Cognition 226 (2022) 105169

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

The relative balance between languages predicts the degree of engagement of global language control

Alba Casado [a,*], Jakub Szewczyk [a,b,c], Agata Wolna [a], Zofia Wodniecka [a,*]

[a] *Psychology of Language and Bilingualism Lab, Institute of Psychology, Jagiellonian University, Krakow, Poland*
[b] *University of Illinois at Urbana-Champaign, USA*
[c] *Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands*

ARTICLE INFO

Keywords:
Global language control
Inhibitory control model
Language balance
Language activation level

ABSTRACT

After naming pictures in their second language (L2), bilinguals experience difficulty in naming pictures in their native language (L1). This phenomenon, the "L2 after-effect", is a lingering consequence of language control mechanisms regulating the activation of L1 and L2 to facilitate L2 production. Building on the Inhibitory Control model proposed by Green (1998), we propose that how much language control is applied depends on the relative balance between the current activation of L1 and L2. In two experiments, Polish-English bilinguals immersed in their L1 performed a blocked picture-naming task. This paradigm provided a continuous measure of the relative balance between the two languages and made it possible to index engagement of control by measuring the L2 after-effect. The results indicate that the higher the activation level of L1 and the lower the activation level of L2, the bigger the L2 after-effect. The results also revealed an enduring down-regulation of L1 activation level in more language-balanced speakers.

1. Introduction

In speakers who know more than one language, all their languages remain active in their mind even when only one language is required in a given context. The phenomenon of language co-activation has been observed in written word recognition (e.g., Dijkstra, 2005; Thierry & Wu, 2007), in spoken word recognition (Marian & Spivey, 2003), in written production (Iniesta, Paolieri, Serrano, & Bajo, 2021), and in spoken production (e.g., Costa, Caramazza, & Sebastian-Galles, 2000; Kroll, Bobb, & Wodniecka, 2006). The remarkable discovery of language co-activation has guided subsequent research in the field that attempts to characterize how bilinguals are able to successfully use contextually appropriate language despite the presence of language non-selectivity at all levels and the consequent interference (Kroll et al., 2006; for alternative arguments see Costa, 2005; La Heij, 2005). Previous research focused on understanding which mechanisms regulate co-activation and interference from the unwanted language(s). The most popular explanatory framework points to language inhibition as the key regulatory process (Green, 1998), although there is a growing notion that more than one control mechanism is probably involved in bilingual language processing (e.g., Gollan, Schotter, Gomez, Murillo, & Rayner, 2014). Still, the precise nature of these language control mechanism(s) is poorly understood. It is also an open question whether all bilinguals use the same mechanisms at all times, or – depending on the particular combination of language experiences of each bilingual (e.g., language pairing, context of language use, proficiency in using the languages, etc.) – different mechanisms of control are involved in different situations (e.g., Declerck, Kleinman, & Gollan, 2020; Van Assche, Duyck, & Gollan, 2013).

Our paper explores engagement of language control as a function of individual differences in activation levels of the native language (L1) and the second language (L2). The methodological approach adopted both here and in previous studies is based on the assumption that involvement of language control can be inferred indirectly via assessment of its side effects (Branzi, Martin, Abutalebi, & Costa, 2014; Costa & Santesteban, 2004; Declerck, Thoma, Koch, & Philipp, 2015; Declerck & Philipp, 2018; Declerck et al., 2020; Degani, Kreiner, Ataria, & Khateeb, 2020; Guo, Liu, Chen, & Li, 2013; Misra et al., 2012; Schwieter & Sunderman, 2008; Van Assche et al., 2013; Wodniecka, Szewczyk, Kałamała, Mandera, & Durlik, 2020). In this vein, the increased difficulty in L1 lexical access following the use of L2 (hereafter the "L2 after-effect") has been interpreted as a consequence of engagement of control

* Corresponding authors at: Psychology of Language and Bilingualism, Lab, Institute of Psychology, Jagiellonian University in Krakow, Ul. Ingardena 6, 30-060, Poland.

E-mail addresses: alba.casado@uj.edu.pl (A. Casado), zofia.wodniecka@uj.edu.pl (Z. Wodniecka).

https://doi.org/10.1016/j.cognition.2022.105169
Received 26 July 2021; Received in revised form 4 April 2022; Accepted 10 May 2022
Available online 13 June 2022

during or after L2 use. In other words, the behaviorally observed cost (slowdown) of accessing L1 after L2 is indicative of the engagement of a regulatory/control process. Although there has been an ongoing discussion over the exact interpretation of the underlying mechanism (see more in the next section), the magnitude of the L2 after-effect is a useful index that can be used to better understand the mechanisms that underlie language control in bilinguals. In the subsequent section, we explain in more detail a paradigm that allows measurement of the L2 after-effect.

1.1. Measurement of the L2 after-effect and theoretical inspirations

The L2 after-effect can be measured in a blocked picture-naming paradigm by comparing the processing costs of production in the native (or stronger) language when it follows longer (blocked) production in a second (or a weaker) language (e.g., Branzi et al., 2014; Wodniecka et al., 2020; see Branzi, Della Rosa, Canini, Costa, & Abutalebi, 2016 for fMRI evidence). In this paradigm, participants have to name pictures in either L1 or L2, and the language of naming changes between blocks (Branzi et al., 2014; Misra et al. 2012; Wodniecka et al., 2020). For instance, in a recent study by Wodniecka et al. (2020), Polish (L1) learners of English (L2) named pictures in L1 following the naming of pictures in either L1 or L2. The results showed a slowdown of response latencies when pictures were named in L1 just after they had been named in L2, compared to when they were named just after L1. Interestingly, the effect persisted for at least 5 min after changing the language. Similar lingering consequences of L2 use on subsequent L1 use have recently been observed for naming pictures in L1 after reading aloud in L2 (Degani et al., 2020).

The L2 after-effect has typically been explained as a consequence of inhibition applied to L1 during L2 production (but see Branzi et al., 2014; Branzi, Martin, Carreiras, & Paz-Alonso, 2020 for alternative explanations). The idea of inhibition as the mechanism that regulates the activation of the unwanted language was first formulated by Green (1998) in the Inhibitory Control Model (IC). This model assumes that in order for the weaker language (L2) to be used, the stronger language (L1) needs to be inhibited. Inhibition applied to L1 lasts for some time after L2 use is finished and consequently leads to a slowdown of L1 retrieval following L2 production. Although inhibition is certainly a valid theoretical framework to explain the L2 after-effect, other explanations cannot be ruled out on the basis of the available evidence. As we argued elsewhere (Wodniecka et al., 2020), the same effect can in fact be interpreted as evidence for alternative mechanisms (e.g., carry-over interference from L2, Branzi et al., 2014). An alternative framework that points to persisting L2 activation as an explanatory mechanism has previously been suggested by Philipp, Gade, and Koch (2007) in reference to similar language after-effects but observed in a smaller timescale (i.e., from trial to trial) in a language-switching paradigm. Under the persisting L2 activation account, when bilinguals speak in their non-dominant language (L2), the L2 has to be strongly activated in order to overcome the activation of the dominant language (L1). Consequently, when bilinguals are required to switch from speaking in L2 to speaking in L1, the increased activation of L2 results in a strong lingering interference between the two languages and hinders the use of L1 (for a similar proposal in which inhibition is not a prerequisite for language selection see Verhoef, Roelofs, & Chwilla, 2009). Similarly to the persisting activation account, a recent model by Blanco-Elorrieta and Caramazza (2021) proposes that the competition between L2 and L1 is modulated by a general selection mechanism which selects the most active of the competing lexical items. This model specifies that the activation of linguistic elements is determined by a combination of factors, such as the lexical frequency of these elements, the proficiency of the speaker, the intended lexical meaning, the communicative context, and the recency of use. As such, the lingering over-activation of L2 that results in retrieval difficulty in L1 can be explained by referring to the recency of L2 use and the communicative context.

Regardless of the controversy over the very nature of the language control mechanism, the IC model makes certain assumptions regarding the origin of the control mechanisms that regulate the concurrent activations of the two languages of a bilingual speaker. The IC model assumes that inhibition is reactive, which means that the amount of inhibition needed to suppress non-target lexical representations (i.e., lemmas with the incorrect language tag) is proportional to their initial activation level.[1] Although it is not explicitly stated in the IC model, this implies that the amount of inhibition applied to the lemmas in the non-target language should vary from speaker to speaker depending on the individual relative balance in the activation levels between L1 and L2 lemmas. The consequence of this assumption is that the strength of inhibition should depend on the relative balance between the languages known by a given speaker. Following Degani et al. (2020), we assume that the relative language balance depends on the baseline activation level of a given language (related to general fluency in this language and current immersion experience) and on more immediate language-activation demands (related, e.g., to recent language use). In speakers who are clearly dominant in one language, the dominant language (L1) should have a higher baseline level of activation than the weaker language (L2), therefore a larger amount of inhibition needs to be applied to L1 lemmas when L2 is used. In speakers who have comparable activation levels in both languages (i.e., balanced bilinguals), the amount of inhibition should be exactly the same for both languages (Green, 1998), or inhibition should not need to be applied to any of the languages at all (Costa, Santesteban, & Ivanova, 2006).

1.2. Degree of language control as a function of relative activation between languages

Here, we build on Green's inhibitory control model and propose an extension that focuses on changes in language activation levels due to requirements in the environment caused by immediate language-activation demands (specifically, recent language use). We make explicit assumptions about the causal relations between theoretical constructs of the relative activation between languages and the magnitude of language control triggered by the recent language use (see Fig. 1 panel A). Fig. 1 panel B provides an exemplary case of an unbalanced speaker. When such a bilingual is immersed in her natural language environment, she has a baseline activation of L1 typically higher than that of L2. Under the IC account, if at some point she is asked to use her L2, the L2 use triggers an inhibitory mechanism that dampens the activation of lemmas in L1. The amount of inhibition depends on the baseline activation of lemmas in both L1 and in L2. The more the L1 lemmas are activated compared to L2 lemmas, the higher the amount of inhibition that is applied to L1. Thus, when the activation of L1 and L2 lemmas is measured after the L2 use, in L1-dominant speakers, the activation level of L1 lemmas should be lower and the activation level of L2 lemmas should be higher than in a situation in which recent language use involved only L1. In more balanced speakers, the baseline activation level of lemmas is similar in both languages, hence the same amount of inhibition should be applied to L1 (when L2 is used) as to L2 (when L1 is used). Thus, in short, the more balanced a speaker is, the smaller the difference between the activation of L1 and L2 lemmas at baseline and after recent language use.

Moreover, we assume that the effects of inhibition will persist for some time after L2 use, thus keeping the activation of L1 lemmas low even if a bilingual speaker returns to using her L1. This assumption is consistent with data indicating that difficulty in accessing L1 lemmas after using L2 is relatively long-lasting: it has been shown to persist for at least a couple of minutes after a language switch (Branzi et al., 2014; Wodniecka et al., 2020). It is worth noting that the predictions from the

[1] More specifically, as stated by Green (1998), "because inhibition is reactive, more active lemmas will be more inhibited" (p. 74).

