Inhibitory control of the dominant language:

Reversed language dominance is the tip of the iceberg

Matthew Goldrick[1] and Tamar H. Gollan[2]

[1]Department of Linguistics, Northwestern University

[2]Department of Psychiatry, University of California, San Diego

Address for Correspondence:
Matthew Goldrick
Department of Linguistics
Northwestern University
2016 Sheridan Rd.
Evanston, IL 60208
Email: matt-goldrick@northwestern.edu

Materials, data, and analysis code available at: https://osf.io/6cghq/

Abstract

Theories of speech production have proposed that in contexts where multiple languages are produced, bilinguals inhibit the dominant language with the goal of making both languages equally accessible. This process often overshoots this goal, leading to a surprising pattern: better performance in the nondominant vs. dominant language, or *reversed language dominance* effects. However, the reliability of this effect in single word production studies with cued language switches has been challenged by a recent meta-analysis. Correcting for errors in this analysis, we find that dominance effects are reliably reduced and reversed during language mixing. Reversed dominance has also consistently been reported in the production of connected speech elicited by reading aloud of mixed language paragraphs. When switching, bilinguals produced translation-equivalent intrusion errors (e.g., saying *pero* instead of *but*) more often when intending to produce words in the dominant language. We show this dominant language vulnerability is not exclusive to switching out of the nondominant language and extends to non-switch words, linking connected speech results to patterns first reported in single word studies. Reversed language dominance is a robust phenomenon that reflects the tip of the iceberg of inhibitory control of the dominant language in bilingual language production.

Keywords: bilingualism, language production, reversed dominance, inhibition

Reversed language dominance is the tip of the inhibitory control iceberg

Over the past several decades, inhibition has served as a cornerstone of theories of bilingual language production. For example, the highly influential proposal of Green (1998) claimed that overlapping semantic representations caused the co-activation of lexical representations in multiple languages (more specifically, of lemmas, which are amodal representations linking the semantic, grammatical, and phonological properties of each lexical item). The resulting cognitive control problem – *how do I speak words only in the language I intend to produce?* – is addressed by using domain-general inhibitory mechanisms (alongside other control mechanisms) to inhibit the activation of representations tagged as belonging to the non-target language. This work was followed by an explosion of behavioral, electrophysiological, and neuroimaging studies seeking signatures of inhibitory control in bilingual language processing (for recent reviews, see Blanco-Elorrieta & Caramazza, 2021; Bialystok & Craik, 2022; Declerck & Koch, in press; de Bruin et al., 2021).

In this work, we focus on a striking empirical pattern that arguably provides one of the strongest motivations for inhibitory language control mechanisms: the phenomenon of *reversed language dominance*. Typically, a bilingual speaker finds it easier to produce words in their dominant vs. nondominant language; for example, a Spanish-English bilingual who has been immersed in an English-dominant environment for the majority of their life might have a larger vocabulary in English, will spend more time speaking English than Spanish, and correspondingly, will retrieve English picture names more quickly than Spanish picture names (Gollan et al., 2008; 2011; for robust language dominance effects in other bilingual populations, see Hanulová et al., 2011; Ivanova & Costa, 2008; Runnqvist et al., 2011). However, as we review in more detail below, studies of a number of bilingual populations, using several paradigms, have found that in mixed language contexts, it is frequently more difficult to retrieve items in the dominant vs. nondominant language — a *reversal* of the typical direction of language dominance.

Critically, reversed language dominance effects are not merely an artifact of asking bilinguals to name single pictures with cued language switching. Reversed dominance has also been observed in studies of connected speech. When bilinguals read aloud mixed language paragraphs, they occasionally produce intrusion errors, in which they avoid switching in their speech by instead producing a translation equivalent word (e.g., when reading *la gente dice that when the…,* producing *que* instead of *that*; Gollan et al., 2014; Kolers, 1966). As we review in more detail below, this paradigm elicits a highly reliable reversed language dominance effect, specifically at points where participants switch languages (i.e., greater rates of intrusion errors when attempting to produce switch words in the dominant vs. nondominant language).

Inhibition – unlike other proposed control mechanisms (as will be explained below) – provides a clear account of such effects. In many models of language production (e.g., Oppenheim et al., 2010), the activation of non-target representations interferes with retrieval and selection of the target word. One mechanism for facilitating target word processing is inhibition of these non-target representations (working in concert with other processes). As proposed by Gollan and Ferreira (2009; see also Declerck et al., 2020; Gollan & Goldrick, 2018; Kleinman & Gollan, 2018), in an attempt to equalize accessibility of the two languages within a mixed language context, speakers use inhibition to reduce the activation of dominant language representations. Reducing the activation of co-activated non-target representations allows nondominant language targets to be more easily accessed and selected. Across individuals and experimental contexts, the use of inhibition is reflected by a continuum of effects reducing the advantage of the dominant language. At the most extreme end of this continuum, speakers "overshoot" the target state of equal accessibility of the two languages, over-inhibiting the language that dominates production in most contexts – producing reversed dominance.

As we discuss in more detail below, alternative approaches to language control, including non-competitive selection (based on Blanco-Elorrieta & Caramazza, 2021), language-specific selection thresholds (Costa & Santestaban, 2004), enhancement of target activation (Branzi et al., 2014; Declerck et al., 2015; Verhoef et al., 2009), and implicit-learning based (Oppenheim et al., 2010, Runnqvist et al., 2019) have great difficulty providing an alternative, comprehensive account for this pattern of results. Reversed dominance is therefore highly informative for distinguishing theories of bilingual language control.

Here we examine two challenges to this theoretically-informative phenomenon. A recent meta-analysis of reversed dominance in picture-naming response times (i.e., slower response times for dominant vs. nondominant language targets) has argued that this effect is neither general nor robust (Gade et al., 2021a). We report a reanalysis that corrects for multiple errors we discovered in the meta-analytic data set, errors in the statistical analysis of the data, and a conceptual misinterpretation of the extant data (i.e., the failure to consider the importance of having an objective measure of dominance effects in single-language testing blocks). Our Bayesian reanalysis estimates a non-zero interaction of language dominance and testing block, such that the standard dominance effect in single language blocks is reversed in mixed language blocks. Distributional analyses of effect sizes are consistent with a continuum of effects, with reversed dominance representing the tip of the iceberg of these effects.

We then report a new study that addresses a missing component of another speech production task that previously revealed robust dominance reversal. In studies of single picture naming language dominance was reversed throughout the mixed language testing block, measured by response times on both switch and non-switch trials – revealing that reversed dominance reflects a proactive control process that applies to the entire testing block, and is not limited to reactive control processes specifically involved in switching. In contrast, studies of connected speech examined the ability to prevent failures of language control (i.e., intrusion errors) and found that these failures very rarely occurred on non-switch words and, as such, showed dominance reversal only at language switch points. To elicit language control failures on both switch and non-switch words we elicited mixed language speech with more frequent language switches than previously reported. This induced more language control failures on both switch and non-switch words and, critically, revealed fully reversed language dominance. This establishes a clearer parallel to previously-reported findings in single word production studies (elicited via naming of single pictures or words), strengthens the claim that reversed dominance arises in more global control mechanisms that operate on both switch and non-switch trials, and reveals that the vulnerability of the dominant language to intrusion errors in previous studies was not merely an artifact of completing a more difficult task (i.e., reading paragraphs written primarily in the nondominant language).

Taken together, these results substantially strengthen the empirical support for reversed dominance effects during bilingual language mixing. We conclude by discussing the challenges these results pose for alternative models of bilingual language control and the broader implications of reversed dominance for theories of language control.

Data availability

Materials, data, and analysis code for the meta-analysis and empirical sections of the paper are available at: https://osf.io/6cghq/ (see https://osf.io/g4kza/wiki/home/ for a guide to the files associated with the re-analysis of Gade et al. (2021a), and https://osf.io/khaxc/wiki/home/ for a guide to the materials and analyses of language control errors in connected speech).

Meta-analytic evidence: A re-analysis of Gade et al. (2021a)

A standard laboratory paradigm for examining bilingual language control is the cued picture naming paradigm, where the language in which the picture should be named is specified by a cue (e.g., color; Meuter & Allport, 1999). While many studies using this paradigm show standard dominance effects in mixed language blocks, reversed dominance effects – longer reaction times to name pictures in the dominant vs. nondominant language – have been observed in a number of studies (see Declerck & Koch, in press, for a review). This includes a range of different bilingual populations, including: Dutch-English (Verhoef et al., 2009); Finnish-English (Jylkkä et al., 2018); German-Dutch (Christoffels et al., 2006); German-English (Heikoop et al., 2016); Mandarin-English (Liu et al., 2019); Spanish-Catalan (Costa & Santestaban, 2004); and Spanish-English bilinguals (Kleinman & Gollan, 2018). Declerck et al. (2020) documented variability in dominance effects within a large (*N* = 286) group of Spanish-English bilinguals performing cued picture naming in both single and mixed language blocks. They found a consistent reduction in dominance effects in mixed vs. single language blocks. Using an objective, validated measure of proficiency in each language (the Multilingual Naming Test, MINT; Gollan et al., 2012), this reduction in dominance was found to be *larger* for less-balanced vs. more-balanced bilinguals, mirroring the larger dominance effects in single language blocks for the former vs. the latter participants.

While reversed dominance effects have frequently been observed in studies analyzing reaction times, as noted above there are many cases in which this pattern is not observed (Declerck & Koch, in press). This pattern of some bilinguals reversing dominance while others do not has also been observed even within individual studies and has been interpreted as a continuum of reduction in dominance effects (Declerck et al., 2020). However, Gade et al. (2021a; see also Declerk & Koch, in press) adopted a different interpretation, instead seeing the dominance reversal as being not “robust” or “replicable.” In support of this interpretation, Gade et al. (2021a) reported a meta-analysis of 73 studies, which they claimed revealed no systematic advantage or disadvantage of the dominant vs. nondominant language in mixed blocks. Note that a correction to the meta-analysis (Gade et al., 2021b) has already clarified that a reverse dominance effect was observed in an analysis considering only studies with a short interval between the language cue and appearance of the target picture.

However, a number of concerns led us to question the bottom line in Gade et al. (2021a). Most critically, the meta-analysis included very few studies with objective, validated measures of language dominance. This is central to any analysis of the reversed dominance effect. As noted above, the claim is that this phenomenon is an extreme example of a more general reduction in dominance effects – supported by previous work showing the importance of baseline differences between single-language blocks (Declerck et al., 2020). An auxiliary analysis by Gade et al. (2021a) attempted to address this by examining data from 19 publications (with a total of 40 experiments) including data from both single and mixed blocks. In this analysis, Gade et al. again found no reliable effect of dominance, i.e., no systematic advantage or disadvantage of the dominant vs. nondominant language in mixed language blocks and puzzlingly also not in single language blocks. The absence of significant language dominance effects — especially in single-language blocks — increased our skepticism of the analysis, given robust reports of dominant language advantages across many studies and bilingual populations (Gollan et al., 2008; Hanulová et al., 2011; Ivanova & Costa, 2008; Runnqvist et al., 2011). Critically, when neither language is clearly dominant in single language blocks, and without an independent measure of dominance, it becomes impossible to say if dominance was reversed or not in mixed language blocks.

Two other aspects of the statistical methods for Gade et al.'s (2021a) auxiliary analysis were also a source of concern. Rather than utilizing a statistical model that explicitly captured the difference between non-switch and switch trials (i.e., where the previous trial was the same vs. different from the current trial) in mixed language blocks, they analyzed the mean reaction time collapsing across switch and non-switch trials in mixed language blocks. This aggregation deprives the statistical model of information about the structure of variation present in the data set, which can distort results (see, e.g., Lachaud & Renaud, 2011, for discussion). Additionally, Gade et al. failed to adjust for the skewed distribution of reaction times. This has a dramatic impact on the power of both frequentist (Baayen & Milin, 2009) and Bayesian analyses (Nicenboim et al., 2018).

Finally, an intrinsic limitation of the meta-analytic data set is its heterogeneity. It includes tasks that are believed to place different demands on language control (e.g., digit naming: Declerck et al., 2012; voluntary switching: Gollan & Ferreira, 2009) and populations that may have reduced control abilities (e.g., older bilinguals: Stasenko et al., 2021). Gade et al. (2021a) presented another auxiliary analysis to address some of these issues by focusing solely on data from cued switching and found similar results. However, this analysis did not simultaneously control for differences across populations; furthermore, it did not consider the substantially heterogeneity of cued switching paradigms. For example, within cued paradigms the rate of switching (Olson, 2016), amount of repetition (Kleinman & Gollan, 2018), and nature of the cue (Lavric et al., 2019) may substantially modulate reaction times in the paradigm. Failure to properly account for these substantial sources of variation (e.g., by attributing such differences to random variation in effects across studies) may substantially reduce the power of the meta-analysis.

In this section, we report a reanalysis of the data from studies including both single and mixed language blocks, and find that (contra Gade et al.) there is robust evidence for a reduction of dominance effects in mixed blocks, with reversed dominance at the extreme end of this continuum.

Data cleaning

Inspection of the data file used in the Gade et al. (2021a)[1] auxiliary analysis revealed an additional, serious issue; in several cases, recorded values did not appear to match the figures

[1] To generate the relevant data set, we downloaded the data file (https://osf.io/69vnx/) and R analysis script (https://osf.io/4vcbx/), executing the code following the comment "*language dominance auxiliary analyses*" (lines 713-743).

in published papers. Both authors therefore reviewed the full data file, independently checking for discrepancies between the original data files and the source papers. For all deviations found, both authors confirmed that the error was present in the original file. Authors of the original papers (see acknowledgements) were then contacted for all cases where recorded data did not appear to match figures, or cases where the recorded values were not available in the paper or figures. We identified errors in 10/41 sets of results (plus 1 additional discrepancy that was off by just 1 millisecond). Incorrect values were used in 7/10 cases; in two cases, dominance effects were recorded as reversed in single language blocks when there was, in fact, a dominant language advantage. One condition from a study was incorrectly omitted, and in 2 cases the original author provided the incorrect data to Gade et al. (the corrected data file with detailed notes on corrections verified by both authors is available at https://osf.io/g4kza/wiki/home/).

