

# The use of a second language enhances the neural efficiency of inhibitory control: An ERP study

Patrycja Kałamała<sup>1</sup> , Jonas Walther<sup>1</sup>, Haoyun Zhang<sup>2</sup>, Michele Diaz<sup>2</sup>, Magdalena Senderecka<sup>3</sup>  and Zofia Wodniecka<sup>1</sup>

## Research Article

**Cite this article:** Kałamała P, Walther J, Zhang H, Diaz M, Senderecka M, Wodniecka Z (2022). The use of a second language enhances the neural efficiency of inhibitory control: An ERP study. *Bilingualism: Language and Cognition* 25, 163–180. <https://doi.org/10.1017/S1366728921000389>

Received: 14 October 2020  
Revised: 22 June 2021  
Accepted: 24 June 2021  
First published online: 5 August 2021

### Keywords:

patterns of language use; inhibition; adaptive control hypothesis; language switching; cognitive training

### Address for correspondence:

Patrycja Kałamała,  
E-mail: [patrycja.kalamala@uj.edu.pl](mailto:patrycja.kalamala@uj.edu.pl);  
Zofia Wodniecka,  
E-mail: [zofia.wodniecka@uj.edu.pl](mailto:zofia.wodniecka@uj.edu.pl)

<sup>1</sup>Psychology of Language and Bilingualism Lab, Institute of Psychology, Jagiellonian University in Krakow, Poland; <sup>2</sup>Social, Life, and Engineering Sciences Imaging Center, the Pennsylvania State University, Pennsylvania, United States and <sup>3</sup>Institute of Philosophy, Jagiellonian University in Krakow, Poland

### Abstract

This study investigated how natural language use influences inhibition in language-unbalanced bilinguals. We experimentally induced natural patterns of language use (as proposed by the Adaptive Control Hypothesis) and assessed their cognitive after-effects in a group of 32 Polish–English bilinguals. Each participant took part in a series of three language games involving real conversation. Each game was followed by two inhibition tasks (stop-signal task and Stroop task). The manipulation of language use in the form of language games did not affect the behavioural measures, but it did affect ERPs. Performance of the inhibition tasks was accompanied by a reduction of P3 and the N450 amplitude differences after games involving the use of L2. The ERP modulations suggest that for bilinguals living in an L1 context the use of L2 enhances neural mechanisms related to inhibition. The study provides the first evidence for a direct influence of natural language use on inhibition.

## 1. Introduction

There is general agreement in the literature that bilingualism involves the workings of cognitive control, i.e., a collection of top-down processes responsible for achieving goal-directed behaviour in the face of distraction (Bialystok, 2017; Green, 1998; Kroll, Dussias, Bogulski & Kroff, 2012). This proposal is supported by prominent models of bilingual language control which posit that bilinguals need to recruit domain-general inhibition mechanisms in order to reduce interference between constantly active languages (*Inhibitory Control Model*; Green, 1998; *Bilingual Interactive Activation Model*; Grainger, Midgley & Holcomb, 2010; Grainger & Dijkstra, 1992; Green, 1998; also see van Heuven & Dijkstra, 2010). This continuous management of multiple languages is proposed to develop and enhance the cognitive control system. However, the exact conditions under which between-language competition recruits cognitive control is a matter of ongoing research because the available evidence is inconsistent (de Bruin, 2019; Leivada, Westergaard, Duñabeitia & Rothman, 2020). Only recently has it been proposed that the engagement of cognitive control in bilinguals may depend on the patterns in which bilinguals use their languages (Abutalebi & Green, 2016; Bak, 2016; Bialystok, 2017; de Bruin, 2019; DeLuca, 2019; Green & Abutalebi, 2013; Pliatsikas, DeLuca & Voits, 2020). In particular, the ADAPTIVE CONTROL HYPOTHESIS (Green & Abutalebi, 2013; ACH hereafter) proposes that different patterns of language use act as a cognitive training, thus triggering different adaptive changes in the cognitive control system.

The current study examined how bilinguals' patterns of language use shape domain-general inhibition. Specifically, we tested the predictions of the ACH that are related to the language-use experiences of single-language and dual-language contexts (for details, see section 1.1). The patterns of language use were experimentally manipulated in a single group of bilinguals via language games. Since the language games involved natural language use (i.e., real conversation), they provided an ecologically valid manipulation of the language-use experience. The study should thereby allow us to assess whether and how natural language use affects inhibition.

### 1.1. The Adaptive Control Hypothesis

The ACH (Green & Abutalebi, 2013) distinguishes between three basic patterns of language use (the so-called interactional context in the ACH): single-language context, dual-language context, and dense code-switching. Importantly, the ACH posits that these patterns engage cognitive control in different ways. Bilinguals who mix elements of two languages, e.g., words, within single utterances (i.e., representing the DENSE CODE-SWITCHING CONTEXT)

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

practically do not engage cognitive control as they utilise whichever language route is most readily available. In contrast, bilinguals who switch but do not mix their languages daily engage much more cognitive control during language use. Such bilinguals operate in either a *SINGLE-LANGUAGE CONTEXT* (SL context), in which the person speaks only one language in each context (e.g., one language at home, another one at work), or a *DUAL-LANGUAGE CONTEXT* (DL context), in which the person speaks two languages in one context but distinct languages are spoken with distinct speakers. Bilinguals representing both SL and DL contexts are hypothesized to engage such cognitive processes as interference control and goal maintenance as this helps them to suppress cross-language interference and maintain fluent use of the target language. However, since interference is more likely to happen when two languages are used in the same situation, bilinguals in a DL context are assumed to engage interference control and goal maintenance to a greater extent. Moreover, since bilinguals in a DL context typically use both languages in one conversation, they also engage additional cognitive processes in their language use, i.e., cue detection, response inhibition, and task engagement/disengagement. These processes enable bilinguals in a DL context to monitor environmental cues that suggest language switches and suppress the currently used language if there is a need to switch.

Overall, bilinguals in both SL and DL contexts are assumed to engage inhibitory processes in their language use. Crucially, however, since bilinguals in a DL context are expected to experience relatively high inhibitory demands on language use, they are assumed to engage and train this mechanism to a greater extent compared with other bilinguals. The available evidence for the ACH comes from two lines of research: cross-sectional studies that assess everyday habits of language use, and studies that manipulate language experience in an experimental setting. The following two sections present relevant findings concerning inhibition (for a detailed summary of the reviewed experiments, see Tables A.3. and A.4. in Wodniecka, Casado, Kałamała, Marecka, Timmer & Wolna, 2020).

### 1.2. Efficiency of inhibition and bilinguals' everyday language-use habits

There are only a few studies that have focused on how variation in everyday language-use habits differentiates bilinguals in terms of their inhibitory skills (Hartanto & Yang, 2020; Kałamała, Szewczyk, Chuderski, Senderecka & Wodniecka, 2020b; Pot, Keijzer & de Bot, 2018; see also Beatty-Martinez, Navarro-Torres, Dussias, Bajo, Guzzardo Tamargo & Kroll, 2019; Gullifer, Chai, Whitford, Pivneva, Baum, Klein & Titone, 2018; Henrard & Van Daele, 2017; Ooi, Goh, Sorace & Bak, 2018). Pot and colleagues (2018) found that greater self-assessed diversity in language use across social contexts (an SL context in this paper) is related to a smaller flanker effect in RTs. This effect was observed in a group of older adults who were highly proficient in L2 and used this language on a daily basis. However, some results do not support the predictions of the ACH. Kałamała and colleagues (2020b) did not find support for the relationship between the self-assessed intensity of the DL context experience and inhibition (assessed by four different tasks) in a group of young adult bilinguals who declared high proficiency of L2 and everyday use of this language. In turn, in a group with similar self-assessed L2 proficiency and daily use of L2 as in Kałamała and colleagues (2020b), Hartanto and Yang (2020)

showed that greater self-assessed exposure to a DL context was related to better task switching (assessed by three different switching tasks), but neither exposure to a DL context nor exposure to an SL context impacted indices of inhibition in this study, which is at odds with the predictions of the ACH (for the effects related to the DCS context, see the text).

There are three potential reasons for the current inconsistency. Firstly, in most studies the patterns of language use were assessed via the participants' self-reports (Hartanto & Yang, 2020; Kałamała et al., 2020b; Pot et al., 2018). However, it is not clear to what extent individuals are able to adequately self-assess their language-use patterns, and studies usually do not report psychometric properties for measures derived from self-reports (for an exception, see Kałamała et al., 2020b). Secondly, even if bilinguals' patterns of language use are adequately assessed, individuals experiencing the same patterns can still differ in other aspects of bilingualism, such as language proficiency (for arguments, see Beatty-Martínez et al., 2019; de Bruin, 2019). Therefore, it is possible that the demands imposed by a pattern of language use could interact with other aspects of bilingualism, and these interactions may confound the measurement of pattern-specific effects (DeLuca, 2019; DeLuca, Rothman, Bialystok & Pliatsikas, 2019; Gullifer et al., 2018; Pliatsikas et al., 2020). Thirdly, the available evidence mostly comes from behavioural studies. However, behavioural measures such as RTs and accuracy reflect not only the cognitive process in question but also other irrelevant processes related to performance (e.g., perceptual processing, memorizing task rules, and so on; the task-impurity problem; Gratton, Cooper, Fabiani, Carter & Karayanidis, 2017; Miyake, Friedman, Emerson, Witzki, Howerter & Wager, 2000). In turn, this might contaminate the measurement of the targeted process (inhibition in this case). In contrast to behavioural measures, neuroimaging methods with a high temporal resolution, such as event-related potentials (ERP), track the neural processes that lead to the behaviourally observed outcomes in real time (Cespón & Carreiras, 2020; DeLuca, 2019; Gratton et al., 2017; Pliatsikas et al., 2020). Although the use of neuroimaging methods enables the investigation of processes that might otherwise be obscured in behavioural measurements, these methods were not used to test the cognitive effects of daily patterns of language use in any of the studies (but for evidence on resting-state brain connectivity, see Bice, Yamasaki & Prat, 2020; Gullifer et al., 2018).

### 1.3. Efficiency of inhibition and short-term manipulation of language experience

A promising approach to circumventing the problems of assessing real-life patterns of language use and controlling the rich diversity of language experiences is to experimentally manipulate patterns of language use within the same group of bilinguals (a so-called within-subject design; Pliatsikas et al., 2020; Wodniecka et al., 2020), which should allow straightforward assessment of cognitive effects related to differences in bilingual patterns of language use. Researchers have only recently begun to experimentally manipulate language experience. Crucially, some of these studies collected both behavioural and neuroimaging data (e.g., Jiao, Liu, Liang, Plummer, Perfetti & Chen, 2019; Wu & Thierry, 2013; see also Cespón & Carreiras, 2020) and thereby provided a detailed measurement of inhibition.

In the seminal research, Wu and Thierry (2013) showed that passive exposure to both languages (i.e., imitating a DL context in this study) improves inhibition. In this study, young adult bilinguals

who declared high-to-moderate proficiency of L2 and everyday use of this language performed a flanker task in which the flanker trials were interspersed with words in either L1 (L1 condition), L2 (L2 condition), or L1 and L2 presented alternately (mixed condition). Participants were more accurate in resolving the flanker conflict when exposed to both languages than when exposed only to L1 or only to L2. Furthermore, their better performance during ongoing exposure to both languages was accompanied by reduced P300 amplitude for flanker-incongruent trials, thus reflecting that participants experienced less interference from incongruent trials (but see Jiao et al., 2019; see also Adler, Valdés Kroff & Novick, 2020; Hofweber, Marinis & Treffers-Daller, 2020).

In addition to trial-by-trial manipulations, language experience can be manipulated with short-term language training, which usually takes the form of a cued picture-naming paradigm. While some studies have shown that short-term language training impacts bilinguals' non-linguistic switching abilities (Prior & Gollan, 2011; Timmer, Calabria & Costa, 2019; Timmer, Christoffels & Costa, 2019) and proactive control (Zhang, Kang, Wu, Ma & Guo, 2015), we are aware of only one study which utilized within-subject, short-term training when testing inhibition (Yang, Ye, Wang, Zhou & Wu, 2018; for evidence in a between-subject design see Liu, Yang, Jiao, Schwieter, Sun & Wang, 2019). In the study by Yang and colleagues (2018), trilingual speakers who were balanced in terms of proficiency and use of L1 and L2 (but not L3) underwent three versions of a blocked picture-naming task (i.e., blocks of pictures requiring the use of one language alternated with blocks of pictures requiring the use of another language), each of which was followed by performance of a flanker task. Additionally, the behavioural measurement was accompanied by fMRI data recording. The language training imitated three different instances of the DL context: L1-L2, L2-L3, L1-L3. The study showed that the flanker effect was substantially reduced after the L1-L2 training compared to both L1-L3 and L2-L3 (but only in terms of accuracy), thus suggesting improved inhibition after a short session of switching between well-known languages. This effect was further confirmed by the fMRI data, which showed reduced neural activation in the prefrontal cortex and some subcortical areas after a session of switching between L1 and L2 compared to the other conditions.