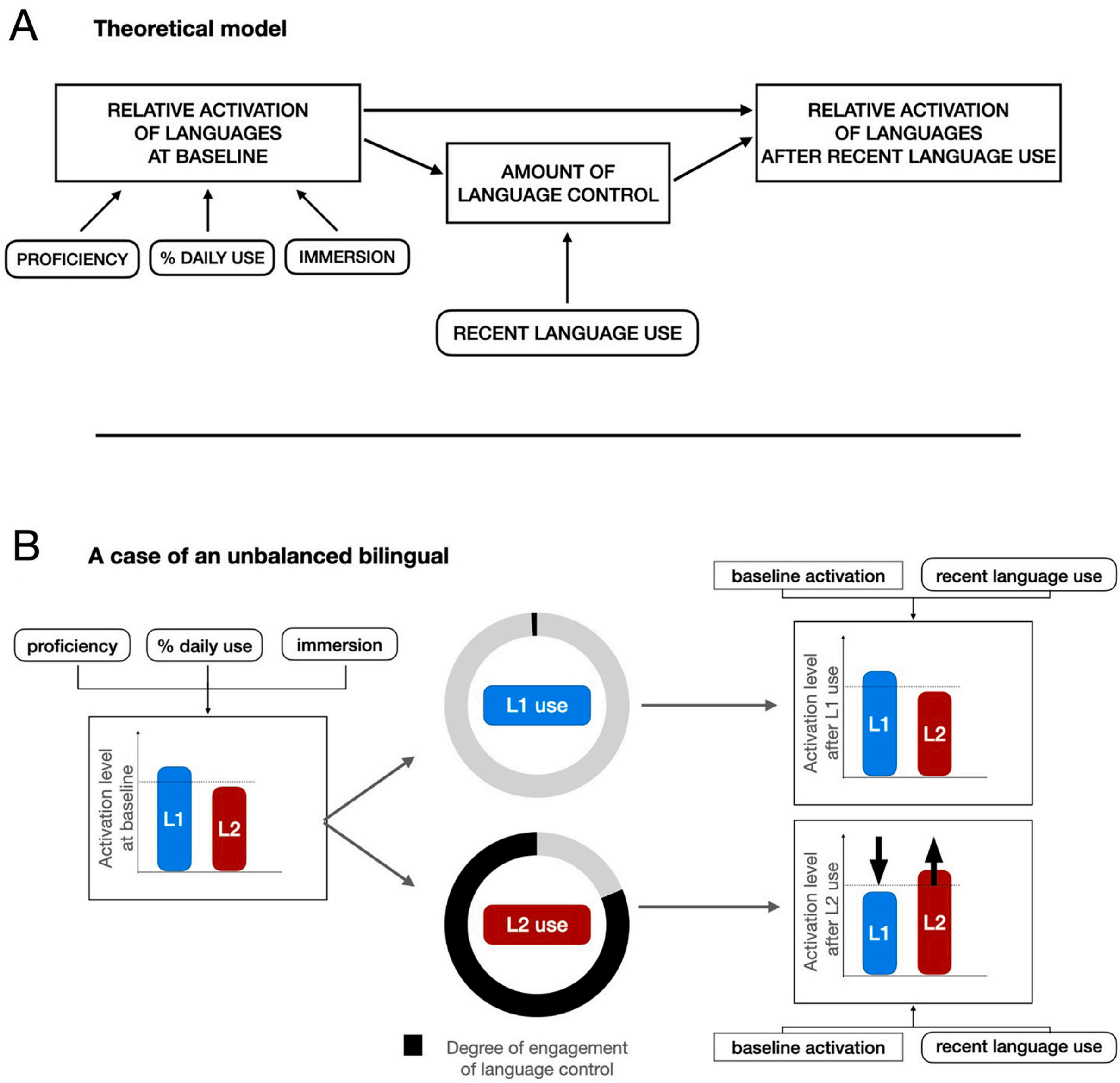

Fig. 1. Panel A) Theoretical model representing how the relative activation of languages at baseline changes after a "recent language use" manipulation. The degree of engagement of language control is determined by the relative activation of languages at baseline and by the requirements of recent language use (e.g., L1 or L2). The relative language activation of languages at baseline is the difference between the activation level of each language, which is a function of language proficiency, percentage of daily language use, and immersion experience. Panel B) A case of an L1-dominant bilingual for whom the level of L1 activation at baseline is higher than that of L2. When asked to use L1, a minimum degree of language control is required, therefore the activation levels of L1 and L2 remain the same. When required to use L2, a high degree of language control will be engaged which will lead to subsequent increase of L2 activation, but also to a decrease in L1 activation.

model would be the same even if the assumed regulatory process did not involve L1 inhibition during L2 use, but rather a different control process (e.g., a boost in L2 activation which then carries over to the subsequent L1 naming and causes interference in lexical access or – as suggested by Green and Abutalebi (2013) – goal maintenance, conflict monitoring, interference suppression or task engagement and disengagement). As such, our model is, in fact, agnostic as to the very process that underlies the predicted effects. Still, because it was directly inspired by the IC model, for the sake of simplicity in the remaining part of the text, whenever talking about the model and its predictions, we will refer to inhibition. We will then comment on the nature of the underlying mechanism in the General Discussion.

We present two experiments which test the specific predictions of our proposal, which was derived from the IC model. Before discussing the experiments in detail, we briefly present the available evidence on the relation between the degree of engagement of language control applied and the individual relative balance between the two languages.

1.3. Empirical evidence on the relation between language activation and language control

Several studies support the idea that the degree of language balance determines how much language control is recruited. These studies primarily used the language-switching paradigm rather than the blocked picture-naming paradigm used in the studies reported here. In the language-switching paradigm, participants are asked to switch from naming in one language to naming in another on a trial-to-trial basis. Language-switching clearly puts different demands on the cognitive system than the blocked picture-naming paradigm, in which participants name pictures in the same language for an entire block that lasts several minutes. Before reviewing the most important findings that have been taken as evidence for the relation between language activation and language control mechanisms, we briefly discuss the indices used in past studies and some limitations of the methodology of these past studies.

Within the language-switching paradigm, two indices of language control have been used to establish the relation between the degree of language balance and the engagement of language control: language-switching cost asymmetry (a larger cost of switching to the stronger language than to the weaker language), and, more recently, the reversed dominance effect, i.e., an overall temporary slowdown of naming in L1 in a language-mixing context (Declerck et al., 2020). The language-switching cost asymmetry has been suggested to reflect the cost of reactivating L1 (reactive inhibition), which was inhibited in the previous naming trial in L2 (for a review see Bobb & Wodniecka, 2013), whereas the reverse dominance effect has been explained with constant (proactive) L1 inhibition (Gade, Declerck, Philipp, Rey-Mermet, & Koch, 2021).[2] However, in real-life settings, bilinguals rarely switch between languages from one word to another, so this paradigm may probe processes that are actually quite distinct from the actual processes going on during bilinguals' language use. The blocked picture-naming paradigm seems to be more ecologically valid, even though it is still implemented in purely laboratory settings.[3]

Regarding the index of language balance, most previous studies have relied on L2 proficiency as a proxy for language balance. An implicit assumption is that high proficiency in L2 implies a greater balance between the languages. Some authors measured L2 proficiency with a competence test (Filippi, Karaminis, & Thomas, 2014), or self-assessment questionnaires (Costa et al., 2006; Costa & Santesteban, 2004; Lu et al., 2019), while others used verbal fluency tasks in L2 (Schwieter & Sunderman, 2008). Altogether, these studies found that the lower the L2 proficiency (or the lower the L2 verbal fluency), the larger the cost of switching to L1. This finding is in line with the idea that the degree of language balance is related to how much control needs to be engaged in L2 use.

An important limitation of this line of research is that language balance is assessed indirectly by relying on L2 proficiency measures (e.g., self-ratings, test of grammar abilities or vocabulary). However, the initial IC proposal (Green, 1998) assumes that the amount of inhibition depends on the relative balance in the activation of both languages. Importantly, language activation is not exclusively related to L2 proficiency (despite the fact that, most likely, it is usually highly and positively correlated); instead, it likely also depends on other contextual factors like immersion or recent language use.[4]

Some exceptions to the line of research described above (where language balance was assessed indirectly based only on L2 proficiency) are two studies which proposed different ways of measuring the language balance. The first is a recent study by Declerck et al. (2020), in which the authors used an index of language balance that relied jointly on both languages (L1 and L2) instead of relying only on L2. More specifically, the authors used the Multilingual Naming Test (Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012), which makes it possible to derive the language-balance index from picture-naming accuracy in L1 and L2. While combining both L1 and L2 accuracy in a picture-naming task preserves the idea that language balance must be measured by focusing on both languages rather than one, we would like to suggest that response times can offer an alternative and possibly more precise measure of differences in the languages' activation as they provide measurement on a scale of milliseconds and thus offer excellent resolution with which to measure the ease of accessing and producing a given word. In other words, we believe that response latencies may provide a sensitive and possibly a more direct quantitative measure of language activation because they relate not only to what participants know but to how quickly they access this knowledge (i.e., how long it takes for a word to be retrieved). In fact, this was the strategy applied by the second study that measured language balance without using L2 proficiency, i.e., the pioneering study by Meuter and Allport (1999). In this study, the authors used response times to measure the balance between activation levels in the two languages. In particular, the authors assessed the baseline language activation by calculating the difference between a participant's speed of naming numerals in L1 and L2. They found that when bilinguals were unbalanced (big differences between L1 and L2 naming speed), the magnitude of the switching cost was bigger for switching to L1 compared to switching to L2; on the other hand, when bilinguals were more language-balanced (small differences between L1 and L2 naming speed), the magnitude of the switching cost was similar for switching to L1 and to L2. In other words, with comparable activation of the two languages, the switching cost asymmetry disappeared. Meuter and Allport interpreted their results as indicating that to successfully use their L2, unbalanced bilinguals need to apply more inhibition to L1 than balanced bilinguals. However, despite the fact that the authors used naming latencies to measure language activation level, their index of balance between the language activations may have been biased because the response latencies were derived from language mixing blocks and from items (digits) that were repeated several times in the course of the experiment.

All in all, previous literature suggests that engagement of the control mechanisms that regulate L1 and L2 activation is bound to individual language balance. However, the available evidence for the relationship between language control and individual language balance is limited due to three methodological issues of past studies: (1) using a rather indirect index of language balance, usually L2 proficiency; (2) using an index of language control that relies on the language-switching

[2] Although L1-inhibition has been argued to be the most parsimonious explanatory account (see Gollan & Goldrick, 2018), there are alternatives that explain the different effects observed in bilinguals:L2 after-effects: Some researchers claim that unbalanced bilinguals experience costs of speaking in L1 after using L2 due to persisting L2-over-activation (Philip et al., 2007). In contrast, balanced bilinguals do not experience costs because their L1 activation level is optimal (Costa & Santesteban, 2004; Schwieter & Sunderman, 2008). Language-switching asymmetries: In the same vein, the language-switching cost asymmetry experienced by unbalanced bilinguals can be explained by assuming persisting L2 over-activation (Philipp et al., 2007).Reverse dominance effect: the overall temporary slowdown of naming in L1 in a language-mixing context experienced by balanced bilinguals can be explained by the language-specific selection criteria (e.g., Costa et al., 2006).

[3] It is worth adding that, in the past, the vast majority of studies used a language-switching paradigm to derive an index of language control. This might not have been ideal because in most previous studies that employed this paradigm a very small set of items was used and was repeated many times (also across the languages used). For example, in Experiment 1 of Costa and Santesteban's (2004) study, 10 items were repeated across 950 trials. Such repetition enhances lexical access in L1 and L2 differently (Kleinman & Gollan, 2018). In consequence, the repetition of items likely impacts the difficulty of returning to L1 after using L2 in a way that is difficult to control (for a review of studies using language switching, see Bobb & Wodniecka, 2013; Declerck & Philipp, 2015; Kleinman & Gollan, 2018). Notably, a recent meta-analysis has shown that the asymmetrical switch costs and the reversed language dominance effect observed in the language-switching paradigm are not always easy to replicate (Gade et al., 2021, see also Bobb & Wodniecka, 2013), which also undermines its utility as a litmus test of language control.

[4] Imagine a bilingual living in the L1 environment who studied L2 for professional purposes and obtained a high level of proficiency but has not used the language for a long time. If we were to test the L2 level, s/he would obtain very high grades in a written exam (i.e., grammatical rules, vocabulary, verbal tenses etc.), but s/he would be unable to have a fluent conversation in L2. Now, imagine a bilingual living in the L2-environment that never took a formal L2 course. If we were to test the L2 level, s/he would obtain very low grades in a written exam, but s/he would be perfectly able to have a fluent conversation with simple grammatical structures and high-frequency vocabulary. The first bilingual is an example of someone who is highly-proficient in L2 but has a low L2 activation level. The second bilingual is an example of someone with low-proficiency in L2 and a high L2 activation level.

paradigm; and (3) a confounded assessment of how broadly language control might affect the language system, i.e., whether it affects the whole language, a specific category of items, or specific items.