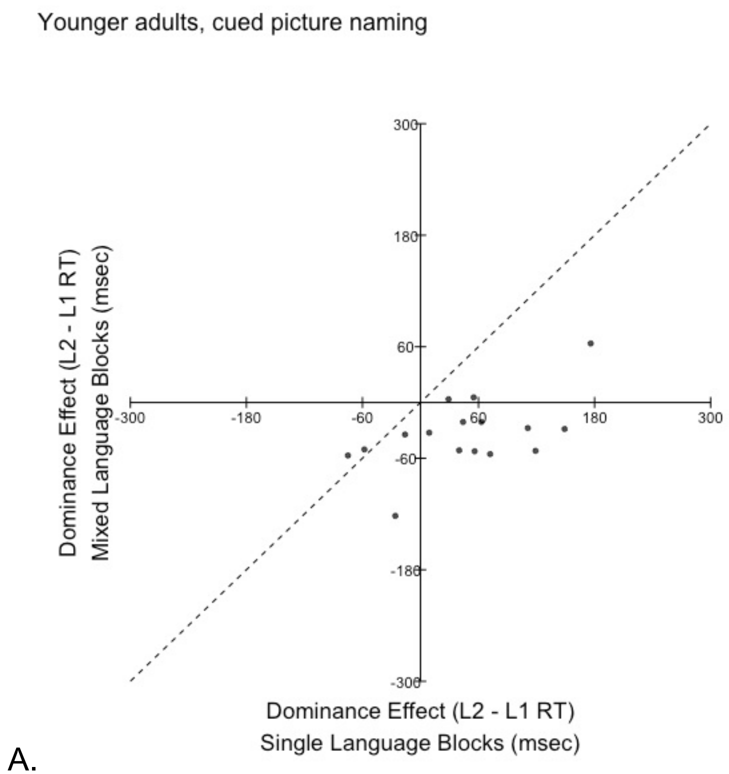
Results

Visualizations

We first visualized the reduction of dominance effects in mixed vs. single language blocks. This is true for the vast majority of studies in this dataset. The most frequent type of study is cued picture naming with younger adults (Figure 1A); almost 90% of these studies show a reduction of dominance in mixed- relative to single-language blocks. The full dataset (adding back various age-groups and task types) shows a similar pattern, although somewhat reduced (Figure 1B; nearly 70% show a reduction in dominance). We should take this latter result with caution, as this data set collapses across several dimensions (populations, tasks) that likely reduce the ability to apply (e.g., aging), or the need to apply (e.g., voluntary switching), inhibitory control (see above for discussion).

Figure 1 reveals not only the reduction of dominance, but the frequent occurrence of reversed dominance (the bottom right quadrant of each scatter plot). Figure 2 situates this phenomenon within the continuum of reduction of dominance focusing on studies that showed normal dominance effects (i.e., a dominant language advantage) in pure blocks. To control for overall differences in RTs across studies, this visualizes the change in dominance across single and mixed blocks relative to the dominance effect in single blocks. Both when limiting the analysis to studies with younger adult bilinguals and cued picture naming (Figure 2A) and in the full dataset (Figure 2B; which included a variety of age groups and different types of switching paradigms), there is clear variability in the degree to which dominance is reduced. The mode of the distribution is slightly greater than 1. This could be viewed as a slight overshoot when the language control system aims to have both languages equally accessible (consistent with the account offered in the introduction; see also Declerck et al., 2020; Gollan & Ferreira, 2009; Gollan & Goldrick, 2018; Kleinman & Gollan, 2018).

It should be noted that in both Figure 1A and in Figure 1B there are a number of studies (25% in the former, roughly 30% in the latter) that fail to show normal language dominance effects in single language blocks. This puzzling pattern raises question as to which (if any) language was in fact dominant in these studies. Inspection of the experimental design and populations used in these studies failed to reveal a consistent source of the origin of these results; we return to this issue in the General Discussion.



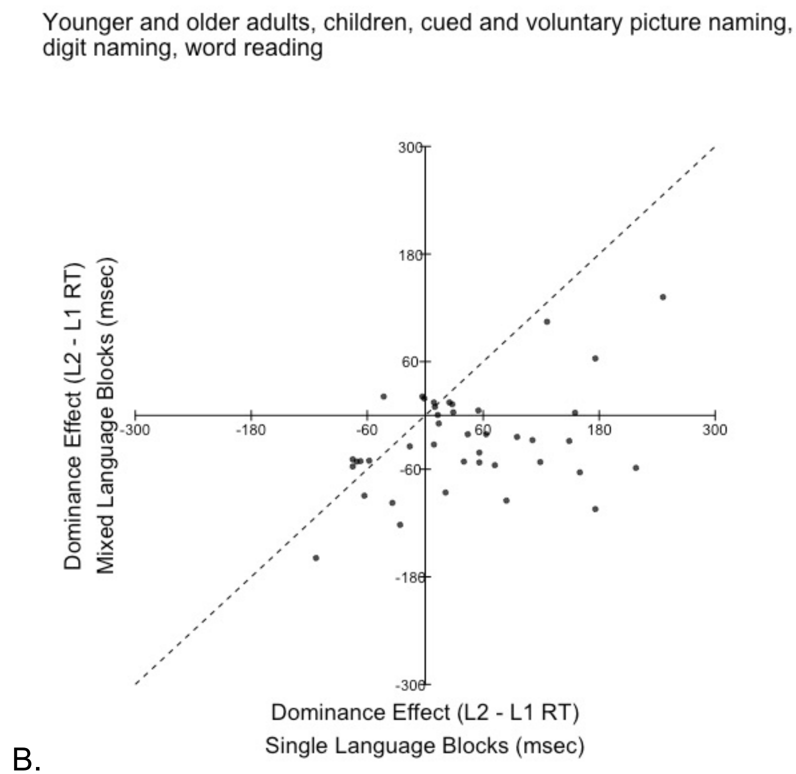
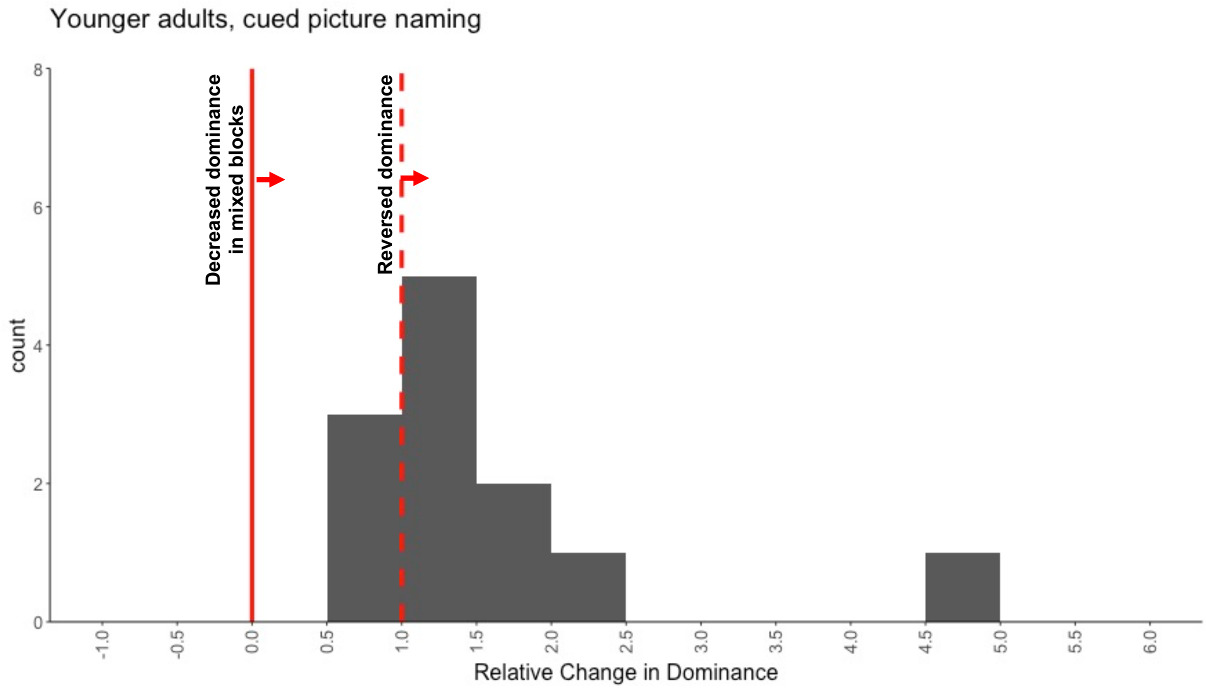
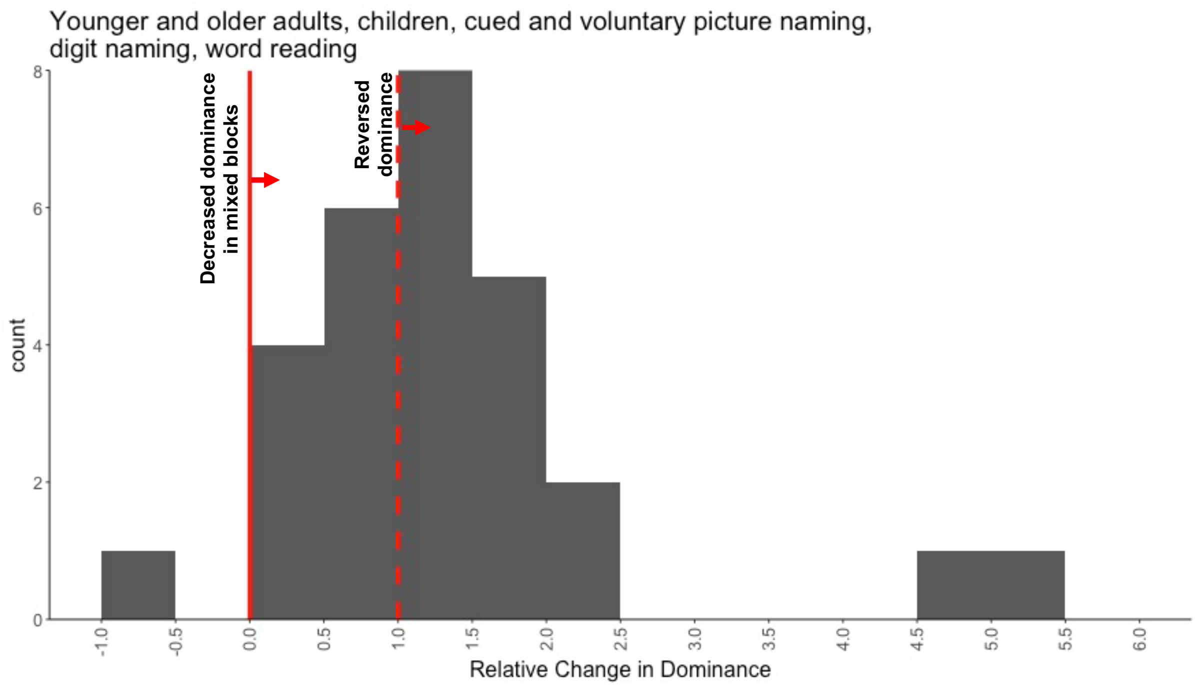


Figure 1. Dominance effects (mean reaction time (RT) for nondominant – dominant trials) in meta-analysis by Gade et al. (2020) (with errors in their data summary from 10/41 observations corrected). Points below the $y = x$ line (dashed) show decreased dominance effects in mixed language blocks. A: Study conditions with younger adults, cued picture naming (N = 16; 75% show a dominant language advantage in single language blocks; 88% show decreased dominance effects in mixed blocks). B: All study conditions included in meta-analysis dataset (N = 41; 68% show a dominant language advantage in single language blocks; 78% show decreased dominance effects in mixed blocks).



A.



B.

Figure 2. Relative change in dominance. *Note*: Graphs only show studies that revealed normal language dominance effects in single-language blocks, i.e., faster RTs for the dominant than for the nondominant language. Dominance = RT for L2 – RT for L1; Relative Dominance = (Dominance Single – Dominance Mixed)/Dominance Single. Zero indicates no change in dominance across blocks. Positive values show decrease in dominance in mixed blocks; values greater than 1 show that dominance is reversed in mixed blocks. Negative values show the one case where dominance effects increased in mixed blocks. A: Study conditions with younger adults, cued picture naming, with faster RTs for dominant language in pure blocks (*N* = 12). B:

All study conditions included in meta-analysis dataset with faster RTs for dominant language in pure blocks (*N* = 28).

Bayesian analysis

To statistically assess whether dominance is reduced in mixed blocks, we analyzed the corrected data using the same framework of Gade et al. (2021a; Bayesian mixed-effects regression with weakly informative priors). However, we adopted a different model structure and transformed our dependent variable; we will examine the impact of these contrasting assumptions (and other differences from Gade et al.) below. Analysis scripts are available at https://osf.io/g4kza/wiki/home/.

Reaction times (RTs) were strongly right skewed; natural log transformation of the RTs reduced this skew, and, relative to a model fit to raw RTs (see below) resulted in residuals that were closer to a normal distribution (particularly in the right tail of the residual distribution) and showed less evidence of heteroscedasticity. The difference between conditions was analyzed by Helmert contrasts: *block type* comparing single language blocks (coded as –2/3) vs. non-switch and switch trials (each coded as 1/3); and a second, *trial type*, contrasting non-switch (–0.5) vs. switch trials (0.5; single language trials were coded as 0). The model included a main effect of language dominance (dominant, –0.5; non-dominant, 0.5) and interactions of dominance with block type and trial type. Following Gade et al. (2021a), we included a random intercept for each study in the meta-analysis (see below for further discussion).

