as possible, overall naming speed (1072ms for *L1 after L1* and 1075ms for *L1 after L2*) was slower than what is usually observed for L1 naming for this particular dataset (897ms; Wolna et al.,

2022). The general slow-down of naming in L1 may have obscured the behavioral effect, which usually is quite small (~50ms).

Discussion

Bilingual language control: domain-general or language-specific?

We explored the neurocognitive mechanisms underlying the L2 after-effect: a difficulty in producing words in the native language (L1) after having used the second language (L2). We compared brain activity corresponding to speaking in L1 after L2 with speaking in L1 after L1 within functionally defined brain networks. Our goal was to test (1) whether the L2 after-effect is related to activity in domain-general or domain-specific brain networks, and (2) whether the L2 after-effect reflects lexical retrieval difficulty, interference resolution, or increased difficulty at the phonological and articulatory stages of speech production. Unlike previous work exploring the neural basis of bilingual speech production, we used individual-level functional localizers (Fedorenko et al., 2010) to provide functional interpretations of brain responses associated with the L2 after-effect.

To address the question of whether the L2 after-effect is associated with increased engagement of language-specific and/or domain-general brain networks, we compared brain responses to speaking in L1 after L2 and L1 after L1 with two well-established, robust functional networks: the Language network (Malik-Moraleda, Ayyash et al., 2022; Fedorenko et al., 2010; Scott et al., 2017) and the Multiple Demand network (Duncan, 2010; Fedorenko et al., 2013). We found that speaking in L1 after L2 is related to increased BOLD responses within the domain-general Multiple Demand network, but not within the domain-specific language network. Furthermore, to explore whether the L2 after-effect is driven by lexical access difficulty, increased interference between language-specific representations, or articulatory difficulty, we used three additional functional tasks as localizers: verbal fluency, the Stroop task, and an articulation task. We found that the L2 after-effect does not overlap with any of these process-specific localizers.

The L2 after-effect reflects neither word retrieval difficulties nor increased interference between language-specific representations

The L2 after-effect has been previously explained within different theoretical frameworks (see Introduction). While these accounts provide plausible explanations of this effect, no univocal evidence in favor or against these interpretations has been provided (Casado et al., 2022; Wodniecka, Szewczyk, et al., 2020; Declerck & Koch, 2022). All available accounts of the L2 after-effect build on a common theoretical assumption that the difficulty in producing words in L1 after using L2 is a consequence of a control mechanism that changes the balance between the relative activation of L1 and L2

representations (Casado et al., 2022; for a discussion see Declerck & Koch, 2022). Importantly, according to these accounts the change in L1-L2 balance results in increased interference between the two languages that persists for some time after the speaker switches back to speaking in L1 (Casado et al., 2022), which, in turn, increases the difficulty of access and selection of (a) lexico-semantic representations (e.g., Casado et al., 2022) and/or (b) phonological representations and articulatory programs of L1 (e.g., Guo et al., 2011).

The results of the current study do not provide evidence supporting the notion that the L2 after-effect is related to increased interference at the lexico-semantic, phonological, or articulatory levels. First of all, we found that the L2 after-effect was not linked to an increased BOLD response within the core language network. Since lexical access and selection mechanisms pertain to core language computations (Hu et al., 2022), difficulties in lexical access should be reflected by an increased BOLD response within the language network. This was, however, not the case in the current study. What is more, we found no increased brain response to speech production in the L1 after L2 condition in either of the functional networks that were identified using verbal fluency and articulation tasks. Even though both localizer tasks produced robust responses and allowed us to identify a number of functional ROIs, they did not overlap with the neural correlates of the L2 after-effect. Furthermore, we did not find any overlap between the neural correlates of the L2 after-effect and the functional network that was identified using the Stroop task; thus, we found no evidence for increased interference between language representations during speech production in L1 after L2.

The L2 after-effect reflects the increased engagement of domain-general cognitive control

By comparing the brain correlates of the L2 after-effect with two large brain networks, we found that the effect was related to an increased response within the Multiple Demand network but not within the Language network. On the most general level, our findings support models that posit a domain-general character of bilingual language control (Abutalebi & Green, 2016; Green & Abutalebi, 2013; Wu et al., 2019).

The Multiple Demand network has been claimed to support cognitive control by integrating different types of information and binding them together to support the current cognitive operation (Assem et al., 2020, 2022; Cole et al., 2013; Koechlin & Summerfield, 2007). Activity within the Multiple Demand network has been linked to mechanisms responsible for the implementation and reconfiguration of task sets (Dosenbach et al., 2006, 2007), voluntary attentional control (goal-directed attention; Asplund et al., 2010; Corbetta & Shulman, 2002), and reorienting attention to novel tasks (Corbetta et al., 2008). An important feature of the Multiple Demand network is that it is more engaged

in less-automatized and untrained tasks (Shashidhara et al., 2019), thus implying that it implements a domain-general mechanism that supports any difficult or not sufficiently automatized activity that requires additional attentional support. Seen from this perspective, speaking in L1 after L2 may recruit additional domain-general resources to support the execution of a language task. However, this leaves open the question of why exactly speaking in L1 after L2 is so much more difficult than speaking in L1 after L1 that it requires additional support from the Multiple Demand network.