Taken together, studies that experimentally manipulate language experience provide promising alternatives for testing how cognitive control in bilinguals can be affected by differences in their language-use experience. Importantly, however, when one attempts to relate experimental manipulations to real-life language-use habits, experimental protocols should involve natural language use. This has not been the case in previous studies, as participants were either passively exposed to language-related stimuli (e.g., Wu & Thierry, 2013) or were engaged in cued picture-naming tasks which require the memorizing of arbitrary associations between cues and languages and artificially force language changes between single words (e.g., Yang et al., 2018). Therefore, these types of language interventions have relatively low ecological validity (for additional arguments, see Blanco-Elorrieta & Pykkänen, 2018; van den Noort, Struys, Bosch, Jaswetz, Perriard, Yeo, Barisch, Vermeire, Lee & Lim, 2019; Wodniecka et al., 2020).

#### 1.4. Present study

In the current study, we tested how natural bilingual language use influences inhibition efficiency. The patterns of language use were experimentally manipulated in a single group of bilinguals via a

series of language games. Since the language games involved real conversations, they provided an ecologically valid manipulation of language-use experience. After each game, the participants performed two inhibition tasks (for an overview of the study design, see Figure 1). This within-subject design allows more straightforward attribution of the observed cognitive after-effects to natural patterns of language use while controlling for individual differences in the participants' background characteristics.

We tested a group of bilinguals who lived in their native-language environment (i.e., Polish, L1) and were relatively homogenous in terms of their background characteristics. Only bilinguals who were proficient in English (L2) and rarely used this language on a daily basis participated in the study (for details, see section 2.1). The games differed in terms of how the languages were used: 1) THE L1 GAME required the use of L1; 2) THE L2 GAME required the use of the non-dominant language, i.e., L2 in an L1 environment; 3) THE DUAL-LANGUAGE (DL) GAME required switching between L1 and L2 depending on the game partner. Since the participants lived in their L1 environment and mostly used L1 on a daily basis, the L1 game did not differ from their typical language use and was considered as the baseline for between-game comparisons. The L2 game and the DL game differed from participants' typical language use and represented the SL context and the DL context, respectively.

We used two well-established inhibition tasks: the stop-signal task and the Stroop task (Diamond, 2013). The former task requires inhibition at the level of manual response, whereas the latter task requires inhibition at the level of speech production. The use of two intrinsically different tasks was intended to provide a more fine-grained measurement of inhibition. At the behavioural level, we focused on the stop-signal reaction time (so-called SSRT) in the stop-signal task and the Stroop effect (in RTs and accuracy) in the Stroop task. We also supplemented the behavioural measurement by recording ERPs. We focused on a set of ERP components whose spatiotemporal characteristics differ and thus reflect the engagement of the inhibition mechanism at various stages of information processing, i.e., N2 and P3 in the stop-signal task (Nieuwenhuis, Yeung, van den Wildenberg & Ridderinkhof, 2003) and N450 in the Stroop task (Liotti, Woldorff, Perez & Mayberg, 2000; see also Cespón & Carreiras, 2020). N2 is a fronto-central negativity that peaks around 200–300 ms after the stimulus onset. More negative N2 amplitudes for unsuccessfully than for successfully inhibited trials in the stop-signal task (i.e., the N2 unsuccessful > successful inhibition effect) are typically interpreted as reflecting detection and/or monitoring of the conflict between go and inhibitory responses (Dimoska, Johnstone & Barry, 2006; Nieuwenhuis et al., 2003; Senderecka, 2016). P3 is a centro-parietal positivity which peaks around 300–350 ms after the stimulus. More positive P3 amplitudes for successfully than for unsuccessfully inhibited trials in the stop-signal task (i.e., the P3 successful > unsuccessful inhibition effect) are assumed to reflect mechanisms involved in successful response inhibition (Berkman, Kahn & Merchant, 2014; Manuel, Bernasconi & Spierer, 2013; Senderecka, 2018; Senderecka, Szweczyk, Wichary & Kossowska, 2018; Spierer, Chavan & Manuel, 2013). N450 is a fronto-central negative deflection peaking around 350–500 ms post-stimulus (Liotti et al., 2000). More negative N450 amplitudes for incongruent than for congruent trials in the Stroop task (i.e., the N450 incongruent > congruent trial effect) are assumed to reflect monitoring and/or suppression of semantic interference between the colour of the ink and the meaning of the word (Hsieh, Huang, Wu, Chang

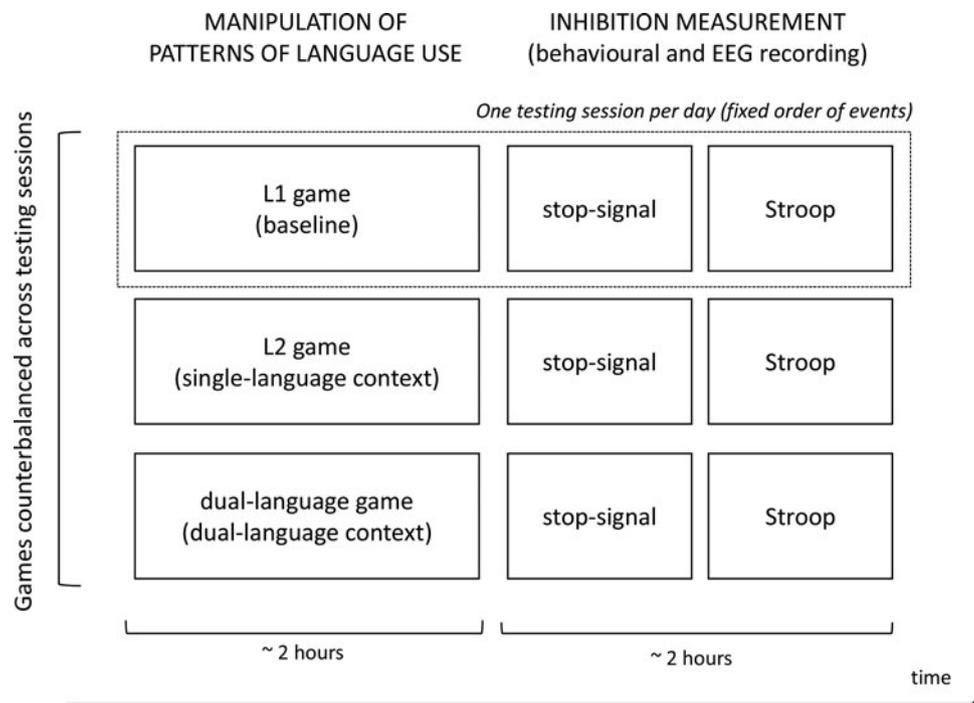


Fig. 1. Overview of the experimental design. Note. L1, Polish; L2, English; dual-language, Polish and English.

& Hung, 2018; Larson, Kaufman & Perlstein, 2009; Liotti et al., 2000; Szűcs & Soltész, 2012).

Based on research which shows cognitive effects of short-term manipulations of language use (Prior & Gollan, 2011; Timmer, Calabria et al., 2019; Yang et al., 2018), we expected that the manipulation of language-use experience would affect the subsequent performance of inhibition tasks. The two predictions were formulated on the basis of the ACH (Green & Abutalebi, 2013). Firstly, if the DL and SL contexts improve inhibition, we should observe more efficient performance in inhibition tasks after playing the DL game and the L2 game compared to after the L1 game. Secondly, if the use of two languages in one context without mixing them (i.e., DL context) benefits inhibition more than the use of these languages separately (i.e., different languages in different contexts; SL context), we should observe more efficient performance in the inhibition tasks after a session of the DL game compared to after the L2 game (see also Figure 1). Regarding the behavioural measures, more efficient inhibition should be reflected in shorter SSRT and smaller Stroop effects in RTs and accuracy. With respect to ERPs, following previous ERP research that examined training effects on inhibitory performance (Chang, Alderman, Chu, Wang, Song & Chen, 2017; Hsieh et al., 2018; Schroder, Dubuson, Dousset, Mortier, Kornreich & Campanella, 2020; see also Cespón & Carreiras, 2020; Wu & Thierry, 2013; Jiao, Grundy, Liu & Chen, 2020), we predicted that improvements in inhibition-related mechanisms after the DL and L2 games would be reflected in a reduction of ERP effects: the N2 unsuccessful > successful inhibition effect, the P3 successful > unsuccessful inhibition effect, and the N450 incongruent > congruent trial effect.<sup>1</sup>

<sup>1</sup>To better isolate the neural activity related to the experimental manipulations, we deliberately focused on ERP amplitude differences between task conditions (i.e., successfully vs. unsuccessfully inhibited trials and incongruent vs. congruent trials) and did not compare absolute (overall) ERP amplitudes across language games. Since the three testing

## 2. Methods

### 2.1. Participants

All participants ( $N = 32$ , 18 female)<sup>2</sup> were right-handed, healthy young adults (mean age 22 years;  $SD = 2.2$  years). They were recruited via an experimental recruitment system at Jagiellonian University, Kraków. Using an online platform, volunteers completed a socio-demographic background questionnaire and two English proficiency tests: the Cambridge General English test (Cambridge Assessment English, 2018; Cambridge test) and the Lexical Test for Advanced Learners of English (Lemhöfer & Broersma, 2012; LexTALE). Only Polish native speakers (L1) in good health (i.e., free of medications and with normal or corrected-to-normal vision) who were relatively proficient in English (L2) (i.e., scored at least 20 out of 25 in the Cambridge test and at least 45 out of 60 in the LexTALE test) were invited to participate in the study. Thirty-one participants completed all three sessions; one participant missed the third session but was included in the analyses where possible.

On average, the participants were highly educated (years of formal education,  $M = 14.56$ ,  $SD = 1.98$ ) and obtained relatively high scores on the fluid intelligence test ( $M = 14.19$  out of 18,  $SD = 3.26$ , on a shortened version of Raven's Advanced Progressive Matrices test; only odd-numbered items and 20 minutes to complete). Their language proficiency and history of language learning were evaluated using a language-background questionnaire based

sessions were conducted on separate days and the language games were counterbalanced across participants (for details, see section 2.2), this approach allowed us to reduce the within-subject error variance related to the experimental design (day-to-day variability of the EEG signal; for similar arguments see Luck & Gaspelin, 2017; Nielsen & Gonzalez, 2020).

<sup>2</sup>The participants' sample size was predetermined based on previous research that adopted a language-training design (i.e., Zhang et al., 2015). An *a priori* power analysis indicated the need for a minimum of 23 participants to detect a language-training effect of size  $d = .80$  for a measure of cognitive control with  $\alpha = .05$ .

on Marian, Blumenfeld and Kaushanskaya (2007) and Li, Zhang, Tsai and Puls (2014). The participants were language-unbalanced Polish–English bilinguals, all of whom had acquired only their native language in early childhood. On average, they started learning English as their L2 at around the age of six ( $SD = 1.87$ ). Table 1 presents self-assessment data concerning the participants' language abilities. The participants rated their overall L1 proficiency higher than their L2 proficiency, which they considered intermediate to high ( $p < .001$ ). Consistent with the self-assessment, they correctly completed 97% ( $SD = 4\%$ ) of the Cambridge test and 85% ( $SD = 7\%$ ) of the LexTALE test, thus indicating their moderate-to-high proficiency in L2. Thirty participants declared they had started learning an additional foreign language (predominantly German, French, or Russian), but their overall self-rated proficiency of these languages ( $M = 3.49$ ,  $SD = 1.36$ ; for a rating scale see Table 1) was lower than their L1 or L2 proficiency ( $ps < .001$ ).

The participants declared that they used their L1 on a daily basis more often than their L2 ( $p < .001$ ). Their language-switching habits were evaluated using two indices: the frequency of intrasentential code-switching and the frequency of intersentential code-switching, both of which are derived from the CODE-SWITCHING AND INTERACTIONAL CONTEXTS QUESTIONNAIRE (Hartanto & Yang, 2016). Overall, participants rarely used two languages in the same situation on a daily basis. When they did, they more frequently switched their languages within single sentences than between sentences ( $p < .001$ ; for details see Table 1).

## 2.2. Measures and procedure

### General procedure

Experimental procedures for data collection and analysis were approved by the Institutional Review Board of Jagiellonian University and the Pennsylvania State University. The participants signed an informed consent form prior to the experiment and were paid PLN 250 (about \$65.75) for their participation in the study.