Table 1 reports the model results. As expected, mixed block trials were reliably slower than trials in single language blocks, and to a smaller degree switch trials were slower than non-switch trials. The credible interval of the main effect of dominance included 0, as did the interaction of dominance and trial type. By contrast, there was evidence for an interaction (the credible interval did not include 0) of language dominance and block type (mixed- vs. single language blocks). This interaction reflected a reversed dominance effect: the estimated marginal means for each condition (from the package *emmeans*, back-transformed from log to raw RTs) show a standard dominance effect in single-language blocks (dominant: 793 msec; non-dominant: 832 msec) and a reversed dominance effect in mixed language blocks (dominant: 911 msec; non-dominant: 886 msec).

Table 1. Posterior mean, standard error, 95% credible interval and $\hat{R}$ statistic (an index of convergence) for parameters in the full model analyzing log RTs. **Bold** shows parameters where the 95% credible interval does not include 0.

Parameter	Mean	S.E.	Lower 95%	Upper 95%	$\hat{R}$
Dominance (L1 vs. L2)	–0.003	0.013	–0.029	0.024	1.000
Block Type (Single, Mixed)	0.101	0.014	0.073	0.129	1.000
Trial Type (Non-Switch, Switch)	0.068	0.016	0.036	0.101	1.001
Interaction: Language (dominant, nondominant) by Block Type (Single, Mixed)	–0.076	0.028	–0.131	–0.019	1.000
Interaction: Language (dominant, nondominant) by Trial Type (Non-Switch, Switch)	–0.006	0.033	–0.071	0.061	1.000

We then considered what properties of our analysis contributed to the divergence of our findings from those of Gade et al. (2021a). Table 2 summarizes the central findings; full model results are reported in Appendix A.

Table 2. Summary of issues with alternative approaches to the meta-analysis. Each alternative model differed from the full model reported in Table 1 by a single property, shown in the first column (see text for details).

Alternative Approaches to Meta-Analysis	**Result**
No single language trials: Regression model does not include single language baseline	Failure to detect reversal of dominance
Collapse stay, switch: Model does not distinguish stay vs. switch trials; data includes mean for all mixed block trials	Less precise estimate of size of reversal of dominance.
Raw RTs: Model failing to account for skewed residuals due to raw RTs	Failure to detect reversal of dominance.

Most importantly, Gade et al.’s (2021) main analysis does not include data from single language blocks. This reflects a key limitation of the source data for their meta-analysis: most studies did not include single language blocks. This is a fundamental flaw: investigation of dominance effects *must* include an objective assessment of performance in single language conditions, as language mixing has a substantial impact on the relative accessibility of the dominant and non-dominant languages (Declerck et al., 2020). The limitations of this analysis can clearly be seen when, mirroring Gade et al.’s main analysis, we limit the main model to log RTs from single language blocks only (holding all other aspects of the model constant). There is no reliable effect of dominance: the credible interval for the main effect and interaction with Trial Type (non-switch vs. switch) both include 0 (main effect of dominance: β = –0.028, s.e. = 0.016, 95% credible interval (–0.058,0.003), $\hat{R}$=1.001; interaction of dominance with Trial Type: β = –0.006, s.e. = 0.031, 95% credible interval (–0.067,0.055), $\hat{R}$=1.000; see Appendix A, Table A1, for full model results). Properly assessing dominance effects requires the inclusion of single-language block data.

Gade et al.(2021a) claimed their auxiliary analysis took into account single language blocks. However, rather than modeling the difference between non-switch and switch trials, they averaged across the two trial types. Repeating our analysis using these averages (while holding all other aspects of the statistical model constant) widens the credible interval for the dominance by block type interaction relative to our full model. The 95% credible interval for the interaction of dominance and block type on this model (β = –0.076, s.e. = 0.033, $\hat{R}$=1.000) has a more negative lower bound than the full model (–0.141 vs. –0.131 in the full model; see table 1) and a less negative upper bound (–0.012 vs. –0.019; see Appendix A, Table A2, for full model results). In order to obtain the best estimate for effects, it's important that a model have access to key information present in the data, rather than a reduced representation that collapses across key data points. Analyses of dominance effect must therefore explicitly separate out both single vs. mixed *and* non-switch vs. switch trials to assess interactions with dominance.

Finally, our model took into account the skew in reaction times through log transformation; Gade et al.'s analyses (both the auxiliary and main analysis) did not. As noted above, the residuals of a model fit to log transformed RTs better fit the assumptions of the linear regression than the model fit to raw RTs. Repeating our analysis, using raw RTs as the dependent measure (holding all other aspects constant) provides suggestive evidence for a reduction in dominance effects, but the credible interval now includes 0 (dominance by block type interaction: β = –7.740, s.e. = 9.548, 95% credible interval (–26.533, 11.216), $\hat{R}$=1.001; see Appendix A, Table A3, for full model results). This shows the importance of checking that regression model assumptions are met.

As noted above, Gade et al. also analyzed a data set with a substantial number of errors. We repeated our analysis, including these errors, and found no major differences. This suggests the patterns we have documented are quite robust in the face of errors in the data.

Finally, as discussed above, modeling differences between study types simply using a random intercept is not ideal, given the heterogeneity of the paradigms used in this meta-analysis. While the data set is too limited to provide a robust examination of this issue, we conducted an exploratory analysis, repeating the main analysis and including a factor contrasting studies of cued picture naming in younger adults (N = 16) vs. other paradigms and populations (N = 25), interacting with all predictors in the main model. There was no robust evidence for a strong effect of study type; credible intervals for interactions with study type included 0.

Discussion

Re-consideration of data from studies included in Gade et al.'s (2021a) meta-analysis reveals a reliable reduction in dominance effects in mixed vs. single-language blocks. The reduction in dominance tends to be slightly greater than the dominance effects observed in single language blocks, yielding reversed dominance effects in a number of studies. Our reanalysis has also highlighted key methodological aspects of assessing reversed dominance; most importantly, studies must explicitly assess dominance effects in single language blocks to properly interpret performance in mixed language blocks.

Reversed language dominance in reading aloud

Given that naturalistic mixed language production is typically not produced one word at a time (the meta-analysis only assessed naming of single pictures, words, or digits), it's important to verify that reversed dominance effects are also found in connected speech. To elicit connected speech with experimental control over the content of speech and the rate of language switches a number of recent studies used reading aloud of full paragraphs with a small number of language switches, another laboratory paradigm for the study of bilingual language control (Gollan et al., 2014; Kolers, 1966). In this task, reversed dominance effects were measured in the form of greater rates of intrusion errors on dominant vs. nondominant language targets at language switch points. Reversed dominance on intrusion errors have been consistently reported in a number of language combinations including Hebrew-English (Fadlon et al., 2019), Mandarin-English (Li & Gollan, 2018; Schotter et al., 2019), and Spanish-English bilinguals (Ahn et al., 2020; Fadlon et al., 2019; Gollan & Goldrick, 2016, 2018; Gollan et al., 2014; Gollan et al., 2017, 2020; Stasenko et al., 2020).

Importantly, an abundance of evidence indicates that errors in this paradigm reflect difficulties in control over language production, rather than failures of visual perceptual processes or lapses of attention. Intrusion errors occur even when the two words look nothing alike (e.g., *pero* and *but* in Spanish-English bilinguals; *or* and *或* in Mandarin-English bilinguals). Additionally, errors are more likely to occur on function vs. content word targets – a pattern also found in spontaneous speech elicited via conversation (Poulisse & Bongaerts, 1994). While the vulnerability of function words in reading aloud might appear to have a ready explanation as a perceptual error (as function words are skipped more often than content words in silent reading; Saint-Aubin & Klein, 2001), in studies with eye-tracking most intrusion errors were produced when bilinguals directly fixated their gaze on the target switch word (or another word in the same language; Gollan et al., 2014), and increased gaze duration did not increase selfcorrection of intrusions for function word targets (Schotter et al., 2019). This suggests intrusions on function words, intrusions in reading aloud, and intrusions in bilingual language production more broadly, reflect difficulties in control over spoken language production.

The read-aloud task has consistently shown reversed dominance at switch points, with more language intrusion errors occurring on switch words in the dominant language (more errors on *few* in *…muy few gente*) vs. the nondominant language (than errors on *poca* in *…very poca people* in English-dominant bilinguals. However, as seen in the previous section, reversed language dominance effects in single word production are observed across both language switch and non-switch trials in mixed language blocks, suggesting that reversed dominance is a general phenomenon of language control in mixed blocks (not specifically a feature of reactive control processes operating on switch trials). To examine whether reversed dominance effects extend to non-switch words, we significantly increased the switching rate, allowing us to observe more targets in each language at both switch and non-switch points. This also allowed us to examine a second limitation of previous work. In these studies, all dominant language switch words— a key data point for observing reversed dominance effects — were presented in paragraphs written mostly in the nondominant language (e.g., there were no English non-switch words in paragraphs written mostly in Spanish). This raises the possibility that reversed dominance effects are not due to properties of the target word but rather reflect the overall difficulty of reading paragraphs in the non-dominant language. By increasing the switching rate, we were able to observe switching on dominant-language words within paragraphs primarily written in the dominant language.

Method

Participants

A total of 48 Spanish-English bilinguals were included in the analyses reported below. Bilinguals were given course credit for their participation through undergraduate classes at the University of California, San Diego (UCSD). Table 2 shows self-reported participant characteristics and ability to name pictures in each language on the MINT Sprint Test (Garcia & Gollan, 2021), which is a fast-administration expanded version of the Multilingual Naming Test (MINT; Gollan et al., 2012). This provided objective and validated measures of English and Spanish proficiency, which we used to determine language dominance. The participants, like most Spanish-English bilinguals at UCSD, have Spanish as their first language but are dominant in English due to extended immersion and English-based schooling. To ensure our analyses below were conducted over a relatively homogeneous group, we examined the performance of 48 English-dominant speakers out of 57 total participants tested (Spanish-dominant bilinguals were excluded and replaced). The number of participants was determined by an a-priori power analysis, which estimated the power to detect differences between dominant and nondominant languages at this sample size as exceeding 0.99 (see Appendix B for details).

Table 2. Participant characteristics for all English-dominant Spanish-English bilinguals

Characteristic	*M*	*SD*
Age	21.13	3.02
Years of Education	13.98	1.44
Age of English acquisition	4.75	2.31
English spoken proficiency self-rated[a]	6.71	0.48
Spanish spoken proficiency self-rated[a]	6.04	0.65
English MINT Sprint score[b]	64.75	4.35
Spanish MINT Sprint score[b]	52.56	8.68
Bilingual Index score[c]	0.81	0.14
Percent English use at home in childhood	22.40	18.93
Current percent English use	65.29	22.79
Current switching frequency[d]	3.79	1.38

Note. MINT = Multilingual Naming Test
[a] Self-rated proficiency level was averaged across ratings for speaking, comprehension of spoken speech, reading, and writing on a scale from 1 (little to no knowledge) to 7 (like a native speaker). [b] Maximum possible score is 80. [c] MINT Sprint score in Spanish/English. [d] Self-rated estimate of how often bilinguals switch languages when speaking with other bilinguals who know the same languages; the 6-point rating scale included the following anchors: 1 (*never*), 2 (*rarely*), 3 (*sometimes*), 4 (*often*), 5 (*most of the time*), and 6 (*almost constantly*).

Materials and Procedure

Two native Spanish-English bilinguals modified 16 English-dominant paragraphs (from Gollan & Goldrick, 2016), aiming to create paragraphs with highly frequent switching while avoiding as much as possible any switches that occurred in positions that seemed to grossly violate naturalistic switching points, and also maintaining a clear *default language*, i.e., the language that the majority of words were written in, and that dominated syntactic structure. Each paragraph appeared in both Spanish- and English-default conditions (counterbalanced across participants; see Appendix C for examples). Relative to previous experiments, these materials created many more opportunities to observe intrusion errors on target words in both languages, at both switch and non-switch points, in paragraphs with either Spanish- or English-default.

This design addresses the two weaknesses of previous work discussed above. Previous studies could only examine dominant-language switch words in paragraphs that were almost entirely written in the non-dominant language. These new materials allowed us to examine dominant-language target words in paragraphs mostly written in the dominant language. Furthermore, by creating more opportunities to observe words in each language at non-switch points, we can, for the first time, examine whether reversed dominance effects extend to non-switch words.

Note that there is one confound inherent to this paradigm. The paragraphs have two switch types – *switch out* of the default language and *switch into* the default language. These are inherently confounded with target word language. For example, in an English-default paragraph, switching on an English target will necessarily involve switching back into the default language, whereas switching on a Spanish target will necessarily involve switching out of the default language. This is critical because in previous studies the default language provides support for retrieval of lexical items in the default language, making switching out much more difficult than switching back to the default language (Goldrick & Gollan, 2018). Therefore, when we test for reversed dominance effects in dominant language paragraphs, support from the default language for selection of targets in the dominant language may reduce the effects we observe.

To confirm that these new materials showed patterns similar to previous results, in addition to target language we manipulated part of speech of switch targets (function vs. content; following the categorization criteria of Bell et al., 2009). Studies with the read aloud task have shown that intrusion errors tend to occur on function words (e.g., Fadlon et al., 2019;

Gollan & Goldrick, 2016, 2018; Gollan et al., 2014, 2020; Schotter et al., 2019). This likely stems from multiple sources. In contrast to content words, function words are highly dependent on grammatical encoding processes for retrieval (e.g., Myers-Scotton & Jake, 2009). The strong activation of syntactic properties of the non-target language at switch points make function word retrieval particularly challenging (Gollan & Goldrick, 2018). Speech output is also more difficult to monitor for function vs. content words (Schotter et al., 2019). In each paragraph, we therefore aimed to have at least two instances of switch words in each one of four conditions, crossing language (English vs. Spanish) with part of speech (content vs. function).

Descriptive statistics for the paragraphs are shown in Table 3. The paragraphs were divided into two lists, such that each bilingual read 16 paragraphs, 8 with English as the default language and 8 with Spanish as the default language. Between participants, each of the two lists was presented an equal number of times in English-default first vs. Spanish-default first, resulting in four groups of participants (with 12 bilinguals in each group). Within each group the 8 paragraphs were always presented in the same fixed order.

Table 3. Properties of English and Spanish default paragraphs.