The results of the localizer-based analyses provide a rather unexpected characterization of the neural correlates of the L2 after-effect. Contrary to the previous interpretations (Branzi et al., 2014, 2016; Guo et al., 2011; Misra et al., 2012; Declerck & Koch, 2022), the current results indicate that the L2 after-effect reflects neither lexical access difficulty, nor articulatory difficulty, nor the additional engagement of language-specific mechanisms that support word retrieval. In other words, our results suggest that the L2 after-effect does not arise from within the language system itself (or from within language-specific representations). So, how can we then account for the increased difficulty in speaking in the native language after using L2? In the following paragraphs, we propose two putative explanations of the difficulty in production in L1 following the use of L2.

Putative explanations of the L2 after-effect

The L2 after-effect as interference between task schemas

As described in the Introduction, most of the evidence for the involvement of cognitive control in bilingual processing comes from the language-switching paradigm and, more specifically, from the asymmetrical language switch cost (i.e., the finding that the cost of switching to the dominant (L1) language is bigger than to the weaker (L2) language) and language mixing cost effects (i.e., the stronger language (L1) suffers more when both languages are mixed within one block of naming) (Costa & Santesteban, 2004; Gollan & Ferreira, 2009; Meuter & Allport, 1999; Verhoef et al., 2009; for review see: Bobb & Wodniecka, 2013; Declerck et al., 2020; Declerck & Koch, 2022; Declerck & Philipp, 2015). Perhaps the most classic explanation of these effects assumes that they result from the proactive language control that is recruited whenever participants have to name a picture in the weaker L2. This control is recruited to “protect” the weaker L2 (less activated or less automatized) from interference with the stronger L1, either by inhibiting L1 (e.g., Green, 1998; Green & Abutalebi, 2013) or by bolstering the activation of L2 (e.g., Blanco-Elloreta & Caramazza, 2021). When, after using L2, a bilingual speaker switches to her L1, naming in L1 is hampered by the consequences of prior control mechanisms: L1 is now either inhibited, or there is an increased interference with L2, which is still strongly activated.

Explanations that assume the involvement of proactive control have been previously used to account for the L2 after-effect (Wodniecka et al., 2020; Declerck & Koch, 2022), and our current findings are generally consistent with this framework. However, they also point to an important caveat regarding the level of the targeted representations. While models of bilingual language processing typically assume that proactive control targets language-specific representations, such as specific lexical units (see, e.g., BIA+ model, Dijkstra & van Heuven 2002 for bilingual language comprehension; Blanco-Elorrieta & Caramazza, 2021 for bilingual language production). The findings of the current study are incompatible with such views. If proactive control operated on language-specific representations, we should see modulations of brain activity in the after-L1 vs after-L2 conditions in regions directly operating on these language-specific representations, namely in the Language network, which was not the case.

Nevertheless, our results support accounts that assume proactive control if one assumes that the control does not target language-specific representations but rather more-abstract ones. Previous literature claims the existence of such representations and refers to them as control representations (Braver et al., 2021) or hierarchical task-schemas (Badre & Nee, 2018). Such representations could be conceived as sets of rules for orchestrating access to language-specific representations – the program we execute when we intend to speak in a given language and which guides all subprocesses involved in subsequent episodes of language production. This may perhaps be best understood in light of the cascade model of executive control (Koechlin & Summerfield, 2007), which model proposes a hierarchy of increasingly abstract control processes. At the sensorimotor control level, a stimulus is simply associated with a reaction (e.g., picking up the phone when it rings). At the contextual control level, the stimulus-response link depends on the current context (not picking up the phone when it rings at your friend's home). Episodic and branching control levels enable exceptions to the rules under which the current contextual rules can be temporarily and conditionally withheld or reinstated (not picking up the phone at a friend's place unless he asked you to). Koechlin and Summerfield note that all dual-task situations require such high-level reconfigurations of rules (i.e., task 2 requires temporarily withholding the rulesets of task 1). Thus, when conceived from the perspective of this framework, the interference between L1 and L2 can occur at the level of rulesets that are managed by these more-abstract levels of cognitive control. The interference occurs between the sets of task rules for speaking in L1 and the sets of task rules for speaking in L2, both of which compete for selection to guide language production. If the interference indeed occurs at the language task-schema level, it would explain why language representations, supported and processed by the Language network, are not affected by a prior use of L2, yet it usually leads to a slowdown of naming (Branzi et al., 2014; Wodniecka et al., 2020, Casado et al., 2022; Degani et al., 2019). We therefore propose that the activation within the

Multiple Demand network may reflect the costs of resolving the interference between L1 and L2 production task schemas.

