Journal of Memory and Language 117 (2021) 104195

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Evaluating theories of bilingual language control using computational models[☆,☆☆]

Mark Lowry[a,*], Chad Dubé[b], Elizabeth Schotter[b]

[a] National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20892, USA[1]
[b] University of South Florida, 4202 East Folwer Avenue, Tampa, FL 33620, USA

ARTICLE INFO

Keywords:
Bilingual language control
Inhibition
Computational models

ABSTRACT

Bilingual language control refers to a bilingual's ability to speak exclusively in one language without the unintended language intruding. It has been debated in the literature whether bilinguals need an inhibitory mechanism to control language output or whether a non-inhibitory mechanism can be used. This paper presents mathematical models instantiating the two accounts. The models explain how participants' reaction times in language production (naming) are impacted by across-trial semantic relatedness and consistency of language (same or different language across trials). The models' predictions were compared to data from an experiment in which participants named semantically-related and -unrelated pictures in their first and second language. Results indicate that within-language facilitation effects are abolished after a language switch, supporting the predictions of the Inhibitory Model. However, within-language facilitation was observed over the course of 'stay' trials in which no language switch was required, contrary to the predictions of both models. A second experiment was conducted to determine the origin of this unexpected facilitation, by separating spreading activation effects from incremental learning effects. The results suggest the facilitation observed in Experiment 1 was due to spreading activation. Together, the modeling and data suggest that language switching abolishes spreading activation effects, but cumulative semantic interference (created by incremental learning) is unaffected by language switching. This suggests that (1) within-language control is non-competitive, (2) between-language language control is competitive and (3) incremental learning plays a role in bilingual language speech production.

Introduction

When bilinguals speak, they must choose their words carefully. Depending on the audience, a bilingual may be free to choose words from either language (e.g., when conversing with other bilinguals), or they may be constrained to only one language (e.g., when conversing with monolinguals). To date, there have been several demonstrations that competition for selection occurs between a bilingual's two languages (for a review, see Kroll, Bobb, Misra, & Guo, 2008). Thus, if two highly active words are candidates for lexical retrieval, how do bilinguals choose the correct one? This study seeks to address whether inhibition is used to control output by examining whether switching languages abolishes short-term spreading activation. Unlike previous studies, it does so in a way that controls for confounding longer-term learning effects.

Much important prior work has sought the answer to this question within the language switching paradigm, in which bilingual participants must name stimuli in one of their two languages depending on a cue which varies over trials (e.g., Costa & Caramazza, 1999; Costa, Santesteban, & Ivanova, 2006; Gollan & Ferreira, 2009; Jackson, Swainson, Cunnington, & Jackson, 2001; Linck, Schwieter, & Sunderman, 2012; Meuter & Allport, 1999; Verhoef, Roelofs, & Chwilla, 2009). Of these studies, some have concluded that inhibition is the mechanism (Green, 1998; Green & Abutalebi, 2013; Meuter & Allport, 1999), while others

☆ The authors would like to thank Dr. Judith Bryant, Dr. Kyna Betancourt and Dr. Geoffrey Potts for their guidance.
☆☆ Experimental data and model code can be found at https://osf.io/qv6w4.
* Corresponding author.
E-mail address: mlowry@mail.usf.edu (M. Lowry).
[1] Dr. Lowry is a Cancer Research Trainee Awardee (CRTA) with the Behavioral Research Program at the National Cancer Institute. The content of this paper is solely the responsibility of the author and does not necessarily represent the official views or policies of the US National Cancer Institute, National Institutes of Health, or Department of Health and Human Services.

https://doi.org/10.1016/j.jml.2020.104195
Received 31 January 2020; Received in revised form 20 November 2020; Accepted 22 November 2020
0749-596X/Published by Elsevier Inc.

posit a non-inhibitory lexical selection mechanism for all bilinguals (e.g., Costa & Santesteban, 2004), or a non-inhibitory mechanism that develops as bilinguals become more fluent (e.g., Costa et al., 2006). The present investigation takes a different approach by instantiating these (predominantly verbal) theories of bilingual control and spreading activation into two computational models (Non-Inhibitory and Inhibitory). We then use the models to predict reaction times on a trial by trial basis for a sample of mostly proficient bilinguals.

Beyond the need for clear formal predictions, there are also issues to address with respect to the diagnosticity of the predominant experimental paradigm. Experiments that use the language switching paradigm have bilingual participants name pictures in either their dominant language (L1) or non-dominant language (L2). The order of the language can be pseudo-randomized or fixed. Analyses focus on switch costs, which measure how much longer it takes participants to switch into a language (i.e., name a picture with a different language than the one used on the previous trial) compared to the time it takes participants to stay in a language (i.e., name a picture using the same language as the previous trial).

According to inhibitory models, if a bilingual's L1 is much stronger than their L2, then the L1 lexicon should be inhibited more than the L2. Overcoming the strong inhibition of L1 when switching back into it should take time. However, less inhibition needs to be overcome when switching into L2. This should lead to asymmetrical switch costs ($[RT_{L1Switch} - RT_{L1Stay}] > [RT_{L2Switch} - RT_{L2Stay}]$). In some studies, this is what has been found (e.g., Meuter & Allport, 1999). However, other studies have found symmetrical switch costs combined with reverse dominance effects (i.e., L1 naming is slower than L2 naming overall; Christoffels, Firk, & Schiller, 2007).

In an ERP study, Verhoef et al. (2009) manipulated the time bilinguals had to prepare to switch languages. When bilinguals were given less time to use inhibition, the authors found asymmetric switch costs and smaller N2 amplitudes (a measure of inhibition) compared to when they were given ample time. They concluded that inhibition can be used, but is not always needed to control bilingual language production. However, when examining n-2 language repetition costs (a variant of the language switching paradigm), Declerck and Philipp (2018) argue that inhibition in trilingual language production cannot be explained without using inhibition. Additionally, Costa et al. (2006) proposed that less proficient bilinguals use an inhibitory mechanism which leads to asymmetrical switch costs, while proficient bilinguals use a non-inhibitory mechanism leading to symmetrical switch costs and reverse dominance effects. The idea that the need for inhibition changes over time is also supported by the results of Jacobs, Fricke, and Kroll (2016), who found that the amount of cross-language activation may depend on a bilingual's proficiency. However, Finkbeiner, Almeida, Janssen, and Caramazza (2006) have argued that asymmetrical switch costs are the result of using "bivalent stimuli" in experiments, and don't reflect language suppression.