	English Default		Spanish Default	
	M	*(Min, Max)*	*M*	*(Min, Max)*
Total Word Length	109.00	(88, 127)	105.69	(86, 126)
Words in Default Language	77.50	(58, 94)	73.19	(57, 93)
Content Switches Back to Default	4.38	(2, 8)	4.44	(2, 7)
Content Switches Out of Default	5.06	(2, 8)	5.31	(2, 10)
Function Switches Back to Default	9.00	(5, 11)	8.94	(6, 12)
Function Switch Out of Default	8.75	(6, 14)	8.56	(6, 13)

Participants signed virtual consent forms prior to meeting with an experimenter via ZOOM for a testing session that was audio and video recorded. During the testing session they completed the tasks in the following order: a language history questionnaire, the paragraphs task, and the MINT Sprint. Paragraphs were presented in on a Power Point display in Calibri Body font 18 (left justified starting at the top left of each slide). Participants either read 8 English-default paragraphs first, or 8 Spanish-default paragraphs first (see below) with each group of 8 paragraphs preceded by a shorter practice paragraph written in the same default language. Participants were instructed as follows: “In this task, you will be reading paragraphs aloud. Please read each paragraph as accurately as you can and at a comfortable pace. You will notice that the paragraphs switch languages sometimes. Just try to read aloud as fluently as you can without making mistakes. Do you have any questions?”

Results

Data and analysis scripts are available at https://osf.io/khaxc/wiki/home/. Building on previous work (Gollan & Goldrick, 2016; Gollan et al., 2014) two native Spanish–English bilingual research assistants transcribed errors. Our analyses (like our a-priori power analysis) focused on intrusions (*N* = 607; e.g., saying *el* instead of *he*; but see Supplemental Materials for an analysis that included intrusions, partial intrusions and accent errors and produced largely the same results). All participants produced at least 2 intrusion errors and as many as 37 (*M*=12.6, *SD*=7.9). A small number of intrusion errors (*N* = 14) were cases in which bilinguals said the correct switch word and then quickly self-corrected to an intrusion error (e.g., the switch word was *de* and the participant said “*de*…*of*…”) and another small set (*N* = 4) were coded as multiple error types (e.g., an intrusion followed by an accent error). These were all coded as single intrusion errors. We also coded partial intrusions (*N* = 202; self-corrected intrusion errors). While these resemble intrusions errors, they have a different distribution (they are less likely to occur on function words: Gollan & Goldrick, 2016; Gollan et al., 2014;

Poulisse & Bongaerts, 1994). A small number of partial intrusions were correct productions that were self-corrected to partial intrusions (*N* = 2). Accent errors (*N* = 152; saying the correct word with the accent of the non-target language) were coded, but are difficult to detect on short function words, and rely on subjective judgments of deviation from canonical pronunciations (see also Kolers, 1966). A small number of accent errors were correct productions that were self-corrected to partial intrusions (*N* = 2). A small number of errors (*N* = 3) were coded as both accent and partial intrusions; these were coded as single partial intrusion errors. Within-language errors, omissions, and insertions were not transcribed.

Switch words

Statistical model structure. Though we tried to avoid cognates as much as possible in the materials it was not possible to avoid them entirely (especially content words). Because cognates likely elicit different control mechanisms in reading aloud (Filippi et al., 2014; Gollan et al., 2014; Li & Gollan, 2018) and cognate status was not controlled in our materials (resulting in just one or two cognate targets in some conditions) we focused our analysis on intrusion errors on non-cognate[2] switch words. To maintain parallelism with all previous studies with this paradigm, we adopted a frequentist statistical approach to analyze the data. Responses were analyzed with a logistic mixed effects regression (Dixon, 2008; Jaeger, 2008), including contrast-coded part of speech (content, –.5 vs. function, .5), switch type (switch back to default, –.5 vs. switch out of default, .5), target language (nondominant Spanish, –.5, vs. dominant English, .5) and all interactions as fixed effects. The by-participant random effect structure, determined by

---

[2] We also excluded an even smaller number of false cognates (e.g., *soy* which means *I am* in Spanish) and proper names.

an iterative procedure aiming for the maximal model appropriate for the data[3], included a random intercept with correlated random slopes for part of speech and target language.

Descriptive statistics. The mean by-participant rate of intrusions is shown in Figure 3, separated by condition. The key result is the greater rate of errors on English targets (dark bars) vs. Spanish targets (light bars). Error bars show confidence intervals as estimated by a bootstrap procedure (where the distribution of a statistic is estimated by repeatedly resampling from the observations with replacement; here, using 1,000 replicates). We used this method throughout, as our primary measure is non-normally distributed proportions (violating the assumptions of measures like standard error).

[3] Starting with the maximal random effect structure (Barr et al., 2013), random effect correlations were removed, followed by removal of random slopes in order of complexity (i.e., three-way interactions were removed, then two-way, then main effects) until the model converged to a non-singular fit. Following Bates et al. (2015) we checked the converged model for over-fitting, excluding random slopes in order of variance magnitude. The parsimonious model was then compared to a model with correlations in the random effects; this was retained if it converged and showed a better fit to the data as assessed by a likelihood ratio test.

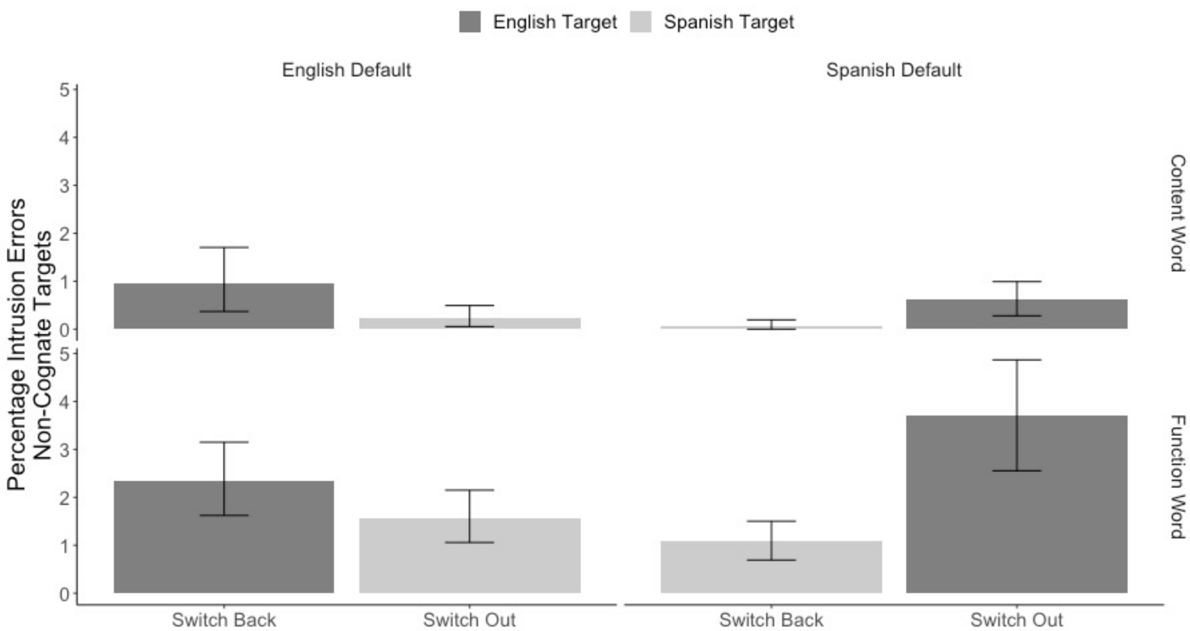


Figure 3. Intrusion errors on non-cognate targets, separated by default language (English, left vs. Spanish, right), part of speech (content, top, vs. function, bottom), and switch type (switch back to the default language vs. switch out of the default). Colors show language of switch word targets; reversed dominance effects can be seen by the relative height of the dark (English) vs. light (Spanish) bars. Error bars show bootstrapped 95% confidence intervals.

Main effect: Reversed dominance. Replicating previous work reviewed above, there was a significant reversed dominance effect on switch targets, a main effect of language such that intrusion errors were more likely to occur on English (dominant language) vs. Spanish (nondominant language) targets (β = 1.37, s.e. β = 0.34, $\chi^2(1)$ = 21.89, $p < .0001$; for all models, p values are based on a likelihood ratio test).

Reversed dominance is found for switching out and switching back. Importantly, language did not interact with switch type (β = –0.77, s.e. β = 0.61, $\chi^2(1)$ =11.86, $p < .18$), suggesting that the difficulty with English word targets held regardless of whether they occurred when switching out of default Spanish (as examined in previous work) or switching back into default English. As shown in Figure 3, in all conditions the dark grey bar is higher than the lighter grey bar. To confirm this important finding, we repeated the analysis within each switch type[4], comparing errors across Spanish- vs. English-default paragraphs. Consistent with previous work, we found a significant reversed dominance effect for switches out of default (β = 0.93, s.e. β = 0.30, $\chi^2(1) = 10.68$, $p < .005$). Extending previous findings, we also found a significant reversed dominance effect also for switches back to default (β = 1.70, s.e. β = 0.53, $\chi^2(1) = 21.55$, $p < .0001$).

Reversed dominance effects are found for both content and function words. Also replicating previous work (Gollan & Goldrick, 2016, 2018; Gollan et al., 2014), there was a significant main effect of part of speech. As shown in Figure 3, intrusion errors were more likely to occur on function words (bottom row) vs. content words (top row; β = 2.06, s.e. β = 0.43, $\chi^2(1) = 40.22$, $p < 0.0001$). This interacted with target language (β = –1.25, s.e. β = 0.65, $\chi^2(1) = 4.71$, $p < 0.03$). Subset models[5] showed that while reversed dominance effects were significant for both parts of speech, the relative gain in probability of making an intrusion error was larger for content words (β = 1.80, s.e. β = 0.60, $\chi^2(1) = 14.92$, $p < 0.0002$) than function words (β = 0.84, s.e. β = 0.13, $\chi^2(1) = 40.22$, $p < 0.0002$). This likely reflects a floor effect; in both English- and Spanish-default paragraphs, Spanish content words show virtually no errors — clear evidence of reversed language dominance.

Switch type effects. Although the focus of our analysis was on reversed dominance effects, the comparison of switch directions (switches back vs. switches out of the default language) provide insight into important aspects of language control._While in the same

[4] Due to convergence issues, these subset models only included random intercepts; random slopes were omitted.
[5] Due to convergence issues, these subset models only included random intercepts; random slopes were omitted.

direction as observed in previous work (Gollan & Goldrick, 2018), switch type effects failed to reach significance (i.e., switches back to the default language elicited as many errors as switches out of the default language – seeming to imply no benefit of default language selection on speech production), neither the main effect (β = 0.41, s.e. β = 0.30, $\chi^2(1)$ = 2.14, $p <$ 0.15), nor the two-way interactions with language (see above) and part of speech (β = 0.02, s.e. β = 0.61, $\chi^2(1) < 1.0$, $p < 0.98$), nor the three-way interaction (β = 1.79, s.e. β = 1.21, $\chi^2(1)$ = 2.57, $p < 0.11$).

To further explore default language effects, we conducted an additional analysis that included all language control errors; in additions to intrusions, we included partial intrusions and accent errors. As shown in Supplemental Materials, this analysis showed a significant main effect of switch type, with more errors occurring when switching out of default vs. switching back. This suggests default-language selection did occur in these materials, but was weaker relative to previous read-aloud studies which had lower switch rates (in those studies, practically no intrusion errors were observed when switching back to default).

Non-switch words

Statistical model structure. Having replicated many of the results in previous studies of the paragraph task, we turned to the non-switch targets. As in previous studies (Goldrick & Gollan, 2016, 2018), the mean intrusion error rate was much lower on non-switch targets (0.32%) vs. switch targets (1.97%). We analyzed these data in two ways. Parallel to the analyses above, we conducted an analysis on intrusion error rates – an *error rate logistic regression model* (as shown in Figure 4), including contrast-coded target word language (nondominant Spanish, –.5, vs. dominant English, .5), contrast-coded default language of the paragraph

(Spanish, –.5, English, .5), and their interaction as fixed effects. A by-participant random intercept was included. In addition, given that the extremely low (or zero) error rates shown in Figure 4 will be difficult for this *error rate* model to accurately fit (Agresti, 2007), we conducted an additional analysis to address this limitation by constructing a similarly-structured *participant error logistic regression model*, which modeled the likelihood that a participant would produce *any* errors in a given condition (as shown in Table 4).

Descriptive statistics. Figure 4 shows the intrusion rate for each participant. Note that there were many participants who produced no errors on non-switch targets. This is highlighted in Table 4, which shows the proportion of participants who did vs. did not produce *any* intrusion errors in a given condition. Consistent with reversed dominance, English words showed a clear disadvantage in Spanish default paragraphs. Consistent with reversed dominance, when English was the default language, participants were roughly *twice as likely* to produce at least one error in English vs. Spanish on a nonswitch word (Table 4, left column), although a small number of participants (11/48) did not show this pattern; they produced a greater number of errors in Spanish than English (Figure 4, left; see below for discussion). Overall, these results provide strong evidence that dominant language was inhibited at the lexical level even on nonswitch targets both when Spanish, and even English, was selected as the default language.

Figure 4. Per-participant percentage intrusion errors on non-switch, non-cognate, targets, separated by target language (English vs. Spanish).