This idea is in line with the Inhibitory Control Hypothesis (Green, 1998), which proposed that language-switching effects are a consequence of competition between language schemas. However, the Inhibitory Control Hypothesis did not clearly define "language schemas": they are sometimes described at a more abstract level as sets of rules that guide speech production in different languages, similarly to what we propose here; however, they are sometimes framed as "regulating the outputs from the lexico-semantic system by altering the activation levels of representations within that system and by inhibiting outputs from that system" (Green, 1998), which implies that they engage in resolution of interference between language-specific representations.

Our proposal is also relevant to a distinction made in the literature regarding bilingual cognitive control, which differentiates between local and global language control. According to this distinction, bilingual language control mechanisms (for example, inhibition) can target not only specific lexical representations (local control) but also the entire language system (global control: Branzi et al., 2014; Van Assche et al., 2013; Wodniecka et al., 2020). The idea that difficulty in bilingual speech production can be driven by increased interference between sets of rules that guide speech production in a given language aligns with the idea of global language control if this "global control" is understood as mechanisms that operate not on language-specific representations (e.g., lemmas, c.f. Van Assche et al., 2013; or nodes linked to a common language node; Dijkstra & van Heuven, 2001; Blanco-Elorrieta & Caramazza, 2021) but on sets of schemas for speaking in a given language.

In sum, while our first tentative explanation of the L2 after-effect builds on a mechanism that has been proposed before (proactive interference during L2 naming that affects the activation level of L1 or L2 representations such that it has negative consequences for subsequent use of L1), it also points to a critical difference. The key novelty of the proposed interpretation is that the affected representations are not language-specific (lexical units, morphosyntactic units, phonological representations, motor programs) but more general and abstract task-schemas that determine the general programs of using L1 or L2.

The L2-after effect as sustained activation of cognitive control

Another possibility that, to the best of our knowledge, has not been proposed in the bilingual language processing literature attributes the after-effects in blocked designs entirely to cognitive control mechanisms without involving any representations at the level of either language or task-schema. This proposal is similar to the first explanation, because it also assumes that a prior speech

production task in L2 requires intensified proactive cognitive control to resolve the interference between stronger L1 and weaker L2. However, unlike the first explanation, here we propose that the prior use of L2 leads to increased activation of cognitive control without impacting language-specific representations. The difficult task of speaking in L2 engages cognitive control so much that the control system remains alert even after the task in L2 concludes. When the speaker changes back to L1, the cognitive system is still "warmed up", which means it is proactively maintaining the L1 production task schema and protecting it from any interference (even though there is none). This may be a similar mechanism to the one driving the Stroop congruency effect, in which the overall proportion of incongruent trials affects the amount of conflict in congruent trials (Tzelgov et al., 1992), with the caveat that the Stroop congruency effect refers to the paradigm in which both the manipulation and the measurement apply to the same block, so it could rely on strategic instead of passive carry-over effects. It could also be argued that support for the explanation that assumes that the L2 after-effect reflects persisting increased engagement of cognitive control comes from the Dual Mechanisms of Control framework (Braver, 2021). According to this framework, proactive control recruits the lateral prefrontal cortex, which, unlike the anterior cingulate cortex, is sensitive not only to errors or conflicts in the previous trial but also to control demands specific to the entire task, or the current communicative context (Blanco-Elorrieta & Caramazza, 2021, Casado et al., 2022). Since we did not observe modulations between the L1-after and L2-after conditions in the anterior cingulate cortex activity (see *Supplementary Materials*) but in the lateral prefrontal cortex, which is part of the Multiple Demand system (by-ROI results in the Multiple Demand network are available in *Supplementary Materials*), it can be argued that the L2 after-effect is a manifestation of proactive control.

In sum, the second tentative explanation of the increased brain response related to the L2 after-effect proposes that it reflects the increased engagement of cognitive control triggered by the previous use of L2 and that this engagement lingers during a subsequent task in L1. However, this hypothesis has never been directly tested and, as such, further experiments are needed to falsify it[10].

Conclusions

The novelty of our contribution is twofold: First, to identify the cognitive mechanisms that correspond to the neural signature of the L2 after-effect, we used the functional localizer approach

[10] One way to verify the idea that the L2 after-effect reflects lingering activation of cognitive control mechanisms would be to conduct a study which would not involve language at all, but instead switching between two nonlinguistic tasks that do not share stimulus and response sets. The critical experimental manipulation would be the difficulty of the first task, while the second, easy and automatic task would be held constant in both conditions. If the hypothesis is true, we should observe worse performance in the second task in the condition in which it follows a difficult task, compared to an easy first task.