In the two experiments presented below, we addressed these methodological shortcomings by using a blocked picture-naming paradigm and focusing on pictures that participants only name once in the course of the experiment. As such, this paradigm allows purer measurement of control operating on the entire language (so-called whole-language or global language control; see Abutalebi & Green, 2016; Van Assche et al., 2013) rather than on a limited set of repeated items (so-called item-level or local language control). It also allows us to extend the findings of the language-switching paradigm to a longer time-course and affords a more direct and continuous test of the degree of an individual's balance between the two languages.

2. The current study

The goal of the two experiments reported in this paper was to test the predictions of the proposal presented above (Fig. 1). In each of the experiments, we tested a different group of unbalanced Polish-English speakers. In both experiments we used a blocked picture-naming paradigm which allowed us to address the methodological issues discussed in the previous section. This paradigm makes it possible to trace within-speaker changes in language-activation levels across different time points. In this paradigm, participants are asked to name pictures in language-consistent blocks (L1 or L2). They first complete a baseline naming block (in L1 or in L2) followed by an experimental block of naming in L1. The critical comparison involves experimental blocks of L1 naming after baseline naming in L1 vs. after baseline naming in L2. In our previous study that employed this paradigm (Wodniecka et al., 2020), we reported an L2 after-effect, i.e., a cost to L1 that is manifested in slower L1 naming latencies after using L2 compared to after using L1. Moreover, the L2 after-effect was observed for completely new items, therefore it appears to reflect control mechanisms applied over the entire language. Therefore, we use the L2 after-effect as a proxy of the engagement of language regulatory mechanisms that operate on the level of the entire language (whole or global language control).

Following the predictions of the model, we assumed that a larger difference between L1 naming after L2 compared to L1 naming after L1 is indicative of a greater cost, i.e., a larger hit to the baseline L1 activation level. To index the relative balance in activation between the two languages for a given individual, we used a continuous measure which was based on the difference between participants' mean baseline naming latencies in L1 and L2 naming. Importantly, both indices were independent because they were derived from different blocks of the task (see details in the Method section below). Predicting the L2 after-effect (i.e., the difference between a block of naming in L1 after L2 and after L1) using the predictors of L1 and L2 baseline naming latencies is free of circularity because the common components related to the measures of L1 and L2 activation cancel each other out when computing the L2 after-effect; these components include picture recognition, concept identification, individual speed of processing (which may vary across days), as well as more general components, such as SES, general verbal ability, literacy level, reading practice, etc. What is left is only the effect of control processes on L1 naming, which in principle could be independent of the baseline L1 and L2 naming times. In an extreme hypothetical case in which all L1 naming times would be identical throughout the entire experiment, the L2 after-effect would be zero and L1 baseline naming latencies (L1 activation) would predict nothing.

Overall, the design of the study allowed us to simultaneously measure the current activation levels of L1 and L2 as well as the cost to L1 as a result of L2 use. Experiment 2 aimed to conceptually replicate the findings of Experiment 1 while using different stimuli materials and refining the measurement of the L2 after-effect by controlling for the potential confounds identified in Experiment 1.

3. Experiment 1

For Experiment 1, we reanalyzed the data from our earlier study (Wodniecka et al., 2020), in which a relatively large group of bilinguals performed a battery of language and cognitive tasks.[5] Here, we reanalyzed the behavioral data from the picture-naming task in which participants named pictures in two testing stages that were separated by 30 min of other tasks completed in L1. Within each testing stage, there were two naming blocks of 60 pictures each. In both testing stages, the second blocks involved naming in L1, while the first blocks differed with respect to the language used in naming pictures: in testing stage 1 it was L1; in testing stage 2 it was L2 (see Fig. 2 for an overview of the design).

As the index of language control we used the L2 after-effect, which is defined as the difference in L1 naming latencies between the second block of the first testing stage (i.e., L1 naming after L1) and the second block of the second testing stage (i.e., L1 naming after L2). Our previous analysis of this dataset (Wodniecka et al., 2020) showed that L1 naming latencies measured in the second block were longer when they were preceded by L2 naming than when they were preceded by L1 naming. In the present analysis, we ask whether the size of the L2 after-effect depends on language balance.

To estimate the balance between L1 and L2 for each participant, we focused on the first blocks, in which participants named pictures in L1 and in L2. We assumed that the mean naming latencies in the first blocks could serve as a proxy for estimating the activation level of L1 and L2 lemmas and, subsequently, the relative language balance of each participant. Importantly, in order to not confound the estimates of language balance with the L2 after-effect, the data from the first blocks were not used in the estimation of the L2 after-effect.

The design of the study (Fig. 2) maps quite straightforwardly to the causal model of L2 after-effects outlined earlier (Fig. 1 panel A). The baseline L1 and L2 activation level is measured in the first blocks of each testing stage (from now on, we will refer to them as the baseline naming latencies). At the same time, naming pictures in the first block of the second testing stage serves as the experimental "recent language use" manipulation, that is, naming in L1 or in L2. We can appreciate its effect on L1 activation after recent language use (which, according to the theory, should be reduced by the engagement of language control) by comparing L1 naming latencies in the second blocks of each testing stage.

3.1. Predictions based on the proposed model

We expected that the smaller the difference in the relative activation of languages between the naming latencies in L1 and L2 at baseline, the smaller the L2 after-effect. Accordingly, we predicted that the magnitude of the L2 after-effect would be negatively associated with baseline naming latencies in L1 and positively associated with baseline naming latencies in L2.

3.2. Methods

3.2.1. Participants

Eighty-four students from three high schools in Kraków, Poland volunteered to participate in the study. Eight participants were removed due to various technical problems or incomplete data, leaving 76 participants who entered the analyses (mean age = 16.4 years, SD = 0.5; 54

[5] The testing was done three times across a span of two years. In the present reanalysis, we focus on the behavioral data from the picture-naming task obtained in the first session. The reanalysis of behavioral data obtained from picture naming in the other two sessions is presented in Appendix B section 1. The pattern of responses in both session 2 and session 3 replicates the findings of session 1 presented here. In the original study, we also collected EEG data, which were reported in the paper by Wodniecka et al. (2020).

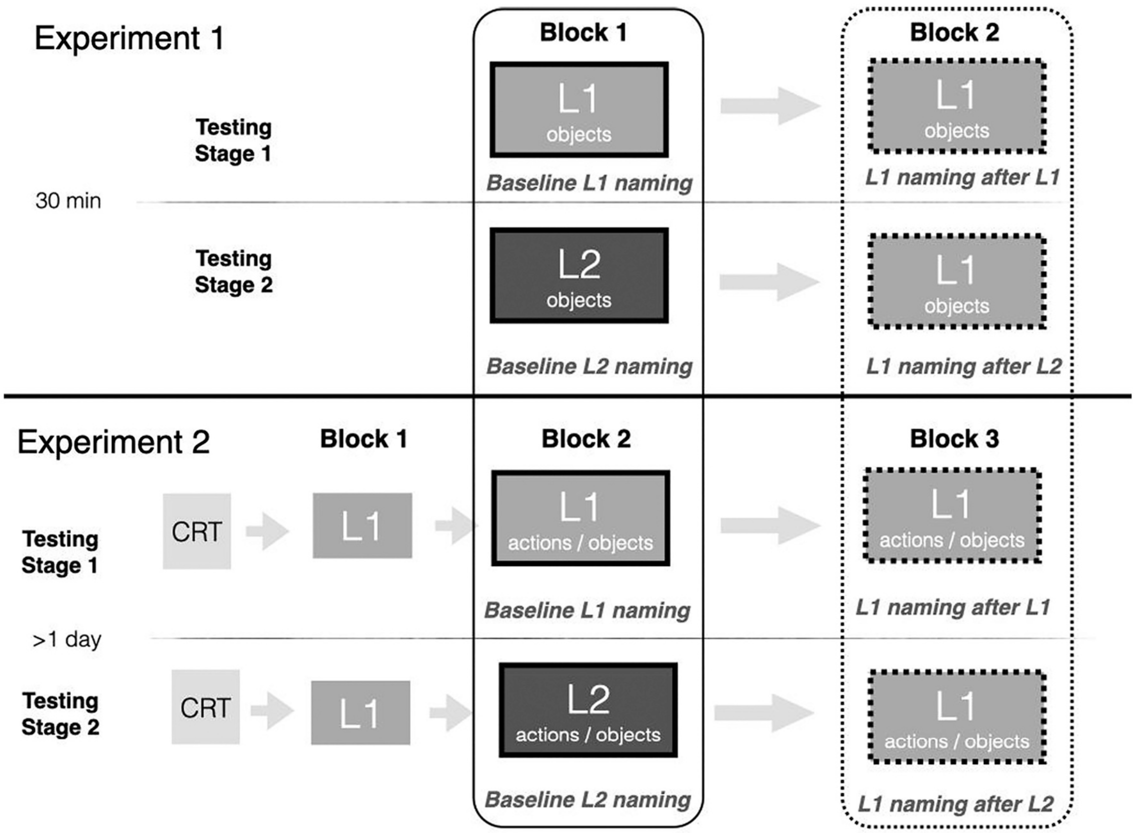

Fig. 2. Schematic representation of the task's design. In Experiment 1, two testing stages were separated by 30 min.; in each stage, participants completed 2 blocks of picture naming in which they named objects. Half of the pictures in the second block were repeated, although the reported analyses focused exclusively on the new (unrepeated) items. In Experiment 2, the two testing stages were separated by at least 1 day; in each stage, participants first performed a Choice Reaction Time task (CRT), followed by three different blocks of picture naming in which they first named actions and then objects inside each block. In Experiment 2, all the pictures were new.

female). The experiment met the requirements of the Ethics Committee of the Institute of Psychology of Jagiellonian University concerning experimental studies with human subjects. Since our participants were minors, written informed consent for participation was given both by the participants and their parents prior to the experiment. After completing the experimental session, all the participants received small gifts and were included in a prize lottery. Background information on the participants' self-assessed language proficiency, learning history and extent of language use were collected with a language background questionnaire based on a language history questionnaire (Li, Sepanski, & Zhao, 2006). For all the participants, Polish was the only language acquired in early childhood. They started learning English by the age of 6 and began using it more intensively when they started attending secondary school at around the age of 13. All the participants declared they had started learning another foreign language (L3; predominantly German or French), but their L3 proficiency was low and they did not speak L3 outside the classroom context (when they were taking formal lessons). All the participants reported using mostly L1 during an average day and speaking L2 only occasionally, mostly during English lessons at school. Similarly, the participants considered their L1 proficiency to be significantly higher than their L2 (for detailed self-assessment data, see Table 1). To obtain an objective L2 proficiency measure, we also asked the participants to fill in the Lexical Test for Advanced Learners of English, LexTALE (Lemhöfer & Broersma, 2012). On average, the participants obtained a score of 67% (SD = 11%), indicating that they were moderately proficient in their L2. Table 1 shows L1 and L2 picture-naming accuracy and naming latencies at baseline, thus indicating the clear relative dominance of their L1. To ensure that faster naming latencies in L2 were associated with higher accuracy scores in L2 naming, we ran a correlation analysis between the two measures at baseline (Block 1); the results showed a negative correlation, $r = -0.24$; $p = .037$, thus revealing that the participants who were faster in L2 were also more accurate in L2. We also assessed their semantic verbal fluency scores for both L1 and L2 (the participants produced as many words as possible in 60 s in each language; the categories were animals, fruits & vegetables, and body parts; the categories were rotated across the participants and languages). The participants produced significantly more words belonging to the target semantic category in L1 (mean = 21.8, sd = 5) than in L2 (mean = 15.3, sd = 5.3).

