Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-022-02168-z

BRIEF REPORT

The temporal dynamics of bilingual language control

Iva Ivanova[1] · Andrea Seanez[1] · Mackenzie Cochran[1] · Daniel Kleinman[2]

Accepted: 10 August 2022

Abstract
Bilinguals need to control interference from the nontarget language, to avoid saying words in the wrong language. This study investigates how often bilinguals apply such control in a dual-language mode, when speaking one language after the other when the two languages cannot be used interchangeably: over and over (every time they say a word), or only once (the first time they use a word or language after a language switch). Three groups of Spanish-English bilinguals named pictures first in their dominant, then in their nondominant, and then again in their dominant language; a fourth control group of bilinguals named pictures in their dominant language throughout. The study targeted language control aftereffects on the dominant language after nondominant naming, typically assumed to reflect recovery from previously applied inhibition. If the dominant language is inhibited every time a nondominant word is produced, subsequent dominant-language naming latencies should increase in proportion to the number of pictures previously named in the nondominant language. We found, however, that the number of nondominant picture-naming trials did not affect subsequent naming latencies in the dominant language, despite ample statistical power to detect such effects if they existed. The results suggest that, in a dual-language mode, bilingual (inhibitory) control is applied over a word's translation upon the word's first mention but not over and over with subsequent repetitions. This conclusion holds true equally for inhibitory and non-inhibitory language control mechanisms.

Keywords Bilingualism · Inhibitory control · Picture naming · Blocked naming · Local inhibition

Introduction

Cognitive and neural adaptations are thought to arise from bilinguals' practice of using their two languages across different situations (Abutalebi et al., 2012; Bialystok et al., 2008; Green & Abutalebi, 2013; but see Paap & Greenberg, 2013). To understand such adaptations, we need to understand the processing dynamics of bilingual language use. Among them are mechanisms that allow bilinguals to control when to speak which language and to avoid using the wrong language (*bilingual language control*). Here, we look into the temporal dynamics of such control, asking how frequently is it exercised: occasionally, or upon saying every word?

In a dual-language mode (in the sense discussed by Green & Abutalebi, 2013), bilinguals use their two languages in close succession in situations in which the languages cannot be used interchangeably (e.g., because one language will not be understood). Wrong-language errors are thus undesirable in such contexts, and, accordingly, are very rarely attested in bilingual production (Gollan et al., 2011; Poulisse, 1999). But this is puzzling given that, in a dual-language mode, both languages are highly activated and hence highly likely to interfere with each other.

The most established theory of how bilinguals prevent other-language interference is the Inhibitory Control Model (Green, 1998). In the model, the attentional system activates "language task schemas" (mental step sequences of different language activities, such as "Speak Language X"), which in turn activate that language's lexico-semantic representations, and inhibit nontarget-language ones (*local inhibition*). An activated language schema can also inhibit a currently irrelevant schema in its entirety – that is, a nontarget language as a whole (*global inhibition*). The existence of local and global loci of control is supported by behavioral (Branzi et al., 2014; Degani et al., 2020; Guo et al., 2011; Kreiner & Degani, 2015) and neuroimaging evidence (Guo et al., 2011; Rossi et al., 2018; Wodniecka et al., 2020).

✉ Iva Ivanova
imivanova@utep.edu

1 Department of Psychology, University of Texas at El Paso, 500 W. University Ave, El Paso, TX 79968, USA

2 Haskins Laboratories, New Haven, CT, USA

Published online: 12 September 2022

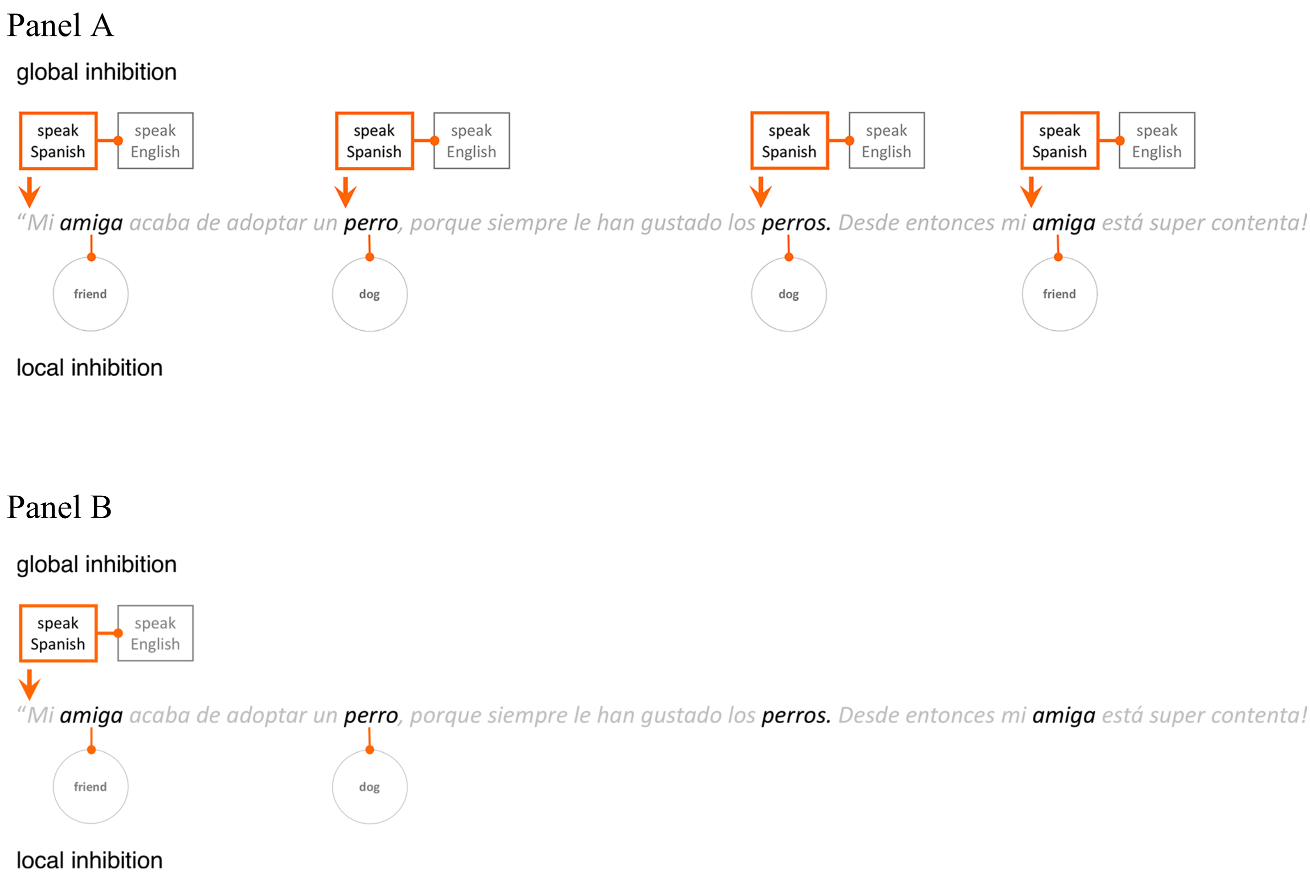

Fig. 1 Schematic representation of how two hypotheses about the timing of bilingual inhibitory control apply to the words *amiga* (Spanish *friend*) and *perro* (Spanish *dog*) in an example utterance. **Panel A:** Over-and-over hypothesis. **Panel B:** Once-only hypothesis. The sentence translates from Spanish as *My friend just bought a dog, because she has always liked dogs. Since then my friend is very happy!* Inhibitory processes are assumed to operate over all other (in Panel B, unique) words in the sentence as well, but those instances of inhibition are not depicted because of the limited space

A core feature of the Inhibitory Control Model is that inhibition is proportional to the strength of the language it acts on. Hence, a dominant language is inhibited more strongly (because it is more likely to interfere) than vice versa (because a weaker nondominant language is not very likely to interfere). Another assumption is that inhibited representations do not become available for production immediately after that language is switched into (because of a passive recovery period; see Wodniecka et al., 2020). This creates a behavioral signature of inhibitory control: Retrieving representations post-inhibition should be harder, and take longer, than pre-inhibition.

The two assumptions predict an asymmetry more strongly disfavoring the dominant language in situations of language mixing. Such an asymmetry has been amply attested in tasks including mixed-language paragraph reading (Gollan et al., 2014), verbal fluency (Van Assche et al., 2013), and mixed-language picture naming (Christoffels et al., 2007; Guo et al., 2011; Meuter & Allport, 1999; see reviews in Bobb & Wodniecka, 2013; Declerck & Philipp, 2015), making it useful to investigate properties of bilingual language control.

The property of interest here is the *time course* of language control. We contrast two broad possibilities. The first possibility, the *Over-and-over hypothesis*, is that inhibition is applied with each repetition of a word (Fig. 1, Panel A). For global inhibition, this would mean that all words from the nontarget language are re-inhibited upon uttering each successive word in the target language. For local inhibition, this would mean that a word's translation equivalent is re-inhibited every time that word is mentioned. The result of either would be that the amount of inhibition of the nontarget language would progressively increase – and be progressively harder to recuperate from – the longer a language is spoken. If only local and not global inhibition is present, only repeated words' translation equivalents would accrue additional inhibition.