(Fedorenko et al., 2010), which to the best of our knowledge has never been used in experiments looking at the brain basis of bilingual speech production. Second, unlike previous studies, our conclusions are based on a design that was optimized to measure the index of language control, i.e., the L2 after-effect, while reducing many possible confounds. We believe that the combination of precise analytical tools, carefully designed tasks, as well as the relatively large number of participants tested (n=42) lead to new insights into the neurocognitive mechanisms of language control in bilinguals.

In contrast to the dominating theories and views in the field, we found that the L2 after-effect is unlikely to arise as a result of interference between language-specific representations in L1 and L2 because it does not overlap with the Language network, or with any task-specific localizer that targets specific language processes (lexical access, articulation, language interference, and inhibition). Instead, we found that the L2 after-effect overlaps with the Multiple Demand network and thus reflects a domain-general mechanism. We propose that the L2 after-effect may manifest as either increased interference between high-level task schemas used to orchestrate language use in L1 and L2, or as a lingering increase in the engagement of domain-general cognitive control mechanisms in speech production.

Acknowledgements

The authors would like to thank all members of our Psychology of Language and Bilingualism Laboratory LangUsta who contributed to the research project by collecting and coding the data, all participants who took part in the study, as well as Mike Timberlake who proofread the text. We are very grateful to Evelina Fedorenko for inspiration and insightful comments on the design and results. This research was possible thanks to an OPUS grant from the National Science Center Poland, awarded to ZW (2017/27/B/HS6/00959).

Data/code availability statement

The raw neuroimaging data collected for this study along with the data on participants' language experience is available at https://openneuro.org/datasets/ds004456.

Data and code necessary to reproduce the presented ROI and behavioral analyses is available at the https://osf.io/59za8/?view_only=866d417f0dd24cf683b37ad025baa369.

Ethics statement

The reported study met the requirements and gained the approval of the Jagiellonian University Institute of Psychology Research Ethics Committee.

Author contributions

Conceptualization: AW, JS, MD, MS, ZW; Data curation: AW; Formal analysis: AW; Funding acquisition: ZW; Investigation: AW, AD; Methodology: AW, JS, MD, MS, AD, ZW; Project administration: ZW; Resources: AD, ZW; Software: AW, AD; Supervision: ZW, JS; Validation: AW; Visualization: AW, JS, ZW; Writing – original draft: AW, JS, ZW; Writing – review & editing: MD, MS, AD

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3), 242–275. https://doi.org/10.1016/j.jneuroling.2006.10.003

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*(4), 689–698. https://doi.org/10.1017/S1366728916000225

Asplund, C. L., Todd, J. J., Snyder, A. P., & Marois, R. (2010). A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. *Nature neuroscience*, *13*(4), 507-512.

Assem, M., Shashidhara, S., Glasser, M. F., & Duncan, J. (2022). Precise topology of adjacent domain-general and sensory-biased regions in the human brain. *Cerebral Cortex*, *32*(12), 2521-2537. https://doi.org/10.1093/cercor/bhab362

Assem, M., Blank, I. A., Mineroff, Z., Ademoğlu, A., & Fedorenko, E. (2020). Activity in the fronto-parietal multiple-demand network is robustly associated with individual differences in working memory and fluid intelligence. *Cortex*, *131*, 1–16. https://doi.org/10.1016/j.cortex.2020.06.013

Assem, M., Glasser, M. F., Van Essen, D. C., & Duncan, J. (2020). A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex. *Cerebral Cortex*, *30*(8), 4361–4380. https://doi.org/10.1093/cercor/bhaa023

Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffmann, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2022). An investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience*, *25*(8), 1014-1019.

Badre, D., & Nee, D. E. (2018). Frontal cortex and the hierarchical control of behavior. *Trends in cognitive sciences*, *22*(2), 170-188.

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*(6), 907–918. https://doi.org/10.1016/j.neuron.2005.07.023

Basilakos, A., Smith, K. G., Fillmore, P., Fridriksson, J., & Fedorenko, E. (2018). Functional Characterization of the Human Speech Articulation Network. *Cerebral Cortex*, *28*(5), 1816–1830. https://doi.org/10.1093/cercor/bhx100

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*. https://doi.org/10.48550/arXiv.1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), Article 1. https://doi.org/10.18637/jss.v067.i01

Benn, Y., Ivanova, A. A., Clark, O., Mineroff, Z., Seikus, C., Silva, J. S., Varley, R., & Fedorenko, E. (2021). *No evidence for a special role of language in feature-based categorization* (p. 2021.03.18.436075). bioRxiv. https://doi.org/10.1101/2021.03.18.436075

Birn, R. M., Kenworthy, L., Case, L., Caravella, R., Jones, T. B., Bandettini, P. A., & Martin, A. (2010). Neural systems supporting lexical search guided by letter and semantic category cues: A self-paced overt response fMRI study of verbal fluency. *NeuroImage*, *49*(1), 1099–1107. https://doi.org/10.1016/j.neuroimage.2009.07.036