3.2.2. *Materials*

Stimuli in the picture-naming task included 180 pictures selected from the online database of the International Picture Naming Project (Szekely et al., 2004). All pictures were black-and-white line drawings sampled from a range of semantic categories; no Polish–English cognates were included (see Appendix A for the complete list of corresponding names in L1 and L2). The pictures were divided into six bins of 30 pictures, each matched in visual complexity, length of the most common names in Polish and English, age of acquisition in English, and the frequency of use in Polish and English (Polish: SUBTLEX-PL; Mandera, Keuleers, Wodniecka, & Brysbaert, 2015; English: SUBTLEX-US; Brysbaert & New, 2009). The six sets were rotated across the experimental conditions such that each picture occurred in each condition across participants.

3.2.3. *Procedure*

Each trial began with the presentation of a blank screen with randomized duration (700–1300 ms), followed by the presentation of a

Table 1
Characteristics of the Polish-English bilinguals in Experiments 1 and 2.

	EXPERIMENT 1		EXPERIMENT 2	
	L1	L2	L1	L2
Self-rated proficiency (%)	99.6 (2.23)	74.3 (10)	98.6 (10)	76.5 (20)
Percentage of daily use (%)	79.7 (15.7)	16 (12.8)	Active	
			80 (26)	16 (13)
			Passive	
			54 (24)	41 (22)
Lextale (mean accuracy in %)	-	66.97 (10.64)	72.8 (7.1)	72.7 (20.4)
Baseline naming accuracy (%)	93.5 (24.6)	75.2 (43.9)	97.1 (16.9)	89.2 (31)
Baseline naming latencies (ms)	889 (249)	1149 (317)	970 (284)	1165 (338)

Note: Standard deviations are given in parenthesis. Self-rated measures based on self-assessment questionnaire: self-rated proficiency (rated on a 1–7 scale in Experiment 1, and 1–10 in Experiment 2); percentage of daily use. Originally, participants were asked to rate the percentage of daily use of L1, L2 and L3. In the present table, we only include the ratings for L1 and L2 (in Experiment 2 separately for active use (top) and passive use (bottom) – this distinction was not available for Experiment 1). Objective measures: Lextale. Experimentally derived measures from the picture-naming task: Baseline naming accuracy and Baseline naming latencies.

fixation point in the center of the screen for 1000 ms. Subsequently, the fixation point was replaced by the target picture until the timeout (3000 ms) or a response occurred. The stimuli were presented using DMDX software (Forster & Forster, 2003) on a 17-in. screen positioned approximately 80 cm from the participant. The participants were instructed to name pictures aloud as quickly and accurately as possible. Reaction times for vocal responses were recorded; naming latencies were automatically measured using a DMDX voice key and subsequently manually screened for any non-speech sounds. The task was administered in two testing stages, separated by other tasks in L1 for approx. 30 min (see Fig. 2 above). Within each testing stage, there were two blocks of picture naming with 60 pictures each. The blocks of L1 naming were preceded by a block of naming in either L1 (first block of the first testing stage) or L2 (first block of the second testing stage; see Fig. 2). Half of the pictures used in the second blocks of each testing stage were new; half were repeated from the preceding block. The information concerning the target language for each block was presented at the onset of the block. The participants completed 8 practice trials before the onset of the first block in each testing stage. Participants were asked to take a short break between the first and the second block of each testing stage. Each of the blocks lasted approximately 10 min. Overall, the participants named 180 unique pictures, 60 of them twice (the repeated conditions). For the purposes of the present study, we analyzed only the new items.[6]

3.3. Data analysis

Prior to running the analyses on naming latencies, for each participant we removed all trials with errors, timeouts, as well as recording failures. All response latencies were filtered to include only the 300–2000 ms range. In total, 17.7% of the data was removed (12.66% due to accuracy, 5.08% due to RT outliers); 66.51% of the excluded data belonged to the L2 naming block. The naming latencies were transformed using a reciprocal transformation (−1000/RT) due to the right-skewed distribution of the data.

3.4. Statistical analyses

All analyses were performed using linear mixed-effects models, as implemented in the lme4 package (version 1.1.21; Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2020), using participants and pictures as crossed random effects.

3.4.1. Indices of L1 and L2 activation at baseline

The indices of L1 and L2 activation at baseline (see Fig. 1, panel B) were estimated based on the mean naming latencies in L1 and L2 in the first blocks of both testing stages (see Fig. 2). To maximize the reliability of these indices, instead of just computing the mean naming latency for each participant, we ran two mixed-effects regression models (one for L1 and another for L2) in which the dependent variable was naming latency in the first blocks. In this model we had no predictors except the intercept. We also included by-participant and by-item random intercepts. Next, we extracted the models' per-participant estimates of naming latency in L1 and L2 (the models' intercept, modified by the appropriate by-participant random intercept; see www.osf.io/27y4p for details). This way of estimating the by-participant naming latencies ensured higher reliability because the mixed-effects model accounted for the regression to the mean. We will refer to these estimates as the Estimate RT of L1 naming and the Estimate RT of L2 naming.

3.4.2. The main model

In the main model, the dependent variable was L1 naming latency in the second blocks of both testing stages (see Fig. 2). As fixed effects we included Preceding Language, Estimate RT of L1 naming, Estimate RT of L2 naming, interaction between the Preceding Language and Estimate RT of L1 naming, and interaction between the Preceding Language and Estimate RT of L2 naming. We also included (log-transformed) Trial number as a predictor to control for plausible effects of familiarization or fatigue or cumulative semantic interference that could accumulate during the experiment (Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Abdel Rahman & Melinger, 2009; Costa, Strijkers, Martin, & Thierry, 2009; Oppenheim, Dell, & Schwartz, 2010; Runnqvist, Strijkers, Alario, & Costa, 2012). The categorical predictor of Preceding Language was deviation coded (preceding L1 = −0.5, preceding L2 = 0.5) and the continuous predictors (Trial number, Estimate RT of L1 naming, and Estimate RT of L2 naming) were centered prior to running the analyses. We fitted the maximal model first (Barr, Levy, Scheepers, & Tily, 2013); in the case of non-convergence or singularities, we simplified it following the recommendations outlined in Bates, Kliegl, Vasishth, and Baayen, (2015). The final model included the by-participant random intercept and uncorrelated random slope for Preceding Language. It also included by-item random intercept and uncorrelated random slope for the interaction between Estimate RT of L1 naming and Preceding Language. We considered as significant any fixed effect with an absolute t-statistic value higher than 2.

3.5. Results

Table 2 shows the accuracy in all blocks of the task. Overall, participants were more accurate when naming in L1 than when naming in L2; the additional analyses presented in Appendix B section 1 show that the accuracy of L1 naming was the same in all blocks and testing stages.

The analysis of naming latencies in the second blocks of both testing stages showed a significant fixed effect of Preceding Language (i.e., the

[6] In the present model, we only take into account the global language level of inhibition assessed by new items. The item-level source of inhibition (which could be assessed by analyzing repeated items) is beyond the scope of the present study.

Table 2
Mean accuracy in picture naming in the baseline and experimental blocks, calculated by participant.

	Baseline block		Experimental block	
Experiment 1	Block 1		Block 2	
Testing Stage	Language	Mean (SD)	Language	Mean (SD)
Stage 1 (L1-L1)	L1	.94 (.24)	L1	.95 (.08)
Stage 2 (L2-L1)	L2	.75 (.44)	L1	.94 (.06)
Experiment 2	Block 2		Block 3	
Testing Stage	Language	Mean (SD)	Language	Mean (SD)
Stage 1 (L1-L1)	L1	.97 (.17)	L1	.97 (.17)
Stage 2 (L2-L1)	L2	.89 (.31)	L1	.96 (.18)

L2 after-effect). Participants were slower to name pictures in L1 when the naming followed a block of naming in L2 than when it followed another block of naming in L1. In addition, there was a significant effect of Estimate RT of L1 naming (the faster participants named pictures in L1 in the first block, the faster they named pictures in L1 in the second blocks of both testing stages); there was also a significant effect of Estimate RT of L2 naming (the faster participants named pictures in L2 in the first block, the slower they named pictures in L1 in the second blocks of both testing stages).

Crucially, Preceding Language entered into two significant interactions: with Estimate RT of L1 naming and with Estimate RT of L2 naming (see Table 3). The L2 after-effect was dependent on naming latencies of both L1 and L2 in the baseline blocks: it was bigger for participants who were faster in L1 naming at baseline (in the first block) and slower in L2 naming at baseline (in the first block). In other words, the L2 after-effect was bigger for participants whose naming latencies in block 1 showed the largest imbalance between L1 and L2 in favor of faster naming in L1 (see Fig. 3). Finally, there was an effect of Trial number, indicating that participants were slower to name the pictures as the experiment progressed.

Table 3
Fixed and random effects from the LME model of naming latencies in the experimental blocks (Block 2 of Experiment 1 and Block 3 of Experiment 2).

EXPERIMENT 1					
Effect	Estimate	SE	t	by-item SD	by-participant SD
Intercept	−1.21	0.08	−63.52	0.12	0.05
Preceding.Lang	0.10	0.01	8.73		0.07
Estimate.RT.L1. naming	0.67	0.06	10.80		
Estimate.RT.L2. naming	0.32	0.08	3.81		
Trial.number	0.09	0.02	5.18		
PrecedingLang: Estimate.RT.L1. naming	−0.33	0.09	−3.43	0.14	–
PrecedingLang: Estimate.RT.L2. naming	0.39	0.13	3.04		–
EXPERIMENT 2					
Effect	Estimate	SE	t	by-item SD	by-participant SD
Intercept	−1.07	0.01	−73.17	0.14	0.05
PrecedingLang	0.03	0.01	2.50	0.03	0.07
Estimate.RT.L1. naming	0.55	0.09	6.06	0.25	–
Estimate.RT.L2. naming	0.24	0.12	2.07		–
Estimate CRT	0.05	0.03	2.07	–	–
Trial.number	−0.13	0.05	−2.77	0.01	0.17
Session	0.01	0.01	0.22		
PrecedingLang: Estimate.RT.L1. naming	−0.23	0.11	−2.11	0.22	–
PrecedingLang: Estimate.RT.L2. naming	0.32	0.14	2.25	0.25	–

3.6. Discussion

As reported previously (Wodniecka et al., 2020), we observed the L2 after-effect: naming new pictures in L1 after a block of L2 naming was slower than naming in L1 after another block of naming in L1. Critically, we found that the L2 after-effect was bigger for participants with high L1 activation level (as indexed by fast baseline L1 naming latencies in the first block) and low L2 activation level (as indexed by long L2 naming latencies in the first block). This pattern of results is in line with our hypothesis, which predicts that the L2 after-effect depends on the relative balance of the participants, where balance is defined as the difference between L2 and L1 activation. However, because Experiment 1 was not originally designed to test this hypothesis, we ran a second study to further explore this novel finding while addressing the limitations of Experiment 1, which we discuss in the following paragraph.

In Experiment 1, the first stage always included two blocks of naming in the L1 blocks; after 30 min of other tasks, the second stage was always composed of a block of naming in L2 followed by a block of naming in L1. The fixed order in that study was intentional. Specifically, we first presented a block of naming in L1 followed by another block of naming in L1 (L1 after L1) to prevent a possible confounding influence of L2 on the L1 baseline block. As it is unknown how long-lasting the effects of L2 on subsequent naming in L1 were, the only way to control for this possibility was to get a baseline measure (L1 after L1) before L2 was first used in the experiment. However, it was possible that the fixed order of testing stages (L1-L1, followed by L2-L1) would confound the measurement of the L2 after-effect with other factors correlated with time, such as fatigue, or cumulative semantic interference (Howard et al., 2006; Runnqvist et al., 2012; see Wolna, Szewczyk, Casado, & Wodniecka, 2021, for discussion). As a consequence, the participants could have been slower to name pictures in L1 in the second testing stage (in which the L2 after-effect was measured) not because it followed a block of naming in L2, but simply because the condition in which they named in L1 after naming in L2 (L1 after L2) was always completed later in the experiment than the condition in which they named in L1 after naming in L2 (L1 after L1). In addition, the fixed order might have induced expectations towards the use of L2 in the second testing stage. In other words, the mere fact that the participants expected that at some point of the task they would need to use their L2 could have put them into "bilingual mode" (Grosjean, 2001) and triggered preparation for using L2. This, in turn, could have resulted in an increase of naming latencies in L1 blocks that preceded the second stage and therefore affected the baseline measurement (first block of L1 naming). Altogether, the participants' expectations that they would use L2 may have modulated the measurement of language balance and perhaps also the magnitude of the L2 after-effect.