Support for the Over-and-over hypothesis was demonstrated by Kleinman and Gollan (2018). These authors analyzed picture-naming latencies from a language-switching task ($N = 416$) and found that the aftereffects of both local and global inhibition accumulated across an entire block of trials. Their results suggest that each time a word is produced in one language, some inhibition is applied both over its translation equivalent and (when naming in the nondominant language) over the whole dominant language.

Of note, these conclusions were drawn from bilinguals' performance in mixed-language blocks comprising a small number of unique stimuli (nine pictures per participant) and frequent language switches (on 50% of trials). Nevertheless, Kleinman and Gollan extrapolated from this finding to situations without frequent switching (such as switching between whole blocks, not individual trials), yielding their prediction that "performance in a dominant-only block should be worse when the preceding nondominant-only block was longer, as more nondominant trials means more opportunities to generate inhibition for the dominant language" (Kleinman & Gollan, 2018, p. 123). We test this prediction here, and propose that this may not be the case: Activating alternative responses to the same stimulus in close succession during trial-level switching may evoke stronger competition, requiring stronger inhibitory control and additional control processes than in situations with less frequent language switching. Of note, in an event-related-potentials (ERPs) trial-level language-switching task with varying run lengths of same-language trials, Zheng et al. (2018, 2020) found results inconsistent with the Over-and-over hypothesis (though also inconsistent with the alternative below).[1]

We consider here an alternative possibility for such situations, that language control is applied only once (*Once-only hypothesis*, Fig. 1, Panel B). Globally, all nontarget lexical representations would be inhibited upon a language switch but not afterwards. Locally, each lexical representation would be inhibited *upon* mention of its translation equivalent (e.g., saying the word *perro* would inhibit the word *dog*), but not reinhibited if the target word is mentioned again. The result of applying both global and local inhibition only once would be that, regardless of how many times words are repeated, the amount of inhibition applied over the other language would remain constant.

The present study

We investigate here the temporal dynamics of language control in longer periods without switching, contrasting two hypotheses (framed around the Inhibitory Control Model but in essence theory-neutral). On the Over-and-over hypothesis, (local or global) inhibition of the nontarget language is applied with every single retrieval of a word. On the Once-only hypothesis, inhibition of nontarget language words is applied only once – globally, upon beginning to speak in the target (here, nondominant) language, and locally, upon each nondominant-language word's first mention. We studied this issue using a blocked naming paradigm, targeting adverse aftereffects on the dominant language from prior nondominant naming, to see if their size would differ when prior nondominant naming differed in length. Three groups of Spanish-English bilinguals dominant in English named pictures first in their dominant language (Phase 1), then in their nondominant language (Phase 2), then again in their dominant language (Phase 3). The three groups differed in the length of nondominant naming (long, medium, and short, operationalized as the number of repetitions of target pictures). An additional group of bilinguals named pictures in English only (including in Phase 2), as a same-language baseline.

We expect that dominant-after-nondominant naming would be slower relative to dominant-before-nondominant naming. Further, if the dominant language is inhibited only once, upon a language switch or upon a word's first mention (Once-only hypothesis), the number of nondominant repetitions should not affect the speed of subsequent dominant naming. However, if inhibition of the dominant language is applied repeatedly (Over-and-over hypothesis), more nondominant repetitions should cause greater slowing of subsequent dominant naming.

Additionally, our experiment included three different item sets (with assignment of pictures counterbalanced across item sets): one repeated in all three phases, one repeated in Phases 2–3, and one introduced in Phase 3. We aimed to separate global effects of language control (affecting all picture naming) from local effects (affecting only pictures that

[1] These studies employed a trial-level language-switching task with varying run lengths of same-language trials. In both studies, switching into naming in the first-and-dominant language incurred a greater cost after shorter (two to three trials) than after longer nondominant-language runs (five to six trials). Further, in event-related potential (ERP) analyses, an N2 component at the switch (assumed at more frontal sites to reflect inhibitory control) was larger after a short same-language run than after a long same-language run. These results are inconsistent with the Over-and-over hypothesis because repeated inhibition over dominant names during nondominant naming would predict more inhibition (thus larger costs and N2) after long than after short nondominant runs. To explain their findings, Zheng and colleagues proposed that speaking a nondominant language soon after switching to it requires considerable top-down control (making it harder to overcome such control if a further switch to the dominant language happens at this point). The need to exercise such control diminishes over time when bottom-up mechanisms take over, and hence a switch to the dominant language at this point is less costly. Conversely, speaking a dominant language does not require much top-down control even after a switch, and therefore switching into the nondominant language has an equivalent cost after a short and a long dominant-language naming run length.

were previously named in a different language), and to test a hypothesis derived from the Inhibitory Control Model that more recently activated dominant-language words would be inhibited more strongly than less recently activated ones.

Method

Participants One hundred and thirty-six Spanish-English bilingual undergraduates (mean age 20.4 years, *SD* = 3.6 years) from the University of Texas at El Paso participated for course credit. Upon arrival, 104 bilinguals were randomly assigned into a Long (*N* = 34), Medium (*N* = 35), or Short nondominant naming group (*N* = 36) that differed in the length of Spanish naming in the second phase of the experiment (explained below). The remaining 32 bilinguals formed the Dominant-only group (for whom there was no Spanish naming) and were tested in a separate session at a later time. The number of participants (set before beginning data collection at 36 participants per group) was chosen as larger than the average of typically used sample sizes in picture-naming studies of bilingual language control (*N* = 12–48; Costa & Santesteban, 2004; Christoffels et al., 2016; Declerck et al., 2012; Li & Gollan, 2018; Kirk et al., 2018; Mosca & Clahsen, 2016; Peeters et al., 2014; Philipp et al., 2007; Verhoef et al., 2009; the design most similar to this one had 18 participants per group, Branzi et al., 2014). This sample size yielded very high statistical power to detect effects of the size that would be expected based on the results of Kleinman and Gollan (2018), as described below and in more detail in the Online Supplementary Material.

The language history characteristics of participants in each of the four groups is reported in Table 1. Dominance in English was a criterion stated on the participant recruitment platform. Participants who did not fulfill this criterion after an objective proficiency test administered after the main experiment were replaced. Objective proficiency in both languages was measured with the Multilingual Naming Test (MINT; Gollan et al., 2012), an untimed productive vocabulary test with 68 pictures of progressive difficulty. Bilinguals in the Long, Medium, or Short nondominant naming groups were considered English-dominant if they named at least four more pictures in English than in Spanish. This number (four) was chosen to satisfy two competing constraints: It needed to be high enough to ensure that participants were actually English-dominant, but low enough to keep necessary participant exclusions to a minimum (see below). The data of five balanced bilinguals (who named on average only two more pictures in English than in Spanish) were retained in the Dominant-only group, who did not speak Spanish in the experiment, because their replacement was impossible during the COVID-19 pandemic and because we aimed to keep group numbers comparable. (Analyses without these five bilinguals produced an identical pattern of results.)

On average, bilinguals named 17 more pictures in English than in Spanish on the Multilingual Naming Test (the difference of English minus Spanish names ranged between 4 and 42 for the Nondominant naming groups, and between −1 and 46 for the Dominant-only group). In addition, bilinguals completed a language history questionnaire. A one-way ANOVA analysis of each language history variable showed that bilinguals in the four groups did not significantly differ on any language characteristic (with the exception of a marginal difference for self-rated English proficiency, which, however, was not reflected in the objective proficiency scores; see Table 1).

The analyses excluded an additional 33 participants whose Multilingual Naming Test scores did not match the criteria specified above. Of these, 17 were balanced (English-Spanish difference: mean = 0.6, range = [−2, 3]) and 16 were Spanish-dominant (English-Spanish difference: mean = -10, range = [−17, −3]). In addition, one participant could not complete the experiment because of technical difficulties, one participant was administered the wrong group for one part of the experiment, one participant provided pilot data, one participant produced the indefinite determiner "a" before each name, and nine participants' Spanish knowledge was not sufficient to complete the task; for eight of them, the experiment was discontinued before the end. In total, 46 additional participants were excluded from analysis.

Materials Forty-eight line drawings were selected from the International Picture Naming database (Bates et al., 2003). Their names were divided into three different lists of 16 items each that were matched on frequency, length in phonemes (see Table 2), and, as much as possible, semantic category. Frequency-per-million values were obtained from the movie-subtitles corpora SUBTLEX-US for American English (Brysbaert & New, 2009; http://expsy.ugent.be/subtlexus/), and SUBTLEX-ESP for Spanish (Cuetos et al., 2011; http://crr.ugent.be/archives/679). (Note that the SUBTLEX-ESP database largely reflects Castilian Spanish use and is thus only our best approximation for the Mexican/border-Spanish-speaking population tested here.) An additional set of 32 pictures to be named in English only were fillers in Phase 1. They were selected in the same way, and no differences were detected between the average frequency and length values of their English names and the average values of the English names of the target items (all *ps* > .25). All picture names are provided in the Appendix.