Blanco-Elorrieta, E., & Caramazza, A. (2021). A common selection mechanism at each linguistic level in bilingual and monolingual language production. *Cognition*, 104625. https://doi.org/10.1016/j.cognition.2021.104625

Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Bilingual Language Control in Perception versus Action: MEG Reveals Comprehension Control Mechanisms in Anterior Cingulate Cortex and Domain-General Control of Production in Dorsolateral Prefrontal Cortex. *Journal of Neuroscience*, *36*(2), 290–301. https://doi.org/10.1523/JNEUROSCI.2597-15.2016

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual Language Switching in the Laboratory versus in the Wild: The Spatiotemporal Dynamics of Adaptive Language Control. *Journal of Neuroscience*, *37*(37), 9022–9036. https://doi.org/10.1523/JNEUROSCI.0553-17.2017

Blank, I., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of Neurophysiology*, *112*(5), 1105–1118. https://doi.org/10.1152/jn.00884.2013

Bobb, S. C., & Wodniecka, Z. (2013). Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology*, *25*(5), 568–585. https://doi.org/10.1080/20445911.2013.792822

Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2016). Language Control in Bilinguals: Monitoring and Response Selection. *Cerebral Cortex*, *26*(6), 2367–2380. https://doi.org/10.1093/cercor/bhv052

Branzi, F. M., Martin, C. D., Abutalebi, J., & Costa, A. (2014). The after-effects of bilingual language production. *Neuropsychologia*, *52*, 102–116. https://doi.org/10.1016/j.neuropsychologia.2013.09.022

Braver, T. S., Kizhner, A., Tang, R., Freund, M. C., & Etzel, J. A. (2021). The dual mechanisms of cognitive control project. *Journal of Cognitive Neuroscience*, *33*(9), 1990-2015.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural Mechanisms of Transient and Sustained Cognitive Control during Task Switching. *Neuron*, *39*(4), 713–726. https://doi.org/10.1016/S0896-6273(03)00466-5

Broos, W. P. J., Duyck, W., & Hartsuiker, R. J. (2018). Are higher-level processes delayed in second language word production? Evidence from picture naming and phoneme monitoring. *Language, Cognition and Neuroscience*, *33*(10), 1219–1234. https://doi.org/10.1080/23273798.2018.1457168

Calabria, M., Costa, A., Green, D. W., & Abutalebi, J. (2018). Neural basis of bilingual language control. *Annals of the New York Academy of Sciences*, *1426*(1), 221–235. https://doi.org/10.1111/nyas.13879

Casado, A., Szewczyk, J., Wolna, A., & Wodniecka, Z. (2022). The relative balance between languages predicts the degree of engagement of global language control. *Cognition*, *226*, 105169.

Chein, J. M., & Schneider, W. (2005). Neuroimaging studies of practice-related change: FMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, *25*(3), 607–623. https://doi.org/10.1016/j.cogbrainres.2005.08.013

Chen, X., Affourtit, J., Ryskin, R., Regev, T. I., Norman-Haignere, S., Jouravlev, O., Malik-Moraleda, S., Kean, H., Varley, R., & Fedorenko, E. (2021). *The human language system does not support music processing* (p. 2021.06.01.446439). bioRxiv. https://doi.org/10.1101/2021.06.01.446439

Coderre, E. L., Smith, J. F., Heuven, W. J. B. V., & Horwitz, B. (2016). The functional overlap of executive control and language processing in bilinguals. *Bilingualism: Language and Cognition*, *19*(3), 471–488. https://doi.org/10.1017/S1366728915000188

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature neuroscience*, *16*(9), 1348-1355.

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage*, *37*(1), 343–360. https://doi.org/10.1016/j.neuroimage.2007.03.071

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, *58*(3), 306-324.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201-215.

Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 308–325). Oxford University Press.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical Selection in Bilinguals: Do Words in the Bilingual's Two Lexicons Compete for Selection? *Journal of Memory and Language*, *41*(3), 365–397. https://doi.org/10.1006/jmla.1999.2651

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491–511. https://doi.org/10.1016/j.jml.2004.02.002

Crittenden, B. M., & Duncan, J. (2014). Task Difficulty Manipulation Reveals Multiple Demand Activity but no Frontal Lobe Hierarchy. *Cerebral Cortex*, *24*(2), 532–540. https://doi.org/10.1093/cercor/bhs333

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human brain mapping*, *8*(2-3), 109-114.