The experiment consisted of three testing sessions that were conducted on separate days with at least a two-day break between them (up to a ten-day break). Each testing session consisted of one language game immediately followed by an electroencephalography (EEG) recording. Both the language games and the EEG sessions were conducted in appropriately adapted laboratory rooms. The participants were always tested individually. After completion of a game, each participant was informed about their overall score in the game and asked to assess the difficulty of the game in terms of speaking effort (from 1 = *extremely easy* to 15 = *extremely difficult*). The assessment of the game's difficulty served as an explicit measure of whether the participants experienced differences in speaking effort across the language games. Then, the participant was immediately directed to an EEG recording room in the company of the main experimenter, seated (approximately 80 cm from the computer screen) and EEG capped, which took up to 20 minutes. Afterwards, two inhibition tasks were performed, i.e., the stop-signal task and the Stroop task, administered in a fixed order using DMDX software (Forster & Forster, 2003). To reduce the number of EEG artifacts, the participant was instructed to restrict their body movements and try to blink only after the response. After completing the experimental tasks in the first session, the participant filled out two language-background questionnaires (both described in section 2.1). At the end of the third session, they completed a shortened form of Raven's Advanced Progressive Matrices test and were informed about the goals of the study. With the exception

**Table 1.** Participants' language experience based on self-assessment questionnaires.

	Polish (L1)	English (L2)
	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )
<b>Self-rated overall proficiency<sup>1</sup></b>	8.91 (0.31)	7.91 (0.73)
Listening exposure	8.97 (0.31)	8.13 (0.87)
Reading proficiency	8.97 (0.31)	8.31 (0.74)
Speaking fluency	8.87 (0.43)	7.53 (0.98)
Writing proficiency	8.84 (0.45)	7.66 (0.83)
<b>Percentage of daily use</b>	61 (18)	39 (23)
<b>Self-rated patterns of language use (Hartanto &amp; Yang, 2016)</b>	<b><i>M</i> (<i>SD</i>)</b>	
Frequency of intersentential code-switching <sup>2,3</sup>	3.72 (1.54)	
Frequency of intrasentential code-switching <sup>2,4</sup>	4.84 (1.87)	

Notes. *M*, mean; *SD*, standard deviation; 1) self-ratings were 1 = "no knowledge of given language" to 9 = "native-like proficiency"; 2) self-ratings were 1 = "never" to 9 = "always"; 3) general tendency to switch languages between sentences; 4) general tendency to switch languages within sentences.

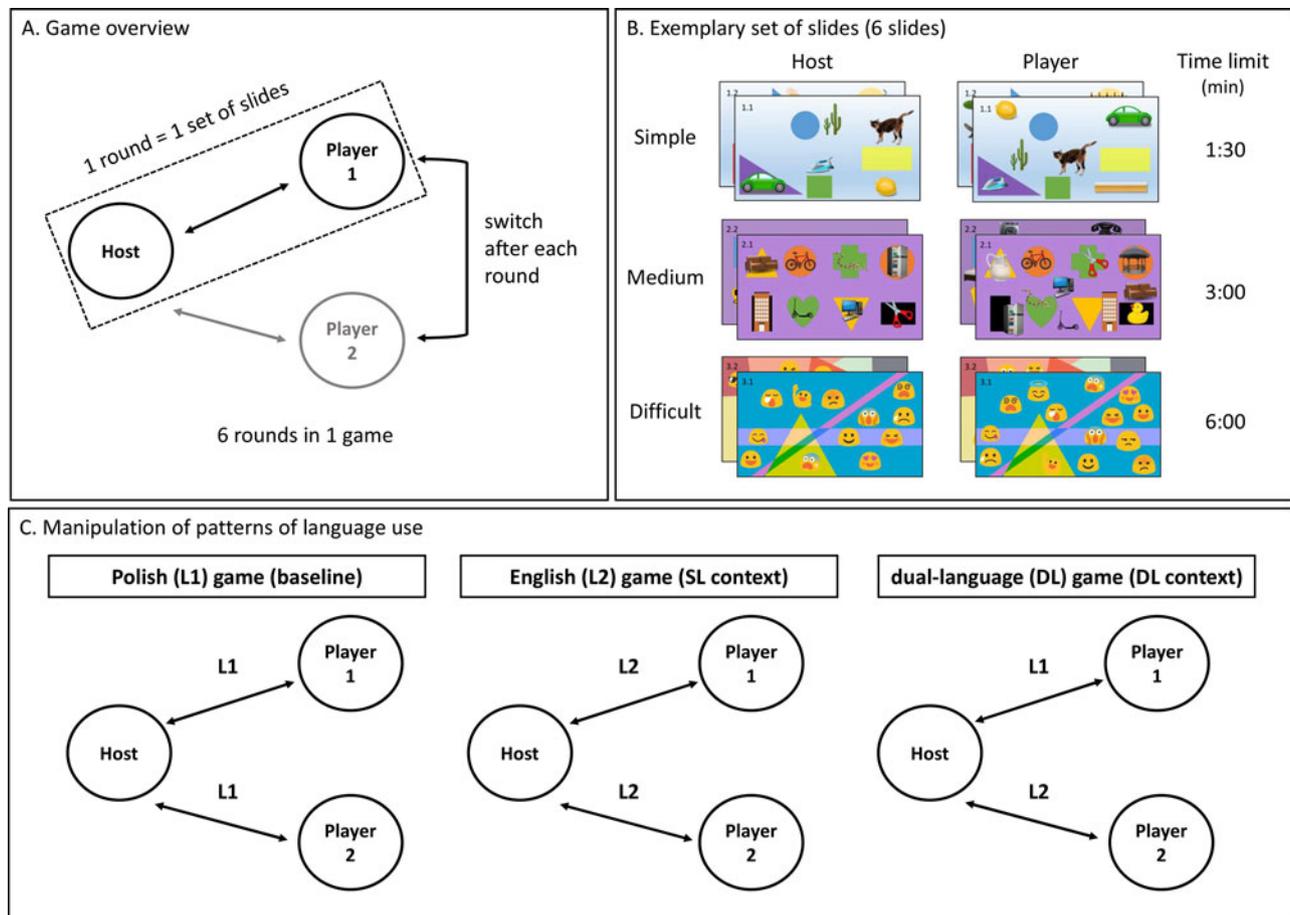
of language games (see below), all instructions and communication were in English (L2 for participants).

### Patterns of language-use manipulation: language games

The language game was based on the Map Task (Brown, Anderson, Shillcock & Yule, 1985). One game involved three players and consisted of six rounds (see also Panel A of Figure 2). One round involved two game partners, each of whom received a set of six picture slides. The slides differed in the number of elements and their arrangement on the slide (see also Panel B of Figure 2). The role of one game partner (the host) was to describe the content of the slides to the other game partner (the confederate), who had to rearrange the elements on their slides and remove unnecessary ones in order to match the host's versions as closely as possible.

The participant was always assigned to the role of the host; the two experimenters who acted as game players switched their roles (confederate and inactive player) between the game rounds. The participants were not aware of the nature of the experimenters who were acting as confederates; instead they were told that all roles in the game were assigned based on a random draw. The experimenters who were acting as confederates were aware of the reasoning behind the experiment. Each game lasted approximately 120 min, out of which around 90 min were used purely for speaking between the host (i.e., a participant) and the confederates. The remaining time was used to clarify the rules and to set up the game and equipment. A detailed description of the stimuli and game procedure can be found in Appendix S1. A fully documented game set-up is available online at <https://osf.io/xy4qg>.

There were three games that differed only in the language-use rules (see also Panel C of Figure 2): 1) all players used Polish; 2) all players used English; 3) one experimenter used Polish and the other used English, therefore the participant had to switch between Polish (L1) and English (L2) between the rounds of the game, depending on the confederate's language (the DL



**Fig. 2.** Overview of the game design. Panel A presents the general setup of the game: every game round involves a host (i.e., a participant) and a confederate (i.e., one of the two experimenters acting as players), each of whom receives one set of picture slides. Panel B shows an exemplary set of slides: based on verbal communication with the host, the confederate re-arranges elements on their slides so that they match the host's versions as closely as possible. The task difficulty increases as the number of items increases and the background becomes more complex. Panel C shows how language-use circumstances are manipulated in the language games: the arrows indicate possible ways of communication. L1 refers to the use of Polish; L2 refers to the use of English.

game). The participants gained points by correctly rearranging elements on the confederate's slides (one point for every correctly completed slide) and lost points if they went over the time limit for the round (minus one point for every thirty seconds of extra time). Additionally, they lost points when incorrect elements were placed on the confederate's slide (minus one point for every incorrect element on a slide) or if they spoke in the wrong language (minus one point for every utterance).

The differences in the language requirements across the three games allowed the patterns of language use to be manipulated, thus constituting the three experimental conditions: the L2 game, the dual-language (DL) game and the L1 game. The L2 game (use of the non-dominant language) was meant to imitate the SL context, while the DL game (switching between L1 and L2 in the same situation) was meant to imitate the DL context. The L1 game (use of the dominant language) was considered as a baseline for between-game comparisons. The order of the games was counterbalanced between participants in a Latin square design; the participants were informed about the version of the game at the beginning of a language-game session.

#### *Measurement of response inhibition: Stop-signal task*

Participants completed a stop-signal task with an auditory stop stimulus (e.g., Senderecka, 2018). Each trial began with the

presentation of a central fixation cross for 1200 ms, immediately followed by the presentation of a black screen for 200 ms. Afterwards, a visual go stimulus was presented for 100 ms in the centre of the screen. The go stimulus consisted of a horizontal arrow pointing to the left or the right with 50% probability for each direction. The stimuli were shown in white against a black background. The length of the arrow in the display was 20 mm (1.71°). The fixation cross was 6 mm (0.51°) in width. Participants were instructed to indicate the direction of the arrow (i.e., left or right) by pressing the corresponding Ctrl key (i.e., left or right, respectively) using their index fingers. In a random sample of 25% of trials, a 1400 Hz tone served as the stop signal. It was presented binaurally over EEG-compatible headphones (Sennheiser HD 429; intensity 60 dB SPL, duration 100 ms) immediately after the presentation of the arrow. The sound prompted participants to inhibit their responses to the primary go task, regardless of the arrow direction.

The interval between the presentation of the go stimulus and the stop signal (i.e., the stop-signal delay, SSD) was varied trial-by-trial using a tracking method: the interval increased or decreased by 50 ms (from 100 to 400 ms) for the next stop-signal trial, depending on whether participants had successfully or unsuccessfully inhibited their response to the go stimulus. There were seven possible SSDs: 100, 150, 200, 250, 300, 350,

and 400 ms. After a successful inhibition, the interstimulus interval became longer; after an unsuccessful inhibition, it became shorter. The initial value of the SSD was set to 150 ms. The tracking method aimed to converge on an SSD at which participants successfully inhibited responses to approximately 50% of the stop-signal trials. The timeout for a trial was 1500 ms.

In each testing session, participants received one practice block of 20 go-trials and six stop-signals. They were instructed to react to the go stimulus as quickly and as accurately as possible, but they also had to try to stop their response during trials that included the stop signal. After the practice runs, they completed five experimental blocks, each consisting of 56 trials with short breaks in between. During the break, the accuracy feedback for go-trials and mean RT were presented centrally on the screen.

#### Measurement of response inhibition: Stroop task

Participants completed a modified version of the Stroop task (Stroop, 1935). Each trial began with a white fixation cross presented centrally for 500 ms, immediately followed by the presentation of a black screen for 300 ms. Afterwards, a coloured word was presented in the centre of the screen until a response was made or time ran out (1600 ms). Then, a blank screen was shown for 180 ms. The stimuli were four Polish words displayed in blue, green, red, or yellow: blue (“niebieski”), green (“zielony”), red (“czerwony”), and yellow (“żółty”). The length of words on the screen was 40 to 70 mm (3.42° to 5.98°). The fixation cross was 7 mm (0.60°) in width. The stimuli were presented against a black background. For the congruent trials, the colour of the ink corresponded to the word’s meaning (e.g., “red” printed in red). For the incongruent trials, the colour of the ink did not correspond to the word meaning (e.g., “red” printed in blue). Participants were instructed to name the colour of the ink aloud as quickly and accurately as possible. RTs for vocal responses were automatically measured using a DMDX voice key and were manually screened for any artifactual sounds. During each experimental session, participants completed two experimental blocks, each consisting of 186 trials with a short break in between. Each block consisted of 30% incongruent and 70% congruent trials presented in random order. In the first testing session, participants first received two practice blocks. The first practice block consisted of 12 trials, and participants named the colour of the ink when a neutral string of letters was presented (i.e., HHHHHH). The second practice block consisted of 12 trials (30% of which were incongruent), and participants named the colour of the ink when the coloured words were presented. In the other testing sessions, they received only the latter practice block in order to remind them of the task’s rules.