Design The pictures were presented in three phases, to be named in English (in Phase 1), then in Spanish (in Phase 2), then again in English (in Phase 3) by the three Nondominant naming groups, and in English throughout by the

Table 1 Participants' language history characteristics

	Long nondominant naming group ($N = 34$)	Medium nondominant naming group ($N = 35$)	Short nondominant naming group ($N = 36$)	Dominant-only group ($N = 32$)	*F, p*
Age of first exposure in years[b]					
English	3.4 (2.5)	[a]4.2 (2.4)	[a]4.3 (2.4)	4.0 (2.7)	$F(3,130) = 0.85$ $p = .47$
Spanish	1.8 (2.2)	[a]1.9 (2.3)	[a]0.8 (2.2)	2.8 (3.5)	$F(3,130) = 1.27$ $p = .29$
Other	15.2 (2.2), $N = 15$	[a]14.0 (2.7), $N = 16$	[a]12.7 (5.2), $N = 14$	12.0 (5.7), $N = 12$	not compared
% daily use now					
English	69% (20%)	[a]70% (16%)	[a]64% (16%)	66% (16%)	$F(3,130) = 0.71$ $p = .55$
Spanish	34% (20%)	[a]30% (17%)	[a]35% (16%)	33% (16%)	$F(3,130) = 0.29$ $p = .83$
% daily use as a child					
English	54% (24%)	[a]50% (28%)	[a]54% (23%)	50% (22%)	$F(3,130) = 0.34$ $p = .80$
Spanish	48% (25%)	[a]50% (28%)	[a]46% (23%)	50% (23%)	$F(3,130) = 0.22$ $p = .89$
Self-rated proficiency[c] (1 = very basic; 10 = native)					
English	9.6 (0.8)	[a]9.3 (1.2)	[a]9.7 (0.5)	[a]9.1 (1.1)	$F(3,130) = 2.30$ $p = .08$
Spanish	6.7 (2.2)	[a]6.6 (2.1)	[a]7.5 (1.5)	[a]6.7 (2.2)	$F(3,130) = 1.56$ $p = .20$
Other	2.7 (2.8), $N = 16$	1.5 (0.7), $N = 12$	2.3 (1.4), $N = 12$	2.9 (2.4), $N = 11$	not compared
Code-switching frequency (1 = never; 6 = a lot or sometimes even constantly)					
	4.2 (1.5)	4.1 (1.6)	4.4 (1.5)	4.3 (1.5)	$F(3,130) = 0.18$ $p = .91$
Productive vocabulary (MINT, of 68)					
English	62 (3)	62 (3)	62 (3)	61 (3)	$F(3,132) = 1.40$ $p = .25$
Spanish	43 (12)	47 (10)	45 (11)	44 (12)	$F(3,132) = 0.70$ $p = .55$

Standard deviations are provided in parentheses

[a]The language history questionnaires of two participants (one in the Medium and one in the Short Nondominant naming group) were missing

[b]For four participants who indicated age of first exposure as a grade level, we applied the following conversion: kindergarten = 5 years; first grade = 6.5 years; second grade = 7.5 years; (beginning of) high school = 14.5

[c]Three participants (one in each of the Long, Medium, and Short nondominant naming groups) seemed to have applied the scale endpoints in reverse (they reported speaking only English and Spanish on an average day, but had self-rated their English and Spanish proficiency as 5 or lower (most ratings were 1–3). We converted these ratings to 6–10, respectively

Table 2 Picture name characteristics

	List 1	List 2	List 3	t-tests
Frequency				
English	67.5 (72.7)	72.6 (68.5)	113.1 (126.5)	all *ps* > .8
Spanish	64.2 (65.7)	69.0 (76.7)	162.2 (330.9)	all *ps* > .7
t-tests	$p = .49$	$p = .61$	$p = .38$	
Length in phonemes				
English	4.4 (1.6)	4.4 (1.8)	4.5 (1.5)	all *ps* > .2
Spanish	5.1 (1.5)	5.3 (1.6)	5.1 (1.2)	all *ps* > .2
t-tests	$p = .13$	$p = .15$	$p = .15$	

Dominant-only group (including in Phase 2; see Fig. 2). There were three different picture sets: Set A was presented in all three phases, Set B was presented in Phases 2 and 3, and Set C was presented in Phase 3 only (and thus contained pictures that had not been named previously). The three picture lists were counterbalanced across the three sets, such that in different experimental versions, each picture list appeared in each set the same number of times.

Phases 1 and 3 contained 96 trials each. Phase 3 included all three picture sets (of 16 pictures each), and each set was repeated twice. Phase 1 included only Set A from the target

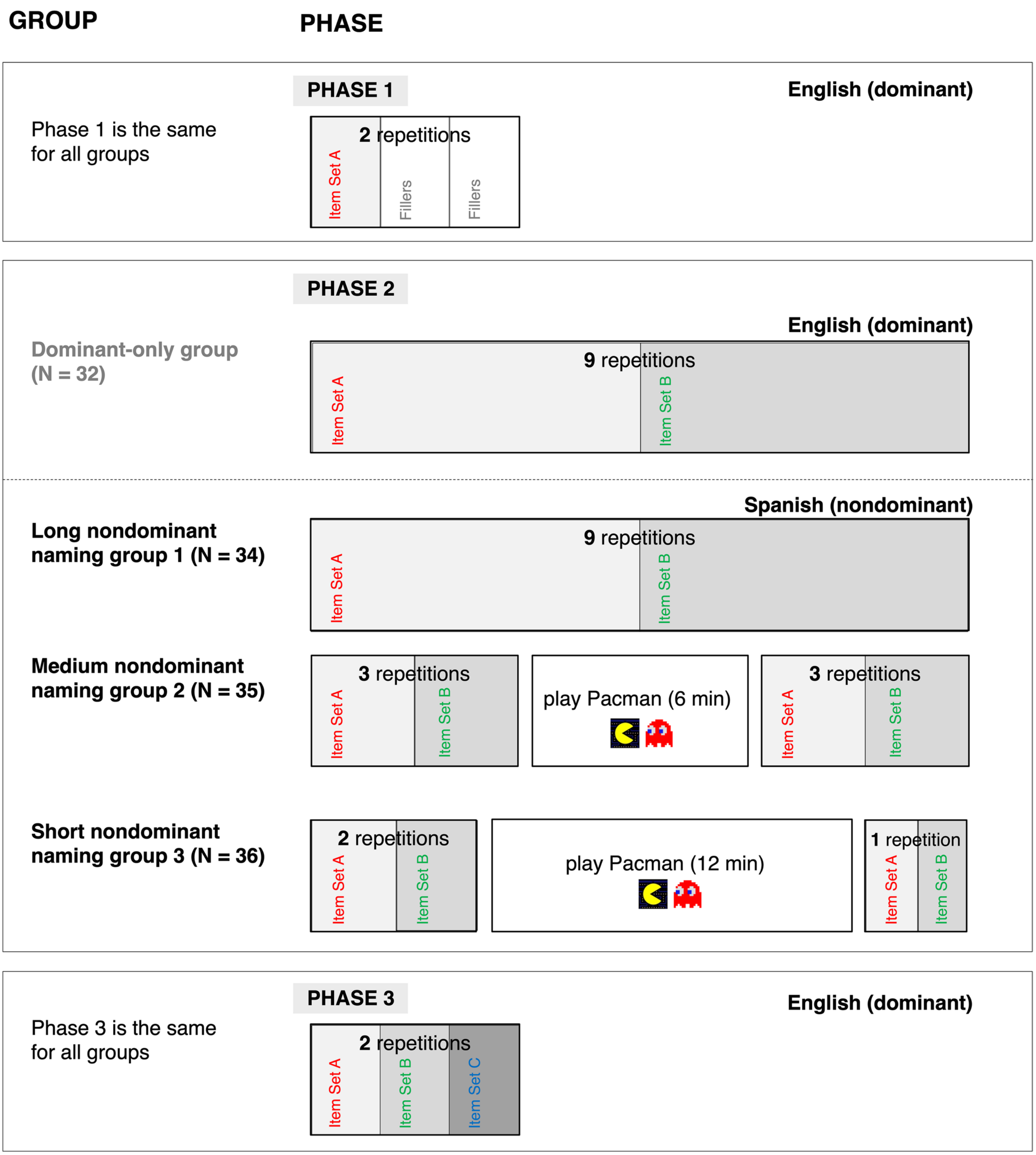

Fig. 2 Study design

items, but the set of 32 filler pictures (also repeated twice and created for this purpose) made it equivalent in length to Phase 3. Phase 2 included Sets A and B (32 pictures in total), which were repeated three times for the Short Nondominant naming group, six times for the Medium Nondominant-naming group, and nine times for the Long Nondominant naming group and the Dominant-only group. In all phases, a new repetition began only when all pictures from a previous repetition had occurred, and the target picture sets (or Set A and fillers for Phase 1) were intermixed within each repetition. Trial order was pseudorandom such that pictures from the same semantic category were apart as much as possible and picture names on consecutive trials did not begin with the same phoneme.