One of the unexpected findings when using the language switching paradigm was that reverse dominance effects were found with symmetrical switch costs. Costa et al. (2006) argue that it reflects a shift in selection threshholds, where bilinguals give preference to the non-dominant language in a switching task. However, the reverse dominance effect was an unexpected finding. Christoffels et al. (2007) have proposed that reverse dominance effects reflect general L1 inhibition, and results from other studies support this idea. For example, Misra, Guo, Bobb, and Kroll (2012) suggest naming in L2 requires sustained inhibition of L1. After bilinguals named a block of pictures in L2, repetition priming was absent when naming a block of pictures in L1. However, the reverse was not true: naming an L1 block first did not eliminate repetition priming when later naming in L2. Additionally, there was a stronger N2 response when naming L1 blocks after L2 blocks but not the other way around. The authors argued that naming in L2 requires inhibition of L1, and the inhibition persists when participants start naming in L1 (although see Branzi, Della Rosa, Canini, Costa, & Abutalebi, 2016; Wodniecka, Szewczyk, Kałamała, Mandera, & Durlik, 2020, for different explanations). Declerck, Kleinman, and Gollan (2020) found reverse dominance effects are correlated with language balance, and concluded that inhibition best accounts for reverse dominance effects. Kleinman and Gollan (2018) found that inhibition of L1 accumulates over the course of a block, eventually causing reverse dominance effects on both stay and switch trials. Taken together, reverse dominance in switching tasks likely indicates that bilinguals use some form of inhibition when controlling languages.

The switching paradigm has been extremely useful for understanding bilingual language control, and the evidence generally points to inhibition being the mechanism bilinguals use. However, the results have been complex, and converging evidence is still needed. In a review of the switching paradigm, Bobb and Wodniecka (2013) state that "other paradigms need to be developed to assess the relative contribution of inhibition to bilingual language control (p.582)." Fortunately, examining semantic effects presents another way to test for inhibition. There are a few studies that examine such effects, but we argue they may have confounded long term incremental learning effects with spreading activation effects. We deal with these issues in the next two sections.

An alternative to the language switching paradigm

Although several studies have examined language switch costs, few studies (if any) have looked at whether spreading activation effects are eliminated after a language switch. Inhibitory and non-inhibitory models make different predictions about what happens to the residual spreading activation in the lexicon of the non-target language. To get a feel for how they differ, we will briefly discuss the idea of activation flow and language selectivity during speech production (for a thorough review, see Costa, 2005).

There are at least three general levels involved in speech production: concepts/semantics, lexicon, and sounds (e.g., Caramazza, 1997; Levelt, Roelofs, & Meyer, 1999). For the most part, inhibitory and non-inhibitory bilingual models agree that conceptual information activates both languages during the initial stages of speech production (although see Costa, Pannunzi, Deco, & Pickering, 2017, for a competing view). The activation flows from the semantic network to both lexicons (e.g., Costa, Miozzo, & Caramazza, 1999; Dijkstra et al., 2019; Hermans, Bongaerts, De Bot, & Schreuder, 1998; Kheder & Kaan, 2019; Santesteban & Schwieter, 2019). The target word is then chosen from the lexicon. How that happens is where the two theories diverge.

Even though both languages seem to be active, non-inhibitory models generally assume that lexical selection is language-specific. Output is controlled by a system that knows what the intended language is without needing inhibition. As Costa (2005, p.313) states, the lexical selection mechanism is "blind to the activation levels of the lexical nodes belonging to the nonresponse language… [and] the level of activation of the target's translation would be irrelevant for the target's selection." For example, if an English-Spanish bilingual intends to say the word dog, conceptual nodes spread activation to both *dog* in the English lexicon and to *perro* in the Spanish lexicon. With no inhibition of the non-target language, activation can theoretically build up over several trials in both lexicons simultaneously. If this is true, spreading activation effects should be unaffected over the course of a picture-naming experiment because there is nothing to counteract the increasing activation.

This process is different for inhibitory models. In such models, lexical selection is language-non-specific and activation of the non-target language must be controlled somehow: after activation flows to both lexicons, the non-target language activation levels act as distractors for the target language. To mitigate this, inhibition is applied to the language not in use. Thus, spreading activation in the non-target language can not build up over the course of a picture-naming experiment. Conversely, if there is a language switch, then the previously activated language gets inhibited. Thus, spreading activation effects are abolished after a language switch. This was an explicit prediction made by Green (1998),

> The controlling schema [can…] reactively inhibit competitors in the non-target language. However, if there is a change of language then any lemmas in the previously active language will become inhibited […] *This should lead to the abolition of both cross-language and within-language competitor priming* [emphasis added]. (p.75)

Consider, for example, an English-Spanish bilingual naming a block of pictures in English that are all semantically related (e.g., mammals). Suppose the first picture is a *dog*, the second is *rabbit* and the third *cow*. Reasoning on the basis of monolingual studies (e.g., Navarrete, Mahon, & Caramazza, 2010; Navarrete, Del Prato, & Mahon, 2012; Navarrete, Prato, Peressotti, & Mahon, 2014), naming might grow faster on average with each successive stimulus. According to inhibitory models, this priming (due to putative spreading activation) should be eliminated after a bilingual participant switches languages (i.e., the bilingual naming *cat* in Spanish; *gato*). Theoretically, it would not matter if the stimulus after the switch was in the same semantic category or in a different one. The facilitation should be abolished. On the other hand, a non-inhibitory model would predict the facilitation to continue even after a language switch because "lexical selection is achieved by a system that does not require the active inhibition of the lexicon-not-in-use" (Costa & Caramazza, 1999, p. 232).

To date, we are unaware of bilingual studies that have looked at naming latencies on a trial by trial basis while manipulating semantic neighbors (i.e., naming semantically related stimuli serially). There have been a few bilingual control studies that examine how language switching affects naming latencies of semantically related stimuli, but those studies did not present semantically-related stimuli one after another. Rather, semantically-related stimuli were separated by filler trials. We argue that filler trials introduce potential confounds through long-term learning effects. We address those confounds in the section *Within-Language Lexical Access: Do Lexical Entries Compete for selection? Implications for Bilingual Lexical Access*.