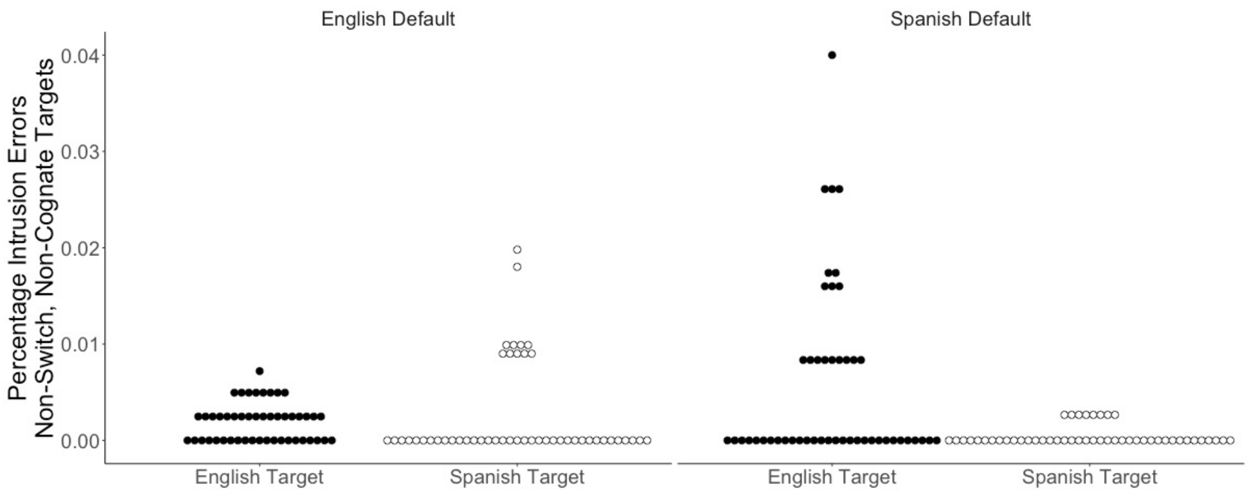


Table 4. Percent of participants who produced at least one intrusion error on non-cognate, non-switch targets, separated by target language (English vs. Spanish) and default language (English vs. Spanish).

	Default Language	
Target Language	English	Spanish
English	56.2%	37.5%
Spanish	22.9%	16.7%

Main effect: Reversed dominance. Critically, both of the error rate and the participant error models[6] showed a significantly higher error rate for English target words (error rate model: $\beta = 1.14$, s.e. $\beta = 0.25$, $\chi^2(1) = 23.96$, $p < 0.0001$; participant error model: $\beta = 1.48$, s.e. $\beta = 0.38$, $\chi^2(1) = 17.86$, $p < 0.0001$). This suggests there is a consistent reversal of language dominance effects on all words in the read aloud task even on non-switch target words.

Standard dominance effects are absent, regardless of default language. Parallel to the analysis at switch points, neither model showed a significant effect of default language,

[6] Similar results were found in an analysis examining all control errors (i.e., intrusions as well as partial intrusions and accent errors; see Supplemental Materials).

although it just missed the significance threshold for the participant error model (error rate model: $\beta = 0.33$, s.e. $\beta = 0.26$, $\chi^2(1) = 1.69$, $p < 0.20$; participant error model: $\beta = 0.67$, s.e. $\beta = 0.36$, $\chi^2(1) = 3.59$, $p < 0.06$).

We next examined whether reversed dominance effects were present in both English and Spanish default paragraphs via the interaction of default language and target word language – note that this analysis pits the power of target language against the power of switch direction (into vs. out of the default language, see Introduction). In the participant error model, both default languages showed reversed dominance effects to a similar extent i.e., the interaction was not significant ($\beta = 0.46$, s.e. $\beta = 0.71$, $\chi^2(1) = 0.42$, $p < 0.52$). However, in the error rate model this interaction was significant ($\beta = -2.87$, s.e. $\beta = 0.51$, $\chi^2(1) = 30.96$, $p < 0.0001$); language dominance effects on nonswitch targets were reversed in Spanish default paragraphs ($\beta = 2.57$, s.e. $\beta = 0.39$, $\chi^2(1) = 57.99$, $p < 0.0001$) but not in English default paragraphs ($\beta = -0.29$, s.e. $\beta = 0.32$, $\chi^2(1) = 0.78$, $p < 0.38$)[7]. In the latter condition, out of the 31 participants that produced at least one error on nonswitch targets, 20 showed reversed dominance (more errors in English vs. Spanish) and only 11 showed standard dominance effects (i.e., more errors on Spanish targets). As measured by the Bilingual Index Score[8] (Gollan et al., 2012) this latter group was more balanced (*M* = 0.87, *SD* = 0.13) as compared to those that showed reversed dominance effects (*M* = 0.76, *SD* = 0.15; $t(23.85) = 2.13$, $p < .05$).

[7] For this subset of the data, the simplest mixed effects regression (with a random intercept for participants) yielded a singular fit.
[8] The ratio of the non-dominant/dominant language scores using our objective proficiency measure, the MINT.

Though dominance was not significantly reversed in this one comparison, particularly for this one subset of individuals, the complete absence of standard dominance effects is consistent with inhibition of dominant language targets across the board in this task. Indeed, even just a reduction of dominance effects would be evidence that the dominant language was inhibited (Declerck et al., 2020), especially when interpreted along the significantly reversed dominance effects overall on nonswitch targets when collapsing across default language (see the description above of the main effect of reversed dominance).

Reading times

Finally, we examined paragraph reading times. On average, bilinguals read English default paragraphs faster (mean = 41.7 seconds; 95% CI (39.9, 43.4)) than Spanish default paragraphs (mean = 44.5 seconds; 95% CI (42.4, 46.8)). To statistically assess this, we used a linear mixed effects regression on reading times[9] including a fixed effect of default language (English, –.5, Spanish .5) with a random intercept by participants (following the regression fitting methods discussed above). The difference across paragraphs was significant $\beta = 2.76$, s.e. $\beta = 0.44$, $\chi^2(1) = 37.97$, $p < 0.0001$). This is consistent with MINT Sprint scores reported above and provides further evidence that bilinguals in the present study were English-dominant.

Discussion

These new data strengthen our confidence that the paragraph reading task consistently shows reversed dominance effects. Similar to studies of reaction times that exhibit reversed dominance, this effect is pervasive for control over production of intrusion errors throughout

[9] The residuals of a model using log-transformed times showed greater violations of normality than the model of raw reading times reported here.

reading aloud of mixed language paragraphs[10]. We replicated the reversed dominance effect on switch words shown in many previous studies. This effect was found across switch types and parts of speech. Critically, by observing performance on a greater number of non-switch targets, we found strong evidence that this effect extends to non-switch words as well. This suggests that reversed dominance is not a manifestation of reactive control processes (contributing to switch trials alone) but reflects a proactive control setting during mixed-language production contexts.

Our results also provide insights into processing in the context of highly frequent language switching. Similar to previous work, we found that function words were more likely to induce language intrusions than content words. As noted above, this likely stems from differences in both language control processes (Gollan & Goldrick, 2018; Myers-Scotton & Jake, 2009) and speech monitoring (Schotter et al., 2019) across word classes. In contrast, the effect of switch type was different under the highly frequent switching conditions studied here. When, as in previous work, switching was less frequent, intrusion errors were much less likely to occur when switching back to the default language of the paragraph vs. switching out of the default language (e.g., Gollan & Goldrick, 2016, 2018). This strong effect of switch type suggests that when the default language is highly active, this provides strong support for the retrieval of target words in the default language. In contrast, when switching is more frequent (as studied here), there are more times when the non-default language must be retrieved; efficient processing would require a reduction of the asymmetry in activation between the default and

[10] But not for overall reading speed as reported above, or for production of within-language errors, or both, as reported in many other reading aloud studies (Gollan et al., 2014; Gollan & Goldrick, 2016; Gollan et al., 2017; 2020; Li & Gollan, 2018; Schotter et al., 2019; Stasenko et al 2020).

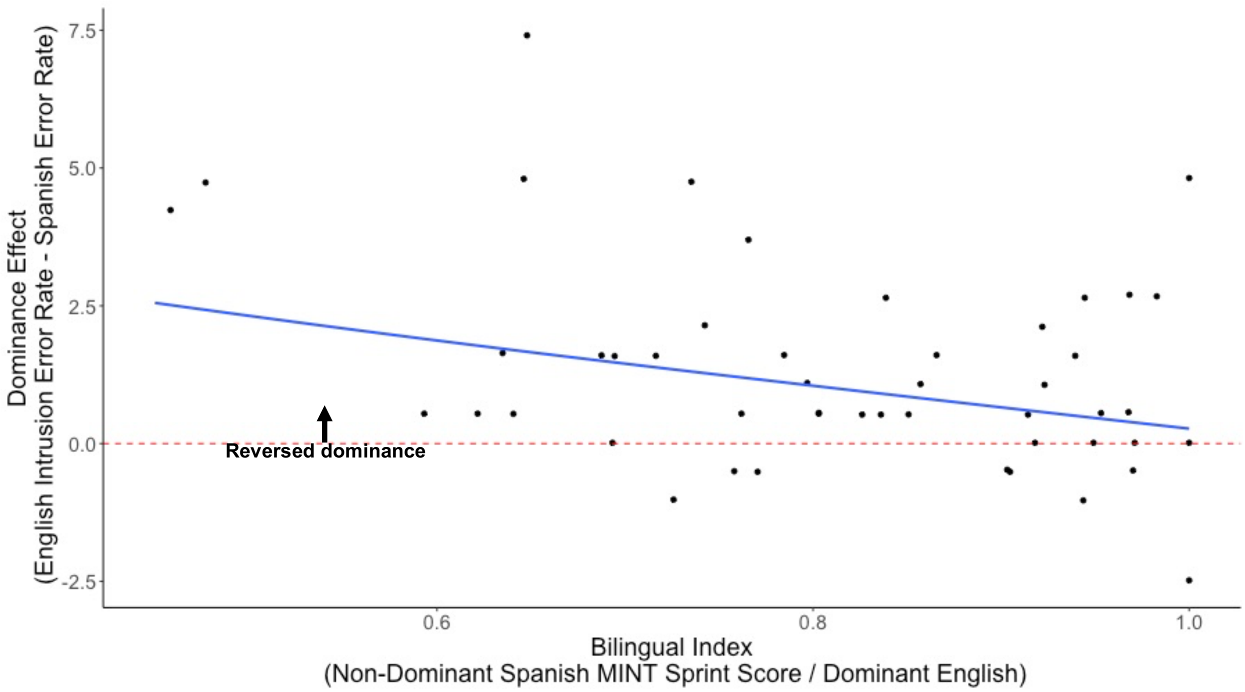
non-default languages, leading to a weaker effect of switch type. As noted above, while intrusion errors showed no significant effect of switch type, an analysis including all language control errors (intrusions, partial intrusions, and accent errors) showed a significant effect of switch type (smaller than observed in previous studies; see Supplementary Materials). This suggests that bilinguals select a default language in both frequent and infrequent switching contexts, but the degree of activation is modulated by the nature of the switching context.

Individual variation in reversed dominance effects

As noted above, there is significant variation in reversed dominance effects. Between-study variation is well documented in the meta-analysis above, and is also found within studies. To explore this within our data, we followed Declerck et al.'s (2020) approach and examined if reversed dominance effects at switch points were smaller in more balanced bilinguals in the present study. To this end, we calculated balance using the Bilingual Index Score (Gollan et al., 2012). More balanced bilinguals have higher Bilingual Index Scores (perfectly balanced bilinguals name the same number of pictures in each language and have a score of 1, bilinguals who name just half as many pictures in the nondominant than in the dominant language would score .5). These data are shown in Figure 5. Note that the majority (83%) of participants show higher intrusion error rates in English vs. Spanish (even though independent measures confirmed English was the dominant language for all bilinguals included in our analyses). However, more balanced bilinguals showed a smaller disadvantage for stopping English over Spanish intrusions. This was confirmed via a mixed-effects logistic regression including target language, Bilingual Index Score, and their interaction, with random intercepts and random slopes for target language by participant. A likelihood ratio test showed a significant interaction

($\chi^2(1) = 9.63$, $p < 0.002$), consistent with a smaller dominance effects in more balanced bilinguals. We interpret this analysis with caution, however, as there are very few unbalanced bilinguals; future work should confirm this pattern with a larger sample. Perhaps most notable in Figure 5 is that nearly all participants produced more intrusion errors with English targets than they did with Spanish targets, even though English was the dominant language.

Figure 5. Individual differences in dominance effects (percentage intrusion errors on non-cognate, English – Spanish targets) as a function of Bilingual Index Score (non-dominant Spanish / dominant English). Red dashed line shows lower limit of dominance reversal (here assuming English should be less error prone based on MINT scores). Blue solid line shows estimated dominance effects from a mixed-effects regression including language, Bilingual Index Score, and their interaction.



General Discussion

Reversed dominance is a striking empirical pattern. In this work, we followed Gollan and Ferreira (2009; Declerck et al., 2020; Kleinman & Gollan, 2018) in arguing that this effect is best understood as one extreme of a continuum of effects arising when dominant language representations are inhibited. Our findings provide clear evidence that reversed language dominance is reliable and replicable across paradigms that measure and elicit different forms of bilingual language production. We reconsidered a recent meta-analysis (Gade et al., 2021a) which questioned this phenomenon. Our review of their method revealed flaws in their data and particularly in their analysis. Our Bayesian re-analysis shows a robust reduction of dominance effects in mixed language blocks. The vast majority of reaction-time-based studies show reduced dominance effects in mixed- vs. single-language blocks. While the distribution of the decrease in dominance (relative to single-language blocks) shows a great deal of variability, dominance tends to be (slightly) reversed in mixed language production. We then established a parallel between these reaction time studies and error data from a reading aloud task in which bilinguals produced hundreds of words per minute. Consistent with previous work, our results showed reversed dominance effects on intrusion errors at switch words and then extended this body of work to show that reversed dominance extends to non-switch words.

Reversed dominance: A challenge for non-inhibitory theories of language control

It is unclear whether alternative approaches to language control can offer a satisfactory account of the reversed dominance effect. Blanco-Elorrieta and Caramazza's (2021) theory of bilingual language production assumes that the activation of non-target representations has no impact on the retrieval and selection of the target word. It's not clear to us how such a theory could account for reversed language dominance. While Blanco-Elorrieta and Caramazza's literature review failed to include discussion of reversed dominance effects – and therefore did not discuss possible mechanisms – they discuss inhibition in the context of a different widely reported effect often attributed to inhibition, the switch-cost asymmetry. This refers to results in which dominance is not reversed but the dominant language exhibits larger switch-costs than the nondominant language (even though responses in the dominant language are faster). They offer a *response exclusion* account of this asymmetry (see also Finkbeiner et al., 2006). According to this account, when a stimulus affords multiple responses (as in a cued picture naming task), the language system makes both options available for response in an articulation buffer. The decision to select one response or the other is made at this articulatory level. When no decision needs to be made (as in a single language production context), responses will be faster. This predicts longer RTs in mixed production contexts, with the lengthening of RT most notable for items that are quickly retrieved in single language contexts (i.e., dominant language responses, Ivanova & Hernandez, 2021). While this mechanism can account for a reduction in dominance effects, particularly on switch trials within mixed blocks, it's not clear how it could account for a reversal. Why would the duration of this selection process *exceed* the advantage of dominant language retrieval, particularly on easier-to-process stay trials? Resolving this question will be critical for developing this account's explanation of reversed dominance effects.