De Baene, W., Duyck, W., Brass, M., & Carreiras, M. (2015). Brain Circuit for Cognitive Control Is Shared by Task and Language Switching. *Journal of Cognitive Neuroscience*, *27*(9), 1752–1765. https://doi.org/10.1162/jocn_a_00817

Declerck, M., Kleinman, D., & Gollan, T. H. (2020). Which bilinguals reverse language dominance and why? *Cognition*, *204*, 104384. https://doi.org/10.1016/j.cognition.2020.104384

Declerck, M., & Koch, I. (2022). The concept of inhibition in bilingual control. *Psychological Review*. https://doi.org/10.1037/rev0000367

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, *22*(6), 1630–1645. https://doi.org/10.3758/s13423-015-0836-1

Degani, T., Kreiner, H., Ataria, H., & Khateeb, F. (2019). The impact of brief exposure to the second language on native language production: Global or item specific? *Applied Psycholinguistics*, 1–31. https://doi.org/10.1017/S0142716419000444

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, *116*(15), 7565–7574. https://doi.org/10.1073/pnas.1811513116

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2020). Duration and extent of bilingual experience modulate neurocognitive outcomes. *NeuroImage*, *204*, 116222. https://doi.org/10.1016/j.neuroimage.2019.116222

Diachek, E., Blank, I., Siegelman, M., Affourtit, J., & Fedorenko, E. (2020). The Domain-General Multiple Demand (MD) Network Does Not Support Core Aspects of Language Comprehension: A Large-Scale fMRI Investigation. *The Journal of Neuroscience*, *40*(23), 4536–4550. https://doi.org/10.1523/JNEUROSCI.2036-19.2020

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(3), 175-197.

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., Fox, M. D., Snyder, A. Z., Vincent, J. L., Raichle, M. E., Schlaggar, B. L., & Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, *104*(26), 11073–11078. https://doi.org/10.1073/pnas.0704320104

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., Burgund, E. D., Grimes, A. L., Schlaggar, B. L., & Petersen, S. E. (2006). A Core System for the Implementation of Task Sets. *Neuron*, *50*(5), 799–812. https://doi.org/10.1016/j.neuron.2006.04.031

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Fedeli, D., Del Maschio, N., Sulpizio, S., Rothman, J., & Abutalebi, J. (2021). The bilingual structural connectome: Dual-language experiential factors modulate distinct cerebral networks. *Brain and Language*, *220*, 104978. https://doi.org/10.1016/j.bandl.2021.104978

Fedorenko, E. (2021). The early origins and the growing popularity of the individual-subject analytic approach in human neuroscience. *Current Opinion in Behavioral Sciences*, *40*, 105–112. https://doi.org/10.1016/j.cobeha.2021.02.023

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428–16433. https://doi.org/10.1073/pnas.1112937108

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-Selective and Domain-General Regions Lie Side by Side within Broca's Area. *Current Biology*, *22*(21), 2059–2062. https://doi.org/10.1016/j.cub.2012.09.011

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*(41), 16616–16621. https://doi.org/10.1073/pnas.1315235110

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194. https://doi.org/10.1152/jn.00032.2010

Friesen, D. C., Edwards, K., & Lamoureux, C. (2021). Predictors of verbal fluency performance in monolingual and bilingual children: The interactive role of English receptive vocabulary and fluid intelligence. *Journal of Communication Disorders*, *89*, 106074. https://doi.org/10.1016/j.jcomdis.2020.106074

Frost, M. A., & Goebel, R. (2012). Measuring structural–functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*, *59*(2), 1369–1381. https://doi.org/10.1016/j.neuroimage.2011.08.035

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost–benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 640–665. https://doi.org/10.1037/a0014981

Gollan, T. H., Kleinman, D., & Wierenga, C. E. (2014). What's easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General*, *143*(6), 2167–2195. https://doi.org/10.1037/a0038006

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(02), 67. https://doi.org/10.1017/S1366728998000133

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: FMRI evidence from Chinese–English bilinguals. *NeuroImage*, *56*, 2300–2309.

Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing Cross-Linguistic Lexical Tasks (CLTs) for Bilingual Preschool Children. In *Assessing Multilingual Children* (pp. 196–240). Multilingual Matters. https://www.degruyter.com/document/doi/10.21832/9781783093137-010/html

Hu, J., Small, H., Kean, H., Takahashi, A., Zekelman, L., Kleinman, D., Ryan, E., Ferreira, V., & Fedorenko, E. (2022). Precision fMRI reveals that the language-selective network supports both phrase-structure building and lexical access during language production. *Cerebral Cortex*, *1*, 21.

Indefrey, P. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Ivanova, A. A., Mineroff, Z., Zimmerer, V., Kanwisher, N., Varley, R., & Fedorenko, E. (2021). The Language Network Is Recruited but Not Required for Nonverbal Event Semantics. *Neurobiology of Language*, *2*(2), 176–201. https://doi.org/10.1162/nol_a_00030

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825-841.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, *5*(2), 143-156.

Juch, H., Zimine, I., Seghier, M. L., Lazeyras, F., & Fasel, J. H. (2005). Anatomical variability of the lateral frontal lobe surface: implication for intersubject variability in language neuroimaging. *Neuroimage*, *24*(2), 504-514.