There were 18 versions of the experiment that varied on Phase 2 length (short, medium, long), which of the three different item lists was in Set A (List 1, List 2, List 3), and which of the remaining two lists was in Set C. The 18 versions were administered to a roughly equal number of participants (on average six): Between four and seven participants completed each version for the three Nondominant naming groups, and five or six participants in the Dominant-only group completed each of the six versions that had a long Phase 2.

Procedure Bilinguals were tested individually, in a single session that lasted approximately 1.5 h (and without face coverings, as data were collected prior to March 2020). The experiment was administered with the DMDX software (Forster & Forster, 2003). The study procedures conformed to Federal guidelines for the protection of human subjects and were approved by the UTEP Institutional Review Board. All participants gave informed consent to participate prior to testing

Nondominant-naming groups. Upon arrival, participants were randomly assigned to a Nondominant naming group (Long, Medium, or Short) and tested individually. After giving informed consent, participants read short instructions in English that asked them to name the images presented on the screen as fast and as accurately as possible without any preceding determiners or disfluencies, and to avoid lip smacks. There were three practice trials with pictures that did not appear in the main experiment. There was no familiarization phase. Phase 2 began with a one-sentence instruction reminder in Spanish, and Phase 3 began with the same in English. On each trial, there was a blank screen for 200 ms, followed by a fixation cross for 300 ms, followed by another blank screen for 200 ms. A picture was then presented for 3,000 ms or until the initiation of a vocal response. Disfluencies, responses different from the intended ones, and voice-key inaccuracies (such as lip smacks and other early or late onsets) were noted down by the trained experimenters in real-time. Naming on each trial was recorded automatically by DMDX.

The procedure also aimed to minimize the confound between nondominant naming length and the time elapsed between the two dominant naming phases. To do this, bilinguals in the Short and Medium nondominant naming groups played the game of Pacman for 12 and 6 min, respectively. This time was estimated by subtracting the time needed to complete Phase 2 in each of these two conditions from the time needed in the Long nondominant naming condition in pilot runs. Further, to ensure that nondominant naming immediately preceded dominant naming in Phase 3 for all three groups, we administered the Pacman game in the middle of the experiment (between the first three and last three stimuli repetitions for the Medium nondominant naming group, and between the first two and the third repetitions for the Short nondominant naming group; see Fig. 2). The Pacman game was retrieved from the Internet and administered on the same computer used for the main experiment. The game of Pacman was chosen instead of solving mathematical problems or a silent period in an attempt to minimize inner speech in a particular language. After the main experiment, participants completed the Multilingual Naming Test and the language history questionnaire.

Dominant-only group. Participants in the Dominant-only group were tested after testing of the Nondominant naming groups was completed. All three phases were completed in English and were preceded by English instructions; the procedure was otherwise identical to that of the Long nondominant naming group.

Coding and data analysis Analyses of naming latencies excluded voice key inaccuracies (1,108 trials, or 2.4% of all data), production errors and different-than-intended names (together, 4,361 trials or 9.3% of all data), and, subsequently, outliers – naming latencies that were slower than 3 standard deviations above each participant's mean or faster than 300 ms (1,197 trials, or 2.6% of all data). In all, latency analyses included 40,054 trials, or 85.7% of all data (Phase 1: $n = 3,862$; Phase 2: $n = 24,507$; Phase 3: $n = 11,685$). Ten percent of the latency data (three subjects chosen at random in each of the four groups) were manually timestamped with CheckVocal (Protopapas, 2007) by the first author. The correlation between the manually and automatically timestamped responses (after exclusion of all responses flagged during online administration as voice key inaccuracies) was $r = .91$, which gives us confidence in the reliability of response time (RT) registration.

The latency data were analyzed with linear mixed-effects models (Baayen, 2008). The hypotheses of this study are tested with analyses of dominant-language naming latencies (Phases 1 and 3), as latency analyses are standard in the literature. Analyses involving Phase 2 latencies (not informative about the research questions of this study) are reported in the Appendix.

We also conducted error analyses, both for Phase 3 and for Phase 2. For these analyses, we classified as production errors wrong-language names, disfluencies, and failures to respond (2,078 trials or 4.4% of all data), but not different-than-intended names (e.g., *cup* instead of *glass*), as long as they were a reasonable name for the given picture. Errors were flagged in real time by the trained experimenters with high accuracy (subsequent comparisons between manual

Table 3 Percentage error rates for all groups

Phase	Short nondominant naming group			Medium nondominant naming group			Long nondominant naming group			Dominant-only group		
	Set A	Set B	Set C	Set A	Set B	Set C	Set A	Set B	Set C	Set A	Set B	Set C
Phase 1	1.8	-	-	2.9	-	-	2.7	-	-	2.1	-	-
Phase 2	8.1	8.4	-	6.9	7.6	-	8.6	9.3	-	1.2	1.1	-
Phase 3	1.2	1.6	2.1	1.3	2.6	1.1	2.1	1.6	3.0	1.3	0.7	0.8
Phase 3 minus Phase 1, Set A												
	-0.6	*-0.3*	*0.3*	*-1.7*	*-0.4*	*-1.9*	*-0.6*	*-1.1*	*0.4*	*-0.9*	*-1.5*	*-1.4*

coding of errors in the timestamped data revealed that online coding omissions were 1.4% of all manually coded data).

Error analyses included all 46,720 trials. Error rates across groups and phases (1.9% for dominant naming; Phase 1: 2.2%; Phase 3: 1.6%) are reported in Table 3. Logistic mixed effects regression modeling of the error-rate data (Jaeger, 2008) are reported in the Appendix.

Trial-level data and analyses codes are publicly available at https://osf.io/serhw/.

To address the predictions of main interest, we compared the influence of length of nondominant naming on naming latencies in Phase 3 relative to Phase 1. Because of the partial nesting of Item Set within Phase (Phase 1 only had items from Item Set A, while Phase 3 had items from all three item sets), we combined the respective levels of Phase and Item Set into a single, within-participant factor (Phase/Item Set) with four levels: Phase 1/Set A, Phase 3/Set A, Phase 3/Set B, and Phase 3/Set C.

A single 4×4 linear mixed-effects model was thus used to estimate condition means in each of the four groups using the *lme4* package (v. 1.1-21; Bates et al., 2015) in R (R Core Team, 2022). The model had fixed effects of Phase/Item Set and Group, which were sum-coded (factor levels set to +/-0.5), and the maximal random effects structure supported by the data. To identify this structure, we followed a three-step procedure (for this model and all other supplementary models). First, we used the *bobyqa* optimizer to fit a model with a maximal random effects structure: random intercepts for participants and items, all within-factor random slopes and their interactions, and correlations between random slopes. If this model did not converge, we removed correlations between random slopes. If the resulting model still did not converge, we identified random slopes accounting for less than 1% of the variance of their associated random factors, then removed all such slopes simultaneously (Bates et al., 2018). This always resulted in convergence. For models with a continuous dependent variable, denominator degrees of freedom were estimated using the Satterthwaite method in the lmerTest package (v. 3.1-3; Kuznetsova et al., 2017). For the main analysis, the equation of the converging model was as follows (note that Group was not allowed to vary by participant as it was a between-participant factor):[2]

```
RT ~ 1 + Group * PhaseAndSet +
    (1 + PhaseAndSet1 + PhaseAndSet2 + PhaseAndSet3 || Participant) +
    (1 + Group1 + Group2 + PhaseAndSet1 + PhaseAndSet2 + PhaseAndSet3 +
        Group1:PhaseAndSet3 + Group2:PhaseAndSet3 + Group3:PhaseAndSet2 +
        Group3:PhaseAndSet3 || Picture)
```

To test hypotheses about differences in the extent to which naming latencies were slower in Phase 3 relative to Phase 1 across different Phase 3 item sets and participant groups, contrasts were applied to the fitted model using the *emmeans* package (v. 1.7.1-1; Lenth, 2021). For each hypothesis that involved comparing groups on a particular (Phase 3 vs. Phase 1) contrast, an *F*-test was performed to determine whether the contrast significantly differed across groups. If so, contrast means were computed separately for each group (with false discovery rate (FDR) controlled via the Benjamini-Yekutieli method) and pairwise comparisons were conducted between group contrast means (with multiple comparisons controlled via the Tukey method).

Statistical power to detect between-group differences After performing the main analyses, we looked into whether the key comparisons had enough statistical power. To do this, we used the data of Kleinman and Gollan (2018) to estimate the expected effect size of each of nine key contrasts: Short versus Medium, Medium versus Long, and Short versus Long, for each of the three item sets. We extracted relevant effect sizes from their statistical models, which quantified how much each naming latency was affected by prior picture-naming trials, then applied them to the sequences of

[2] Random effects were numerically coded variables, with *k*-1 terms representing each *k*-level factor; these terms, which are represented by factor names followed by numbers (e.g., PhaseAndSet2), were treated independently when removing random slopes. Due to the use of sum coding (with factor levels set to -0.5, 0 and +0.5), it is not easy to interpret these terms individually. The double-bar notation (||) indicates that correlations between random effects were removed from the model.