Costa and Santesteban (2004) proposed that highly proficient and balanced bilinguals can set language-specific selection thresholds, in effect selectively disadvantaging one language or the other. However, the same pattern has been observed in lower-proficiency speakers (see Declerck & Koch, in press, for a review), suggesting this particular proposal is not a viable account. If we extend this account to allow it to apply regardless of proficiency, it becomes a 'notational variant' of the inhibition theory – i.e., the same mechanism, simply stated in different terms. In connectionist systems, thresholds are typically represented by subtracting a constant value from the net input to a unit. This can be equivalently represented by constant negative input – i.e., inhibition – to the unit. For example, each unit can have an inhibitory connection to a constantly activated bias unit (e.g., Smolensky, 2005). The equivalence between these mechanisms suggests threshold and inhibitory accounts are notational variants (a possibility acknowledged by Costa et al., 2006: 1069).

Other proposals follow the inhibition account in assuming that the activation of non-target representations interferes with target processing, but attribute reversed dominance to an increase in activation of representations in the nondominant language (such that it is more active than the dominant language: Declerck et al., 2015; see Branzi et al., 2014, Verhoef et al., 2009, for related accounts assuming persistence of facilitation of the nondominant language). This produces a result similar to inhibition (greater activation of the nondominant language) but does so by 'lifting up' the nondominant language rather than suppressing the dominant language. However, as pointed out by Gollan and Goldrick (2018), if speakers are able to facilitate access to the nondominant language in this challenging, mixed language processing context, why are they unable to do so in an easier, pure language context (where cross-language competitors are assumed to be activated, albeit to a lesser degree than in mixed contexts)? Furthermore, if, parallel to inhibition accounts, we assume that similar lexical selection mechanisms are used during monolingual production, why are participants unable to boost activation of low frequency words, eliminating the disadvantage relative to high frequency words? To address these challenges, one must assume that this facilitation effect is specific to language mixing. While logically possible, it is not clear what would motivate such an exception to the rule (e.g., given that lower frequency words are harder to access across in most if not all contexts, why couldn't this exception mechanism extend to these words to reduce or abolish within-language frequency effects?). Such an account is clearly not parsimonious.

If we turn to the literature in speech production outside of bilingualism, Oppenheim et al. (2010) propose that competition effects between semantically related words within a single language arise due to a competitive learning process. When attempting to produce a target, semantically related non-target words become co-activated due to overlapping semantic features (similar to translation equivalents). Oppenheim et al. proposed that after retrieval, incremental learning processes will strengthen connections from these features to the target and weakening features' connections to the non-target items. On subsequent trials, when attempting to name these previously unnamed semantically related words, retrieval will be impaired (an effect known to extend across languages; Runnqvist et al., 2012). Such a mechanism could provide a means for reducing dominance effects (Runnqvist et al., 2019); however, it is not clear why a dominance *reversal* would be found in mixed language contexts where the two languages occur with equal frequency (as is typical of the design of cued switching studies). (See Lowry et al. (2021) for additional experimental and computational modeling data supporting theories incorporating both incremental learning and inhibition.)

Other support for inhibitory mechanisms of language control

The reversed dominance pattern presents a significant challenge to several theories of bilingual language control. In our view, inhibition provides the clearest, most parsimonious account of these findings. Critically, as summarized in Declerck and Koch's (in press) recent review, there are other data that provide strong support for inhibitory mechanisms. Specifically, the *N–2 language repetition cost* (Babcock & Vallesi, 2015; Branzi et al., 2016; Declerck et al., 2021; Declerck, et al., 2015; Declerck & Philipp, 2018; Guo et al., 2013; Phillip et al., 2007; Philipp & Koch, 2009; Timmer et al., 2018) is most clearly accounted for by inhibition. This is the cost required when switching back to the language of the penultimate trial. For example, if a trilingual is switching between English, Turkish, and German, there are slower reaction times to the third trial in the sequence *German*-Turkish-*German* vs. English-Turkish-*German*. This is expected if inhibition is used to enable the switch from German to Turkish; such inhibition would be absent when switching from another language (English) to Turkish. In contrast, a facilitatory mechanism that supported switching by boosting activation of the target language (see e.g., Branzi et al., 2014; Declerck et al., 2015; Verhoef et al., 2009, for related proposals) would predict N–2 repetition benefits. These clearly contrasting predictions suggest that inhibition is required to explain this phenomenon.

While our analysis did not focus on the switch cost asymmetry (Meuter & Allport, 1999), it has frequently been argued to support inhibitory language control theories. In this case, it is carry-over of inhibition from one trial to the next that increases RTs, an effect that is stronger for the dominant vs. nondominant language due to the greater inhibition of the former. However, note that our re-analysis of Gade et al. (2021a) replicated their failure to find evidence of a reliable change in dominance across non-switch vs. switch trials. Most critically, this effect is clearly amenable to non-inhibitory accounts (e.g., persistence of activation of the previous target; see Bobb & Wodniecka, 2013; Declerck & Koch, in press, for discussion). Reversed dominance presents a more compelling empirical and theoretical argument for inhibition.

Advancing theories of language control

Incorporating inhibition into a more comprehensive theory of language control is still an open challenge. Theories incorporating inhibition have always made clear that it is part of a broader network of mechanisms. For example, in the seminal work of Green (1998) inhibition worked in concert with attentional and monitoring mechanisms to enable task-appropriate bilingual language processing. More broadly, it is clear that other aspects of language experience and the context of language production must influence language control. Practice and experience are critical (e.g., Gollan et al., 2008; 2011); inhibiting the dominant language simply cannot help you retrieve lexical items you do not know – i.e., nothing can replace retrieval practice and exposure to establish proficiency in the nondominant language. The social and discourse context in which language production occurs can facilitate the retrieval of lexical items in a specific language (e.g., Blanco-Elorrieta & Pylkkänen, 2017, 2018; Liu et al., 2013; Woumans et al., 2015); speaking to an English monolingual about a topic that is almost exclusively discussed in English will make it easier to recall English lexical items from memory for production. Similarly, language-specific syntactic structures can support retrieval of the grammatically appropriate lexical items in the target language (e.g., Gollan & Goldrick, 2016, 2018); planning to produce a gender-marked noun phrase in Spanish (*el niño* 'the boy' vs. *la niña* 'the girl') will facilitate retrieval of gender-marked determiners in Spanish vs. unmarked determiners in English. Outside of language production specifically, recent evidence suggests that some specific aspects of the diverse array of language experiences across different bilinguals likely impact aspects of cognitive control more broadly (Beatty-Martínez et al., 2020; Bialystok, 2021; Bialystok & Craik, 2022; Gullifer & Titone, 2021; Zhang et al., 2021). Inhibition is clearly not sufficient to account for all aspects of bilingual language production; however, evidence like the reversed dominance effect suggest it is part of the toolkit proficient bilingual speakers bring to bear on the problem of language control (c.f. Blanco-Elorrieta & Caramazza, 2021 and Costa & Santesteban, 2004, who argued that only nonproficient bilinguals rely on inhibitory control). A key challenge is articulating how these multiple mechanisms interact to facilitate fluent production.

The reading aloud task highlights one such critical challenge: we have hypothesized simultaneous inhibition of representations in the dominant language (producing reversed dominance effects) and activation of representations in the default language (facilitating access to default language targets; Gollan & Goldrick, 2018; Li & Gollan, 2021). In other words, when the dominant language is selected as the default language it is both activated and inhibited at the same time. The simultaneous impact of these two mechanisms places a crucial constraint on theories of language control: there must be some means to both inhibit *and* enhance representations associated with a given language *at the same time*. More specifically, the body of the data from the reading aloud task suggests that in dominant-language-default paragraphs speakers can both enhance the activation of lexical representations from the dominant language (via boosting from the syntactic representations from the dominant language) *and, at the same time,* inhibit the activation of the very same dominant language lexical representations (via a language node or language task schema).

What type of control model allows for the separation of control mechanisms across levels of representation, allowing a single language to be subject to a mixture of excitatory and inhibitory processes across levels? These issues have been partially addressed by previous proposals. Control theories have previously claimed that processes that enhance activation can be involved in control at the same time as inhibition. For example, Green and Abutalebi's (2013:518) Adaptive Control Model, building on Green (1998) and extended in Green and Wei (2014)[11], allowed for proactive activation of language task schemas, competition between different schemas, and (simultaneously) "reactive inhibition of representations that trigger selection of these competing task schemas." Note that this mixture of mechanisms targets *different languages* (e.g., the target language is activated by the task schema, but reactive inhibition effects the non-target language) and there is no specification of effects across multiple levels of linguistic representation (which would be needed to both select one language as default while simultaneously inhibiting it at or through a different processing level). Other proposals have noted that coactivation of multiple languages occurs at many levels of linguistic representation (e.g., Kroll et al., 2006) motivating the need for inhibitory control mechanisms throughout the production system (Kroll et al., 2008). However, such proposals have not considered the possibility that distinct processes (activation and inhibition) could impact representations from the same language at different levels of representation and processing.

To accommodate these phenomena, we propose the Adaptive Control Model must be extended to specify how different control mechanisms influence processing at multiple levels of linguistic representation. Consider an English-dominant participant producing a mixed-language paragraph with English as the default language. Within the production system, the default

[11] In certain types of codeswitching, Green and Wei (2014) propose that language task schemas can interact in different ways. Of particular relevance here inhibitory interactions can be replaced by cooperative control mechanisms. Because the read aloud task forces participants to switch languages, we would characterize the code-switching in this task as competitive (participants *must* unexpectedly access a different language).

language status of English is reflected by activation of English-specific syntactic structures (Gollan & Goldrick, 2018; Li & Gollan, 2021). At the same time, because the paragraph reading task forces participants to mix Spanish and English, task schemas associated with both languages are co-activated. The Spanish task schema globally inhibits the activation of dominant-language English lexical items (with the goal of equalizing accessibility of English and Spanish targets at appropriate points in the paragraph).

Focusing specifically on the nature of inhibitory mechanisms, the perspective adopted in the current paper is underspecified in several important ways. What underlies the variation in inhibition across participants and production contexts? There is clearly a continuum of effects, but what are the underlying sources of this variation? It could reflect differences across participants (e.g., the absolute size of the reduction in dominance effects is correlated with the baseline difference in single-language blocks; Declerck et al., 2020; as shown above and in Declerck et al., 2020, the extent to which dominance effects are reduced is smaller for more balanced bilinguals), differences across tasks (e.g., voluntary vs. cued switching; Gollan & Ferreira, 2009), differences across populations (e.g., Mandarin-speaking students in the US tend to mix languages less frequently than Spanish-speaking students; Prior & Gollan, 2011), and differences across age groups (e.g., older bilinguals do not reverse dominance, which could reflect an inhibitory deficit or a better ability to avoid overshooting relative to young bilinguals; Stasenko et al., 2021[12]).

[12] But see puzzlingly intact reversed dominance effects in aging bilinguals and even in bilinguals with Alzheimer's disease in the reading aloud task (Gollan & Goldrick, 2016; Gollan et al., 2017; 2020).

Relatedly, there are a number of studies in the Gade et al. (2021a) meta-analysis that show a pattern of reversed dominance in single language blocks. This counter-intuitive pattern was also observed for a small subset of the hundreds of participants in Declerck et al. (2020). As Declerck et al. noted, this could partially reflect challenges of measurement of dominance with small numbers of items, as well as counterbalancing of condition order and interference in dominant language productions from effects of blocked language order effect (where single-language block performance is worse after performing a single-language block in another language: Branzi et al., 2014; Degani et al., 2020; Guo et al., 2011; Kreiner & Degani, 2015; Misra et al., 2012; Runnqvist et al., 2019; Van Assche et al., 2013; Wodniecka et al., 2020). However, such findings have been consistently observed in at least one bilingual population. Specifically, studies conducted in the Basque Autonomous Community, Spain, have typically found faster RTs for picture naming in the nondominant language, Basque, relative to the dominant language, Spanish, despite finding evidence of Spanish dominance in other objective measures such as verbal fluency tasks (e.g., de Bruin et al., 2017; de Bruin et al., 2018; Jevtovic et al., 2019). Additional work is needed to establish a gold standard measure of language dominance and proficiency (Garcia & Gollan, 2021) that will be most sensitive to mechanisms of bilingual language control even in bilinguals without one clearly dominant language.

Greater specification of how inhibitory mechanisms are acquired and adapt over time (see Filippi et al., 2014 for one approach) might help constrain investigations into the sources of variation in reversed dominance effect. In considering this issue, we note that as far as we are aware, the reversed dominance pattern has never been documented in any other domain of human performance or nonlinguistic task switching[13] (in contrast to the switch-cost asymmetry which is widely reported in both linguistic and nonlinguistic switching tasks; e.g., Finkbeiner et al., 2006; Meuter & Allport, 1999). The apparently specific nature of reversed language dominance may reflect the unique demands imposed by interactional contexts where unintended switching could prevent successful communication. When communicating with a speaker who does not know their dominant language, the bilingual speaker *cannot* switch out of the nondominant language (if they want to be understood). Ensuring there are *no* intrusions from the dominant language in such contexts, given noise in language control and lexical access process, might require significant inhibition, perhaps providing the seed for over-inhibition in mixing contexts.