The continuous naming paradigm: incremental learning effects or a lack of inhibition?

Sometimes, monolingual studies have tried to examine within-language competition during speech production using the continuous naming paradigm. Under the continuous naming paradigm, participants name stimuli from several semantic categories. The stimuli are presented so that pictures of a similar semantic category are never presented one after another. Naming latencies for semantically-related trials tend to increase by 10–30 ms with each presentation of a semantic neighbor (e.g., Damian & Als, 2005). This increase has been termed *cumulative semantic interference* (i.e., CSI).

Initially, researchers thought that semantic activation persists over time, creating interference. For example, Wheeldon and Monsell (1994) found that when a semantic prime was presented before a stimulus, naming latency was longer compared to an unprimed stimulus. Interestingly, short lags produced less inhibition than longer lags. They explained this result by proposing two types of activation during speech production. The first occurs within the semantic network and is relatively short-lived. It has a small facilitatory effect. The second occurs at the lexical level and has a stronger inhibitory effect. This led to the idea that activation builds up over time in the lexicon and that semantic neighbors compete for selection. CSI was thought to be a reflection of this.

However, such an explanation has been questioned in more recent experiments because the number of filler trials does not affect the magnitude of the cumulative semantic interference (Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Navarrete et al., 2010). One might expect the residual activation to dissipate after several filler trials. Additionally, Navarrete et al. (2012) found facilitatory effects trial by trial when participants named semantic neighbors without filler trials, but interference effects when filler trials were introduced. Critically, when there were no filler trials between semantic neighbors, each successive trial was named faster than the previous trial. Based on the Wheeldon and Monsell (1994) results, one would expect interference after the second or third presentation of a semantic neighbor. It seems unlikely that CSI is the result of persisting activation. Rather, incremental learning is a more likely candidate.

Incremental learning is an idea inspired by neural network models (e.g., Oppenheim, Dell, & Schwartz, 2010). Within the context of lexical access, it posits that connections between semantic and lexical nodes are constantly updated, albeit incrementally. When a picture is presented (e.g., *dog*), and a participant names it, the neural connections between the concept and word become strengthened. This strengthening is long lasting, and is different than just temporary activation. When the picture is presented a second time later on in an experiment, the word is retrieved more quickly. This type of facilitation is commonly referred to as repetition priming, but according to incremental learning, it is a function of memory, not activation. There is a cost associated with this incremental learning. After the picture is named, the connections between the target word and the semantic network get stronger, but the trade-off is that the connections between semantically-related neighbors and the semantic network become weaker. For example, naming a picture of a bat will strengthen the connections between *bat*'s conceptual and lexical nodes, but it will weaken the connections between *whale*'s conceptual and lexical nodes. When *whale* must be named, it takes longer to retrieve it from the lexicon due to the weakened connections. However, at the end of the trial, *whale*'s connections are strengthened, and its neighbors' are further weakened. When a third neighbor (e.g., *dog*) is named, naming takes even longer than it did for *whale*. If incremental learning is what is responsible for the cumulative semantic interference in bilingual naming experiments using variants of the continuous naming paradigm, then claims about inhibition and/or spreading activation should be met with skepticism.

Within-language lexical access: do lexical entries compete for selection? Implications for bilingual lexical access

Evidence that lexical entries compete for selection within a language comes from a few different experimental approaches. The first source of evidence comes from the continuous naming paradigm (the CSI effects discussed previously) the second source is from distractor tasks, and the third is from the blocked-naming paradigm.

We should note that conclusions from all three experimental methods have been challenged with respect to whether competition exists during lexical access. We think those challenges are important, and they bring up potential issues with construct validity. We have briefly discussed whether CSI effects under the continuous naming paradigm reflect competition or incremental learning. We now turn our attention to distractor tasks and the blocked-naming paradigm.

Experiments that use distractor tasks (Damian & Bowers, 2003; Hermans et al., 1998) have participants name pictures while a distractor word is presented orally or visually. If semantic neighbors compete at the lexical stage, then artificially increasing the activation of a semantically related distractor (e.g., through semantic relatedness) should increase naming latencies of a target word. This seems to be the case. For example, a distractor like *mouse* slows naming of semantically related target like *dog* compared to an unrelated target like *airplane*. The increase in naming latency has been attributed to competition between lexical entries.

This explanation is supported by results of Shitova, Roelofs, Schriefers, Bastiaansen, and Schoffelen (2016). They compared a Stroop task to a picture word inteference task, a type of distractor task. The authors wanted to determine when exactly the interference happens in picture-word interference tasks. The authors argued that if the interference happens during the word planning stage, then an N400 wave should be observed in picture-word interference tasks (i.e., negativity 400 ms after stimulus onset; this is the time that lexical selection is thought to happen). That is what they found: a similar N400 for both the Stroop and Picture Word Interference tasks. The authors concluded that the negative ERP found for the incongruent stimuli and semantically related distractors arises during the word-planning stage.

Distractor tasks have been important for understanding how the lexical and semantic networks connect. However, there are potential pitfalls when using them to make conclusions about lexical competition. For example, in picture word interference tasks, participants name a picture stimulus while trying to ignore a distractor word presented at the same time. In general, semantically related words slow naming compared to semantically unrelated words. But how much of this is due to competition within the lexicon vs. inhibition from another process? A semantic distractor may share visual features with the target. As the participant reads the distractor word, those visual features are also activated. This activation happens in bottom-up fashion (i.e., the distractor is exogenous to the participant), and must be ignored/inhibited. The process may not reflect normal lexical access.