Finally, another aspect of the design of Experiment 1 that could have influenced the results is the inclusion of repeated items. Although in the presented analyses we excluded the repeated items, it is unclear how the mixing of repeated and new items may affect lexical access to the names of new pictures. Therefore, in Experiment 2, to allow purer assessment of the global language control, we included only new items in the task design.

4. Experiment 2

In order for it to be a conceptual replication of Experiment 1, Experiment 2 was specifically designed to explore how the balance between languages modulates the magnitude of the L2 after-effect. In order to overcome the limitations caused by the fixed order of sessions, in Experiment 2 we divided the testing into two sessions that were completed on separate days. To further control for possible confounds

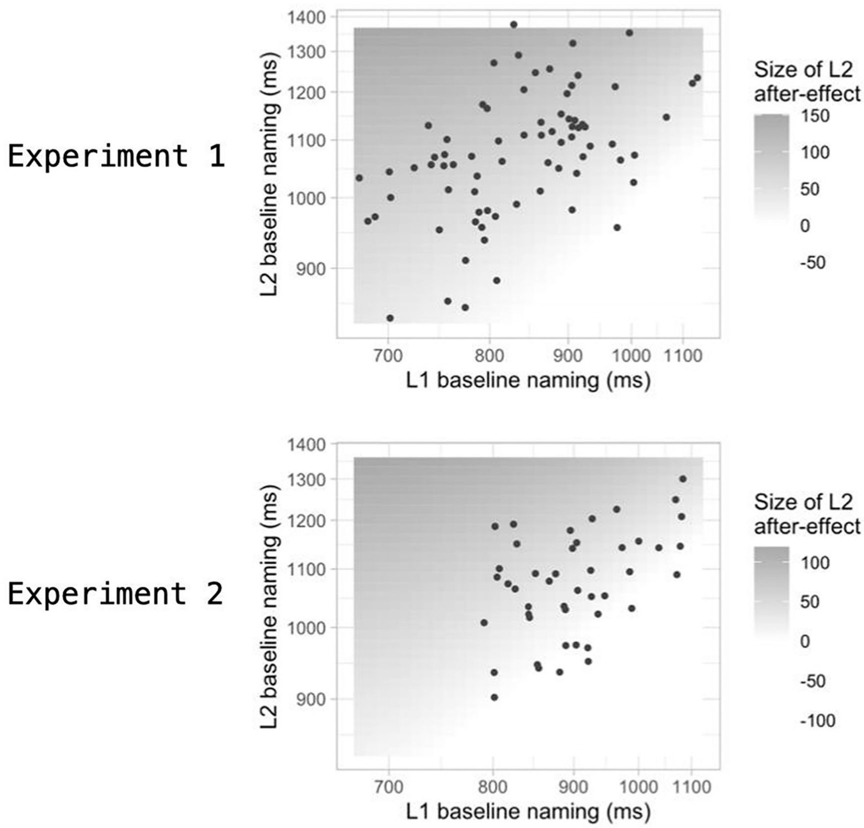

Fig. 3. Model-based predictions of the magnitude of the L2 after-effect as a function of L1 and L2 baseline naming speed in Experiment 1 and Experiment 2. The x-axis corresponds to L1 baseline naming latencies (transformed back to milliseconds) measured in block 1, while the y-axis corresponds to L2 baseline naming latencies (transformed back to milliseconds). The black points show the L1 and L2 baseline naming latencies of the participants. The L2 after-effect was calculated by subtracting the predicted naming latencies in L1 after L1 from the predicted naming latencies in L1 after L2; the result was then transformed back to the millisecond scale for visualization purposes. The magnitude of the L2 after-effect is represented by the greyscale. The darker it is, the bigger the L2 after-effect. White color corresponds to the absence of the L2 after-effect.

related to expectations regarding L2 use, the order of testing stages was counterbalanced, and we introduced a double-blinded procedure so that neither the participant nor the experimenter were aware of the appearance of L2 in the ongoing testing session (for details, see Procedure).

To make it possible to determine whether the L2 after-effect operates on the level of global language, each trial included a unique, never-repeated item. Finally, we used a different database of pictures than in Experiment 1 that included pictures of objects and actions in order to generalize the findings not only across participants but also across study materials.

However, running the two experimental sessions on separate days required us to control for another potential confound: day-to-day variability in the processing speed of each participant. It has been shown that participants' reaction times fluctuate from day to day (e.g., Rabbitt, Osman, Moore, & Stollery, 2001), which could confound the measurement of the L2 after-effect. First, we included the session number as a covariate in the main statistical model. Moreover, each session began with a non-linguistic choice-reaction task. This provided us with an estimate of mean reaction time in the choice-reaction task on a given day which we could subsequently include as a covariate in the main statistical model and statistically adjust for day-to-day variability in processing speed.

All in all, the aim of Experiment 2 was to conceptually replicate the results of Experiment 1 and to provide a purer measurement of the L2 after-effect. In Experiment 2, we avoided possible confounds related to the fixed order of blocks (like in Experiment 1) by dividing the experiment into two sessions: one aimed to measure L1 after L1 naming, and another one to measure L1 after L2 naming. This allowed us to control for the ordinal effects as well as to counterbalance the order of sessions between participants and control for their expectation of the upcoming use of L2. We also controlled for day-to-day variability in the processing speed of each participant and only used unrepeated items to be able to make stronger inferences about control mechanisms operating on the level of global language.

4.1. Participants

Fifty-five Polish-English bilinguals (mean age = 22.96, SD = 3.91) recruited via an online advertisement took part in the experiment and received monetary remuneration for their participation. Following the preregistered criteria, to qualify for the study the participants needed to have a minimum level of proficiency in L2 (English), as assessed by online tests: at least 60% on LexTALE score (Lemhöfer & Broersma, 2012) and 18/25 points or higher in General English Test (Cambridge assessment: https://www.cambridgeenglish.org/test-your-english/general-english/). To ensure that the participant sample included only individuals who had high naming accuracy, we excluded 5 participants due to low naming accuracy in L1 (1 participant, less than 90% across all L1 blocks) or in L2 (4 participants, less than 60% in the L2 block). The final number of participants was established using a stopping rule (see www.osf.io/27y4p), according to which the precision of the L2 after-effect measurement (whether it had reached a desired standard, i.e. SE < 0.012) was checked after 30 participants and repeated after every other 5 participants. The data collection was set to stop after the stopping rule criteria were met or the data for 50 participants had been collected. The criterion for the stopping rule was based on the results of our previous experiment (Wodniecka et al., 2020) and was preregistered (see the preregistration protocol for further information: www.osf.io/27y4p). The final sample included 50 participants, each of which was tested in two sessions. Background information on participants' self-assessed language proficiency, learning history, and extent of language use were collected with a language background questionnaire based on the L2 Language History Questionnaire (Li et al., 2006). All the participants were language-unbalanced Polish–English bilinguals whose only language acquired in early childhood was Polish (mean age of acquisition of L2 = 7.47 years old). They started using English more intensively in secondary school at around the age of 12. Additionally, they had learned another foreign language (L3; predominantly German, French or Spanish), but their L3 proficiency was low and they did not speak L3 outside the classroom context (when they were taking formal lessons). All the participants reported mostly speaking L1 during an average day and speaking L2 only occasionally. Similarly, the participants considered their L1 proficiency as significantly higher than L2. To obtain an objective L2 proficiency measure, we also asked participants to fill in the Lexical Test for Advanced Learners of English, LexTALE (Lemhöfer & Broersma, 2012). On average, the participants obtained a score of 72% (SD = 20%), indicating that they were proficient in their L2 (for more details see Table 1).

4.2. *Choice reaction time task*

In order to control for the baseline reaction speed in a given session, the participants performed a Choice Reaction Time task (CRT, adapted from: https://www.cambridgecognition.com/cantab/cognitive-tests/choice-reaction-time-crt/) at the beginning of each experimental session. During each trial, an arrow appeared on the computer screen, the direction of which always corresponded with its position in the screen (e.g., an arrow pointing to the right on the right side of the screen). The participants' task was to indicate the direction and position of each arrow by pressing the response buttons on the keyboard as soon as possible. The task consisted of 10 training stimuli followed by 200 target stimuli. Each trial started with a blank screen displayed for 1000 ms, followed by 500 ms of a fixation point. Subsequently, an arrow was presented for a maximum of 3000 ms for the participants to respond. The position and orientation of the arrows was randomized, with an equal number of arrows pointing to the right and to the left side. The total duration of the task was approximately 10 min.

4.3. *Picture-naming task*

Right after the CRT task, the participants performed the picture-naming task. To control for the possible effect of the participants' expectations, a double-blind procedure was implemented at the level of the experimental instructions. The experimenter explained to the participants that the task was composed of 3 experimental blocks in each session; in each block they would see various pictures in the center on the screen. Their task was to name the pictures aloud; depending on the instructions displayed on the screen before each experimental block, they would have to use either Polish (L1) or English (L2). The participants were told that the language of naming was randomly assigned to each block by a computer. In such a way, the languages to be used as well as their order within the experiment were unpredictable for the participants. Additionally, the experimenter did not know which of the two sessions would be performed. We selected a total of 342 colored pictures from the two databases: the CLT database and an additional set of colored pictures (Haman, Łuniewska, & Pomiechowska, 2015; Haman et al., 2017), including pictures representing objects (as in Experiment 1) and pictures representing actions; pictures whose names were Polish-English cognates were excluded. We divided the pictures into 6 blocks of items. Each block included 20 pictures representing actions and 37 pictures representing objects. Within each block, actions were always displayed before objects; pictures within each category were presented randomly. Pictures were matched on name agreement, lexical frequency and the length (in phonemes) of the dominant name in Polish between blocks. We also selected 16 additional pictures to create four practice blocks that were presented before the first and the second picture-naming block of each experimental session (see Fig. 2). Each practice block consisted of 4 pictures presented in a fixed order: 2 pictures representing actions and 2 pictures representing objects. For a complete list of stimuli, see Appendix A. The pictures were presented using DMDX software (Forster & Forster, 2003) in the middle of the screen for 2000 ms and were preceded by a fixation cross displayed for 1000 ms. In each session, the first and the third block required naming in L1 (Polish). The second block required naming in L1 or in L2, depending on the experimental session (see Fig. 2).

4.4. *Data analysis*

All trials in the picture-naming task were scored for naming accuracy. Only trials with correct responses were used in the analyses. Reaction times were extracted using Chronset, which is an automated tool for detecting speech onset (Roux, Armstrong, & Carreiras, 2017). Items in which naming latencies were shorter than 300 ms or longer than 2000 ms were rejected. A total of 15.81% of the data was rejected (10.56% due to accuracy, 5.65% due to RT outliers). 50.26% of the excluded data belonged to the L2 naming block. The analysis was performed using linear mixed-effects models, as implemented in the lme4 package (version 1.1.21; Bates, Maechler, et al., 2015) in R (R Core Team, 2020), using participants and pictures as crossed random effects. For the purpose of the present experiment, we analyzed the data from blocks 2 and 3.

4.4.1. *Indices of L1 and L2 activation at baseline*

We calculated the indices in a similar way as in Experiment 1 using the estimates of the mean latencies of naming in L1 and L2 in the second blocks of both experimental sessions.

4.4.2. *Index of CRT estimate*

Additionally, we calculated the estimates of response latencies in the CRT task for each session in order to control for the effect of baseline reaction speed (CRT estimate). We derived the CRT estimate separately for each session from a linear mixed model fitted for the reaction times in the CRT task with Session as a fixed factor.