Kim, C., Chung, C., & Kim, J. (2012). Conflict adjustment through domain-specific multiple cognitive control mechanisms. *Brain Research*, *1444*, 55–64. https://doi.org/10.1016/j.brainres.2012.01.023

Kim, C., Cilles, S. E., Johnson, N. F., & Gold, B. T. (2012). Domain general and domain preferential brain regions associated with different types of task switching: A Meta-Analysis. *Human Brain Mapping*, *33*(1), 130–142. https://doi.org/10.1002/hbm.21199

Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching: A review. *Psychonomic Bulletin & Review*, *17*(1), 1–14. https://doi.org/10.3758/PBR.17.1.1

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences*, *11*(6), 229-235.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*(2), 119–135. https://doi.org/10.1017/S1366728906002483

La Heij, W. (2005). Selection process in monolingual and bilingual lexical access. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). Oxford University Press.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, *44*, 325-343.

Liljeström, M., Kujala, J., Stevenson, C., & Salmelin, R. (2015). Dynamic reconfiguration of the language network preceding onset of speech in picture naming. *Human Brain Mapping*, *36*(3), 1202–1216. https://doi.org/10.1002/hbm.22697

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2012). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes*, *27*(10), 1479–1488. https://doi.org/10.1080/01690965.2011.613209

Mackie, M. A., Van Dam, N. T., & Fan, J. (2013). Cognitive control and attentional functions. *Brain and cognition*, *82*(3), 301-312.

Malik-Moraleda, S., Cucu, T., Lipkin, B., & Fedorenko, E. (2021). The Domain-General Multiple Demand Network Is More Active in Early Balanced Bilinguals Than Monolinguals During Executive Processing. *Neurobiology of Language*, *2*(4), 647–664. https://doi.org/10.1162/nol_a_00058

Martin, A., Wiggs, C. L., Lalonde, F., & Mack, C. (1994). Word retrieval to letter and semantic cues: A double dissociation in normal subjects using interference tasks. *Neuropsychologia*, *32*(12), 1487–1494. https://doi.org/10.1016/0028-3932(94)90120-1

Mendoza, M. N., Blumenfeld, H. K., Knight, R. T., & Ries, S. K. (2021). Investigating the Link Between Linguistic and Non-Linguistic Cognitive Control in Bilinguals Using Laplacian-Transformed Event Related Potentials. *Neurobiology of Language*, *2*(4), 605–627. https://doi.org/10.1162/nol_a_00056

Meuter, R. F. I., & Allport, A. (1999). Bilingual Language Switching in Naming: Asymmetrical Costs of Language Selection. *Journal of Memory and Language*, *40*(1), 25–40. https://doi.org/10.1006/jmla.1998.2602

Michon, K. J., Khammash, D., Simmonite, M., Hamlin, A. M., & Polk, T. A. (2022). Person-specific and precision neuroimaging: Current methods and future directions. *NeuroImage*, *263*, 119589. https://doi.org/10.1016/j.neuroimage.2022.119589

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, *24*(1), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, *67*(1), 224–237. https://doi.org/10.1016/j.jml.2012.05.001

Naylor, L., Stanley, E., & Wicha, N. (2012). Cognitive and Electrophysiological Correlates of the Bilingual Stroop Effect. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00081

Nee, D. E., Jonides, J., & Berman, M. G. (2007). Neural mechanisms of proactive interference-resolution. *NeuroImage*, *38*(4), 740–751. https://doi.org/10.1016/j.neuroimage.2007.07.066

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(2), 241–268. https://doi.org/10.3758/s13415-011-0083-5

Nozari, N., & Novick, J. (2017). Monitoring and Control in Language Production. *Current Directions in Psychological Science*, *26*(5), 403–410. https://doi.org/10.1177/0963721417702419

Pallier, C., Dehaene, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first?. *Cerebral cortex*, *13*(2), 155-161.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

R Core Team. (2020). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Reverberi, C., Kuhlen, A. K., Seyed-Allaei, S., Greulich, R. S., Costa, A., Abutalebi, J., & Haynes, J.-D. (2018). The neural basis of free language choice in bilingual speakers: Disentangling language choice and language execution. *NeuroImage*, *177*, 108–116. https://doi.org/10.1016/j.neuroimage.2018.05.025

Rossi, E., Newman, S., Kroll, J. F., & Diaz, M. T. (2018). Neural signatures of inhibitory control in bilingual spoken production. *Cortex*, *108*, 50–66. https://doi.org/10.1016/j.cortex.2018.07.009

Runnqvist, E., Strijkers, K., Alario, F.-X., & Costa, A. (2012). Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language*, *66*(4), 850–869. https://doi.org/10.1016/j.jml.2012.02.007

Scott, T. L., Gallée, J., & Fedorenko, E. (2017). A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cognitive Neuroscience*, *8*(3), 167–176. https://doi.org/10.1080/17588928.2016.1201466

Shain, C., Paunov, A., Chen, X., Lipkin, B., & Fedorenko, E. (2022). No evidence of theory of mind reasoning in the human language network. *bioRxiv*, 2022-07.