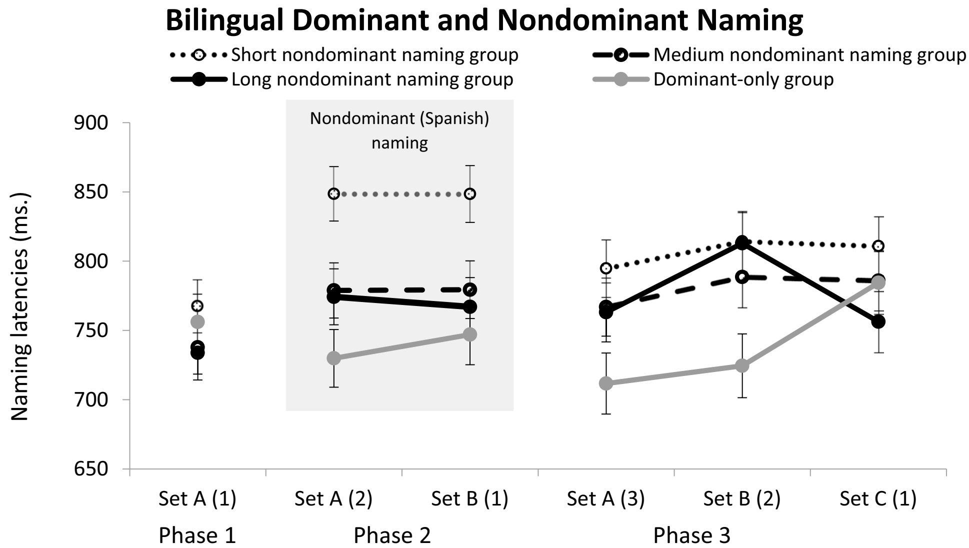

Fig. 3 By-participant mean naming latencies for each phase and item set. The numbers next to each item set on the x-axis indicate how many phases that item set has occurred in until and including the current one. Error bars represent standard error

pictures that were presented to participants in the present experiment and averaged across all trials included in latency analyses (see the Online Supplementary Material for more details). Using the resulting (very large) effect sizes, which ranged from 121 ms to 290 ms, we performed power calculations with the R package simR (v. 1.0.5; Green et al., 2015). For each contrast, 1,000 simulations were run to estimate power to detect an effect of the expected size. These simulations established that the present experiment was sufficient to detect all nine key contrasts with 100% power.

Of course, these power calculations are only useful insofar as the trial-level effect sizes would be expected to generalize between experiments. Bilingual language proficiency is known to affect the magnitude of reversed dominance effects (Declerck et al., 2020), but bilinguals in this experiment and the ones in Kleinman and Gollan's (2018) study had very similar mean proficiency scores (English MINT: 61.8 vs. 60.6; Spanish MINT: 44.7 vs. 45.7; scores out of 68). Another potential difference is that some participants in the present experiment took a break in the middle of the Spanish block, but reasonable assumptions about how that could have affected performance would not have decreased the expected effect sizes (see the Online Supplementary Material for details).

Results

Figure 3 plots by-participant mean naming latencies, and Fig. 4 plots latency differences of Phase 3 with Phase 1. Table 4 reports results for all statistical tests.

Did nondominant naming engage language control mechanisms?

To test our hypotheses, we need an index that language control was applied in our task: slower latencies in Phase 3 relative to Phase 1 for all Nondominant naming groups (but not for the Dominant-only group). To test this, we compared latencies in Phase 3 (averaged across item sets A, B, and C) to those of Phase 1 (Set A was the only item set in that phase). All Nondominant naming groups were significantly slower in Phase 3 than in Phase 1 (38–44 ms), while the Dominant-only group showed a nonsignificant speedup (21 ms), and was significantly different from each of the other groups. This pattern indicates that the slowdowns observed for the Nondominant naming groups cannot be attributed to overall experiment length or other factors unrelated to language mixing.

Did aftereffects on dominant naming differ with length of prior nondominant naming?

Pairwise comparisons (conducted as part of an analysis across all four groups, as described above) revealed that the slowdown magnitude did not differ between the Short, Medium, and Long nondominant naming groups. These results provide no support for the key prediction of the *Over-and-over hypothesis* that the slowdown should increase with the length of nondominant naming in Phase 2. They are instead consistent with the Once-only hypothesis.

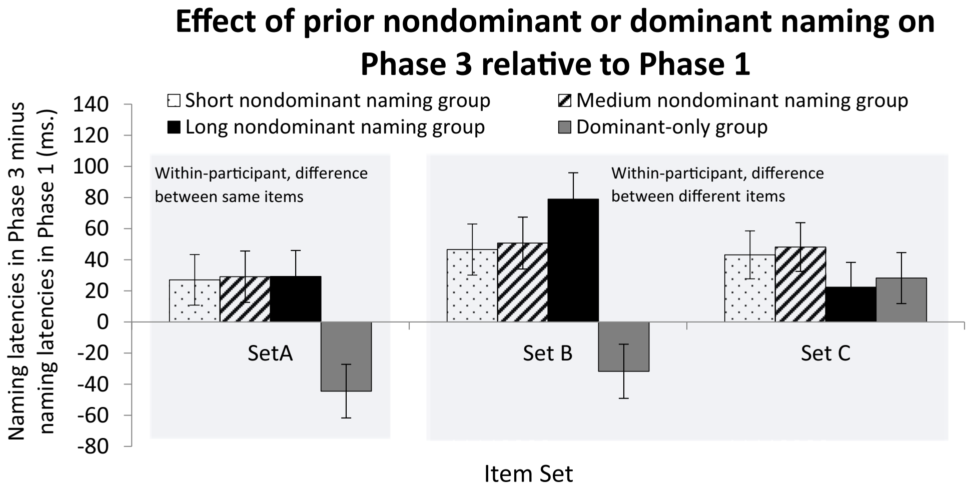

Fig. 4 Latency differences of Phase 3 with Phase 1. Error bars represent standard error

Did the aftereffects on dominant naming differ for repeated and new items?

Before examining effects of nondominant naming on different item sets, we first confirmed that groups significantly differed in their patterns of Phase 3 naming latencies across item sets, as indicated by a significant two-way interaction between group and item set on those latencies. Below, we compare groups separately for each item set, then compare item sets separately for each group.

Item Set A Across Nondominant naming groups, latencies in Phase 3 were significantly slower (24 ms) for Item Set A than in Phase 1 (also for Item Set A),[3] and did not differ in pairwise comparisons, while the Dominant-only group showed the reverse pattern (a significant 49-ms speedup), and differed from each other group. That is, for pictures that were named in all three phases, latencies sped up when there was no language change across phases, but modestly slowed down in Phase 3 when there was a switch to the nondominant language in Phase 2.

Item Set B Across all Nondominant naming groups, naming latencies in Phase 3 were significantly slower (58 ms) for Item Set B than in Phase 1 (for Item Set A), and did not differ in pairwise comparisons, while the Dominant-only group showed the reverse pattern (a significant 38-ms speedup), and differed from each of the other groups. We take the effects for both Sets A and B to index the classic aftereffects of inhibition applied to individual items (local inhibition).

Item Set C Across all groups, naming latencies in Phase 3 were significantly slower (35 ms) for Item Set C than in Phase 1 (for Item Set A), and the size of this effect did not significantly differ across groups, with all four groups showing non-significant slowdowns (23–49 ms). That is, naming new items was delayed to a similar extent in Phase 3 for the group that named pictures in the dominant language throughout as for the groups that changed languages, indicating that we did not detect aftereffects of prior global inhibition.

Differential effects of item sets in Phase 3 across groups We also compared, for each group, the effect of item set in Phase 3 only (as licensed by the significant interaction between group and Phase 3 item set). For the Long nondominant naming and Dominant-only groups, this effect was significant; further pairwise contrasts indicated that the Dominant-only group named pictures in Set C more slowly than pictures in Sets A and B, while the Long nondominant naming group named pictures in Set B more slowly than pictures in Sets A and C. For the Short and Medium nondominant naming groups, the effect of item set was marginal; both groups named pictures in Set B more slowly than pictures in Set A.

The Dominant-only group effect is easily explained with repetition priming. The Long nondominant naming group effect, in contrast, shows that when participants named pictures (Set B) many times in Spanish, they were subsequently slower to name those same pictures in English (but this slowdown did not generalize to new pictures, Set C) – possibly the effects of local inhibition. However, this pattern

3 For Item Set A, this contrast did not reach significance when Nondominant naming groups were considered independently due to a combination of (lack of) pooled variance and the correction for multiple comparisons (all n.s. effects between 22 and 28 ms). For Item Set B, the contrast was significant for each Nondominant naming group when considered individually (all effects between 47 and 73 ms).