4.4.3. *Main model*

In the main model, the dependent variable was the naming latencies in the third blocks of both experimental sessions (both involved naming in L1, see Fig. 2). As fixed effects we included Preceding Language, Estimate of CRT, Estimate RT of L1 naming, Estimate RT of L2 naming, interaction between the Preceding Language and Estimate RT of L1 naming, and the interaction between the Preceding Language and Estimate RT of L2 naming. We also included a log-transformed Trial number to control for effects of familiarization, Session order, and the fatigue or cumulative semantic interference that could accumulate during the experimental session. The categorical predictors were deviation coded: Preceding Language (preceding L1 = −0.5, preceding L2 = 0.5), and Session order (first = −0.5, second = 0.5). The continuous predictors (Trial number, Estimate of CRT, Estimate RT of L1 naming, and Estimate RT of L2 naming) were centered prior to running the analyses. In addition, the RTs were transformed into the inverse score (−1000/RT) due to the right-skewed distribution of the data. We fitted the maximal model first (Barr et al., 2013); when non-convergence or singularities occurred, we simplified them following the recommendations outlined in Bates et al. (2015). The final model included by-participant random intercept and uncorrelated random slopes by Preceding Language and Session order. It also included by-item random intercept; uncorrelated by-item random slopes for Preceding Language; Estimate RT of L1 naming; the interaction between Preceding Language and Estimate RT of L1 naming; and the interaction between Preceding Language and Estimate RT of L2 naming. We considered as significant any factor with an absolute t-statistic value higher than 2.

4.5. *Results*

Table 2 shows accuracy in all blocks of the task. Similarly to Experiment 1, participants were more accurate when naming in L1 than in L2, and the accuracy of L1 naming was the same in blocks 2 and 3 in testing stages 1 and 2 (see Appendix B section 2 for detailed analyses on accuracy measure). The analysis of naming latencies in the third blocks of both experimental sessions showed a significant effect of Preceding Language. The participants were slower to name pictures in L1 when it followed a block of naming in L2 than when it followed a block of naming in L1. In addition, there was a significant effect of Estimate RT of L1 naming (the faster the participants named pictures in L1 in the second block, the faster they named pictures in L1 in the third blocks of both experimental sessions) and a significant effect of Estimate RT of L2 naming (the faster the participants named pictures in L2 in the second block, the slower they named pictures in L1 in the third blocks of both experimental sessions). Crucially, Preceding Language entered into two significant interactions: with Estimate RT of L1 naming and with Estimate RT of L2 naming (see Table 3). Both L1 and L2 baseline latencies contributed to the L2 after-effect: the faster the participants' naming in L1 (in the second block) and the slower their naming in L2 (in the second block), the bigger the L2 after-effect. In other words, the L2 after-effect was bigger for participants whose naming latencies in block 2 showed the largest imbalance (i.e., they were L1 dominant; see Fig. 3 for a visualization of this effect). Finally, there was an effect of Trial number, indicating that participants were slower to name the pictures as the experiment progressed. Most likely this effect reflects the fact that it is slower to name pictures of actions (which were always presented in the second part of a block) than pictures of objects (which were presented in the first part of each block, see Fig. 2). As such, the Trial effect in Experiment 2 is not exactly comparable with the effect of Trial number observed in Experiment 1 (which only included pictures of objects).

To further explore the Trial effect in Experiment 2, we run an additional analysis including Picture type (actions vs. objects) as a fixed-effect factor (see Table 4). As expected, Picture type was significant, revealing that actions took longer to name than objects; however, when Picture type was taken into account, Trial number was no longer significant. Importantly, the interaction between Picture type and Preceding language was also not significant, indicating that the magnitude of the L2 after-effect was not different for pictures of objects and actions. Interestingly, we also found a significant interaction between Trial number and Preceding Language. Further exploration of this interaction revealed that naming latencies increased with Trial number after the L1 use, whereas they decreased with Trial number after L2 use (see Fig. 4).

Table 4
Fixed and random effects from the LME model of naming latencies in the experimental block (Block 3).

Effect	Estimate	SE	t	by-Item SD	by-Participant SD
Intercept	−1.09	0.01	−79.13	0.13	0.06
PrecedingLang	0.07	0.02	3.82		0.07
Estimate.RT.L1.naming	0.56	0.09	6.24	0.21	
Estimate.RT.L2.naming	0.24	0.12	2.05		
Estimate.Choice.RT	0.05	0.02	1.97		
Trial.number	0.01	0.01	0.16		
Session	0.01	0.01	0.18		
Picture.type	−0.12	0.02	−6.33		0.04
PrecedingLang:Estimate.RT.L1.naming	−0.24	0.11	−2.13	0.21	–
PrecedingLang:Estimate.RT.L2.naming	0.34	0.14	2.31	0.21	–
PrecedingLang:Picture.type	0.04	0.02	1.80		
PrecedingLang:Trial.number	−0.06	0.02	−2.8		

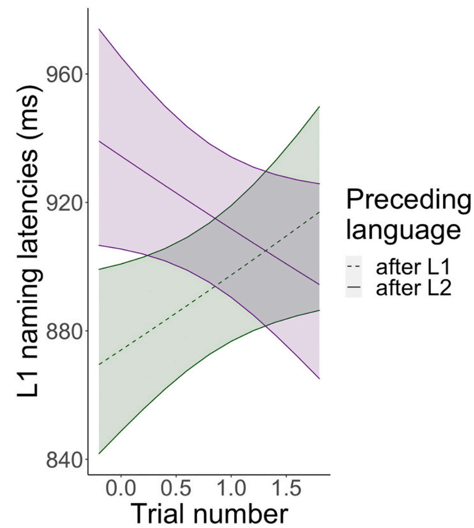

Fig. 4. Interaction between Trial number and Preceding language in Experiment 2. Trial number is log-transformed. The naming latencies are transformed back to millisecond for visualization purposes.

4.6. Discussion

Similarly to Experiment 1 and other previous studies (e.g., Branzi et al., 2014), the results showed the L2 after-effect, i.e., longer naming latencies when naming in L1 followed naming in L2 than when naming in L1 followed naming in L1. Crucially, the magnitude of the L2 after-effect was modulated by the relative balance between L1 and L2. The results of the present experiment replicated Experiment 1: the bigger the difference between the current L2 and L1 activation levels, the bigger the L2 after-effect (see Fig. 3). Note that the naming latencies of the experimental block in Experiment 1 are shorter than naming latencies in Experiment 2 (see Fig. 3). This difference might have arisen simply because in Experiment 2 we included pictures of actions, which lead to longer naming latencies than pictures of objects. This prompted us to check if the picture type (actions vs. objects) affects not only overall naming latencies but also the magnitude of the L2 after-effect. However, we found that the interaction between the preceding language and the picture type was not significant (see Table 4). Another possible source of differences in naming latencies across the two experiments was that participants of Experiment 1 were younger (mean age = 16.4 years, SD = 0.5) than those of Experiment 2 (mean age = 22.96, SD = 3.91). Further research would be needed to explore whether teenagers are systematically faster to name pictures than young adults. The differences in naming latencies between the experiments could also result from the fact that some items were repeated in Experiment 1. Stasenko, Kleinman, and Gollan (2021) recently suggested that repetition can offset the effects of inhibition, and this attenuation of L1 inhibition can mask the inhibition applied to new items. Although our analysis focused only on new (never repeated items), the results of Experiment 1 might still have been biased by the inclusion of repeated items in the design. Importantly, Experiment 2 was free of similar bias or contamination.

Finally, we conducted exploratory analyses of the interaction between Trial number and Preceding language which revealed changes in the L2 after effect (see Fig. 4). Each language followed a different dynamic: after L1 use, there was an increase of naming latencies as trial number increased. In contrast, after L2 use, there was a decrease of naming latencies as trial number increased. As we argued earlier, the increase in L1 naming latencies after L1 use may be explained by fatigue or cumulative semantic interference (Howard et al., 2006; Runnqvist et al., 2012). The decrease in L1 naming latencies after L2 use indicates that the initial difficulty of retrieving words in L1 after L2 use diminishes with time and speakers regain their initial L1 activation levels.

5. General discussion

In the two presented experiments, we explored the connection between 1) the relative balance between L1 and L2 and 2) the magnitude of L1 inhibition applied during L2 production. We tested Polish-English bilinguals with varied levels of current activation in L1 and L2. We used novel indices of an individual's engagement of language control and language balance. Both indices were derived from a blocked picture-naming task in which participants named a set of pictures in L1 following a set of completely new pictures in either L1 or L2. To index language control, we used the L2 after-effect, i.e., a slow-down of L1 naming after L2 use. To index the language balance, we used a continuous measure based on baseline picture-naming latencies in L2 and in L1.

The results of both experiments provide clear evidence for the relationship between the relative balance between the two languages and the degree of language control involved during L2 use. Although most participants experienced the L2 after-effect, the effect was larger in those who were characterized by a greater imbalance between languages. These findings are in line with previous research that investigated differences in the degree of engagement of language control in bilingual speakers, but the previous research used different indices (for language control, the indices used in the past were derived from a language-switching paradigm; for language balance, the indices were typically based on L2 proficiency or language production accuracy, e.g., Costa & Santesteban, 2004; Costa et al., 2006; Declerck et al., 2020; Filippi et al., 2014; Lu et al., 2019; Schwieter & Sunderman, 2008). Although the converging evidence supports the idea that recruitment of language control during L2 use depends on the relative language balance, we believe that the current study provides a number of important insights into the theoretical accounts of bilingual language control as well as into the methodology used to study its underlying mechanisms.

5.1. Implications for theorizing on the mechanism of bilingual language control

Initially, Green proposed (1998) that whenever bilinguals want to use their second language, they have to inhibit their first language in order to prevent it from interfering with L2. Building on the assumptions of the original IC model, we extended the theory to propose that the amount of inhibition applied to L1 depends on the relative balance activation level of the languages known by a bilingual at baseline. In the proposed model, we assume that language-unbalanced bilinguals have L2 lemmas that are less activated than L1 lemmas by default, therefore they are much more prone to interference from L1. The proposed model further assumes that more inhibition is needed when there is a large difference between L1 and L2 baseline activation levels, namely when L1 activation level is high (i.e., L1 baseline naming latencies are relatively faster) and L2 activation level is low (i.e., L2 baseline naming latencies are relatively slower), as it is in the case of unbalanced bilinguals. In contrast, when there are small differences between L1 and L2 baseline activation levels (the case of balanced bilinguals), less inhibition is needed to overcome the interference.

The first finding of the current study is that the amount of inhibition needed to suppress L1 during L2 use is proportional to the relative balance in the activation levels of L1 and L2 in a given speaker. Because we modeled L1 and L2 activation separately, we found that they both determine the amount of inhibition applied to L1 during the use of L2, but in opposing directions. That is, the higher the activation level of L1 and the lower the activation level of L2, the bigger the L2 after-effect. The observed influence of L1 activation on the magnitude of the L2 after-effect directly confirms the assumption stated in the IC model (Green, 1998) that inhibition is reactive, i.e., the amount of inhibition applied to L1 is proportional to the baseline activation of L1. The fact that the amount of inhibition applied to L1 also depends on the activation level of L2 is, however, less obvious and extends Green's proposal to imply that the reactivity of language also depends on the activation level of the other language(s) the bilingual knows; for instance, in speakers with a high baseline activation level of L2, less inhibition is applied to L1. Importantly, as indicated several times throughout the manuscript, a similar pattern would be expected if the underlying control mechanism were not L1 inhibition but increased L2 activation that persists for some time and carries over to the subsequent L1 use. Under this account, L1-dominant speakers would activate L2 more intensively than more-balanced speakers, and the cost of this increased L2 activation for subsequent L1 use would be more pronounced for them. In other words, also under this alternative account, L1-dominant speakers would demonstrate larger L1 sensitivity to L2 use. We believe that the nature of the control mechanism is unresolvable with the currently available evidence.