To further complicate the picture, there may in fact be multiple inhibitory mechanisms (Declerck & Philipp, 2015; Gollan et al., 2014); the same computational principle may be instantiated by both domain-specific and domain-general mechanisms (see Nozari & Novick, 2017, for discussion of these issues in the context of monitoring and control). One means of teasing these apart has relied on the presence vs. absence of intercorrelations in individual differences in inhibitory effects across cognitive domains; low correlations within individuals are taken to imply independent mechanisms (e.g., Rey-Mermet et al., 2018). However, this is a challenging approach, as accurately measuring such individual differences is difficult (for recent discussions, see Draheim et al., 2021; Nicosia & Balota, 2020; Segal et al., 2019; Segal et al., 2021), and could easily lead to prematurely dispensing with inhibition as an explanatory

[13] Ivanova and Hernandez (2021) argue that parallel effects are observed in monolingual production (switching between basic-level and subordinate naming). However, their monolingual data do not show a *reversal* of effects; basic-level names are associated with longer reaction times in both pure and mixed blocks.

mechanism (as advocated by, e.g., Rey-Mermet & Gade, 2018; Rey-Mermet, Gade, & Oberauer, 2018). Another approach is to look for cross-domain transfer of training (Wu et al., 2021) or to look for modulation of control in one task on another (e.g., Adler et al., 2020; Hofweber et al., 2020; Jiao et al., 2022). If training or recent experience utilizing inhibition in one domain enhances performance in the other, this supports the use of a domain-general control mechanism. The absence of transfer is taken to indicate that there is no domain-general mechanism (Bialystok & Craik, 2022). However, if we allow for the possibility that there are both domain-specific *and* domain-general mechanisms, the predictions for transfer are not warranted; practice may engage only the domain-specific mechanism. Developing new techniques for assessing this complex array of possibilities is a key area for future work exploring the rich empirical and theoretical landscape (as advocated for in bilingualism research more broadly by Navarro-Torres et al., 2021).

Conclusions

A complex set of mechanisms have been proposed to explain the amazing ability of bilinguals to control the language of production. We should adopt a healthy skepticism of such complex models, while taking care to avoid favoring overly simplistic models of this advanced skill. In our view, reversed language dominance effects and the reduction of dominance effects across single-language versus mixed-language blocks reflect the same underlying cognitive mechanism and provide striking, often replicated, reliable, and strong evidence that inhibition is a requisite part of any theory of language control.

Author Note

Supported by grants from the National Institute of Health (DC011492; AG076415) and the National Science Foundation (BCS 1923065) to THG and the National Science Foundation (DRL 2219843) to MG. Thanks to Mayra Murillo for data collection and coding, and to Elma Blom, Marco Calabria, Angela de Bruin, Megan Gross, Dan Kleinman, Barbara Köpke, Emilie Massa, and Mona Timmermeister for help in cleaning data from the meta-analysis. In working on this paper, we thought often of our departed colleague, Albert Costa; we deeply miss arguing with him about theories of speech production. Albert was a fluent multilingual who nonetheless reported feeling exhausted by extended time spent speaking a nondominant language, English. At the end of a long evening, he would sometimes insist on switching out of English, even when MG, his monolingual interlocutor, could not understand him. In notes from a last conversation between TG and Albert, he mentioned a need to determine what cognitive mechanisms and what types of practice would "prevent massive interference" when learning a new language. The fact that this master of multilingual communication recognized the enormous effort required to resist non-target-language competitor forms serves, for us, as a powerful source of intuition for the role of inhibition in bilingual speech production. Though this inspiration might have irritated Albert, he surely would have derived great pleasure from arguing with us about it.

References

Adler, R. M., Valdés Kroff, J. R., & Novick, J. M. (2020). Does integrating a code-switch during comprehension engage cognitive control?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(4), 741-759.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Ahn, D., Abbott, M. J., Rayner, K., Ferreira, V. S., & Gollan, T. H. (2020). Minimal overlap in language control across production and comprehension: Evidence from read-aloud versus eye-tracking tasks. *Journal of Neurolinguistics*, *54*, 100885.

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *Journal of the Acoustical Society of America, 149*, 1271-1284.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28.

Babcock, L., & Vallesi, A. (2015). Language control is not a one-size-fits-all languages process: Evidence from simultaneous interpretation students and the n−2 repetition cost. *Frontiers in Psychology, 6,* 1622.

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, 255-278.

Bates, B., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models*. arXiv preprint available at https://arxiv.org/pdf/1506.04967.pdf

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 1022-1047.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92-111.

Bialystok, E. (2021). Bilingualism as a slice of Swiss cheese. *Frontiers in Psychology*, 5219.

Bialystok, E., & Craik, F. I. M. (2022). How does bilingualism modify cognitive function? Attention to the mechanism. *Psychonomic Bulletin & Review, 29,* 1246–1269.

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

Blanco-Elorrieta, E. & Pylkkänen, L. (2017) Bilingual language switching in the lab vs. in the wild: the spatio-temporal dynamics of adaptive language control. *Journal of Neuroscience, 37*, 9022–9036.

Blanco-Elorrieta, E., Pylkkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences, 22*, 1117–1126.

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*, 568-585.

Branzi, F. M., Calabria, M., Boscarino, M. L., & Costa, A. (2016). On the overlap between bilingual language control and domain-general executive control. *Acta Psychologica, 166(3), 21–30*.

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

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

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*, 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), 1057-1074.

De Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology*, *8*, 522.

De Bruin, A., Dick, A. S., & Carreiras, M. (2021). Clear theories are needed to interpret differences: Perspectives on the bilingual advantage debate. *Neurobiology of Language*, *2*(4), 433-451.

De Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, *103*, 28-43.

Declerck, M. (2020). What about proactive language control? *Psychonomic Bulletin & Review*, *27*, 24-35.

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

Declerck, M., & Koch, I. (in press). The concept of inhibition in bilingual control. *Psychological Review*.

Declerck, M., Koch, I., & Philipp, A. M. (2012). Digits vs. pictures: The influence of stimulus type on language switching. *Bilingualism: Language and Cognition*, *15*, 896-904.

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

Declerck, M., Schuch, S., & Philipp, A. M. (2021). Conflict adaptation during multilingual language production as evidenced by the n-3 effect. *Bilingualism: Language and Cognition, 24*, 211–214.

Declerck, M., Thoma, A., Koch, I., & Philipp, A. M. (2015). Highly proficient bilinguals implement inhibition – Evidence from N-2 repetition costs when switching between three languages. *Journal of Experimental Psychology: Learning, Memory & Cognition, 41*, 1911-1916.

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*, 153–183.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, *59*(4), 447-456.

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, *150*(2), 242-275.

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.

Finkbeiner, M., Almeida, J., Janssen, N., & Caramazza, A. (2006). Lexical selection in bilingual speech production does not involve language suppression. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1075-1089.

Gade, M., Declerck, M., Philipp, A. M., Rey-Mermet, A., & Koch, I. (2021a). Assessing the evidence for asymmetrical switch costs and reversed language dominance effects–a meta-analysis. *Journal of Cognition*, *4*, 55. http://doi.org/10.5334/joc.186

Gade, M., Declerck, M., Philipp, A. M., Rey-Mermet, A., & Koch, I. (2021b). Correction: Assessing the evidence for asymmetrical switch costs and reversed language dominance effects–a meta-analysis. *Journal of Cognition*, *4*, 60.

Garcia, D. L., & Gollan, T. H. (in press). The MINT Sprint: Exploring a fast administration procedure with an expanded Multilingual Naming Test. *Journal of the International Neuropsychological Society*.

Gavino, M. F., & Goldrick, M. (in press). Consequences of mixing and switching languages for retrieval and articulation. *Bilingualism: Language and Cognition.*

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

Gollan, T., & Goldrick, M. (2016). Grammatical constraints on language switching: Language control is not just executive control. *Journal of Memory and Language, 90*, 177-199.

Gollan, T., & Goldrick, M. (2018). A switch is not a switch: Syntactically-driven bilingual language control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 143-156.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58, 787–814. http://dx.doi.org/10.1016/j.jml.2007.07.001

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,* 585– 595. http://dx.doi.org/10.1177/0956797613512661

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*, 186.

Gollan, T. H., Stasenko, A., Li, C., & Salmon, D. P. (2017). Bilingual language intrusions and other speech errors in Alzheimer's disease. *Brain and Cognition, 118*, 27–44. https://doi.org/10.1016/J.BANDC.2017.07.007.

Gollan, T. H., Li, C., Stasenko, A., & Salmon, D. P. (2020). Intact reversed language-dominance but exaggerated cognate effects in reading aloud of language switches in bilingual Alzheimer's disease. *Neuropsychology*, *34*(1), 88-106.

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*, 594–615. https://doi.org/10.1017/ S1366728911000332.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67–81. DOI: https://doi.org/10.1017/S1366728998000133

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

Green, D. W., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition and Neuroscience, 29*, 499-511.

Gullifer, J. W., & Titone, D. (2021). Bilingualism: A neurocognitive exercise in managing uncertainty. *Neurobiology of Language*, *2*, 464-486.

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*, 277–283.

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

Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes, 26*, 902–934.

Heikoop, K. W., Declerck, M., Los, S. A., & Koch, I. (2016). Dissociating language-switch costs from cue-switch costs in bilingual language switching. *Bilingualism: Language and Cognition*, *19*, 921-927.

Hofweber, J., Marinis, T., & Treffers-Daller, J. (2020). Experimentally induced language modes and regular code-switching habits boost bilinguals' executive performance: Evidence from a within-subject paradigm. *Frontiers in Psychology*, 2897.

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

Ivanova, I., & Hernandez, D. C. (2021). Within-language lexical interference can be resolved in a similar way to between-language interference. *Cognition*, *214*, 104760.

Jiao, L., Timmer, K., Liu, C., & Chen, B. (2022). The role of language switching during cross-talk between bilingual language control and domain-general conflict monitoring. *Cognitive Science*, *46*(8), e13184.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Jevtović, M., Duñabeitia, J. A., & de Bruin, A. (2020). How do bilinguals switch between languages in different interactional contexts? A comparison between voluntary and mandatory language switching. *Bilingualism: Language and Cognition*, *23*(2), 401-413.

Jylkkä, J., Lehtonen, M., Lindholm, F., Kuusakoski, A., & Laine, M. (2018). The relationship between general executive functions and bilingual switching and monitoring in language production. *Bilingualism: Language and Cognition*, *21*, 505-522.

Kleinman, D., & Gollan, T. H. (2018). Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition, 173*, 115–132. https://doi.org/10.1016/j.cognition.2018.01.009

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.

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, *128*, 416-430.

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

Kolers, P. A. (1966). Reading and talking bilingually. *The American Journal of Psychology, 79,* 357–376. http://dx.doi.org/10.2307/1420877

Lachaud, C. M., & Renaud, O. (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics*, *32*(2), 389-416.

Lavric, A., Clapp, A., East, A., Elchlepp, H., & Monsell, S. (2019). Is preparing for a language switch like preparing for a task switch?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(7), 1224-1233.

Li, C., & Gollan, T. H. (2018). Cognates interfere with language selection but enhance monitoring in connected speech. *Memory & Cognition*, *46*(6), 923-939.

Li, C., & Gollan, T. H. (2021). What cognates reveal about default language selection in bilingual sentence production. *Journal of Memory and Language*, *118*, 104214.

Li, Y., Yang, J., Sherf, K. S., & Li, P. (2013). Two faces, two languages: an fMRI study of bilingual picture naming. *Brain and Language, 127*, 452–462.

Liu, C., Timmer, K., Jiao, L., Yuan, Y., & Wang, R. (2019). The influence of contextual faces on bilingual language control. *Quarterly Journal of Experimental Psychology*, *72*, 2313-2327.

Lowry, M., Dubé, C., & Schotter, E. (2021). Evaluating theories of bilingual language control using computational models. *Journal of Memory and Language*, *117*, 104195.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 503-535.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*, 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*, 224–237.

Myers-Scotton, C., & Jake, J. (2009). A universal model of code-switching and bilingual language processing and production. In B. Bullock & A. Jacqueline Toribio (Eds.), *The Cambridge* *handbook of linguistic code-switching* (pp. 336–357). New York, NY: Cambridge University Press.

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

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science*, *42*, 1075-1100.

Nicosia, J., & Balota, D. (2020). The consequences of processing goal-irrelevant information during the Stroop task. *Psychology and Aging*, *35*(5), 663.

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, *26*, 403-410.

Olson, D. J. (2016). The gradient effect of context on language switching and lexical access in bilingual production. *Applied Psycholinguistics*, *37*(3), 725-756.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*, 227-252.

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*, 395-416.

Philipp, A. M., & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching between languages in naming tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1187–1195.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15,* 36–57.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society*, *17*(4), 682-691.

Ratiu, I., & Azuma, T. (2017). Language control in bilingual adults with and without history of mild traumatic brain injury. *Brain and Language, 166*, 29–39.

Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, *25*, 1695-1716.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 501–526.

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*, 850-869.

Runnqvist, E., Strijkers, K., & Costa, A. (2019). Error-based learning and lexical competition in word production: Evidence from multilingual naming. *PloS One*, *14*, e0213765.

Runnqvist E., Strijkers K., Sadat J., & Costa A. (2011). On the temporal and functional origin of L2 disadvantages in speech production: A critical review. *Frontiers in Psychology, 2*, 1–8.

Saint-Aubin, J., & Klein, R.M. (2001). Influence of parafoveal processing on the missing-letter effect. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 318-334.

Schotter, E. R., Li, C., & Gollan, T. H. (2019). What reading aloud reveals about speaking: Regressive saccades implicate a failure to monitor, not inattention, in the prevalence of intrusion errors on function words. *Quarterly Journal of Experimental Psychology*, *72*(8), 2032-2045.

Segal, D., Stasenko, A., Gollan, T.H. (2019). More evidence that a switch is not (always) a switch: Binning nilinguals reveals dissociations between task and language switching. *Journal of Experimental Psychology: General*, 148, 501-519.

Segal, D., Prior, A., & Gollan, T. H. (2021). Do all switches cost the same? Reliability of language switching and mixing costs. *Journal of Cognition*, 4(1), 3.