The forgoing suggests it may be informative to compare such a procedure with a flanker task (Eriksen & Eriksen, 1974) where a target stimulus (e.g., an arrow) is flanked by visually similar distractor stimuli (e.g., arrows pointing in the opposite direction). In picture-word interference tasks, semantically related distractors take on the role of the incongruent flankers, and the "competition" found in such experiments is taken as an inverse measure of attentional filtering or inhibition, as in classical Flanker logic. Using this procedure, Dell'Acqua et al. (2010, p. 8) found that the target stimulus "initiates ultra-fast access to semantic representations," which then influences the orthographic processing of the distractor word. Additionally, if words compete within the lexicon, then semantically near distractors (e.g., *zebra*) should produce more interference with a target (e.g., *horse*) compared to a semantically distant distractor (e.g., *whale*). However, Mahon, Costa, Peterson, Vargas, and Caramazza (2007) found the opposite pattern of results, and they concluded that lexical selection among monolinguals is non-competitive. These considerations make it difficult to identify the locus of competition in distractor tasks.

The third and final source of evidence is the blocked naming paradigm. This paradigm has similar methodology to the continuous naming paradigm. Under the blocked naming paradigm, participants name stimuli one after another in semantically related or semantically unrelated blocks. Depending on the experiment, words sometimes repeat within a block. Sometimes semantically related blocks contain stimuli from more than one category. Usually, care is taken to make sure unrelated blocks have no stimuli from the same category. For example, Damian, Vigliocco, and Levelt (2001) used 25 pictures from 5 semantic categories. In related blocks, 5 stimuli from the same category were named multiple times. In unrelated blocks, 5 stimuli from different categories were named multiple times. Kroll and Stewart (1994) were among the first to use this paradigm. Their results showed faster naming latencies in unrelated blocks compared to related blocks. This result is consistent with the idea that lexical entries compete for selection. The logic being that in related blocks, the activation from previously named semantic neighbors makes it more difficult to name the current stimulus.

However, Navarrete et al. (2014) have argued that blocked naming studies do not account for incremental learning effects. More specifically, they take issue with the fact that blocked naming studies often repeat stimuli more than once per block (e.g., Damian et al., 2001). According to theories of incremental learning, repeating a stimulus strengthens the stimulus' connections to the semantic network and causes repetition priming. However, such repetition also has the unintended consequence of weakening the connections of semantic neighbors. In a block of related stimuli, this weakening is a problem and will increase overall reaction times when semantic neighbors are named. In an unrelated block, this is not a problem: semantic neighbors with weakened connections are never named. In fact, because stimuli are repeated more than once, unrelated blocks receive an overall benefit from repetition priming. According to Navarrete et al. (2014), related blocks are slower because of these weakened connections, not because of competition/interference. Other studies have also pointed to incremental learning as the cause of semantic interference in monolingual studies (e.g., Damian & Als, 2005; Navarrete et al., 2012). Our perusal of the prior literature reviewed thus far suggests to us there are at least 3 options for providing a clear and decisive test of the inhibitory and non-inhibitory models.

In this section, we have explored three influential and useful experimental methods that examine whether lexical entries compete for selection. There are three potential confounding effects that may lead to an incorrect conclusion that competition exists in the lexicon: effects related to (1) attention, (2) weakening of semantic connections due to incremental learning and (3) strengthening of semantic connections due to incremental learning.

Specifically, To eliminate attentional effects (1), one should choose a task that is free of verbal distractors within a trial. To mitigate incremental learning effects that weaken connections (2), one should present semantically related stimuli serially one after another. Doing so gives spreading activation the best chance for n-1 to affect trial n before the residual activation decays. To mitigate incremental learning effects that both strengthen and weaken connections (2 and 3), one should control for repetition effects. One way to do this is not repeat a stimulus in the experiment until all other stimuli have been named. Effects from incremental learning should be canceled out (i.e., all semantic neighbors connections have been strengthened and weakened equally). For example, Navarrete et al. (2014) divided stimuli in half for unrelated and related blocks and counterbalanced their presentation across participants. Another option is to alternate between unrelated and related sub-blocks while not repeating a stimulus in the experiment until all have been presented. We chose the latter option. Either way, this should prevent stimuli within an unrelated block from benefiting from repetition priming (i.e., within a block, there is no repetition).

The question naturally arises: Do these three potential confounding effects also apply to bilingual production studies? The answer is most certainly yes. For example, Kleinman and Gollan (2018) observed repetition priming in a bilingual switching experiment.[2] Additionally, three bilingual control experiments have used a paradigm similar to continuous naming in order to test for inhibition (Lee & Williams, 2001; Hong & Macwhinney, 2011; Runnqvist, Strijkers, Alario, & Costa, 2012), but some of their results can be explained by incremental learning effects.

For example, Lee and Williams (2001) had participants name target words after a language switch in English or French. Before filler trials, targets were either primed or unprimed. In primed sequences followed by a language switch (e.g., *snow [English prime], cinema [English filler], toit[French filler], rain [English target]*), target words were named at similar speeds as unprimed sequences with a language switch (e.g., *heart [English non-prime], cinema [English filler], toit[French filler], rain [English target]*). However, if the sequence was primed without a language switch (e.g., *snow [English prime], cinema [English filler], roof [English filler], rain [English target]*), naming of the target slowed down. Lee and Williams (2001) argued that switching languages abolishes semantic interference because inhibition suppresses activation of semantic competitors. Hong and Macwhinney (2011) tried to extend these results to Chinese-English bilinguals. For fluent bilinguals, they did not replicate the results of Lee and Williams (2001), and switching languages did not abolish semantic interference effects.

In a series of experiments, Runnqvist et al. (2012) found that

[2] Kleinman and Gollan (2018) argue that the L2 repetition priming they found was caused by a lack of inhibition compared to L1, but they acknowleged that some of their results could be due to incremental learning and proposed further research to work out the details.

switching languages did not affect semantic interference when naming under the continuous naming paradigm. In one of their experiments, participants named semantically related stimuli separated by filler trails in one language only (e.g., *car* followed by two filler items in the same language, *airplane*, two fillers, *bus* etc.). With each presentation of a semantic neighbor, interference increased by about 12 ms (e.g., naming *airplane* took 12 ms longer than naming *car*; naming *bus* took 12 ms longer than naming *airplane*). In another experiment, the filler items were in another language, forcing participants to switch languages. The interference pattern was the same: with each semantic presentation, naming latencies increased 12 ms. There was no interaction between language switching and experiment. This was true for both L1 and L2, and the authors concluded that bilinguals do not use inhibition to control language output.