### 2.3. Data pre-processing

#### Accuracy and reaction times: Stop-signal task

We focused on SSRT, which provides an estimate of the latency of the inhibitory process (Verbruggen & Logan, 2008). It was calculated following the standard procedure by Logan (1994). RTs from go responses in which no stop signal occurred were rank ordered. The  $n$ th RT was selected, where  $n$  was obtained by multiplying the number of no-signal RTs in the distribution (210) by the probability of responding (e.g., 0.5 if the inhibition rate in the task was 50%) for each participant separately. The SSRT was calculated by subtracting the average SSD from the  $n$ th RT (for details, see Logan & Cowan, 1984; Verbruggen & Logan, 2008). The SSRT scores were normally distributed.

#### Accuracy and reaction times: Stroop task

We focused on accuracy and RTs for congruent trials and incongruent trials. For the accuracy measure, timeouts were taken as erroneous responses. For RTs, only correct trials were included. Also, extremely short RTs (< 300 ms) and RTs that were three standard deviations above or below the condition mean for a participant were discarded from the analysis (2.8% of all trials). Due to the skewed distribution of RTs, the data was log-transformed.

#### ERP data

The continuous scalp EEG was recorded from 32 Ag/AgCl active electrodes (with preamplifiers) using the BioSemi ActiveTwo system. The electrodes were secured in an elastic cap according to the extended 10–20 international electrode placement system. The zero-reference principal voltage values (each site was quantified relative to the DRL and CMS loop) were digitized at a sampling rate of 256 Hz. The horizontal and vertical electro-oculograms were monitored using additional electrodes placed above and below the right eye and in the external canthi of both eyes in order to control for ocular artifacts.

EEG data were pre-processed using the BrainVision Analyzer 2 (Brain Products, Munich, Germany). All channels were re-referenced to the average of the two mastoid electrodes. The data were filtered with a 0.05 Hz high-pass filter (slope 24 dB/oct) and a 45 Hz low-pass filter (slope 12 dB/oct). The EEG data were then segmented relative to stimulus onset into -100–700 ms segments. Ocular artifacts were corrected using the Gratton and Coles method (Gratton, Coles & Donchin, 1983). After ocular correction, contaminated trials exceeding amplitudes of  $\pm 75 \mu\text{V}$  were rejected by a semi-automatic procedure.

Stimulus-locked segments were subsequently checked separately for each trial type (i.e., a successful and an unsuccessful stop in the stop-signal task, and correct-congruent and correct-incongruent in the Stroop task). Afterwards, ERPs were aligned to the pre-stimulus baseline from -100 ms to 0 ms. The mean number of artifact-free epochs per participant included in the ERP analysis for the stop-signal task was as follows: successful stop,  $M = 35$  ( $SD = 4$ ; min = 18); unsuccessful stop,  $M = 33$  ( $SD = 3$ ; min = 23). For the Stroop task: correct-congruent,  $M = 256$  ( $SD = 14$ ; min = 171); correct-incongruent,  $M = 103$  ( $SD = 5$ ; min = 78).

Appropriate electrode clusters and time windows for the targeted ERP components, i.e., N2, P3 and N450, were selected a priori based on previous studies using the same or similar experimental paradigms: a pronounced negativity around the fronto-central electrodes (Fz, Cz, FC1, FC2) in the 220–270 ms time window for N2 (Dimoska et al., 2006; Senderecka, 2016); a pronounced positivity around the centro-parietal electrodes (Cz, Pz, CP1, CP2) in the 270–400 ms time window for P3 (Senderecka, 2018; Senderecka et al., 2018); and a pronounced negativity at the fronto-central electrodes (Fz, FC1, FC2, Cz) in the 350–500 ms time window for N450 (Kałamała, Ociepka & Chuderski, 2020a; Larson, Clayson & Clawson, 2014; Rey-Mermet, Gade & Steinhauser, 2019). Mean voltage amplitudes in the pre-specified electrode clusters and time windows for each trial were used for statistical analysis. The distributions of the ERP data did not differ from the normal distribution.

### 2.4. Statistical analyses

Linear mixed effects (LME) regression models were used to establish how prior language-use manipulation affected the estimates

of inhibition: the SSRT, N2 and P3, Stroop RT and N450. Initially, we planned to also analyse the accuracy of the Stroop task; however, we noted very high accuracy scores (see Table 2), therefore no further analysis was conducted. All models were fitted using the lme4 package in R (version 1.1-13; R Core Team, 2019) with the BOBYQA optimizer included (Bates, Mächler, Bolker & Walker, 2015). The fixed effects were coded using *a priori* contrasts, as recommended by Schad, Vasishth, Hohenstein and Kliegl (2020). We tested the two models against each of the outcome variables.

Model 1 tested the first prediction, i.e., whether the language experiences that arise in the DL and SL contexts enhance inhibition compared to the use of L1 in an L1 environment. The model for the SSRT included Prior language use (i.e., L1 game, L2 game, DL game) as the participant-related fixed effect. The model for the other outcome variables included two additional participant-related fixed effects: Trial type (successful and unsuccessful stops in the stop-signal task; congruent and incongruent in the Stroop task) and the interaction between Prior language use and Trial type. Prior language use was coded using TREATMENT CONTRAST with the L1 game as the reference level so that the estimated model parameters reflected differences between the L2 game (SL context) and the L1 game (i.e., L2-L1 game contrast) and between the DL game (DL context) and the L1 game (DL-L1 game contrast). Trial type was coded using SUM CONTRAST such that the model parameters reflected the difference between the trial types (i.e., successful vs. unsuccessful stops in the stop-signal task, and congruent vs. incongruent in the Stroop task). Model 2 tested the second prediction, i.e., whether the DL context improves inhibition more than the SL context. The model included the same participant-related fixed effects but differed in the levels of TREATMENT CONTRAST for Prior language use. Here, the L2 game (SL context) was taken as the reference for the DL game (DL context), and the L1 game was excluded (i.e., DL-L2 game contrast). Models 1 and 2 were supplemented with direct tests for each type of Prior language use separately, i.e., DL game, L2 game, L1 game, each of which included Trial type as defined above.

For each model, we started with the maximal structure of random effects. If the model did not converge, we first removed correlations between random effects; in the next step, we removed the random effects with the smallest unique variance, following the recommendation by Bates and colleagues (2015). Only models for the SSRT needed trimming in the structure of the random effects. Absolute *t* values greater than the conventional level of two were considered significant. The data and the R scripts are available at <https://osf.io/xy4qg>.

Of primary interest were two types of effects. The interaction of Trial type and Prior language use in Models 1 and 2 served to assess differences in the behavioural and neural efficiency of inhibition due to prior language-use manipulation. The main effect of Trial type after each game separately served to assess sensitivity to a task manipulation after a language-use manipulation.

### 3. Results

#### 3.1. Missing data and data exclusion

We excluded data from participants with < 90% accuracy. For the stop-signal task, we removed two participants' data and one other participant's first testing session data. For the Stroop task, we removed two participants' data and one participant's third testing

session data. In the LME analyses, participants with missing data from only one session were included. In total, 30 participants were included in the analyses for the stop-signal and Stroop tasks.

#### 3.2. Behavioural data

The L2 game and the DL game were assessed as similarly difficult in terms of speaking effort ( $M = 7.93$ ,  $SD = 2.34$  and  $M = 7.68$ ,  $SD = 2.76$ , respectively;  $p > .05$ ) and both were assessed as more difficult than the L1 game ( $M = 5.45$ ,  $SD = 2.56$ ;  $t(29) = 5.46$ ,  $p < .001$  and  $t(29) = 5.24$ ,  $p < .001$ , respectively). Table 2 presents behavioural data from the inhibition tasks across the three language-game conditions; Table 3 presents the results of the LME analyses for behavioural outcomes.

#### Stop-signal task (SSRT)

The analysis of the SSRT did not show any effects of Prior language use. The effects of the L2-L1 comparison, the DL-L1 comparison in Model 1, and the DL-L2 comparison in Model 2 were non-significant.

#### Stroop task (RT)

Both Model 1 and Model 2 revealed a main effect of Trial type. Consistent with this, the direct tests showed a Trial type effect in each language-game condition separately ( $ts \geq 13.74$ ). RTs were slower for incongruent trials compared to congruent ones. None of the other effects were significant in Models 1 and 2.

#### 3.3. ERP data

Figure 3 presents grand averages of stimulus-locked ERPs in the stop-signal task; Figure 4 presents grand averages of stimulus-locked ERPs in the Stroop task. The targeted ERP components, i.e., N2, P3 and N450, were identified and all demonstrated their classic spatiotemporal characteristics. Consistent with previous research that used auditory stop-signal stimuli (Dimoska et al., 2006; Ramautar, Kok & Ridderinkhof, 2006; Skippen, Fulham, Michie, Matzke, Heathcote & Karayanidis, 2019), the N2 component did not show a clearly distinguished peak as it partially overlapped in time with a broadly distributed positivity, which smeared the N2 peak out.

#### Stop-signal task: N2 (220–270 ms)

Both models revealed a main effect of Trial type. The N2 was more negative for unsuccessful than for successful stop trials. Consistent with this, the direct tests showed a Trial type effect in each language-game condition ( $ts \geq 4.10$ ). Neither Model 1 nor Model 2 revealed any effects of Prior language use (see Table 4).

#### Stop-signal task: P3 (270–400 ms)

Table 5 presents the estimates of the LME models. Model 1 showed a significant main effect of Trial type. P3 was more positive for successful than for unsuccessful stop trials. The Trial type  $\times$  DL-L1 game interaction effect revealed a trend toward significance, which suggested that the P3 successful > unsuccessful inhibition effect was reduced after the DL game compared to after the L1 game. Model 2 did not show any significant effects. The direct tests for each language-game condition separately showed that the main effect of Trial type was significant after the L1 game ( $t = -2.98$ ) but was non-significant after the DL and L2 games ( $t = -1.03$  and  $t = -1.34$ , respectively). The analyses

**Table 2.** Behavioural data of the Stroop task (left) and the stop-signal task (right).

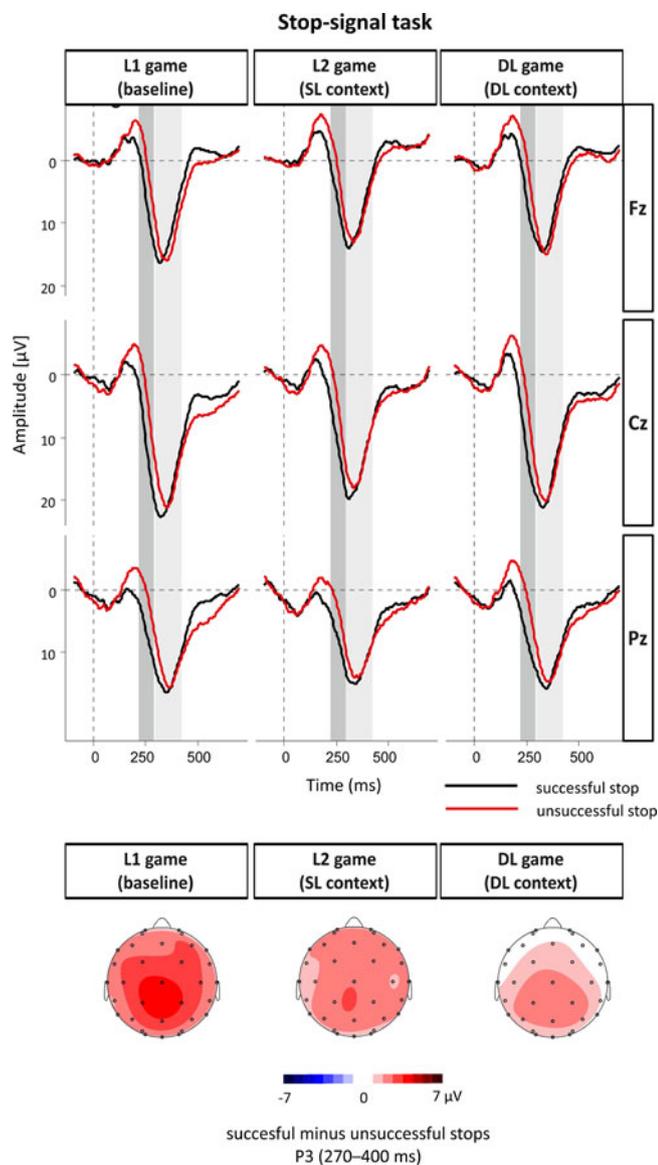
Prior language use	Trial type in the Stroop task	Stroop task				Stop-signal task	
		Accuracy (%)		RTs (ms)		SSRT (ms)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<b>L1 game (baseline)</b>	Congruent	100	0	582	87	211	36
	Incongruent	98	3	727	90		
<b>L2 game (SL context)</b>	Congruent	100	1	579	96	208	39
	Incongruent	98	3	713	117		
<b>DL game (DL Context)</b>	Congruent	100	1	574	102	214	43
	Incongruent	98	3	711	123		

Notes. L1, Polish; L2, English; DL, dual-language; *M*, mean; *SD*, standard deviation; SSRT, stop-signal reaction times; RTs, response latencies.