Smolensky, P. (2006). Optimization in neural networks: Harmony maximization. In P. Smolensky & G. Legendre *The harmonic mind: From neural computation to optimality-theoretic grammar* (vol.1: Cognitive architecture, pp. 345-392). Cambridge, MA: MIT Press.

Stasenko, A., Hays, C., Wierenga, C. E., & Gollan, T. H. (2020). Cognitive control regions are recruited in bilinguals' silent reading of mixed-language paragraphs. *Brain and Language*, *204*, 104754.

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*, 806-821.

Timmer, K., Calabria, M., Branzi, F. M., Baus, C., & Costa, A. (2018). On the reliability of switching costs across time and domains. *Frontiers in Psychology, 9,* 1032.

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*, 1781–1792.

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

Woumans, E., Martin, C. D., Vanden Bulcke, C., Van Assche, E., Costa, A., Hartsuiker, R. J., & Duyck, W. (2015). Can faces prime a language? *Psychological Science*, *26*, 1343-1352.

Wu, Y. J., Chen, M., Thierry, G., Fu, Y., Wu, J., & Guo, T. (2021). Inhibitory control training reveals a common neurofunctional basis for generic executive functions and language switching in bilinguals. *BMC Neuroscience, 22*, 36.

Zhang, H., Diaz, M. T., Guo, T., & Kroll, J. F. (2021). Language immersion and language training: Two paths to enhanced language regulation and cognitive control. *Brain and Language, 223*, 105043.

Appendix A: Full Results for Alternate Models of Gade et al. (2021a) Data

Table A1. Posterior mean, standard error, 95% credible interval and R̂ statistic for parameters in a model analyzing log RTs in mixed language blocks only.

Parameter	Mean	S.E.	Lower 95%	Upper 95%	R̂
Dominance (L1 vs. L2)	–0.028	0.016	–0.058	0.003	1.001
Trial Type (Non-Switch, Switch)	0.069	0.015	0.037	0.098	1.000
Interaction: Dominance by Switch	–0.006	0.031	–0.067	0.055	1.000

Table A2. Posterior mean, standard error, 95% credible interval and R̂ statistic for parameters in a model analyzing log RTs, with mixed language blocks represented by the average of stay and switch trial RTs.

Parameter	Mean	S.E.	Lower 95%	Upper 95%	R̂
Dominance (L1 vs. L2)	0.010	0.016	–0.023	0.042	1.000
Block Type (Single, Mixed)	0.102	0.017	0.069	0.134	1.000
Interaction: Dominance by Block Type (Single, Mixed)	–0.076	0.033	–0.141	–0.012	1.000

Table A3. Posterior mean, standard error, 95% credible interval and R̂ statistic for parameters in a model analyzing raw RTs.

Parameter	Mean	S.E.	Lower 95%	Upper 95%	R̂
Dominance (L1 vs. L2)	–1.898	8.149	–18.127	13.942	1.000
Block Type (Single, Mixed)	29.036	8.448	12.402	45.516	1.000
Trial Type (Non-Switch, Switch)	18.046	8.641	0.733	35.066	1.001
Interaction: Dominance by Block Type (Single, Mixed)	–7.740	9.548	-26.533	11.216	1.001
Interaction: Dominance by Trial Type (Non-Switch, Switch)	–1.013	9.564	-19.745	17.781	1.000

Appendix B: Power analysis

Prior to running the experiment, a Monte Carlo analysis was used to estimate the power for detecting the effect of language (i.e., to test for a reversed dominance effect). (Code is available at https://osf.io/khaxc/wiki/home/.) A logistic mixed effects model with contrast-coded factors default language (English vs. Spanish), part of speech (Function vs. Content), their interaction, and a random intercept by participant, was fitted to the rate of intrusion errors observed for single word switches out of the default language by Gollan and Goldrick (2018). This fitted model was used to simulate a novel data set, using the number of function and content words in the current study, and a variable number of participants. The fixed effects were drawn from a multivariate normal distribution; this distribution had means set to the effect size estimates from the fitted model and a variance-covariance matrix equal to the fitted model. The random intercepts were drawn from a normal distribution based on the random effect variance estimate of the fitted model. The model was then re-fit to the simulated data, and the significance of the default language effect for the simulated data was assessed via model comparison. Runs were only retained if the model successfully converged without warnings. This process was repeated 1,000 times to estimate power. At 20 participants, power for detecting the difference between English and Spanish target words was estimated at 0.911; 30 participants, 0.957; 44 participants (the sample size of Gollan & Goldrick, 2018), .998.

Appendix C: Example of Paragraph Variation Across Default Languages

An example paragraph with English and Spanish default variants (presented between subjects). <u>Switch</u> words are underlined here, but no underlining was shown to participants. The full set of materials is available in the Open Science Foundation repository for this paper: https://osf.io/khaxc/wiki/home.

English default: Well, they were gemelas. They were always juntas. If they would go to do an errand, lo hacían juntas. If they would do a trabajo, lo hacían together. The ropa that they wore siempre era the same. If the dress era blanco, then the two had to wear lo mismo. The people say que cuando las twins looked at other people, it caused their skin to itch. Pero algunas personas believe it, and some do not. These gemelas siempre slept together in one room. Después, when they were growing up una de las twins fell in love with a muchacho.

Spanish default: Bueno, ellas eran twins. Siempre estaban together. Si iban a hacer un mandado, they did it together. Si hacían algún job, they did it juntas. La clothes que se ponían was always la misma. Si el vestido was white, entonces las dos tenían que llevar the same thing. La gente dice that when the gemelas miraban a otras personas, hacían que les picara la piel. But some people lo creen, y otros no. Estas twins always dormían juntas en un cuarto. Later, cuando fueron creciendo one of the gemelas se enamoró de un boy.

Supplemental materials: All control errors on non-cognates

Switch words

Partial intrusions and accent errors differ from intrusions in some ways, motivating our main analysis' focus on intrusions. However, these errors do index failures of language control, and occurred at high rates in this study. Therefore, in a post-hoc analysis, we consider all of these control errors together. Participants produced at least 6 control errors and as many as 62 (*M*=20, *SD*=11.2). We assessed these data in a mixed-effects logistic regression with fixed effects identical to the model above. Our iterative procedure resulted in an intercept-only random effects structure for participants.

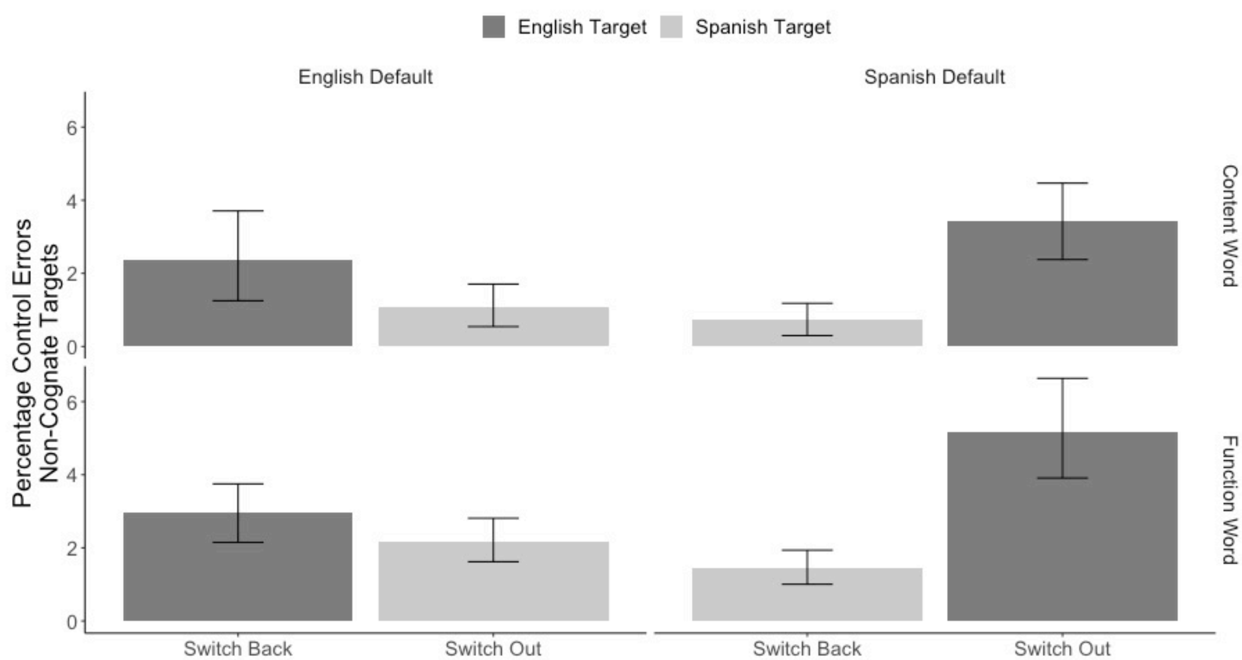


Figure S1. Percent intrusions on non-cognate targets, separated by default language (English, left vs. Spanish, right), part of speech (content, top, vs. function, bottom), and switch type (switch back to the default language vs. switch out of the default). Colors show language of switch word targets. Error bars show bootstrapped 95% confidence intervals.

There was again a significant effect of target language, with control errors occurring significantly more often on English vs. Spanish targets (β = 1.02, s.e. β = 0.13 , $\chi^2(1)$ = 74.11, p <

0.0001). This did not interact with switch location (β = 0.07, s.e. β = 0.25, $\chi^2(1) < 1$, $p < 0.80$). Parallel to the analysis of intrusions, the effect of part of speech was significant, with more errors on function vs. content words (β = 0.54, s.e. β = 0.13, $\chi^2(1) = 19.71$, $p < 0.0001$). Parallel to the main analysis, this did not interact with switch type (β = 0.09, s.e. β = 0.25, $\chi^2(1) < 1$, $p <$ 0.73); the three way interaction was also not significant (β = 0.20, s.e. β = 0.50 , $\chi^2(1) < 1$, $p <$ 0.70). In contrast to the main analysis, part of speech did not interact with target language (although the trend was in the same direction β = –0.40, s.e. β = 0.25, $\chi^2(1) = 2.48$, $p < 0.12$).

There was one key difference from the main analysis; consistent with previous work (Gollan & Goldrick 2018), there was a significant effect of switch location, with more errors occurring when switching out of the default language vs. switch back to the default (β = 0.46, s.e. β = 0.13, $\chi^2(1) = 13.55$, $p < 0.0005$). By considering a greater number of errors, our analysis has greater power, allowing the effect of the default language on errors to emerge.

Non-switch words

As in the main dataset, English words shows a consistent disadvantage, particularly in Spanish default paragraphs.

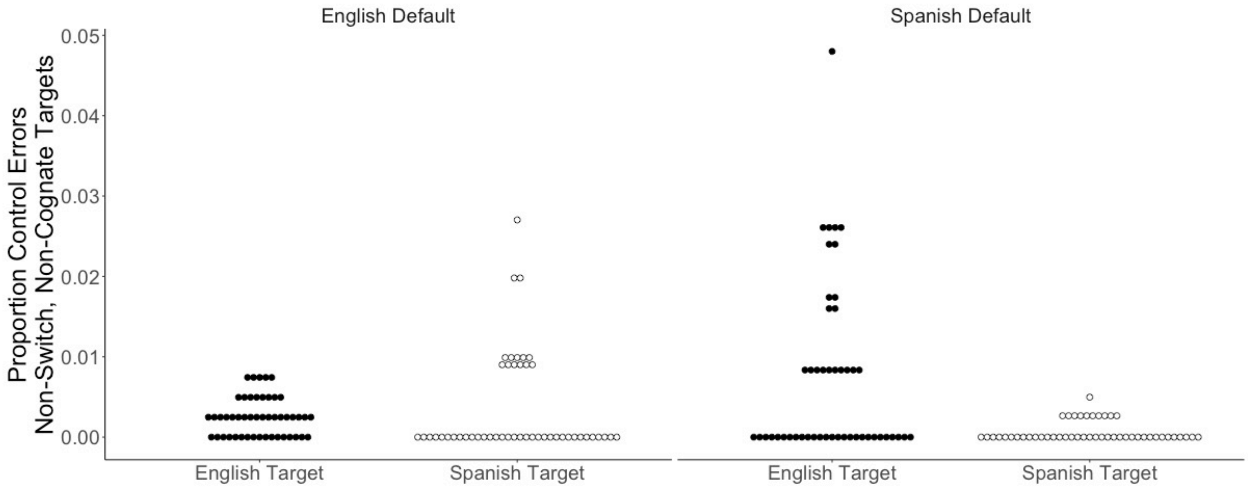


Figure S2. Per-participant percentage control errors on non-switch, non-cognate, targets, separated by target language (English vs. Spanish).

Table S1. Total number of participants producing at least one control error on non-cognate, non-switch targets, separated by target (English vs. Spanish) and default language (English vs. Spanish).

	Default Language	
Target Language	English	Spanish
English	64.6%	43.8%
Spanish	29.2%	22.9%

Analysis structure followed that of the analysis reported in the main body of the paper. As in the main analysis, both the models showed a significantly higher error rate for English target words (error rate model: $\beta = 0.10$, s.e. $\beta = 0.02$, $\chi^2(1) = 27.12$, $p < 0.0001$; participant error model: $\beta = 1.55$, s.e. $\beta = 0.38$, $\chi^2(1) = 19.51$, $p < 0.0001$).

As in the main analysis, other effects were less consistent. With more errors to analyze, the effect of default language reached the significance threshold in the participant error model ($\beta = 0.75$, s.e. $\beta = 0.36$, $\chi^2(1) = 4.52$, $p < 0.04$) but it was still not significant in the error rate model: $\beta = 0.03$, s.e. $\beta = 0.02$, $\chi^2(1) = 2.10$, $p < 0.15$). As in the main analysis, the error rate model showed a significant interaction (β = –0.03, s.e. β = 0.004, $\chi^2(1) = 39.80$, $p < 0.0001$) but this was not significant in the participant error model (β = 0.70, s.e. β = 0.71, $\chi^2(1) = 0.98$, $p <$ 0.33).