The three studies assumed, based on Wheeldon and Monsell (1994), that a prime stimulus stays active for some time after it is named. If true, then the long lasting activation can affect a target long after it is named. The activation is also sensitive to inhibition. If bilinguals use inhibition when controlling language output, then a language switch should abolish activation associated with the prime stimulus.

We are worried that testing for cumulative semantic interference using the continuous naming in these three studies does not elicit ever-increasing spreading activation that stays elevated for “some time” (Runnqvist et al., 2012, p. 851). Rather, the CSI is due to an incremental learning mechanism. As initially proposed by Collins and Loftus (1975), it was assumed that spreading activation decays over time. This decay should reduce cumulative interference when semantically related neighbors are separated by filler trials. This does not seem to happen in experiments with monolinguals or bilinguals. But, if the cumulative interference is the result of trade-offs associated with incremental learning, then inhibition caused by a language switch would have no effect on the observed increase of reaction times. In other words, a different method is needed that examines priming over a short time (i.e., naming semantic neighbors without filler trials).

The current study

Because of the inconsistent results obtained with the language switching paradigm (see Bobb & Wodniecka, 2013), and the potential confounds from semantic competitor studies (Lee & Williams, 2001; Hong & Macwhinney, 2011; Runnqvist et al., 2012), we suspect that modifying the methods of the blocked naming paradigm may provide more diagnostic tests of inhibitory control accounts. To make clear predictions, we instantiated two models of bilingual language control to understand how reaction times might change trial by trial in a picture-naming experiment. One of the models assumes that inhibition is used, while the other does not. Consistent with the bilingual literature, both models assume competition exists within a language. However, only the Inhibitory Model assumes competition between languages. We use the models to simulate reaction times over two types of sub-blocks: semantically uniform and semantically mixed. Semantically mixed blocks are analogous to unrelated blocks. Semantically uniform blocks are analogous to semantically related blocks. In the semantically uniform blocks, naming latencies of six semantically related stimuli are estimated. The first three trials are stay trials, trials four and five are switch trials, while trial six is also a stay trial. Semantically mixed blocks are identical in terms of trial type (i.e., stay/switch), and semantic relatedness on trials one through four. However, the semantic category changes on trial five. Comparisons on trials five and six between the mixed and uniform sub-blocks should help explain whether switching languages abolishes semantic interference effects. Additionally, it is assumed that a stimulus is never repeated until all other stimuli have been named. Thus, the overall design is similar to the blocked naming paradigm with three exceptions: (1) Mixed sub-blocks contain semantically related stimuli to produce spreading activation effects that can be influenced by language switching, (2) reaction times are analyzed trial by trial within a sub-block (not averaged by sub-block) and (3) repetition effects are controlled for. The models were derived in order to make concrete, *a priori* predictions, and estimate the magnitude of semantic interference effects due to spreading activation.

The predictions of the two models were tested in Experiment 1. For example, if there is no difference in naming latencies after a language switch between the semantically mixed and semantically uniform blocks (i.e., on trial five of a sub-block), then this would indicate inhibition is used by bilinguals to control language output. However, if naming latencies after a language switch are longer during semantically uniform blocks compared to mixed blocks, this would indicate inhibition is not used. Note that Experiment 1 controlled for the possible confounds mentioned in the previous section. There were no distractor items to control for attention effects. To mitigate incremental learning effects, stimuli in the experiment did not repeat until all other stimuli had been named, and there were no filler trials between semantic neighbors in order to allow semantic activation to spread before it decays. Additionally, we primed each target with four semantically-related neighbors (three stay followed by one switch). If the conclusions of Wheeldon and Monsell (1994) are correct, then it may take “ a delay of only 3–4 s” for the competition in the lexicon to offset and overpower the transitory facilitation from the semantic network (p.345). Any facilitation seen in the second prime should be counteracted by the presentation of the third prime.

For the second experiment, we thought it was important to replicate results from the continuous naming paradigm. Consider, for example, the possibility in Experiment 1 that there is no difference in naming latencies between mixed and uniform blocks after a language switch. This would provide evidence that bilinguals use inhibition when switching languages. But, one might argue that the stimuli used in the current study were flawed (e.g., the stimuli in this study were not sufficiently related to produce interference). We therefore conducted the second experiment where the same stimuli were used, but this time semantically related pictures were separated by filler trials (i.e., stimuli from other semantic categories). If, with each presentation of a semantic neighbor, naming latencies increase even after a language switch, then this would indicate that the stimuli were sufficiently related to produce semantic interference. Additionally, if in Experiment 1, no difference is found between uniform and mixed blocks after a language switch, but in Experiment 2 semantic interference is found, this would strongly support the idea that the semantic interference found in the continuous naming paradigm is the result of incremental learning, not spreading activation.

Deriving quantitative predictions from the inhibitory and non-inhibitory accounts

Because there have been inconsistencies in the literature in trying to determine whether bilingual speakers rely on inhibition to control language output using switch costs, we have chosen to examine another aspect of bilingual language control: how language switching affects spreading activation. To derive predictions *a priori*, two computational models were developed: one that assumes inhibition, and one that does not. The former is inspired by the Inhibitory Control Model (Green, 1998). We refer to this as the Inhibitory Model. The second is based on research by Costa and colleagues (Costa & Caramazza, 1999; Costa & Santesteban, 2004; Costa et al., 2006). We refer to this as the Non-Inhibitory Model.

Both models are structurally similar. They are partly inspired by the Dual Route Cascaded Model (DRC; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) in that the model architecture has been specified beforehand rather than using a learning algorithm (e.g., backpropogation). Because of this, specific predictions based on the models were made *a priori* before the experiments were conducted, which allowed for clear predictions based on the verbal theories.

Both models use three inputs to estimate naming latencies trial by trial. Those inputs are language (L1 or L2), type of trial (SWITCH or STAY) and semantic relatedness of trial n to trial n-1 (TRUE or FALSE). On a trial, a target word receives activation from the semantic network.[3] Once the target word's activation reaches some threshold that is greater than its distractors, it is chosen. Its activation (and the distractors' activations) then decays and the next trial starts. Notice that two of the three inputs (i.e., type of trial and semantic relatedness) for a given trial rely on characteristics of the previous trial. For example, if the current trial was in the same semantic category as the previous trial, then the semantic-relatedness input would be set to *true*. This tells the model that any residual activation from the previous trial affects the current trial. In this way, effects of spreading activation (or lack of it) can be accounted for. This will be explained partially in later sections entitled *The Inhibitory Model* and *The Non-Inhibitory Model*.