Table 4 Statistical results for all tests involving Phase 3 naming latencies

Phases and sets being compared	Contrast performed		*B* (ms)	95% CI (ms)	Test	*df*	Statistic	*p*
[Phase 3/Sets A,B,C] vs. [Phase 1/Set A]	Significant when pooled across groups?		25	[10, 40]	*t*	107	3.33	.001
	Different between groups?				*F*	3, 126	5.83	< .001
	By group	Short	38	[12, 64]	*t*	142	2.94	.011
		Medium	44	[18, 70]		145	3.31	.010
		Long	40	[14, 67]		142	3.04	.011
		Dominant-only	-21	[-48, 6]		142	-1.56	.252
	Pairwise comparisons between groups	Dominant-only vs. each Nondominant	[59, 65]		*t*		> 3.32	< .007
		Within Nondominant	[-3, 6]		*t*		< 0.32	> .989
Phase 3: Sets A vs. B vs. C	Different between groups?				*F*	6, 143	4.27	< .001
[Phase 3/Set A] vs. [Phase 1/Set A]	Significant when pooled across Non-dominant groups?		24	[9, 40]	*t*	215	3.03	.006
	Different between groups?				*F*	3, 245	9.11	< .001
	By group	Short	22	[-2, 47]	*t*	275	1.79	.156
		Medium	27	[2, 52]		280	2.12	.145
		Long	24	[-1, 49]		271	1.90	.156
		Dominant-only	-49	[-75, -24]		276	-3.77	.002
	Pairwise comparisons between groups	Dominant-only vs. each Nondominant	[72, 76]		*t*		> 4.21	< .001
		Within Nondominant	[-3, 5]		*t*		< 0.29	> .991
[Phase 3/Set B] vs. [Phase 1/Set A]	Significant when pooled across Non-dominant groups?		58	[42, 74]	*t*	211	7.22	< .001
	Different between groups?				*F*	3, 238	15.46	< .001
	By group	Short	47	[22, 72]	*t*	274	3.75	< .001
		Medium	55	[30, 80]		281	4.27	< .001
		Long	73	[47, 99]		279	5.56	< .001
		Dominant-only	-38	[-64, -12]		275	-2.90	.008
	Pairwise comparisons between groups	Dominant-only vs. each Nondominant	[86, 111]		*t*		> 4.94	< .001
		Within Nondominant	[8, 26]		*t*		< 1.49	> .448
[Phase 3/Set C] vs. [Phase 1/Set A]	Significant when pooled across Non-dominant groups?		39	[13, 65]	*t*	128	3.00	.006
	Different between groups?				*F*	3, 115	0.49	1.000
	By group	Short	45	[6, 84]	*t*	139	2.25	.107
		Medium	49	[9, 89]		139	2.43	.107
		Long	24	[-16, 65]		142	1.18	.532
		Dominant-only	24	[-17, 65]		138	1.14	.532
	Pairwise comparisons between groups	Dominant-only vs. each Nondominant	[0, 25]		*t*		< 0.93	> .791
		Within Nondominant	[-25, 4]		*t*		< 0.93	> .792

Table 4 (continued)

Phases and sets being compared	Contrast performed		*B* (ms)	95% CI (ms)	Test	*df*	Statistic	*p*
Phase 3: Sets A vs. B vs. C	By group	Short			*F*	2, 176	3.14	.096
		Set B > Set A	25	[4, 45]	*t*	405	2.39	.045
		Medium			*F*	2, 165	3.51	.089
		Set B > Set A	28	[7, 49]	*t*	399	2.59	.027
		Long			*F*	2, 171	10.61	< .001
		Set B > Sets A & C	[49, 49]		*t*		> 2.85	< .014
		Dominant-only			*F*	2, 177	9.95	< .001
		Set C > Sets A & B	[62, 73]		*t*		> 3.70	< .001

The order of table rows matches the order in which tests are reported in the text. Analyses described as "Within Nondominant" refer to pairwise comparisons between the three Nondominant naming groups. Where multiple tests are reported in a single table row, confidence intervals (CIs) and *df* values are not provided; beta estimates, *t*-/*F*-values, and *p*-values are provided for the range of tests. As *p*-values are corrected for multiple comparisons, thereby lowering the true alpha level for individual tests below .05, some tests are non-significant (corrected *p* > .05) even though corresponding 95% CIs do not include 0

is not entirely consistent with the predictions of the Over-and-over hypothesis, which predicts smaller effects in the same direction for the other Nondominant naming groups. However, naming latencies were slower in Set B than Set C by only 6 ms for the Medium group and by 2 ms for the Short group (both n.s.), versus 48 ms for the Long group. Instead, the results are more consistent with a stronger local inhibition effect applied beyond a certain threshold of non-dominant language repetition.

Finally, all three Nondominant naming groups named Phase 3 pictures in Set B significantly slower than in Set A. Given that both item sets were named in the nondominant language in Phase 2 but only Set A was previously named in the dominant language in Phase 1, this suggests that using names in the dominant language first may actually have a protective effect from the adverse aftereffects of language control on the dominant language, contrary to our hypothesis.

Discussion

We investigated if bilingual (inhibitory) control over a currently irrelevant language is applied only once (Once-only hypothesis) or all the time (Over-and-over hypothesis). Three groups of bilinguals named pictures first in their dominant, then in their nondominant, and then again in their dominant language (and a control group named pictures in the dominant language only). If the dominant language is inhibited upon naming every picture in the nondominant language, the number of nondominant repetitions should be proportional to the subsequent dominant disadvantage. In accordance with the Inhibitory Control Model and much empirical evidence, nondominant naming caused a subsequent dominant naming delay. Of most interest, we found little evidence for differential effects of the number of prior non-dominant repetitions on subsequent dominant naming speed. These results are consistent with the Once-only hypothesis, and do not support the Over-and-over hypothesis.

Looking particularly at item-specific versus whole-language effects, we found a naming delay for dominant-after-nondominant naming. This delay was present both for names first produced in the dominant language (Set A) and, and to a greater extent, for names first produced in the nondominant language (Set B). However, for new items (those introduced in the last dominant naming phase) the delay was equivalent to that incurred by dominant-only naming throughout the experiment. That is, there was no sign of global inhibition. We thus interpret our results as supporting the Once-only hypothesis in its "local" version: Inhibition is applied only once, upon the first mention of every word, and does not accumulate with each subsequent repetition.

We are unsure why we did not see effects of global inhibition as detected in prior studies (Casado et al., 2022; Degani et al., 2020; Kreiner & Degani, 2015; Wodniecka et al., 2020). It could be that the engagement of global control is stronger with a greater imbalance between the languages (Casado et al., 2022) – and for most of our participants, it is common in daily life to use both languages. Our methodology may also not have been sensitive enough to detect (the likely more subtle) global effects. These may be more easily detectable by measuring tip-of-the-tongue states for low-frequency words (Kreiner & Degani, 2015) or cross-language intrusion errors in cases where nontarget-language borrowings are habitually preferred over target-language words (Degani et al., 2020). The discrepancy may also lie in the slightly different assumptions and methodologies across studies, including that prior studies measuring production latencies (Branzi et al., 2014; Guo et al., 2011; Wodniecka et al., 2020) did not include a single-language group. We used this group as baseline because single-language naming latencies for nonrepeated names (as with the first presentation of Item Set C) gradually slow down throughout an experiment (Székely et al., 2003).

Our results contrast with those of Kleinman and Gollan (2018), as we found no evidence for inhibition accumulation, despite substantial power to detect it. We think this is primarily because the engagement and type of language control mechanisms differ as a function of task demands. In a task where participants switch languages on 50% of trials and name a small set of pictures many times in both languages,[4] adjusting the balance of word and language-wide activation on every trial may be a reasonable response to task demands. Also, other control processes may be at play, such as conflict monitoring, attending to cues, or maintaining readiness of two competing responses. In contrast, in a task in which participants rarely if ever need to switch languages – and they do so at expected times (between blocks) – adjustments to word and language-wide activation can be made infrequently.

Separately, pictures that were previously named in the dominant language (Set A) were subsequently named *faster* in the dominant language compared to pictures that were not (Set B). This suggests that preactivating dominant names' lexical forms has a protective effect against the adverse inhibitory aftereffects on the dominant language induced by nondominant naming. This result seems inconsistent with the core assumption of the Inhibitory Control Model that most highly activated names (such as those belonging to the dominant language) would be inhibited most strongly. However, the two may not be incompatible. For example, recent prior dominant naming may confer phonological or articulatory facilitation, counteracting the negative effects of inhibition.

But how long does inhibition applied "only once" persist? Our results suggest that inhibition over the dominant language is applied only upon each nondominant-language word's first mention (though it might be reapplied over longer stretches of speaking). We assume that such inhibition over a language is recovered from (in a passive process, see Wodniecka et al., 2020) when the dominant language is spoken again, with a different conversation partner or different situation. For many bilinguals, that would be within the same day – but, if not, we think that it is possible that inhibitory effects (conceptualized as (unnaturally) lower activation levels) persist for weeks and even months. Such situations may explain cumulative adverse effects on the dominant language after immersion in the nondominant language (Baus et al., 2013; Linck et al., 2009).