**Table 3.** Estimates of the LME models for behavioural measures.

Outcome variable	Predictors	Model 1			Model 2			
		Estimates	CI	<i>t</i> value	Estimates	CI	<i>t</i> value	
<b>SSRT</b>	(Intercept)	<b>211.1</b>	<b>196.64–225.55</b>	<b>28.62</b>	<b>208.36</b>	<b>193.45–223.28</b>	<b>27.38</b>	
	L2-L1 game	-2.73	-15.76–10.30	-0.41				
	DL-L1 game	2.69	-10.34–15.72	0.41				
	DL-L2 game				5.43	-8.15–19.00	0.78	
	<b>Random effect (by subject)</b>	<b>SD</b>			<b>SD</b>			
	(Intercept)	30.36			31.37			
	Residual	25.01			26.38			
	<b>Predictors</b>	<b>Estimates</b>	<b>CI</b>	<b><i>t</i> value</b>	<b>Estimates</b>	<b>CI</b>	<b><i>t</i> value</b>	
	<b>Stroop RT</b>	(Intercept)	<b>6.46</b>	<b>6.41–6.50</b>	<b>282.57</b>	<b>6.44</b>	<b>6.38–6.50</b>	<b>220.17</b>
		Trial type [1]	<b>0.11</b>	<b>0.10–0.13</b>	<b>14.02</b>	<b>0.10</b>	<b>0.09–0.11</b>	<b>17.10</b>
	L2-L1 game [2]	-0.01	-0.04–0.02	-0.97				
	DL-L1 game [3]	-0.02	-0.05–0.01	-1.23				
	[1] : [2]	-0.01	-0.02–0.00	-1.72				
	[1] : [3]	-0.01	-0.02–0.01	-0.83				
	DL-L2 game [4]				-0.01	-0.04–0.02	-0.41	
	[1] : [4]				0.00	-0.01–0.04	1.07	
	<b>Random effect (by subject)</b>	<b>SD</b>			<b>SD</b>			
	(Intercept)	0.12			0.16			
	Trial type [1]	0.04			0.03			
	L2-L1 game [2]	0.08						
	DL-L1 game [3]	0.09						
	[1] : [2]	0.03						
	[1] : [3]	0.03						
	DL-L2 game [4]				0.08			
	[1] : [4]				0.01			
	Residual	0.15			0.15			

Notes. CI, Confidence Intervals; SD, standard deviation; SSRT, stop-signal reaction time; RT, reaction time; significant effects bolded.

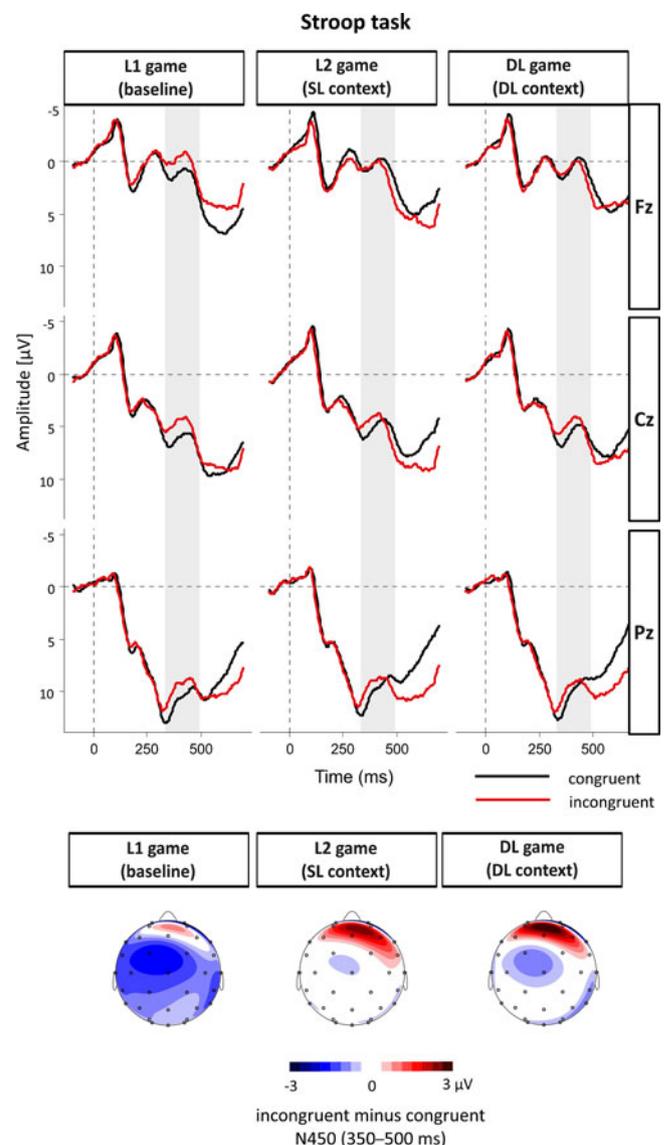


**Fig. 3.** ERP waveforms from the stop-signal task. Stop-signal-locked grand-averaged waveforms at midline electrode sites (top) and scalp potential difference maps for the P3 component (bottom) broken down by prior language-use manipulation. The component-specific windows examined in this study are highlighted.

indicated that successfully inhibited stop signals evoked a more pronounced P3 than unsuccessfully inhibited ones after the L1 game but not after the DL and L2 games.

#### Stroop task: N450 (350–500 ms)

Table 6 presents the estimates of the LME models. Model 1 revealed the main effect of Trial type: N450 was more negative for incongruent trials than for congruent trials. Moreover, Trial type interacted with the L2-L1 game and the DL-L1 game comparisons in Model 1: the N450 incongruent > congruent trial effect was reduced after the DL and L2 games compared to after the L1 game. Model 2 did not show any effects. The direct tests showed that the main effect of Trial type was significant after the L1 game ( $t = 2.52$ ) but was non-significant after the DL and L2 games ( $t = 0.07$  and  $t = 0.27$ , respectively). The analyses indicated that the N450 amplitudes were sensitive to the



**Fig. 4.** ERP waveforms from the Stroop task. Stimulus-locked, grand-averaged waveforms at midline electrode sites (top) and scalp potential difference maps for the N450 component (bottom) that were broken down by prior language-use manipulation. The N450 time-window examined in this study is highlighted.

congruency manipulation after the L1 game but not after the DL or L2 games.

## 4. Discussion

### 4.1. Results summary

This study investigated how natural patterns of language use shape inhibition efficiency in L1-dominant bilinguals living in an L1 environment. Rather than identifying the patterns of language use via the lifelong language experiences of bilinguals (Hartanto & Yang, 2020; Kałamała et al., 2020b) or imitating them via artificial experimental tasks (e.g., Prior & Gollan, 2011; Timmer, Calabria et al., 2019; Yang et al., 2018; Wu & Thierry, 2013), we induced these patterns in a series of language games involving natural language use. Moreover, in order to provide a fine-grained measurement of inhibition, we used two well-

**Table 4.** Estimates of the LME models for the N2 component.

Predictors	Model 1			Model 2		
	Estimates	CI	t value	Estimates	CI	t value
(Intercept)	4.39	-0.65-9.43	1.71	<b>4.68</b>	<b>0.42-8.94</b>	<b>2.15</b>
Trial type [1]	<b>-3.41</b>	<b>-4.58 - -2.23</b>	<b>-5.67</b>	<b>-2.69</b>	<b>-3.98 - -1.41</b>	<b>-4.10</b>
L2-L1 game [2]	0.28	-5.42-5.98	0.10			
DL-L1 game [3]	1.29	-3.58-6.16	0.52			
[1] : [2]	0.74	-0.91-2.39	0.88			
[1] : [3]	-0.38	-1.76-0.99	-0.55			
DL-L2 game [4]				1.15	-3.03-5.32	0.54
[1] : [4]				-1.18	-2.85-0.50	-1.37
<b>Random effect (by subject)</b>	<b>SD</b>			<b>SD</b>		
(Intercept)	13.48			11.52		
Trial type [1]	2.67			3.04		
L2-L1 game [2]	15.12					
DL-L1 game [3]	12.66					
[1] : [2]	3.74					
[1] : [3]	2.77					
DL-L2 game [4]				10.82		
[1] : [4]				3.84		
Residual	13.67			13.87		

Notes. CI, Confidence Intervals; significant effects bolded.

established inhibition tasks, the stop-signal task and the Stroop task (Diamond, 2013), and supplemented the behavioural measurement by recording ERPs (Cespón & Carreiras, 2020; Pliatsikas et al., 2020).

Drawing on the ACH (Green & Abutalebi, 2013) and previous research showing cognitive improvements after the short picture-naming sessions (Liu et al., 2019; Prior & Gollan, 2011; Timmer, Calabria et al., 2019; Timmer, Christoffels et al., 2019; Yang et al., 2018; Zhang et al., 2015), we formulated two predictions regarding the relationship between patterns of language use and inhibition. More efficient inhibition after the L2 and DL games compared to after the L1 game (baseline) would indicate beneficial roles of both the DL and SL contexts (compared to the use of L1 in the L1 environment). More efficient inhibition after the DL game compared to after the L2 game would indicate that the DL context benefits inhibition to a greater extent than the SL context.

Overall, we replicated classic behavioural and ERP effects in the inhibition tasks. In the stop-signal task, the SSRT fell within the standard range (from 150 to 300 ms in young, healthy participants, Wessel & Aron, 2015), whereas in the Stroop task faster responses were observed for congruent than for incongruent trials (Stroop, 1935). The targeted ERPs, i.e., N2, P3, and N450, demonstrated spatiotemporal characteristics consistent with expectations based on previous ERP reports (for evidence on N2 and P3, see Berkman et al., 2014; Dimoska et al., 2006; Manuel et al., 2013; Nieuwenhuis et al., 2003; Senderecka, 2018; for evidence on N450, see Hsieh et al., 2018; Larson et al., 2009; Liotti et al., 2000). Importantly, faster RT was related to smaller amplitude

differences in P3 and N450 (i.e., a smaller P3 successful > unsuccessful inhibition effect and a smaller N450 incongruent > congruent trial effect; for details, see Appendix S2). This indicates that smaller ERP effects were associated with more efficient cognitive processing, which is in line with previous research (Chang et al., 2017; Hsieh et al., 2018; Schroder et al., 2020). At the same time, RT was unrelated to the N2 unsuccessful > successful inhibition effect. However, we found this measure unreliable and therefore excluded it from the interpretation in the study (for details, see Appendix S3).

With regards to the prior language-use manipulation, the participants perceived the games as being different in terms of speaking effort. While the L1 game was assessed as very easy, both the L2 game and the DL game were assessed as difficult. The results suggest that the games involving the use of L2 indeed imposed demands on the participants' language use. Notably, however, the DL game, which on the basis of the ACH was assumed to induce the highest demands on language use, was judged to be as difficult as the L2 game (imitating the SL context).

With regards to the behavioural data, we did not observe any effects of the prior language-use manipulation. The latency of the response inhibition mechanism (indexed by SSRT) and the efficiency of interference resolution (indexed by the Stroop effect in RT) were similar regardless of how the participants used their languages in the preceding language game. In fact, the absence of behavioural effects in this study corroborates the findings from our recent latent-variable study (Kałamała et al., 2020b), in which the behavioural measures of inhibition were unrelated to the self-assessed patterns of language use in a large

**Table 5.** Estimates of the LME models for the P3 component.

Predictors	Model 1			Model 2		
	Estimates	CI	t value	Estimates	CI	t value
(Intercept)	<b>14.94</b>	<b>12.54–17.35</b>	<b>9.74</b>	<b>12.61</b>	<b>10.78–15.55</b>	<b>10.37</b>
Trial type [1]	<b>-1.46</b>	<b>-1.76 - -0.67</b>	<b>-3.05</b>	-0.78	-1.83-0.28	-1.44
L2-L1 game [2]	<b>-2.37</b>	<b>-2.58 - -1.02</b>	<b>-2.05</b>			
DL-L1 game [3]	-0.39	-0.70-0.86	-0.44			
[1] : [2]	0.71	-0.29-0.24	1.31			
[1] : [3]	1.01	-0.12-1.41	1.90			
DL-L2 game [4]				1.90	-0.10-4.21	1.68
[1] : [4]				0.31	-0.99-1.20	0.59
<b>Random effect (by subject)</b>	<b>SD</b>			<b>SD</b>		
(Intercept)	8.17			6.38		
Trial type [1]	2.12			2.50		
L2-L1 game [2]	5.71					
DL-L1 game [3]	4.05					
[1] : [2]	2.04					
[1] : [3]	1.97					
DL-L2 game [4]				5.34		
[1] : [4]				2.13		
Residual	11.43			11.55		

Notes. CI, Confidence Intervals; significant effects bolded.

group of bilinguals derived from the same population as in this study. However, in contrast to previous studies (e.g., Kałamała et al., 2020b; Pot et al., 2018; Liu et al., 2019), this study employed not only behavioural but also ERP measurements. The P3 component in the stop-signal task and the N450 component in the Stroop task showed the effects related to prior language use. As predicted, differences in the N450 amplitudes between incongruent and congruent trials were reduced after the DL and L2 games compared to after the L1 game. Similarly, differences in the P3 amplitudes between successfully and unsuccessfully inhibited trials were reduced after the DL game compared to after the L1 game (but this effect was marginal). In contrast to our predictions, however, we did not find differences in P3 and the N450 amplitudes after the DL and the L2 games. Crucially, the direct tests showed that both P3 and N450 were sensitive to the inhibition demands imposed by the tasks (i.e., stop signals and interference, respectively) after the L1 game but were insensitive to these after the L2 and the DL games. The modulation of ERPs in the inhibition tasks consistently suggests that it is less effortful to implement inhibition when prior language use involved language-switching or the exclusive use of a non-dominant language.