General modeling approach

We begin by making the fewest assumptions possible in deriving a general model which can be easily adapted to instantiate the inhibitory and non-inhibitory accounts. We start by assuming, as the verbal accounts do, that performance latencies reflect changes in activation over time within a lemma, with these changes influenced by inputs from the semantic network. Then simple expressions for activation $a(t)$ in the lemma and input $\sigma(t)$ from the semantic network can be written down as follows:

$$\begin{aligned} &\frac{da}{dt} + c_0 a(t) = pf(t) \\ &\frac{d\sigma}{dt} + c_0\sigma(t) = f(t) \end{aligned} \tag{1}$$

This system states that the change in activation over time ($\frac{da}{dt}$) in a lemma beyond its prior or resting level is due to a positive contribution proportional (p) to its input from the semantic network (expressed as a function of time or $f(t)$) as well as a negative contribution ($c_0\tilde{}a(t)$) representing decay and/or inhibitory loss proportional ($-c_0$) to the prior or resting level activation ($a(t)$). A suitable and relatively uncontroversial choice can also be made for the form of the activation function $\sigma(t)$, the sigmoid, in which case the second equation is absorbed into the first. Regrouping terms we have

$$da = -c_0(a(t) - p\sigma(t))dt + pd\sigma \tag{2}$$

This is a simple linear first-order ordinary differential equation, which can be solved using an integrating factor:

$$\begin{aligned} a(t) &= e^{-c_0 t}\int pf(t)e^{c_0 t}\mathrm{d}t + ce^{-c_0 t} \\ &= pe^{-c_0 t}\int\left(\frac{c_0 e^{c_0 t}}{1+e^{-t}} - \frac{e^{c_0 t - t}}{(1+e^{-t})^2}\right)\mathrm{d}t + ce^{-c_0 t} \\ &= pe^{-c_0 t}\frac{e^{c_0 t}}{1+e^{-t}} + ce^{-c_0 t} \\ &= ce^{-c_0 t} + p\frac{1}{1+e^{-t}} \end{aligned} \tag{3}$$

The final equation shows that activation of a lemma at a given time point ($a(t)$) is equal to the amount of inhibition applied to a lemma ($ce^{-c_0 t}$) plus the amount of activation from the semantic network ($p\frac{1}{1+e^{-t}}$) where p controls the proportion of total activation the lemma receives. In what follows, we use this basic functional form as a skeleton for instantiation of the Inhibitory and Non-Inhibitory models.

[3] The model makes no claims about the semantic network's structure (e.g., whether it is decompositional or non-decompositional)

The inhibitory model

The Inhibitory Model tries to represent activations of a target word, its semantic neighbors, and unrelated words on both stay and switch trials. The total activation of a word ($a_{j,k,l,m}$) is calculated by adding activation from the semantic network, or removing activation through decay or inhibition. The subscripts j, k, l and m stand for type of trial (j; stay [$j = 1$], switch [$j = 2$]), type of word (k; target [$k = 1$], previous target [$k = 2$], and other distractors [$k = 3$]), language (l; dominant language [$l = 1$], non-dominant language [$l = 2$]), and m refers to whether the word is in the intended language or unintended language (intended [$m = 1$], unintended [$m = 0$]). The following equation is used to calculate a word's activation level on stay trials at any given point in time:

$$a_{1,k,l,m} = a_0 e^{-\epsilon h_l t} + p_{k,m}\frac{1}{1+L_{j,l}e^{-t}} \tag{4}$$

where a_0 represents a word's initial activation at the beginning of a trial, t represents the total time activation is applied to a word, h_l is an inhibition parameter whose value depends on the relative strength of a bilingual's language, and p represents the proportion of activation a word receives from the semantic network based on whether the word is a target and in the intended language. Target words in the intended language receive most of the activation (i.e., $p_{1,1}$=0.75 or 75%). Distractors in the intended language split the remaining activation. Distractors in the unintended language receive no activation (e.g., $p_{1,2}$=0). This is a simplifying assumption. It is unlikely the non-target language receives no activation. However, inhibitory models do assume that the net effect is inhibition. This assumption ensures activation decreases and that "lemmas in the previously active language [become] inhibited" in order to "[abolish] both cross-language and within-language competitor priming" (Green, 1998, p.75). ϵ determines whether the word is inhibited on a given trial. Specifically, ϵ is set equal to 1-m, and the exponential acts as an indicator, turning off the inhibitory coefficient when the word is in the intended language ($m = 1$), and turning it on if the word is in the unintended language ($m = 0$). L determines the rate at which a word receives activation from the semantic network. The larger L is, the slower the rate of activation. We assume that the rate a word receives activation is independent of inhibition. For example, *L2* words could receive activation at a slower rate because their connections to the semantic network are weaker, and not because inhibition is being applied to them. We conceptualize the L parameter as an activation parameter and not an inhibition parameter.

Inhibition is controlled by the p parameter and the $e^{-\epsilon ht}$ portion of the equation. Note that for words in the intended language (i.e., when ϵ=0), there is no inhibition (i.e., e^0 leaves only $a_0$in the left term of RHS of the activation equation), and activation from the semantic network is added to the initial activation at the beginning of the trial. Conversely, words in the unintended language receive no activation (i.e., $p_{k,1}$=0) from the semantic network, but they are inhibited based on their initial activation levels at the start of the trial. On switch trials, inhibition of the non-target language occurs in parallel with activation of the target language, but only after the target language has been reactivated. This means that the language activation parameters must be stronger than in a non-inhibitory model to reactivate the target language. Theoretically, the more proficient a bilingual is, the more activation they need to remain efficient. These simulations show that strong activation is also needed to offset strong inhibition.

Consistent with previous experimental results, it is also assumed that the non-target language remains inhibited until a language switch (Misra et al., 2012). In other words, the model assumes that there is global inhibition of the non-target language.