Salig, L. K., Valdés Kroff, J. R., Slevc, L. R., & Novick, J. M. (2021). Moving from bilingual traits to states: Understanding cognition and language processing through moment-to-moment variation. *Neurobiology of Language*, *2*(4), 487-512.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00772

Shashidhara, S., Mitchell, D. J., Erez, Y., & Duncan, J. (2019). Progressive Recruitment of the Frontoparietal Multiple-demand System with Increased Task Complexity, Time Pressure, and Reward. *Journal of Cognitive Neuroscience*, *31*(11), 1617–1630. https://doi.org/10.1162/jocn_a_01440

Smith, D. M., Perez, D. C., Porter, A., Dworetsky, A., & Gratton, C. (2021). Light through the fog: Using precision fMRI data to disentangle the neural substrates of cognitive control. *Current Opinion in Behavioral Sciences*, *40*, 19–26. https://doi.org/10.1016/j.cobeha.2020.12.004

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Sulpizio, S., Del Maschio, N., Fedeli, D., & Abutalebi, J. (2020). Bilingual language processing: A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *108*, 834–853. https://doi.org/10.1016/j.neubiorev.2019.12.014

Tahmasebi, A. M., Davis, M. H., Wild, C. J., Rodd, J. M., Hakyemez, H., Abolmaesumi, P., & Johnsrude, I. S. (2012). Is the Link between Anatomical Structure and Function Equally Strong at All Cognitive Levels of Processing? *Cerebral Cortex*, *22*(7), 1593–1603. https://doi.org/10.1093/cercor/bhr205

Tomaiuolo, F., MacDonald, J. D., Caramanos, Z., Posner, G., Chiavaras, M., Evans, A. C., & Petrides, M. (1999). Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: an in vivo MRI analysis. *European Journal of Neuroscience*, *11*(9), 3033-3046.

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, *20*(6), 727–735. https://doi.org/10.3758/BF03202722

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2011). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *Quarterly Journal of Experimental Psychology*, *64*(3), 447–466. https://doi.org/10.1080/17470218.2010.505292

Van Assche, E., Duyck, W., & Gollan, T. H. (2013). Whole-language and item-specific control in bilingual language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1781–1792. https://doi.org/10.1037/a0032859

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*, *110*(1), 84–99. https://doi.org/10.1016/j.cognition.2008.10.013

Wehbe, L., Blank, I. A., Shain, C., Futrell, R., Levy, R., von der Malsburg, T., Smith, N., Gibson, E., & Fedorenko, E. (2021). Incremental Language Comprehension Difficulty Predicts Activity in the Language Network but Not the Multiple Demand Network. *Cerebral Cortex*, *31*(9), 4006–4023. https://doi.org/10.1093/cercor/bhab065

Weissberger, G. H., Gollan, T. H., Bondi, M. W., Clark, L. R., & Wierenga, C. E. (2015). Language and task switching in the bilingual brain: Bilinguals are staying, not switching, experts. *Neuropsychologia*, *66*, 193–203. https://doi.org/10.1016/j.neuropsychologia.2014.10.037

Wodniecka, Z., Casado, A., Kałamała, P., Marecka, M., Timmer, K., & Wolna, A. (2020). Chapter Eight—The dynamics of language experience and how it affects language and cognition. In K. D. Federmeier & H.-W. Huang (Eds.), *Psychology of Learning and Motivation* (Vol. 72, pp. 235–281). Academic Press. https://doi.org/10.1016/bs.plm.2020.02.005

Wodniecka, Z., Szewczyk, J., Kałamała, P., Mandera, P., & Durlik, J. (2020). When a second language hits a native language. What ERPs (do and do not) tell us about language retrieval difficulty in bilingual language production. *Neuropsychologia*, *141*, 107390. https://doi.org/10.1016/j.neuropsychologia.2020.107390

Wolna, A., Łuniewska, M., Haman, E., & Wodniecka, Z. (2022). Polish norms for a set of colored drawings of 168 objects and 146 actions with predictors of naming performance. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01923-3

Wu, J., Yang, J., Chen, M., Li, S., Zhang, Z., Kang, C., Ding, G., & Guo, T. (2019). Brain network reconfiguration for language and domain-general cognitive control in bilinguals. *NeuroImage*, *199*, 454–465. https://doi.org/10.1016/j.neuroimage.2019.06.022

Xu, J., Moeller, S., Auerbach, E. J., Strupp, J., Smith, S. M., Feinberg, D. A., Yacoub, E., & Uğurbil, K. (2013). Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *Neuroimage, 83,* 991-1001. https://doi.org/10.1016/j.neuroimage.2013.07.055

Zhang, Y., Wang, T., Huang, P., Li, D., Qiu, J., Shen, T., & Xie, P. (2015). Free Language Selection in the Bilingual Brain: An Event-Related fMRI Study. *Scientific Reports*, *5*(1), Article 1. https://doi.org/10.1038/srep11704