#### 4.2. Evidence for the Adaptive Control Hypothesis

Evidence on how patterns of language use shape inhibition comes from studies that assessed everyday habits of language use (e.g., Beatty-Martínez et al., 2019; Kałamała et al., 2020b) and studies that experimentally manipulated language experience (e.g., Liu et al., 2019; Wu & Thierry, 2013; Zhang et al., 2015). However,

the lifelong experience of bilingualism can be challenging to measure, and the low ecological validity of language-production tasks complicates inferences about how patterns of language use shape cognitive control. In contrast to previous research, this study induced natural language use in an experimental setting in order to investigate the direct effects of natural language use on inhibition.

Although the behavioural measures of inhibition were not modulated by the prior language-use manipulation, the ERP results provide the first evidence for a direct relationship between natural patterns of language use and inhibition. The reduction of the P3 and N450 effects after the DL and L2 games suggests that it is less effortful to implement inhibition when prior language use involved language-switching or the exclusive use of a non-dominant language. These findings are in line with the ACH and suggest that the DL and SL contexts train inhibitory mechanisms, which translates into less effortful implementation of inhibition in a subsequent task. At the same time, the fact that the magnitudes of P3 and N450 did not differ from each other after the DL and L2 games contradicts the ACH's prediction. The absence of differences for the two conditions suggests that, regardless of whether bilinguals switch languages in a context (i.e., operates in a DL context) or are restricted to use only one language (i.e., operates in an SL context), their neural mechanisms of inhibition are trained in a similar way (but for an alternative interpretation, see section 4.3). Interestingly, the effects observed for the DL and SL contexts correspond to the participants' perception of how difficult the games were: the L2 game and the DL game were assessed as being similarly difficult in terms of speaking effort, but both were assessed as more difficult than

**Table 6.** Estimates of the LME models for the N450 component.

Predictors	Model 1			Model 2		
	Estimates	CI	t value	Estimates	CI	t value
(Intercept)	<b>3.49</b>	<b>1.75–5.23</b>	<b>3.93</b>	<b>2.90</b>	<b>0.45–5.34</b>	<b>2.33</b>
Trial type [1]	<b>–0.67</b>	<b>–1.16 – –0.18</b>	<b>–2.69</b>	0.10	–0.64–0.84	0.26
L2-L1 game [2]	–0.60	–2.69–1.50	–0.56			
DL-L1 game [3]	–0.27	–1.70–1.17	–0.37			
[1] : [2]	<b>0.77</b>	<b>0.07–1.46</b>	<b>2.17</b>			
[1] : [3]	<b>0.65</b>	<b>0.05–1.25</b>	<b>2.11</b>			
DL-L2 game				0.33	–1.16–1.81	0.43
[1] : [4]				–0.12	–0.62–0.39	–0.46
<b>Random effect (by subject)</b>	<b>SD</b>			<b>SD</b>		
(Intercept)	4.72			6.78		
Trial type [1]	1.09			1.91		
L2-L1 game [2]	5.68					
DL-L1 game [3]	3.76					
[1] : [2]	1.59					
[1] : [3]	1.26					
DL-L2 game				4.00		
[1] : [4]				0.88		
Residual	13.06			13.36		

Note CI, Confidence Intervals; significant effects bolded.

speaking during the L1 game. In summary, the current pattern of results indicates that the cognitive training that bilinguals receive during their everyday language use of a non-dominant language in an L1 environment affects the neural implementation of inhibitory control.

### 4.3. Limitations and future directions

The study provides direct evidence for cognitive effects related to the language-use patterns. However, some findings are limited by the nature of the data and therefore require further investigation. The lack of evidence in the ERP data for better inhibition after the DL game compared to the L2 game suggests that the language experiences of the DL and SL contexts impact inhibition to the same extent. However, this finding may not generalize to the entire bilingual population as it may be a consequence of the specific language-dominance profile of the tested population. Since we tested a group of L1-dominant bilinguals living in an L1 environment, we speculate that the effect of inhibiting L1 in the L2 game (mimicking the SL context) was disproportionately large (compared to what was originally proposed in the ACH), which might have translated into the absence of differences between the SL and DL contexts in this study. This in turn suggests that L1-dominant bilinguals living in an L1 environment recruit inhibition processes in a unique way (for a similar argument, see Hofweber et al., 2020; see also Goral, Campanelli & Spiro, 2015). Future research should thoroughly examine the interactions between language dominance and language-use patterns. An alternative explanation for the absence of differences in the

ERP data between the DL and L2 games is that we encountered a floor-effect in measuring ERPs. Since the P3 and N450 amplitude differences were not sensitive to inhibition demands after the DL and L2 games (as reflected by the absence of task-manipulation effects after these games), no further cognitive improvement related to the DL game could have been captured by our ERP measurement (Bialystok, Poarch, Luo & Craik, 2014).

In terms of the study design, it is important to note that the use of EEG required a short break between the language and the EEG sessions (dedicated to EEG capping). During the break, conversation was kept to a minimum, but if something required an explanation, the experimenter always used English (L2 for participants). Therefore, one may argue that the additional use of L2 could have interfered with the language-use patterns induced during the games. It is worth noting, however, that the aim of this study was to test how prior cognitive training in the form of a language game affects inhibition. Therefore, while the additional use of L2 may have led to slight deviations in the induced language-use pattern, it should not have removed the cognitive effects of the two-hour training sessions. Nevertheless, in order to provide a more methodologically rigorous design, future research should limit language use between subsequent sessions. Relatedly, since we used a vocal Stroop task, it could be argued that naming in L1 in the Stroop task interfered with the preceding language-use pattern and thereby contaminated the measurement of pattern-specific effects in this task. We consider this scenario unlikely because the Stroop task data was consistent with the data from the stop-signal task, which did not include any linguistic material. However, in order to obtain quantitative

evidence, we performed an additional analysis. Assuming that the preceding language-use manipulation indeed interfered with performance on the Stroop task, we should observe some changes during the ongoing performance of this task after games involving the use of L2. The analysis clearly showed that performance on the Stroop task did not differ across trials after the DL and L2 games, which suggests that the use of L1 in the Stroop task did not contaminate the measurement of pattern-specific effects.<sup>3</sup> Importantly, the evidence for the stability of the pattern-specific effect in the task requiring overt language production suggests an important property of the language-induced effects on cognitive control: although induced by speaking, they did not dissipate despite subsequent language production. Future research should address this issue more thoroughly.

Finally, the pattern of the neural activation did not translate into the behaviourally observed outcomes. We see three possible explanations for this discrepancy. The first explanation might be that the inhibition tasks were performed at the upper limit, which made it impossible to observe the inhibition benefits of the SL and DL contexts at the behavioural level (a so-called ceiling effect). This is a likely explanation as accuracy in the Stroop task was close to 100% and the SSRT was shorter than in several previous studies using the same or similar versions of the stop-signal task (e.g., Greenhouse & Wessel, 2013; Senderecka, 2018; Wagner, Wessel, Ghahremani & Aron, 2017). The specificity of the participants' sample additionally supports this explanation as we tested young bilingual adults who are often argued to be at the peak of cognitive efficiency, therefore they are susceptible to ceiling effects (Bialystok, 2017; Bialystok, Martin & Viswanathan, 2005; but see Samuel, Roehr-Brackin, Pak & Kim, 2018). The second explanation is related to the specificity of the experimental manipulation. Since we incorporated a relatively short-term language-use manipulation, it can be speculated that behavioural effects would be observed with regards to the longer-lasting language use. Since the ACH does not define a time frame for the cognitive effects of language use, this issue requires further research. The third explanation is that the behavioural and ERP measurements are to some extent dissociable and therefore the differences in the neural activation patterns do not always translate into behaviourally observed effects (Gratton, Sun & Petersen, 2018; van den Noort et al., 2019). While behavioural measures reflect not only a specific cognitive process targeted in the study but also different peripheral processes involved in performance, ERPs reflecting temporal changes in the activity of specific brain processors are more precise manifestations of specific cognitive processes. In the ERP literature, the P3 and the N450 are well-established markers of inhibition (Larson et al., 2014; Pires, Leitão, Guerrini & Simões, 2014). Building on this logic, P3 and the N450 likely reflected the engagement of inhibition in this study, but the behavioural measures did not.

## 5. Conclusions

This study shows how inhibition efficiency can be modulated by bilinguals' language-use experience. By adopting a within-subject design and a multiple-measure approach, the study is the first to

<sup>3</sup>To test whether the performance of the Stroop task had changed over time, we used LME. RTs after the L2 game and RTs after the DL game were regressed on the three participant-related fixed effects: Trial number (scaled), Trial type, and their interaction. The fitting procedure is described in section 2.4.3. Neither the Trial number nor the interaction affected RTs in the Stroop task after the DL game ( $t = 1.50$  and  $t = 1.23$ , respectively) and the L2 game ( $t = 1.65$  and  $t = 1.02$ , respectively).

test the direct effects of a relatively natural and ecologically valid language-use manipulation (i.e., conversation) on inhibition on both behavioural and electrophysiological levels. The study increases our knowledge about the specific conditions in which language use can benefit inhibition. Specifically, we observed a more neurally efficient implementation of inhibition after prior use of L2. Crucially, the study suggests that the exclusive use of L2 and the alternate use of L1 and L2 might be comparable in enhancing inhibitory control when bilinguals reside in an L1 environment.

The study is also timely with respect to the ongoing discussion regarding the ecological validity of manipulating language use in an experimental setting (Blanco-Elorrieta & Pyllkänen, 2018; van den Noort et al., 2019). Evidence for the cognitive effects of language training mostly comes from studies utilizing artificial cued language-production paradigms. Our findings indicate that natural patterns of language use can be successfully induced in well-controlled experimental settings and may affect the workings of the cognitive control system. The study should thereby inspire future research to use more ecologically valid manipulations.

**Acknowledgments.** The research was funded by a National Science Centre grant to Z.W. [2015/18/E/HS6/00428] and by a National Science Foundation grant [OISE-1545900]. During work on the paper, M.S. was supported by a National Science Centre grant [2015/19/B/HS6/00341] and P.K. was supported by National Science Centre grants [2017/27/N/HS6/01029; 2020/36/T/HS6/00363] and Foundation for Polish Science (START). The authors gratefully acknowledge the help of Maria Badanova with experiment preparation and data collection; Jakub Szewczyk with preparing experimental tasks and analysis; and research assistants who contributed to data collection as confederates in language games.