The Inhibitory Model assumes a fully competitive system, both between and within languages. In order for the target word to be chosen on a stay trial, its activation level must be some ratio (V; the competition parameter) of the sum of all other distractor activations in both languages. In other words, for a target to "win" its activation must be greater than or equal to a fraction V of the sum of all the activations for the distractors in the target language plus the sum of the activations for all the translated words in the non-target language. Denoting a particular combination of conditions on the target activation subscripts as x and that of all other non-target words as d, this response rule can be written

$$V \leqslant \frac{a_x}{\sum a_d} \tag{5}$$

which amounts to another fairly uncontroversial and widely-used decision function in mathematical modeling, the Luce ratio decision criterion. [4]

Once V is less than or equal to the ratio of the target word and the sum of the distractors, the target is selected. Until this happens, activation or inhibition is applied to each word. If $V = 0.50$, then one can calculate the time needed by replacing a_x and $\sum a_d$ with their respective equations, and solve for t. t is then converted to milliseconds and is added to a noise parameter. The noise parameter changes with each trial and is randomly selected from an ex-Gaussian distribution, which has three parameters: μ, σ, and τ (Luce, 1986).

The target word and distractor activation levels then decay based on an exponential decay function. If the next trial's stimulus is from the same semantic category as the current trial, then the initial activation levels of the next trial's words are set to the final decayed activation of the current trial. If not, the activation levels reset to the initial activation of a language's words on trial one of the simulation (this is the resting activation parameter, or R_l). This process allows spreading activation of semantically related trials to affect each other, while eliminating spreading activation once a semantic category changes.

On switch trials, lexical selection happens in two steps. First, the intended language's words must be reactivated. This is represented in the following equation by T_0:

$$T_{0,k,l,m} = a_0 e^{-\epsilon h_l t_s} + \epsilon \frac{-R_l Y_l}{e^t + 1}\Big|_0^{t_s} \tag{6}$$

where the rate of reactivation (Y) depends on how strongly the intended language's words were inhibited (h) and the overall strength of the language (L) on the previous trial. Thus, Y is proportional to the sum of the language strength parameter on stay trials plus the inhibition parameter (i.e., $Y \propto Lh$). t_s is the switch cost, and represents how much time has passed in this reactivation stage. Once T_0 is greater than or equal to R, the lexical selection stage begins and T_0 acts like the initial activation a_0. The equation then becomes

$$a_{2,k,l,m} = T_0 e^{-\epsilon h_l t_s} + p_{k,m} \frac{1}{1 + Y_l e^{-t}} \tag{7}$$

t can then be calculated in a similar manner to how it is found in stay trials. However, in order to find the total time it takes to select a word on a switch trial, t_s must be added to t. Then, it is added to the noise parameter. Note also, that the language strength parameter (L) has been replaced with the rate of reactivation parameter (Y). It is assumed that the rate of activation from the semantic network is still affected by inhibition, which is why Y is used instead of L.

The non-inhibitory model

The Non-Inhibitory Model is similar to the Inhibitory Model, with three major exceptions. First, there is no between-language competition (i.e., between L1 and L2), meaning inhibition is not needed. Thus, on both stay and switch trials, the total activation of a word is represented by the following equation:

$$a_{1,k,l,m} = a_0 + p_{k,m} \frac{1}{1 + L_{j,l} e^{-t}} \tag{8}$$

Second, having no between-language competition also affects how words are selected: the denominator in the decision rule, T_d, is now restricted to only the set of within-language competitors d^w[5]:

$$V \leqslant \frac{a_x}{\sum a_{d^w}} \tag{9}$$

Finally, on switch trials, there is no need to reactivate an inhibited language, and there is no inhibition parameter h. Switch costs are determined by manipulating the L parameter. L determines how fast a word receives activation from the semantic network. This then determines how quickly the ratio between target word and distractors changes. To make the model more equivalent to the inhibitory model, the L_{switch} parameter is also proportional to the L_{stay} parameter. This is also consistent with Costa et al. (2006), who argue that selection criteria threshholds are related to a bilingual's proficiency in a language. Theoretically, one could manipulate the L_{stay} parameters to get reverse-dominance effects (i.e., making the $L_{1,\ stay}$ parameter larger than $L_{1,\ switch}$ would also increase the $L_{1,\ switch}$ parameter relative to the $L_{2,\ switch}$, slowing naming for L1 in general), but that is beyond the scope of this study and was not tested here.

A priori simulations to predict effect size

Overview. Before gathering any data, simulations based on the parameters chosen *a priori* were conducted. This was done to constrain the models' predictions and estimate effect size. Note, that after we gathered data in Experiment 1, we fit the models to the data. We report those results under the section titled *Fitting the Models to the Data of Exp. 1*. Two-hundred experiments were simulated for each model. Each simulated experiment consisted of 40 "participants" naming 768 trials. Trials were grouped into two types of sub-blocks: mixed and uniform. Each sub-block type consisted of six trials. In uniform sub-blocks, all stimuli came from the same semantic category. In mixed sub-blocks, semantic category changed on Trial 5. All sub-blocks had the same trial type order: stay, stay, stay, switch, switch, stay.

Because uniform sub-blocks are all semantically related, the Non-Inhibitory Model should predict that there will be semantic interference on trials five and six, even after the language switches on trial four. However in mixed sub-blocks, changing semantic categories on trial five should abolish these effects. Thus, the Non-Inhibitory Model should predict longer naming latencies for uniform sub-blocks on trials five than for mixed sub-blocks. On the other hand, the Inhibitory Model should predict that inhibition applied to the non-target language on trial four will abolish interference effects on trials five during uniform sub-blocks. In other words, the Inhibitory Model should predict similar reaction times on trials five and six for both mixed and uniform sub-blocks.

It is assumed that noise in the model follows an ex-Gaussian distribution, and each simulated participant was assigned a μ, σ, τ parameter. A random value from a participant's distribution was selected each trial. Thus, random variation was added to reaction times based on an exponentially-driven distribution.

Parameter Values. Parameter values were chosen by hand. Two considerations were made when choosing parameter values. First, we wanted to be as faithful to the verbal theories as possible. Second, we wanted to make realistic predictions based on previous research.