Interestingly, there are some reasons to think that L1 and L2 baseline activations are yoked together in the lifetime of a bilingual speaker, such as with accumulated experiences of using L2 (for example, through immersion) bilinguals gradually become faster in accessing L2 at the expense of increasing the speed of accessing L1. This assumption seems in line with studies comparing bilinguals and monolinguals which show that bilinguals are overall slower to access L1 than monolinguals (for evidence from picture-naming tasks, see Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Ivanova & Costa, 2008). Even though no longitudinal study on a bilingual population has directly shown that L2 gets more activated in L2 learners at the expense of L1 activation level, the results reported here allow us to speculate on such a relationship. That is, as an L2 learner becomes more language-balanced, her L2 becomes faster, and her L1 becomes proportionally slower. Such a pattern would be consistent with Kroll and colleagues' recent “desirable difficulty” hypothesis (Bogulski, Bice, & Kroll, 2019), which proposes that L2 learners need to regulate L1, and this regulation creates difficulties in L1 processing but benefits L2 learning.

Fig. 5 visualizes a comparison across participants with different levels of balance between L1 and L2 in our data that support the above hypothesis (see also the figure in Appendix C). The most prominent effect visible in Fig. 5 is the pattern we have already discussed above: the individual language balance modulates the magnitude of the L2 after-effect (i.e., the difference between naming in L1 after L2 versus after L1). However, the figure also reveals a new aspect of the effect, namely that recent use of L2 slows down L1 naming latencies to the same extent, regardless of the speaker's balance between the languages. This observation points to an important and curious feature of bilinguals' language system: the slow-down in L1 access observed after using L2 is likely a downstream effect of the language regulation that helps to achieve a level of L1 that is optimal for the L2 use for a given speaker. In the case of an L1-dominant bilingual, language control during or after L2 use needs to be engaged more intensively in order to achieve optimal level to access L2. In the case of a more balanced bilingual, however, no or very little language control during L2 use is needed because the baseline L1 activation is *already* at the optimal level to allow L2 access. In other words, more-balanced individuals appear to have the baseline level of L1 set lower, which allows them to easily access L2 at any point and prevent them from experiencing the L2 after-effect. However, on the basis of the available evidence it is impossible to determine which exact mechanism drives this reduced L1 activation: it could be constant global L1 inhibition or simply lower baseline activation of L1, which allows the two languages to be kept equally accessible. This lower baseline activation of L1 may be a reflection of lower frequency of use of each

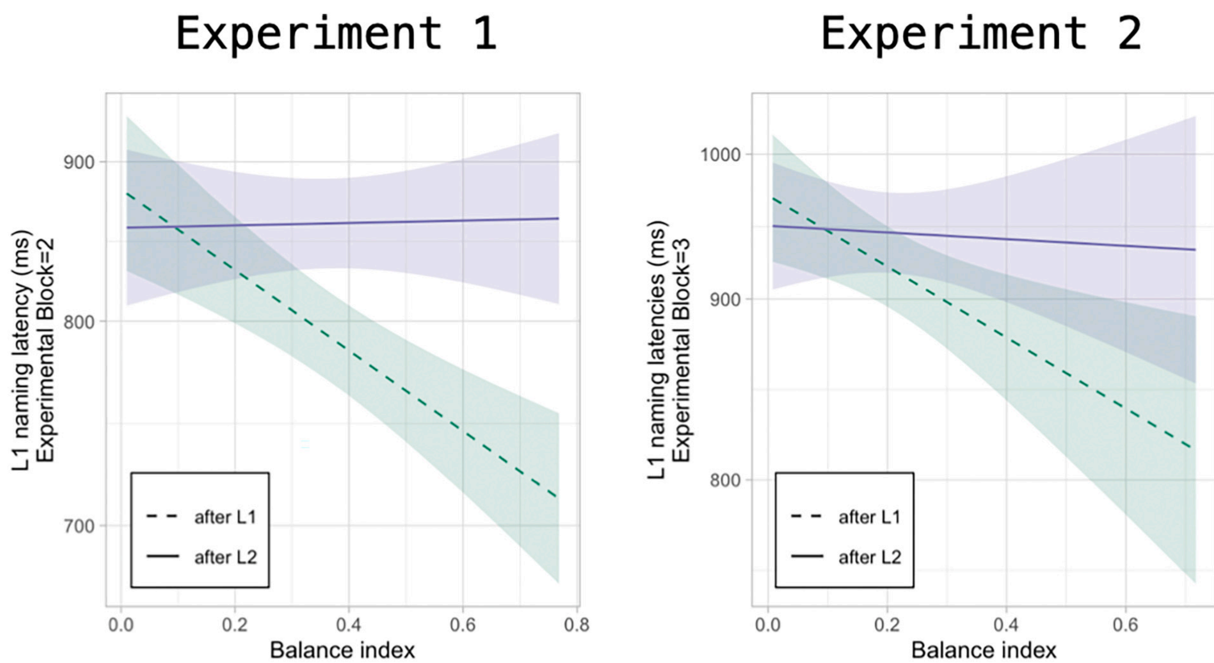

Fig. 5. Predicted naming latencies in L1 after L1 and in L1 after L2 as a function of participants' language balance. The x-axis represents the balance index, which was calculated by subtracting the predicted baseline naming latencies in L1 from the predicted baseline naming latencies in L2. The lower the value, the greater the balance between languages. The y-axis represents L1 naming latencies in the experimental blocks (Experiment 1: Block 2; Experiment 3: Block 3). Note that since language balance is defined as the difference between L1 and L2 baseline naming latencies, the same language balance can be obtained with different combinations of baseline naming latencies in L1 and L2. See Appendix C for the model's predictions for all possible combinations of L1 and L2 activation. For this graph, we selected a cross-section through these combinations such that L1 and L2 activations are in a constant trade-off (e.g., when the value of L1 baseline naming latencies decreases, the value of L2 baseline naming latencies increases by the same amount). Such a cross-section corresponds to an idealized bilingual whose baseline L2 activation increases at the expense of L1 baseline activation.

Table 5
Dynamics of language activation in L1-dominants and balanced bilinguals due to short-term L2 use.

	L1 activation **AT BASELINE**	Presence of language control **DURING L2 USE**	L1 activation **AFTER L2 USE**
L1-dominant bilinguals	High	YES	Different → Lower
Balanced bilinguals	Low	NO	Same → Low

language (in line with the frequency lag hypothesis proposed by Gollan, Montoya, Cera, & Sandoval, 2008; Gollan et al., 2011) or simply a result of constant co-activation of both languages (in line with the selection-by-activation proposal of Blanco-Elorrieta & Caramazza, 2021). In any case, the enduring down-regulation of L1 activation or up-regulation of L2 activation allows balanced bilinguals to use their two languages more efficiently on an everyday basis by better adjusting their language system to the requirements of the bilingual environment. Table 5 schematically presents differences in the assumed processes at baseline during L2 use and after L2 use for bilinguals with a different language balance. L1-dominant bilinguals are susceptible to L2 use, which leads to a decrease in L1 activation level and/or an increase in L2 activation level. This suggests that in these bilinguals the use of L2 engages some form of language control mechanism. More-balanced bilinguals, on the other hand, are not susceptible to L2 use: their L1 activation level is at the same (low) level both after and before L2 use (see Table 5). This suggests that for balanced bilinguals L2 use does not trigger or require a special language control mechanism; however, some regulatory process is responsible for their lower L1 activation level at baseline. The possibilities include permanent L1 inhibition (an active process), lower frequency of L1 use (passive process), or constant co-activation of both languages (passive process). However, the available evidence is agnostic as to the exact nature of the process.[7]

The third important theoretical insight of the current study relates to the ongoing debate over the scope of language control, which refers to the issue of how broadly language control affects the language system, that is, whether it affects the language as a whole, a specific category of items, or translation equivalents. A wide scope of language control is usually referred to as “whole-language” (Van Assche et al., 2013) or “global” control (e.g., Abutalebi & Green, 2016), while a narrow scope of language control relates to translation equivalents and is referred to as “item-level” (Van Assche et al., 2013) or “local” control (e.g., Abutalebi & Green, 2016). Based on their findings, Van Assche and colleagues suggested that the ability to exert global control may only develop for some types of bilinguals, whereas item-level control is present in all speakers. Based on the pattern observed in a series of verbal fluency tasks, the authors claimed that global language control was present in early Chinese-English bilinguals living in the US, but it was not present in late Dutch-English bilinguals in the Netherlands. This finding led Van Assche and colleagues to speculate that global language control is more likely to occur when bilinguals 1) speak structurally distinct languages; 2) acquired both languages in early childhood and reversed language dominance in the course of their lives; and 3) live in the L2 context. However, Degani et al. (2020) recently tested the two scopes of language control (i.e., global language level and item-level) in a group of late

[7] As a side note, we would like to propose that it is exactly the comparable accessibility of L1 and L2 words in balanced bilinguals that might drive the higher rate of voluntary switches that is observed in balanced vs. unbalanced bilinguals in a voluntary language switching task (see Gollan and Ferreira, 2009 for the results).

speakers of two similar languages (Arabic and Hebrew) and still observed global language control in these speakers. The authors concluded that – contrary to the speculations of Van Assche et al. (2013) – global language control is not limited to bilinguals speaking two languages as structurally distinct as Chinese and English; therefore, the language dissimilarity and the early age of L2 acquisition are unlikely to be crucial factors that drive the engagement of global language control. Because the Arabic-Hebrew participants were partially immersed in L2 and frequently changed languages in their daily lives, Degani and colleagues suggested that the factors that engage global language control are probably L2 immersion and frequent changes of languages. In contrast to this conclusion, our data indicate that global language control can be systematically observed even in bilinguals whose life environment primarily involves L1. Importantly, although both our study and the one by Degani and colleagues employed a similar task to measure language control (i.e., picture naming), Degani and colleagues derived their balance measures on the basis of only error rates, whereas we used an index based on response latencies. Moreover, unlike the investigation by Degani and colleagues, our measurement of global language control, especially in Experiment 2, was free from additional influences of item-level control as we did not repeat any items. Furthermore, we found that the magnitude of the L2 after-effect was not affected by the different picture types (actions vs. objects), meaning that the effect is not restricted to certain word categories. All in all, we provide the first clear evidence that the use of a second language leads to global language control, even in speakers who live in an L1-dominant environment.[8] Based on the empirical evidence discussed above, we propose that our data reveal language control on the global language level. Whether similar effects can be observed on the item-level remains an open empirical question.

Finally, our results bring some new insights into the time-course of L2 after-effects. Previous studies showed that the effect is relatively long lasting (Branzi et al., 2014; Misra, Guo, Bobb, & Kroll, 2012; Wodniecka et al., 2020): it was observed even after approximately 5 min (Branzi et al., 2014; Wodniecka et al., 2020) or up to two blocks of naming in L1 (Misra et al., 2012) after L2 use. This suggests that activation of L1 lemmas remains low for quite some time after a bilingual speaker returns to using her L1. We performed a follow up analysis which revealed that the magnitude of the L2 after effect decays over time. The slow-down of L1 naming latencies after recent L2 use was greater right after using L2 and decreased with time (see Table 4 and Fig. 4). This suggests that the L2 after-effect is temporary and diminishes even within a few minutes. Consistent naming in L1 helps participants to recover their ability to retrieve L1 words with an ease similar to when they retrieve L1 without prior exposure to L2.

5.2. *Methodological implications for studying language inhibition in production*

We proposed and effectively used two novel indices that seem to be more ecologically valid than those used in previous research. The first index assessed a speaker's balance between the two languages known. Inspired by a solution used by Meuter and Allport (1999), we proposed that this balance can be conceptualized as the difference in the activation of L1 and L2, and it is operationalized by comparing picture-naming latencies in L1 and L2. The proposed index of balance seems to be more directly related to the levels of language activation (both at baseline and after the recent language use) than the indices based on proficiency that were used in most previous studies.[9]

Second, we were able to demonstrate that the L2 after-effect that was derived from a blocked picture-naming paradigm can be successfully used as an index of language control. It proved to be a robust measure which survives statistical adjustments for possible confounds in the experimental design as well as in daily variations in participants' baseline speed of processing. While controlling for these covariates is certainly important (see Experiment 2 for details), we propose that the L2 after-effect can be used as an effective index of individual strength of language control in bilingual speech production. This index is sensitive enough to detect variability across speakers and to allow investigation of individual differences. Moreover, the consequences of the language control applied over L1 during L2 use (reflected by the L2 after-effect) are relatively long-lasting. As such, the L2 after-effect better reflects processes that take place during natural language use than the language-switching paradigm that has been used in most studies so far and which is arguably rather artificial. The L2 after-effect is also free from some other limitations of indices derived from a language-switching paradigm, e.g., assessment of inhibition on a trial-to-trial basis (as such tapping the transient rather than long-lasting process); intense item repetition, which makes inferences about global language control impossible; or frequent goal-switching demand, which complicates isolation of language control processes from the more domain-general demands imposed by the switching requirement.