## References

- Abutalebi J and Green DW (2016) Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition* 19(04), 689–698. <https://doi.org/10.1017/S1366728916000225>
- Adler RM, Valdés Kroff JR and Novick JM (2020) Does integrating a code-switch during comprehension engage cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46(4), 741–759. <https://doi.org/10.1037/xlm0000755>
- Bak TH (2016) Cooking pasta in La Paz: Bilingualism, bias and the replication crisis. *Linguistic Approaches to Bilingualism*. <https://doi.org/10.1075/lab.16002.bak>
- Bates D, Mächler M, Bolker B and Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beatty-Martínez AL, Navarro-Torres CA, Dussias PE, Bajo MT, Guzzardo Tamargo RE and Kroll JF (2019) Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000770>
- Berkman ET, Kahn LE and Merchant JS (2014) Training-Induced Changes in Inhibitory Control Network Activity. *Journal of Neuroscience* 34(1), 149–157. <https://doi.org/10.1523/JNEUROSCI.3564-13.2014>
- Bialystok E (2017) The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin* 143(3), 233–262. <https://doi.org/10.1037/bul0000099>
- Bialystok E, Martin MM and Viswanathan M (2005) Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism* 9(1), 103–119. <https://doi.org/10.1177/13670069050090010701>
- Bialystok E, Poarch G, Luo L and Craik FIM (2014) Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging* 29(3), 696–705. <https://doi.org/10.1037/a0037254>
- Bice K, Yamasaki BL and Prat CS (2020) Bilingual Language Experience Shapes Resting-State Brain Rhythms. *Neurobiology of Language* 1(3), 288–318. [https://doi.org/10.1162/nol\\_a\\_00014](https://doi.org/10.1162/nol_a_00014)

- Blanco-Elorrieta E and Pykkänen L (2018) Ecological Validity in Bilingualism Research and the Bilingual Advantage. *Trends in Cognitive Sciences* 22(12), 1117–1126. <https://doi.org/10.1016/j.tics.2018.10.001>
- Brown G, Anderson A, Shillcock R and Yule G (1985) *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.
- Cambridge Assessment English (2018, April 10) <https://www.cambridgeenglish.org/test-your-english/general-english/?fbclid=iwar06awplojesuczrre1jvvl4apu8t77xnjgxbvssqgxcrc8fqm0ihcr3ig4>
- Cespón J and Carreiras M (2020) Is there electrophysiological evidence for a bilingual advantage in neural processes related to executive functions? *Neuroscience & Biobehavioral Reviews*, S0149763420305005. <https://doi.org/10.1016/j.neubiorev.2020.07.030>
- Chang Y-K, Alderman BL, Chu C-H, Wang C-C, Song T-F and Chen F-T (2017) Acute exercise has a general facilitative effect on cognitive function: A combined ERP temporal dynamics and BDNF study: Acute exercise, BDNF, ERPs, and cognition. *Psychophysiology* 54(2), 289–300. <https://doi.org/10.1111/psyp.12784>
- de Bruin A (2019) Not All Bilinguals Are the Same: A Call for More Detailed Assessments and Descriptions of Bilingual Experiences. *Behavioral Sciences* 9(3), 33. <https://doi.org/10.3390/bs9030033>
- DeLuca V (2019) Future Directions in Examining Neurological Adaptation to Bilingual Experiences. *Journal of Experimental Neuroscience* 13, 117906951987659. <https://doi.org/10.1177/1179069519876597>
- DeLuca V, Rothman J, Bialystok E and 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>
- Diamond A (2013) Executive Functions. *Annual Review of Psychology* 64(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dimoska A, Johnstone SJ and Barry RJ (2006) The auditory-evoked N2 and P3 components in the stop-signal task: Indices of inhibition, response-conflict or error-detection? *Brain and Cognition* 62(2), 98–112. <https://doi.org/10.1016/j.bandc.2006.03.011>
- Forster KI and Forster JC (2003) DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers* 35(1), 116–124. <https://doi.org/10.3758/BF03195503>
- Goral M, Campanelli L and Spiro A (2015) Language dominance and inhibition abilities in bilingual older adults. *Bilingualism: Language and Cognition* 18(1), 79–89. <https://doi.org/10.1017/S1366728913000126>
- Grainger J and Dijkstra T (1992) On the Representation and Use of Language Information in Bilinguals. In RJ Harris (ed.), *Advances in Psychology*. North-Holland, Vol. 83, pp. 207–220. [https://doi.org/10.1016/S0166-4115\(08\)61496-X](https://doi.org/10.1016/S0166-4115(08)61496-X)
- Grainger J, Midgley K and Holcomb PJ (2010) Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). *Language Acquisition across Linguistic and Cognitive Systems* 52, 267–283.
- Gratton C, Sun H and Petersen SE (2018) Control networks and hubs. *Psychophysiology* 55(3), e13032. <https://doi.org/10.1111/psyp.13032>
- Gratton G, Coles MG and Donchin E (1983) A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology* 55(4), 468–484.
- Gratton G, Cooper P, Fabiani M, Carter CS and Karayanidis F (2017) Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology*. <https://doi.org/10.1111/psyp.13016>
- Green DW (1998) Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1(2), 67–81. <https://doi.org/10.1017/S1366728998000133>
- Green DW and 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>
- Greenhouse I and Wessel JR (2013) EEG signatures associated with stopping are sensitive to preparation. *Psychophysiology* 50(9), 900–908. <https://doi.org/10.1111/psyp.12070>
- Gullifer JW, Chai XJ, Whitford V, Pivneva I, Baum S, Klein D and Titone D (2018) Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia* 117, 123–134. <https://doi.org/10.1016/j.neuropsychologia.2018.04.037>
- Hartanto A and Yang H (2016) Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition* 150, 10–19. <https://doi.org/10.1016/j.cognition.2016.01.016>
- Hartanto A and Yang H (2020) The role of bilingual interactional contexts in predicting interindividual variability in executive functions: A latent variable analysis. *Journal of Experimental Psychology: General* 149(4), 609–633. <https://doi.org/10.1037/xge0000672>
- Henrard S and Van Daele A (2017) Different Bilingual Experiences Might Modulate Executive Tasks Advantages: Comparative Analysis between Monolinguals, Translators, and Interpreters. *Frontiers in Psychology* 8, 1870. <https://doi.org/10.3389/fpsyg.2017.01870>
- Hofweber J, Marinis T and Treffers-Daller J (2020) How different code-switching types modulate bilinguals' executive functions: A dual control mode perspective. *Bilingualism: Language and Cognition* 23(4), 909–925. <https://doi.org/10.1017/S1366728919000804>
- Hsieh S-S, Huang C-J, Wu C-T, Chang Y-K and Hung T-M (2018) Acute Exercise Facilitates the N450 Inhibition Marker and P3 Attention Marker during Stroop Test in Young and Older Adults. *Journal of Clinical Medicine* 7(11), 391. <https://doi.org/10.3390/jcm7110391>
- Jiao L, Grundy JG, Liu C and Chen B (2020) Language context modulates executive control in bilinguals: Evidence from language production. *Neuropsychologia* 142, 107441. <https://doi.org/10.1016/j.neuropsychologia.2020.107441>
- Jiao L, Liu C, Liang L, Plummer P, Perfetti CA and Chen B (2019) The contributions of language control to executive functions: From the perspective of bilingual comprehension. *Quarterly Journal of Experimental Psychology* 72(8), 1984–1997. <https://doi.org/10.1177/1747021818821601>
- Kałamała P, Ociepka M and Chuderski A (2020a) ERP evidence for rapid within-trial adaptation of cognitive control during conflict resolution. *Cortex* 131, 151–163. <https://doi.org/10.1016/j.cortex.2020.07.012>
- Kałamała P, Szewczyk J, Chuderski A, Senderecka M and Wodniecka Z (2020b) Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis. *Cognition* 204, 104373. <https://doi.org/10.1016/j.cognition.2020.104373>
- Kroll JF, Dussias PE, Bogulski CA and Kroff JRV (2012) Juggling Two Languages in One Mind. In *Psychology of Learning and Motivation* Vol. 56. Elsevier, pp. 229–262. <http://linkinghub.elsevier.com/retrieve/pii/B9780123943934000078>
- Larson MJ, Clayson PE and Clawson A (2014) Making sense of all the conflict: A theoretical review and critique of conflict-related ERPs. *International Journal of Psychophysiology* 93(3), 283–297. <https://doi.org/10.1016/j.ijpsycho.2014.06.007>
- Larson MJ, Kaufman DAS and Perlstein WM (2009) Neural time course of conflict adaptation effects on the Stroop task. *Neuropsychologia* 47(3), 663–670. <https://doi.org/10.1016/j.neuropsychologia.2008.11.013>
- Leivada E, Westergaard M, Duñabeitia JA and Rothman J (2020) On the phantom-like appearance of bilingualism effects on neurocognition: (How) should we proceed? *Bilingualism: Language and Cognition*, 1–14. <https://doi.org/10.1017/S1366728920000358>
- Lemhöfer K and Broersma M (2012) Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Li P, Zhang F, Tsai E and Puls B (2014) Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition* 17(03), 673–680. <https://doi.org/10.1017/S1366728913000606>
- Liotti M, Woldorff MG, Perez R and Mayberg HS (2000) An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia* 38(5), 701–711. [https://doi.org/10.1016/S0028-3932\(99\)00106-2](https://doi.org/10.1016/S0028-3932(99)00106-2)
- Liu C, Yang C-L, Jiao L, Schwieter JW, Sun X and Wang R (2019) Training in Language Switching Facilitates Bilinguals' Monitoring and Inhibitory Control. *Frontiers in Psychology* 10, 1839. <https://doi.org/10.3389/fpsyg.2019.01839>
- Logan GD (1994) On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). Academic Press.
- Logan GD and Cowan WB (1984) On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review* 91(3), 295.