Also, the Once-only hypothesis is not incompatible with bilingual adaptations. Bilingual language use involves many processes beyond (inhibitory) language control, such as goal maintenance, monitoring, or cue detection (see Green & Abutalebi, 2013). Also, practice applying inhibition upon words' first mention can still accrue with more frequent switches, and over a longer period than tested here (and more words will be used over time).

We adopted the Inhibitory Control Model framework to explain the study's logic, but our research question and conclusions can instead be made about other proposed language control mechanisms. The only necessary assumption in our study is that language control applied over the nondominant language subsequently has adverse effects on the dominant language (more than vice versa). However, instead of inhibition of the dominant language, language control may entail "hyper-activating" the nondominant language (e.g., by lowering its lexical selection threshold; Branzi et al., 2014), which then creates increased interference during dominant production.[5] Whichever the mechanism, our conclusions remain the same: In a context with relatively minimal language switching, language control is applied upon a word's first mention over its translation equivalent, not over and over. Future research should determine how the timing of language control unfolds for language use beyond single naming, in different contexts of language use.

Appendix

The Appendix is divided into two sections. Section 1 contains a list of stimuli used in the experiment. Section 2 contains three sets of additional analyses: One each for Phase 3 error rates, Phase 2 naming latencies, and Phase 2 error rates.

[4] The tension between effects of repetition priming and inhibition, which was central to Kleinman and Gollan's (2018) account, may also play out differently in experiments with larger set sizes: Participants in their experiments named nine unique pictures, versus 80 in the present experiment (48 critical pictures + 32 filler pictures); this difference likely affected the expected effect sizes as well.

[5] For a more detailed discussion of different bilingual language control mechanisms, see Ivanova and Hernandez (2021).

1 Stimulus List

#	Type	Name (English)	Length in phonemes (English)	Frequency (English)	Semantic category	Name (Spanish)	Length in phonemes (Spanish)	Frequency (Spanish)
List 1								
1	exp	**dog**	3	192.8	animal	**perro**	4	166.2
2	exp	**butterfly**	9	5.5	animal	**mariposa**	8	6.7
3	exp	**hand**	4	279.6	body	**mano**	4	256.0
4	exp	**foot**	3	64.9	body	**pie**	3	103.4
5	exp	**church**	4	69.7	building	**iglesia**	7	82.9
6	exp	**shoe**	3	30.4	clothing	**zapato**	6	17.1
7	exp	**bread**	5	28.3	food	**pan**	3	48.2
8	exp	**cheese**	3	39.0	food	**queso**	4	47.0
9	exp	**chair**	4	49.2	furniture	**silla**	4	55.5
10	exp	**waiter**	6	13.2	human	**mesero**	6	4.8
11	exp	**star**	4	81.4	nature	**estrella**	7	61.1
12	exp	**cloud**	5	11.7	nature	**nube**	4	9.6
13	exp	**pencil**	6	9.9	object	**lapiz**	5	14.8
14	exp	**knife**	5	46.8	object	**cuchillo**	6	41.5
15	exp	**ring**	4	92.7	object	**anillo**	5	59.7
16	exp	**tree**	3	65.0	plant	**arbol**	5	52.5
List 2								
17	exp	**cat**	3	66.3	animal	**gato**	4	54.0
18	exp	**chicken**	5	61.7	animal	**pollo**	4	47.3
19	exp	**eye**	2	111.8	body	**ojo**	3	64.8
20	exp	**lightbulb**	8	NA	furniture	**foco**	4	3.5
21	exp	**door**	3	292.1	building	**puerta**	6	328.9
22	exp	**bridge**	5	45.7	building	**puente**	6	40.3
23	exp	**dress**	4	87.2	clothing	**vestido**	7	80.9
24	exp	**strawberry**	8	5.5	food	**fresa**	5	3.5
25	exp	**apple**	4	23.7	food	**manzana**	7	19.0
26	exp	**table**	4	105.6	furniture	**mesa**	4	109.6
27	exp	**nurse**	4	45.0	human	**enfermera**	9	35.9
28	exp	**moon**	3	50.0	nature	**luna**	4	66.0
29	exp	**suitcase**	7	13.4	object	**maleta**	6	18.8
30	exp	**clock**	4	58.6	object	**reloj**	5	112.5
31	exp	**box**	3	89.7	object	**caja**	4	78.0

Table (continued)

#	Type	Name (English)	Length in phonemes (English)	Frequency (English)	Semantic category	Name (Spanish)	Length in phonemes (Spanish)	Frequency (Spanish)
32	exp	**ball**	3	105.0	object	**pelota**	6	40.5
List 3								
33	exp	**bird**	4	45.5	animal	**pajaro**	6	23.1
34	exp	**horse**	4	92.9	animal	**caballo**	6	62.2
35	exp	**finger**	6	36.7	body	**dedo**	4	45.8
36	exp	**heart**	4	244.2	body	**corazon**	7	252.8
37	exp	**window**	6	86.0	building	**ventana**	7	73.2
38	exp	**house**	4	514.0	building	**casa**	4	1378.3
39	exp	**shirt**	4	46.4	clothing	**camisa**	6	46.3
40	exp	**cake**	4	45.1	food	**pastel**	6	54.6
41	exp	**icecream**	8	0.7 (obtained from N-Watch)	food	**nieve**	5	39.3 (obtained from B-PAL)
42	exp	**bed**	3	187.1	furniture	**cama**	4	178.2
43	exp	**king**	4	129.3	human	**rey**	3	114.3
44	exp	**rain**	4	48.9	nature	**lluvia**	5	33.1
45	exp	**present**	7	89.5	object	**regalo**	6	95.0
46	exp	**book**	3	177.0	object	**libro**	5	159.7
47	exp	**glass**	4	60.7	object	**vaso**	4	26.1
48	exp	**leaf**	3	5.2	plant	**hoja**	4	13.2
Fillers (used in Phase 1)								
101	filler	**backpack**	6	3.6	object	not used	not used	not used
102	filler	**boy**	3	529.8	human	not used	not used	not used
103	filler	**key**	2	86.9	object	not used	not used	not used
104	filler	**carrot**	5	3.8	food	not used	not used	not used
105	filler	**cow**	3	25.5	animal	not used	not used	not used
106	filler	**ear**	3	32.0	body	not used	not used	not used
107	filler	**desk**	4	43.9	furniture	not used	not used	not used
108	filler	**drum**	4	8.5	object	not used	not used	not used
109	filler	**fire**	4	215.5	nature	not used	not used	not used
110	filler	**fork**	4	8.8	object	not used	not used	not used
111	filler	**hat**	3	64.2	clothing	not used	not used	not used
112	filler	**snake**	5	22.4	animal	not used	not used	not used
113	filler	**tire**	4	12.4	vehicle	not used	not used	not used
114	filler	**road**	4	111.9	building	not used	not used	not used
115	filler	**mushroom**	6	2.1	food	not used	not used	not used
116	filler	**wallet**	5	22.8	object	not used	not used	not used

Table (continued)

#	Type	Name (English)	Length in phonemes (English)	Frequency (English)	Semantic category	Name (Spanish)	Length in phonemes (Spanish)	Frequency (Spanish)
117	filler	**arm**	3	65.4	body	not used	not used	not used
118	filler	**bag**	3	94.0	object	not used	not used	not used
119	filler	**bear**	4	57.4	animal	not used	not used	not used
120	filler	**candle**	5	8.0	object	not used	not used	not used
121	filler	**egg**	2	26.0	food	not used	not used	not used
122	filler	**shower**	5	41.1	furniture	not used	not used	not used
123	filler	**sun**	3	69.7	nature	not used	not used	not used
124	filler	**flag**	4	17.5	object	not used	not used	not used
125	filler	**spoon**	4	7.6	object	not used	not used	not used
126	filler	**fish**	3	83.5	animal	not used	not used	not used
127	filler	**sock**	3	9.0	clothing	not used	not used	not used
128	filler	**girl**	4	557.1	human	not used	not used	not used
129	filler	**balloon**	5	8.7	object	not used	not used	not used
130	filler	**mirror**	5	24.2	object	not used	not used	not used
131	filler	**pig**	3	39.1	animal	not used	not used	not used
132	filler	**umbrella**	7	7.5	object	not used	not used	not used

2 Additional analyses

Phase 3 error rates

Across the four groups, error rates were lower in Phase 3 (1.6%) than in Phase 1 (2.4%), B = -1.19, 95% CI = [-1.60, -0.79], z = -5.79, p < .001 (F1 means are shown in percentiles for interpretability, but model parameters are in logit space). This improvement did not significantly differ between groups, $X^2(3)$ = 5.05, p = .168, and every group showed a significant or marginally significant improvement in Phase 3 (between 0.2% and 1.3%), all $|z|$s > 2.11, all ps < .071. Similarly, all three item sets showed a significant improvement in Phase 3 (between 0.6% and 0.9%), all $|z|$s > 3.43, all ps < .002. The extent to which this improvement varied across groups did not itself vary across item sets, as indicated by a non-significant interaction between group and Phase 3 error rates, $X^2(6)$ = 5.39, p = .495. Thus, consistent with the results from the RT analyses, the length of the nondominant naming block in Phase 2 did not affect Phase 3 performance.