It should be noted that the design used in the reported experiments did not allow us to investigate a possible mechanism of L2 regulation that occurs during L1 production. Based on the model, such language control would indeed be expected in bilinguals with more dominant L2. However, here we intentionally limited our investigation to L1 to avoid contamination of the results by the previously observed effects of language testing order (for details, see, e.g., Guo et al., 2013; Van Assche et al., 2013). Future research should definitely explore the possibility of reverse language effects.

6. Conclusions

In the current paper, we explored whether individual variation in the relative balance between two languages impacts the degree of engagement of control during language production. We proposed a causal model that explicitly outlines the assumptions of the original inhibitory account with respect to the baseline activation level of a bilingual's languages. In the reported experiments, we assessed the degree of engagement of global language control by using the "L2 after-effect", an index derived from a blocked picture-naming task. To investigate individual variation in languages' activation, we used a continuous measure

[8] Previous studies by Misra et al. (2012) and Guo et al. (2013) suggested that inhibition in unbalanced bilinguals living in an L1 environment affects the dominant language as a whole, but the reported effects were confounded by repetition of the same pictures across experimental blocks (i.e., most likely involving item-level inhibition). In a different design, Branzi et al. (2014) distinguished between new and repeated pictures. They reported similar behavioral and ERP responses for new and repeated items, which they took as evidence that the L2 after-effect was observed on both the whole-language and the item-level levels, despite engaging different brain regions (see Branzi et al., 2016 for fMRI evidence). However, recent research has shown that mixing repeated and new items in the design can mask the measurement of inhibition engaged at the whole-language level (Stasenko et al., 2021). Importantly, the design of our Experiment 2 is free from similar impurities of the inhibition index.

[9] To explore the possibility brought up by one of the Reviewers that our measure of relative language balance, which relies only on picture naming latencies and not accuracy, may lead to misidentification of balanced bilinguals as those who were quick but not necessarily accurate, we ran additional analyses that are available as supplementary material from OSF (www.osf.io/27y4p). In these analyses, we explored whether the items that were excluded from the response time calculation in L2 were harder to name. To this end, we compared L2 accuracy in participants from the two most extreme groups: the most-balanced bilinguals and least-balanced bilinguals. We found that the items excluded from calculation of response latencies in fact had similar difficulty (as measured by the lexical frequency) for both the most and the least balanced bilinguals. As such, excluding items which participants failed to name did not bias our index of balance between languages.

of language balance that was based on naming latencies in the picture-naming task. Importantly, we kept many other potential confounding factors constant.

Our results demonstrate that individual differences in speakers' language-activation levels indeed affect the magnitude of the L2 after-effect. The more language-balanced bilinguals are, the lower the involvement of global control caused by L2 production. In addition, we have also provided some tentative evidence that language-balanced bilinguals appear to have a lower baseline activation level of L1 than unbalanced speakers. Future research should attempt to establish the source of this overall lower L1 activation in more balanced speakers. It is possible that the overall slower L1 performance in balanced speakers is driven by a different mechanism than language control applied due to L2 production. The observed dynamics in the language system may rely on multiple mechanisms of language control that are applied at different time points of language use.

Authors' note

All raw data and the code needed to reproduce all results is available at https://osf.io/9gv7u/.

Credits

- **Conceptualization** (Ideas; formulation or evolution of overarching research goals and aims): AC, JS, ZW
- **Data curation** (Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use): AC, JS, AW
- **Formal analysis** (Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data): AC, JS, AW
- **Funding acquisition** (Acquisition of the financial support for the project leading to this publication): ZW
- **Investigation** (Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection): AW
- **Methodology** (Development or design of methodology; creation of models): AC, JS, AW, ZW
- **Project administration** (Management and coordination responsibility for the research activity planning and execution): AC, ZW
- **Resources** (Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools): ZW
- **Software** (Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components): AC, JS, AW
- **Supervision** (Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team): ZW
- **Validation** (Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs): AC, JS, AW
- **Visualization** (Preparation, creation and/or presentation of the published work, specifically visualization/data presentation): AC, JS, AW
- **Writing - original draft** (Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation): AC, JS, AW, ZW
- **Writing - review and editing** (Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages): AC, JS, AW, ZW

Declaration of Competing Interest

none.

Acknowledgements

This research was possible thanks to a National Science Centre grant (2015/18/E/HS6/00428) awarded to Z.W. Data collection for Experiment 1 was partially supported by National Science Centre Poland grant N106 366 340 awarded to Z.W and the Foundatation for Polish Science subsidy awarded to Z.W.. The authors gratefully acknowledge the help of all members of our Psychology of Language and Bilingualism Laboratory, LangUsta, who contributed to the research project by collecting and coding the data, and Michael Timberlake for proofreading. This article is available to the public with the financial support of the Jagiellonian University as part of the Excellence Initiative - Research University program (at the Faculty of Philosophy).

Supplementary data (Appendices)

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105169.

References

Abdel Rahman, & Melinger. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes, 24*(5), 713–734. https://doi.org/10.1080/01690960802597250

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, Kliegl, Vasishth, & Baayen. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967. arXiv preprint*.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(1–7), 2014.

Blanco-Elorrieta, E., & Caramazza, A. (2021). A common selection mechanism at each linguistic level in bilingual and monolingual language production. *Cognition, 104625*.

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

Bogulski, C. A., Bice, K., & Kroll, J. F. (2019). Bilingualism as a desirable difficulty: Advantages in word learning depend on regulation of the dominant language. *Bilingualism: Language and Cognition, 22*(5), 1052–1067.

Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2016). Language control in bilinguals: Monitoring and response selection. *Cerebral Cortex, 26*(6), 2367–2380.

Branzi, F. M., Martin, C. D., Abutalebi, J., & Costa, A. (2014). The after-effects of bilingual language production. *Neuropsychologia, 52*, 102–116.

Branzi, F. M., Martin, C. D., Carreiras, M., & Paz-Alonso, P. M. (2020). Functional connectivity reveals dissociable ventrolateral prefrontal mechanisms for the control of multilingual word retrieval. *Human Brain Mapping, 41*(1), 80–94.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.

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

Costa, Caramazza, & Sebastian-Galles. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

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

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1057.

Costa, Strijkers, Martin, & Thierry. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences, 106*(50), 21442–21446. https://doi.org/10.1073/pnas.0908921106

Declerck, M., Kleinman, D., & Gollan, T. H. (2020). Which bilinguals reverse language dominance and why? *Cognition, 204*, Article 104384.

Declerck, M., & Philipp, A. M. (2015). A sentence to remember: Instructed language switching in sentence production. *Cognition, 137*, 166–173.

Declerck, M., & Philipp, A. M. (2018). Is inhibition implemented during bilingual production and comprehension? n-2 language repetition costs unchained. *Language, Cognition and Neuroscience, 33*(5), 608–617.

Declerck, M., Thoma, A. M., Koch, I., & Philipp, A. M. (2015). Highly proficient bilinguals implement inhibition: Evidence from n-2 language repetition costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(6), 1911.

Degani, T., Kreiner, H., Ataria, H., & Khateeb, F. (2020). The impact of brief exposure to the second language on native language production: Global or item specific? *Applied PsychoLinguistics, 41*(1), 153–183.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. *Handbook of bilingualism: Psycholinguistic approaches, 178*, 201.

Filippi, R., Karaminis, T., & Thomas, M. S. (2014). Language switching in bilingual production: Empirical data and computational modelling. *Bilingualism: Language and Cognition, 17*(2), 294–315.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116–124.

Gade, M., Declerck, M., Philipp, A. M., Rey-Mermet, A., & Koch, I. (2021). Assessing the evidence for asymmetrical switch costs and reversed language dominance effects–a meta-analysis. *Journal of Cognition, 4*(1).

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*, 787–814.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition, 33*(7), 1220–1234.

Gollan, T. H., Schotter, E. R., Gomez, J., Murillo, M., & Rayner, K. (2014). Multiple levels of bilingual language control: Evidence from language intrusions in reading aloud. *Psychological Science, 25*(2), 585-595.1.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General, 140*(2), 186.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism (Cambridge, England), 15*(3), 594.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67–81.

Green, & Abutalebi. (2013). Language control in bilinguals: the adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5). https://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (2001). The bilingual's language modes. *One mind, two languages: Bilingual language processing, 122*.

Guo, T., Liu, F., Chen, B., & Li, S. (2013). Inhibition of non-target languages in multilingual word production: Evidence from Uighur–Chinese–English trilinguals. *Acta Psychologica, 143*(3), 277–283.

Haman, E., Łuniewska, M., Hansen, P., Simonsen, H. G., Chiat, S., Bjekić, J., … Gagarina, N. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from cross-linguistic lexical tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics, 31*(11−12), 818–843.

Haman, Łuniewska, & Pomiechowska. (2015). Designing cross-linguistic lexical tasks (CLTs) for bilingual preschool children. *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 196–240). Multilingual Matters.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100*(3), 464–482.

Iniesta, A., Paolieri, D., Serrano, F., & Bajo, M. T. (2021). *Bilingual writing coactivation: Lexical and sublexical processing in a word dictation task. Bilingualism: Language and Cognition*.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica, 127*(2), 277–288.

Kleinman, D., & Gollan, T. H. (2018). Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition, 173*, 115–132.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism, 9*(2), 119.

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

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods, 38*(2), 202–210.

Lu, A., Wang, L., Guo, Y., Zeng, J., Zheng, D., Wang, X., … Wang, R. (2019). The roles of relative linguistic proficiency and modality switching in language switch cost: Evidence from Chinese visual unimodal and bimodal bilinguals. *Journal of Psycholinguistic Research, 48*(1), 1–18.

Mandera, P., Keuleers, E., Wodniecka, Z., & Brysbaert, M. (2015). Subtlex-pl: Subtitle-based word frequency estimates for polish. *Behavior Research Methods, 47*(2), 471–483.

Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied PsychoLinguistics, 24*(2), 173.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25–40.

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

Oppenheim, Dell, & Schwartz. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*(2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007

Philipp, A. M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching: evidence from switching language-defined response sets. *European Journal of Cognitive Psychology, 19*(3), 395–416.

R Core Team. (2020). A language and environment for statistical computing. *Foundation for Statistical Computing*.

Rabbitt, P., Osman, P., Moore, B., & Stollery, B. (2001). There are stable individual differences in performance variability, both from moment to moment and from day to day. *The Quarterly Journal of Experimental Psychology Section A, 54*(4), 981–1003.

Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods, 49*(5), 1864–1881.

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

Schwieter, J. W., & Sunderman, G. (2008). Language switching in bilingual speech production: In search of the language-specific selection mechanism. *The Mental Lexicon, 3*(2), 214–238.

Stasenko, A., Kleinman, D., & Gollan, T. H. (2021). Older bilinguals reverse language dominance less than younger bilinguals: Evidence for the inhibitory deficit hypothesis. *Psychology and Aging, 36*(7), 806.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., … Federmeier, K. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51*(2), 247–250.

Thierry, & Wu. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences, 104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

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

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

Wodniecka, Z., Szewczyk, J., Kałamała, P., Mandera, P., & Durlik, J. (2020). When a second language hits a native language. What ERPs (do and do not) tell us about language retrieval difficulty in bilingual language production. *Neuropsychologia, 141*, Article 107390.

Wolna, A., Szewczyk, J., Casado, A., & Wodniecka, Z. (2021). *Now you see, now you don't. How does the experimental design impact the measurement of L2 after-effect in L1 production?* [manuscript in preparation]. Institute of Psychology, Jagiellonian University.