- Luck SJ and Gaspelin N** (2017) How to get statistically significant effects in any ERP experiment (and why you shouldn't): How to get significant effects. *Psychophysiology* 54(1), 146–157. <https://doi.org/10.1111/psyp.12639>
- Manuel AL, Bernasconi F and Spierer L** (2013) Plastic modifications within inhibitory control networks induced by practicing a stop-signal task: An electrical neuroimaging study. *Cortex* 49(4), 1141–1147. <https://doi.org/10.1016/j.cortex.2012.12.009>
- Marian V, Blumenfeld HK and Kaushanskaya M** (2007) The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research* 50(4), 940–967.
- Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A and Wager TD** (2000) The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology* 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Nielsen K and Gonzalez R** (2020) Comparison of Common Amplitude Metrics in Event-Related Potential Analysis. *Multivariate Behavioral Research* 55(3), 478–493. <https://doi.org/10.1080/00273171.2019.1654358>
- Nieuwenhuis S, Yeung N, van den Wildenberg W and Ridderinkhof KR** (2003) Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective & Behavioral Neuroscience* 3(1), 17–26. <https://doi.org/10.3758/CABN.3.1.17>
- Ooi SH, Goh WD, Sorace A and Bak TH** (2018) From Bilingualism to Bilingualisms: Bilingual experience in Edinburgh and Singapore affects attentional control differently. *Bilingualism: Language and Cognition* 21(4), 867–879. <https://doi.org/10.1017/S1366728918000020>
- Pires L, Leitão J, Guerrini C and Simões MR** (2014) Event-Related Brain Potentials in the Study of Inhibition: Cognitive Control, Source Localization and Age-Related Modulations. *Neuropsychology Review* 24(4), 461–490. <https://doi.org/10.1007/s11065-014-9275-4>
- Pliatsikas C, DeLuca V and Voits T** (2020) The Many Shades of Bilingualism: Language Experiences Modulate Adaptations in Brain Structure. *Language Learning* 70(S2), 133–149. <https://doi.org/10.1111/lang.12386>
- Pot A, Keijzer M and de Bot K** (2018) Intensity of Multilingual Language Use Predicts Cognitive Performance in Some Multilingual Older Adults. *Brain Sciences* 8(5), 92. <https://doi.org/10.3390/brainsci8050092>
- Prior A and Gollan TH** (2011) Good Language-Switchers are Good Task-Switchers: Evidence from Spanish–English and Mandarin–English Bilinguals. *Journal of the International Neuropsychological Society* 17(04), 682–691. <https://doi.org/10.1017/S155617711000580>
- R Core Team** (2019) European Environment Agency. Retrieved April 22, 2020, from <https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006>
- Ramautar JR, Kok A and Ridderinkhof KR** (2006) Effects of stop-signal modality on the N2/P3 complex elicited in the stop-signal paradigm. *Biological Psychology* 72(1), 96–109. <https://doi.org/10.1016/j.biopsycho.2005.08.001>
- Rey-Mermet A, Gade M and Steinhauser M** (2019) Sequential conflict resolution under multiple concurrent conflicts: An ERP study. *NeuroImage* 188, 411–418. <https://doi.org/10.1016/j.neuroimage.2018.12.031>
- Rouder JN and Haaf JM** (2019) A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review* 26(2), 452–467. <https://doi.org/10.3758/s13423-018-1558-y>
- Samuel S, Roehr-Brackin K, Pak H and Kim H** (2018) Cultural Effects Rather Than a Bilingual Advantage in Cognition: A Review and an Empirical Study. *Cognitive Science* 42(7), 2313–2341. <https://doi.org/10.1111/cogs.12672>
- Schad DJ, Vasishth S, Hohenstein S and Kliegl R** (2020) How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language* 110, 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- Schroder E, Dubuson M, Dousset C, Mortier E, Kornreich C and Campanella S** (2020) Training Inhibitory Control Induced Robust Neural Changes When Behavior Is Affected: A Follow-up Study Using Cognitive Event-Related Potentials. *Clinical EEG and Neuroscience* 51(5), 303–316. <https://doi.org/10.1177/1550059419895146>
- Senderecka M** (2016) Threatening visual stimuli influence response inhibition and error monitoring: An event-related potential study. *Biological Psychology* 113, 24–36. <https://doi.org/10.1016/j.biopsycho.2015.11.003>
- Senderecka M** (2018) Emotional enhancement of error detection—The role of perceptual processing and inhibition monitoring in failed auditory stop trials. *Cognitive, Affective & Behavioral Neuroscience* 18(1), 1–20. <https://doi.org/10.3758/s13415-017-0546-4>
- Senderecka M, Szewczyk J, Wichary S and Kossowska M** (2018) Individual differences in decisiveness: ERP correlates of response inhibition and error monitoring. *Psychophysiology* 55(10), e13198. <https://doi.org/10.1111/psyp.13198>
- Skippen P, Fulham WR, Michie PT, Matzke D, Heathcote A and Karayanidis F** (2019) Reconsidering electrophysiological markers of response inhibition in light of trigger failures in the stop-signal task [Preprint]. *Neuroscience*. <https://doi.org/10.1101/658336>
- Spierer L, Chavan CF and Manuel AL** (2013) Training-induced behavioral and brain plasticity in inhibitory control. *Frontiers in Human Neuroscience* 7. <https://doi.org/10.3389/fnhum.2013.00427>
- Stroop JR** (1935) Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18(6), 643–662. <https://doi.org/10.1037/h0054651>
- Szűcs D and Soltész F** (2012) Functional definition of the N450 event-related brain potential marker of conflict processing: A numerical stroop study. *BMC Neuroscience* 13(1), 35. <https://doi.org/10.1186/1471-2202-13-35>
- Timmer K, Calabria M and Costa A** (2019) Non-linguistic effects of language switching training. *Cognition* 182, 14–24. <https://doi.org/10.1016/j.cognition.2018.09.001>
- Timmer K, Christoffels IK and Costa A** (2019) On the flexibility of bilingual language control: The effect of language context. *Bilingualism: Language and Cognition* 22(3), 555–568. <https://doi.org/10.1017/S1366728918000329>
- van den Noort M, Struys E, Bosch P, Jaswetz L, Perriard B, Yeo S, Barisch P, Vermeire K, Lee S-H and Lim S** (2019) Does the Bilingual Advantage in Cognitive Control Exist and If So, What Are Its Modulating Factors? A Systematic Review. *Behavioral Sciences* 9(3), 27. <https://doi.org/10.3390/bs9030027>
- van Heuven WJB and Dijkstra T** (2010) Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews* 64(1), 104–122. <https://doi.org/10.1016/j.brainres-rev.2010.03.002>
- Verbruggen F and Logan GD** (2008) Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences* 12(11), 418–424. <https://doi.org/10.1016/j.tics.2008.07.005>
- Wagner J, Wessel JR, Ghahremani A and Aron AR** (2017) Establishing a Right Frontal Beta Signature for Stopping Action in Scalp EEG: Implications for Testing Inhibitory Control in Other Task Contexts. *Journal of Cognitive Neuroscience* 30(1), 107–118. [https://doi.org/10.1162/jocn\\_a\\_01183](https://doi.org/10.1162/jocn_a_01183)
- Wessel JR and Aron AR** (2015) It's not too late: The onset of the frontocentral P3 indexes successful response inhibition in the stop-signal paradigm. *Psychophysiology* 52(4), 472–480. <https://doi.org/10.1111/psyp.12374>
- Wodniecka Z, Casado A, Kałamała P, Marecka M, Timmer K and Wolna A** (2020) The dynamics of language experience and how it affects language and cognition. In *Psychology of Learning and Motivation* (Vol. 72, pp. 235–281). Elsevier. <https://doi.org/10.1016/bs.plm.2020.02.005>
- Wu YJ and Thierry G** (2013) Fast Modulation of Executive Function by Language Context in Bilinguals. *Journal of Neuroscience* 33(33), 13533–13537. <https://doi.org/10.1523/JNEUROSCI.4760-12.2013>
- Yang J, Ye J, Wang R, Zhou K and Wu YJ** (2018) Bilingual Contexts Modulate the Inhibitory Control Network. *Frontiers in Psychology* 9. <https://doi.org/10.3389/fpsyg.2018.00395>
- Zhang H, Kang C, Wu Y, Ma F and Guo T** (2015) Improving proactive control with training on language switching in bilinguals: *NeuroReport* 26(6), 354–359. <https://doi.org/10.1097/WNR.0000000000000353>

## Appendices

### Appendix S1. Language games: stimuli and game procedure

#### Game stimuli

Each picture slide included five to twelve elements on a geometrically shaped background. In total, 54 unique slides were prepared, which allowed the creation of nine unique sets, each including six slides (one set of slides per game round). The host's and confederate's slides shared elements and backgrounds but differed in the number of elements (up to five additional elements were given to the confederate, i.e., these were to be removed) and their arrangement on the slide. In each set, the slides were split into three pairs that differed within the following dimensions: 1) the number of elements per slide (5, 8 or 12 elements on the host's slide); 2) complexity of the background (basic geometric shapes, e.g., cylinders and rectangles, vs. complex geometric shapes, e.g., overlapping lines); 3) semantic category (elements from distinct semantic categories, e.g., cat, car, lemon, etc. vs. elements from the same semantic category, e.g., lion, tiger, cat, etc.). These dimensions allowed the difficulty of the game to be manipulated (i.e., more elements, more complex background, and more semantically related elements were assumed to make the slide description more difficult), and thus constituted the three levels of difficulty within a game round: simple, moderate, and difficult. When two slides had been completed, the level of difficulty for the game increased. All slides were 1280 x 720 pixels in bitmap image format and were presented using Microsoft PowerPoint on laptop computers with a screen resolution of 1366 x 768 pixels. The host was presented with the slides in presentation mode, while the confederate was presented with the slides in editing mode, which enabled them to rearrange the slide elements using a computer mouse. All slides used in the experiment and the materials required for the games are available online at <https://osf.io/xy4qg>.

#### Game rules

The full game consisted of six rounds. Each round was time-limited and its duration depended on the difficulty of the slide. The time limits for the simple, moderate, and difficult sets were 1:30 min, 3:00 min, and 6:00 min, respectively. If the confederate finished rearranging their slides on a given difficulty level before the time limit, they received additional slides that contained the same elements as the finished slides but in a different arrangement so that they matched the difficulty level. The host and the confederate were allowed to communicate freely with each other; however, they could not use gestures, show each other the laptop screens, or use any other communication tools.

#### Game set-up

In total, four individuals were present during a language game session: three game players and the main experimenter, who was responsible for explaining and enforcing the game rules and monitoring the course of the game so that each language game was similar in terms of duration and sequence of events. At the beginning of the game, the three players (i.e., the two experimenters acting as players and the participant) were seated at a table with two laptops. Then, the main experimenter explained the rules of the game in English (L2 for participants). Once the procedure was clear to everyone, the three players engaged in the game. During the game, the main experimenter kept time and gave the sound signals for the start of each round and thirty seconds before the end of the time limit. This approach allowed a natural and voluntary end to the conversation between the host and the confederate. Moreover, the experimenters who acted as players were instructed to engage the participant during the game to enable a fluid and natural conversation. For example, if the participant could not remember the correct name of an object, the confederate would ask them questions about its visual features to allow the game to continue smoothly and to use the available time effectively. After six rounds of the game had been completed, the main experimenter scored the slides for accuracy.

### Appendix S2. Associations between behavioural and ERP data

In order to test the functional interpretation of the targeted ERP effects, i.e., the better the efficiency of inhibition-related mechanisms, the smaller the

ERP amplitude differences between the task conditions, we used LME regression models. N2, P3 and N450 were regressed on the following variables: RTs (i.e., the SSRT for N2 and P3, and the log-transformed Stroop RTs for N450; both scaled), Trial type (i.e., successful and unsuccessful stops for N2 and P3; congruent and incongruent for N450) and their interaction. All models included Trial type and Prior language use as random effects. The fitting procedure and contrasts were as described for Model 1 in section 2.4. None of the models needed trimming. We expected that RTs would interact with Trial type so that faster RTs would be related to smaller ERP effects, i.e., the N2 unsuccessful > successful inhibition effect, the P3 successful > unsuccessful inhibition effect and the N450 incongruent > congruent trial effect.

The model for N2 showed the main effects of Trial type and SSRT ( $t = -7.47$  and  $t = -8.64$ , respectively). However, Trial type and SSRT did not interact with each other ( $t = -0.39$ ). Contrary to the prediction, the magnitude of N2 (i.e., the N2 unsuccessful > successful inhibition effect) was not related to behavioural performance in the stop-signal task. The model for P3 showed the main effect of Trial type ( $t = -2.48$ ), but it did not show the main effect of SSRT ( $t = 0.41$ ). Crucially, it showed an interaction between Trial type and SSRT ( $t = -4.39$ ). SSRT positively predicted the P3 amplitude for successful stop trials ( $t = 2.01$ ) but did not predict the P3 amplitude for unsuccessful stop trials ( $t = -1.24$ ). This indicates that the P3 successful > unsuccessful inhibition effect was smaller for faster SSRT, which is in line with the prediction. The model for N450 did not show a main effect of Trial type ( $t = 0.82$ ) but it revealed the main effect of Stroop RTs and an interaction between Trial type and Stroop RTs ( $t = -2.20$  and  $t = -2.91$ , respectively). The Stroop RTs negatively predicted the N450 amplitude for incongruent trials ( $t = -3.64$ ) and did not predict N450 for congruent trials ( $t = -0.62$ ). As predicted, the N450 incongruent > congruent trial effect was smaller for faster Stroop RTs.

### Appendix S3. Test-retest reliability analysis

The test-retest reliabilities of the inhibition measures (i.e., the SSRT, N2 and P3, Stroop RTs and N450) were computed to verify whether the study had sufficient psychometric properties to detect intra-individual variation within the experimental manipulation, i.e., prior language use. Since SSRT is a single value for a participant, the estimates were assessed using only the classic Pearson's correlation coefficient ( $r$ ). The estimates for the other measures were additionally assessed using hierarchical models, as recommended by Rouder and Haaf (2019).

The classic  $r$  was computed for each pair of language games. The hierarchical models resembled those presented in Rouder and Haaf (2019). The effects of Trial type (i.e., successful vs. unsuccessful stops in the stop-signal task; incongruent vs. congruent in the Stroop task) for each language game (i.e., L1 game, L2 game, DL game) were taken as the fixed effects, whereas the overall Trial type effect (i.e., an effect regardless of a language game) and the idiosyncratic deviations within an individual (i.e., differences in the Trial type effect within a participant) were taken as random effects. The correlation coefficients derived from the hierarchical models were expressed by the multivariate distribution (for more details, see Rouder & Haaf, 2019).

Table S3 presents the test-retest reliabilities. All of the measures except for the N2 component demonstrate acceptable test-retest reliabilities in our study. For the SSRT, the Stroop RTs, and the P3 successful > unsuccessful inhibition effect, the estimates indicate excellent reliability. For the N450 incongruent > congruent trial effect, the estimates differ depending on the between-game comparison, but overall they are considered acceptable. Non-significant reliability estimates of the N2 unsuccessful > successful inhibition effect indicate that the N2 data was not stable over time and as such should be excluded from the interpretation in our study. At the same time, sufficient reliabilities of both the behavioural and the P3-N450 data imply that the current discrepancy between the behavioural and the ERP findings cannot be easily explained by the idiosyncratic properties of the study. While sufficient reliability for SSRT and the Stroop effect in RT shows that the absence of effects in the behavioural data is not a measurement error, the satisfactory reliabilities for the P3 and N450 data further support the presence of the reported effects.

**Table S3.** Test-retest reliability of inhibition measures across language-game sessions.

Task	Measure	Language game 1	Language game 2	Pearson's <i>r</i> correlation coefficient	<i>p</i> -value for Pearson's correlation coefficient	Model estimated correlation coefficient
<b>Stop-signal task</b>	SSRT	L1	L2	0.61	0.00	-
	SSRT	L1	DL	0.62	0.00	-
	SSRT	L2	DL	0.60	0.00	-
	N2	L1	L2	0.12	0.59	0.18
	N2	L1	DL	0.34	0.11	0.52
	N2	L2	DL	0.12	0.59	0.17
	P3	L1	L2	0.48	0.02	0.68
	P3	L1	DL	0.46	0.03	0.71
	P3	L2	DL	0.55	0.01	0.79
<b>Stroop task</b>	RTs	L1	L2	0.77	0.00	0.80
	RTs	L1	DL	0.66	0.00	0.69
	RTs	L2	DL	0.88	0.00	0.92
	N450	L1	L2	0.43	0.02	0.51
	N450	L1	DL	0.41	0.03	0.53
	N450	L2	DL	0.73	0.00	0.85