Phase 2 naming latencies

To compare Phase 2 naming latencies (or, below, Phase 2 error rates) between groups and item sets, we adopted the same modeling approach as for the main analyses: We combined all combinations of Phase and Item Set into a single, within-participant factor (Phase/Set) with three levels: Phase 1/Set A, Phase 2/Set A, and Phase 2/Set B; then we used a single 4x3 linear mixed-effects model to estimate condition means for each of the four groups.

First, as a sanity check, we examined whether naming latencies in Phase 2 were slower than naming latencies in Phase 1 for the three groups (Short, Medium, Long) that named Phase 2 pictures in their nondominant Spanish. Across the four groups, naming latencies were on average 39 ms slower in Phase 2 than in Phase 1, 95% CI = [23, 55] ms, $t(109)$ = 4.89, p < .001. This slowdown differed significantly between groups, $F(3, 125)$ = 13.05, p < .001: Naming latencies were significantly slower in Phase 2 than in Phase 1 for the Short group, B = 90 ms, 95% CI = [63, 117] ms, $t(158)$ = 6.61, p < .001; the Medium group, B = 46 ms, 95% CI = [20, 73] ms, $t(148)$ = 3.43, p = .003; and the Long group, B = 42 ms, 95% CI = [15, 69] ms, $t(144)$ = 3.09, p = .007; but not the Dominant-only group, which showed a non-significant effect in the direction of Phase 2 facilitation, B = -23 ms, 95% CI = [-51, 4] ms, $t(142)$ = -1.68, p = .200. Pairwise comparisons between groups revealed that the Phase 2 slowdown was significantly larger for all three groups that named Spanish in Phase 2 compared to the group that named English in Phase 2, all $|t|$ > 3.57, all $p < .003$. Between groups who named Spanish in Phase 2, the slowdown was marginally larger for the Short group (who had the fewest Spanish trials) compared to the Medium group, $t(134) = -2.44$, $p = .075$, and significantly larger compared to the Long group, $t(123) = -2.62$, $p = .048$, though the Medium and Long groups did not differ, $t(127) = -0.23$, $p = .996$. These results demonstrate that (as expected) participants named pictures in Spanish slower than they named pictures in English, but that overall, this Spanish slowdown decreased as the number of Spanish trials increased.

Next, we determined, for each item set, whether the slowdown was present in Phase 2 and whether it varied across groups. The extent to which slowdowns varied across groups itself did not vary across item sets, as indicated by a non-significant interaction between group and Phase 2 naming latencies, $F(3, 111) = 0.77$, $p = .513$. Descriptively, we note that for all groups, Phase 2 naming latencies for Item Set A and Item Set B differed by at most 16 ms (vs. between-group differences in Phase 2 slowdowns that were as large as 113 ms).

Phase 2 error rates

Across the four groups, error rates were higher in Phase 2 (6.4%) than in Phase 1 (2.4%), $B = 0.43$, 95% CI = [0.03, 0.84], $z = 2.09$, $p = .037$ (note that parameter estimates are in logit space). Similarly to the Phase 2 RT analyses, error rates were significantly lower in Phase 2 than in Phase 1 for the Dominant-only naming group (by 1.0%), $B = -1.46$, 95% CI = [-2.27, -0.65], $z = -3.53$, $p = .002$, but significantly higher in Phase 2 than in Phase 1 for all three Nondominant naming groups (by 4.3% to 6.5%), reflecting more difficulty with naming in the nondominant language: Short, $B = 1.32$, 95% CI = [0.62, 2.02], $z = 3.71$, $p = .002$; Medium: $B = 0.77$, 95% CI = [0.18, 1.36], $z = 2.57$, $p = .021$; Long: $B = 1.11$, 95% CI = [0.42, 1.79], $z = 3.15$, $p = .004$. Consistent with this, the difference between Phase 1 and Phase 2 error rates differed significantly between groups, $X^2(3) = 36.10$, $p < .001$, with pairwise comparisons revealing that the effect was significantly different between each of the Nondominant naming groups and the Dominant-only group, all $|z| > 4.68$, all $p < .001$, but not between the three Nondominant naming groups, all $|z| < 1.28$, all $p > .580$. The extent to which these effects varied across groups did not itself vary across item sets, as indicated by a nonsignificant interaction between group and Phase 2 error rates, $X^2(3) = 0.46$, $p = .928$.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-022-02168-z.

Acknowledgements Preparation of this article was funded by R21HD109797 from NICHD, and award 2021124 from NSF-PAC, awarded to Iva Ivanova. Many thanks to Ana-Paula Anaya, Aziz Atiya, Margo Berumen, Graciela Favela, Carolina Hernandez, Victoria Jeffress, Stephanie Ojeda, and Lluvia Mendiola for help with data collection.

Declarations

Competing interests The authors have no competing interests to declare.

References

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., ... & Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, *22*, 2076-2086.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., ... & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344-380.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). *Parsimonious mixed models*. Available from arXiv:1506.04967v2 (stat.ME)

Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica, 142*(3), 402–409.

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 859–873.

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.

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

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*, 977–990.

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

Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research, 1147*, 192–208.

Christoffels, I. K., Ganushchak, L., & La Heij, W. (2016). When L1 suffers: Sustained, global slowing and the reversed language effect in mixed language context. In J. W. Schwieter (Ed.), *Cognitive Control and Consequences of Multilingualism* (pp. 171–192). John Benjamins.

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.

Cuetos, F., Glez-Nosti, M., Barbon, A., & Brysbaert, M. (2011). SUBTLEX-ESP: frecuencias de las palabras espanolas basadas en los subtitulos de las peliculas. *Psicológica, 32*(2), 133–144.

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review, 22*(6), 1630–1645.

Declerck, M., Koch, I., & Philipp, A. M. (2012). Digits vs. pictures: The influence of stimulus type on language switching. *Bilingualism: Language and Cognition, 15*(4), 896–904.
Declerck, M., Kleinman, D., & Gollan, T. H. (2020). Which bilinguals reverse language dominance and why? *Cognition, 204*, 104384.
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.
Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116–124.
Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science, 22*(9), 1155–1164.
Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *15*(3), 594-615.
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.
Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67–81.
Green, P., MacLeod, C. J., & Nakagawa, S. (2015). SIMR: an R package for power analysis of generalized mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504
Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530.
Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese–English bilinguals. *NeuroImage, 56*(4), 2300–2309. https://doi.org/10.1016/j.neuroimage.2011.03.049
Ivanova, I., & Hernandez, D. C. (2021). Within-language lexical interference can be resolved in a similar way to between-language interference. *Cognition, 214*, 104760. https://doi.org/10.1016/j.cognition.2021.104760
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.
Kirk, N. W., Kempe, V., Scott-Brown, K. C., Philipp, A., & Declerck, M. (2018). Can monolinguals be like bilinguals? Evidence from dialect switching. *Cognition, 170*, 164–178.
Kleinman, D., & Gollan, T. H. (2018). Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition, 173*, 115–132.
Kreiner, H., & Degani, T. (2015). Tip-of-the-tongue in a second language: The effects of brief first-language exposure and long-term use. *Cognition, 137*, 106–114.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13
Lenth, R. V. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means. [R package]. URL: https://CRAN.R-project.org/package=emmeans
Li, C., & Gollan, T. H. (2018). Cognates facilitate switches and then confusion: Contrasting effects of cascade versus feedback on language selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(6), 974.
Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507–1515.
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.
Mosca, M., & Clahsen, H. (2016). Examining language switching in bilinguals: The role of preparation time. *Bilingualism: Language and Cognition, 19*(2), 415–424.
Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*(2), 232–258.
Peeters, D., Runnqvist, E., Bertrand, D., & Grainger, J. (2014). Asymmetrical switch costs in bilingual language production induced by reading words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(1), 284–292.
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.
Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production*. Benjamins.
Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39*(4), 859–862. https://doi.org/10.3758/BF03192979
R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
Rossi, E., Newman, S., Kroll, J. F., & Diaz, M. T. (2018). Neural signatures of inhibitory control in bilingual spoken production. *Cortex, 108*, 50–66.
Székely, A., D'amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., ... & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 621-633.
Van Assche, E., Duyck, W., & Gollan, T. H. (2013). Whole-language and item-specific control in bilingual language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1781–1792.
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*, 107390.
Zheng, X., Roelofs, A., & Lemhöfer, K. (2018). Language selection errors in switching: language priming or cognitive control? *Language, Cognition and Neuroscience, 33*(2), 139–147.
Zheng, X., Roelofs, A., Erkan, H., & Lemhöfer, K. (2020). Dynamics of inhibitory control during bilingual speech production: An electrophysiological study. *Neuropsychologia, 140*, 107387.

Open practices statement Trial-level data and analyses code are publicly available at https://osf.io/serhw/. The stimuli are provided in the Appendix. The experiment was not preregistered.