It should be noted that various combinations of parameter values can give similar results. Adding a new parameter (e.g., one related to inhibition) often necessitates changes in other parameter values in order to

[4] Reordering the formula, shows the activation of the target relative to the distractors must be larger than the threshold value $a_x \geqslant V \cdot \sum a_d$

[5] Rearranging terms gives $a_x \geqslant V \cdot \sum a_d$

obtain the same numerical predictions. Therefore, we chose not to make the L1 and L2 parameters equal between the models because we wanted their architectures to remain true to the verbal theories they were built to instantiate. We then set the values to make reasonable predictions before gathering data. The problem with setting the parameter values between models exactly equivalent at this stage is that inhibition has a strong effect on the simulations because more activation is required to overcome inhibition and reach the threshold needed to name the picture. In other words, a model that contained all the parameter values of the non-inhibitory model and simply added inhibition on top would result in wildly different overall naming times, which would make the comparisons between the models' predictions to the data more difficult to carry out.

The language strength parameters on stay trials, L_{stay}, represent how quickly the target word or distractors receive activation from the semantic network. The lower the number, the faster the word is activated. For the Inhibitory model, we chose 0.5 for L1 and 8 for L2 to capture the assumption that L1 is stronger than L2.

The Non-Inhibitory model is more balanced ($L1 = 6$ and $L2 = 8$) for two reasons: (1) It has fewer competitors (i.e., no between-language distractors) and a value of 0.5 for L1 would cause lexical selection to happen too quickly without much interference and as initially proposed (2) Costa and colleagues theorized that balanced bilinguals are more likely to use a non-inhibitory mechanism to control language output.[6] We also assumed that very few bilinguals are completely "balanced," so we made L1 slightly stronger. It should be noted that making the non-inhibitory model's L1 and L2 parameters less balanced can still produce similar reaction times in the simulations.

The fact that the language strength parameters are different across models and languages is theoretically driven. But, it actually has little influence on reactions times. The language strength parameters reflect how fast a word receives activation from the semantic network. Because the non-inhibitory model has fewer competitors, and the inhibitory model uses inhibition that happens in parallel with the activation on switch trials, the models can have relatively similar naming latencies. Both models can be equally efficient. Additionally, we assume that a larger language strength parameter does not necessarily mean inhibition. Many things can affect how fast the semantic network activates words, including the strength of the connection to the lexicon, the lexical selection mechanism etc.

The next two parameters are the h parameters. These parameters are unique to the Inhibitory Model. They determine how quickly words in the non-target language are inhibited. Theoretically, this should happen quickly. The larger h is, the greater the inhibition. We chose the values 2.5 and 1.5 for L1 and L2 respectively. Both accomplish the goal of quick inhibition. L1 is chosen to be greater because that is one of the main assumptions Green (1998) makes (i.e., more inhibition is applied to the stronger language).

Next are the Y parameters. These parameters are also unique to the Inhibitory model. They determine how quickly a word is reactivated and activated after being inhibited. Y is thus analogous to the language strength parameter on stay trials.

We assumed that the rate of reactivation is reflected in bilinguals' naming latencies on stay trials, which is the core theoretical assumption in the literature. Additionally, latencies should depend on how much a language was inhibited. Paraphrasing Green (1998) the greater the inhibition that is applied to a language, the more effort it takes to overcome that inhibition.

Thus, we chose to set the reactivation parameter $Y = Lh$. Since this is in the bottom of the logistic equation, larger values of Y will reduce the rate at which a word is reactivated (i.e. the slope of the sigmoid). Thus, a bigger h (i.e., more inhibition) slows down the reactivation, as Green proposed. It also means that this is not a free parameter but is fully determined by the choice of L and h.

The L_{switch} parameters only apply to the Non-Inhibitory model. Mechanisms controlling language output in the model need time to carry out the operation. Like the L_{stay} parameter, the larger L_{switch} is, the slower the rate at which a word is activated. L_{switch} must be quite a bit larger than the L_{stay} parameters in order to create switch costs that are realistic (80–100 ms). After trial and error, these values were set to 40.5 and 54.

The larger the c parameter is, the more rapidly word activations decay. c was set to 0.01 on the assumption that decay happens somewhat gradually over a couple thousand milliseconds. If c were too high, then there would be no interference from one trial to the next (i.e., all the activations would decay back to resting activation by the start of the next trial). To help guide this choice, we chose values that were slightly higher than the resting activation parameters (3.2–4.0), applied the decay function to them with various c values, and then plotted the decay functions over time to see what values of c appeared most reasonable. 0.01 worked well in that it allowed activation from the previous trial to affect the current trial without causing too much interference.

The resting activation parameter, R, captures what a word's activation level should be in the absence of semantic input or inhibition. For simplicity we assume R is the same for all words (i.e., targets and distractors) within a language.

We chose 3 for L1 and 1.5 for L2 to indicate that the first language is stronger than the second and should have greater resting activation. The chosen values are somewhat arbitrary, but they cannot be too big (i.e., 10) because in such a case the target activation measured over time (based on the logistic function) would never become large enough to overcome the competition of the distractors.

As mentioned earlier, the p parameter reflects the proportion of semantic activation a word receives, and we started with the assumption that the target word gets the bulk of the activation. 75% strikes us as a reasonable, though somewhat arbitrary, starting assumption. But, because the Inhibitory model has more potential distractors, we increased the value to 80% to give the target a boost in that model. The other distractors split the remaining activation.

To choose the value of the V parameter, it was necessary to consider what the words' activation levels are ($V =$ target activation/distractor activation). There are three main activations to focus on in this model: target, main distractor, and other distractor. In the Non-Inhibitory model, all 3 types of words start with roughly the same activation (≈ 3 for each). If $V =$ Target Activation/Distractor Activation, then at the very beginning of the trial $V/D = 0.5(i.e., 3/(3+3))$.

Based on these initial values, and the values of the R and L, 0.566 produced realistic interference effects within a language (≈20–25 ms with each presentation of a semantically related neighbor).

This value was used to inform the Inhibitory model as well [$V = 3/(3 + 3 + 1.5*3)$, where 1.5 is the initial activation of L2 distractors]. The inhibition and activation occur in parallel, and the between-language distractors' activation levels quickly approach a fraction of their initial activation. Setting V to 0.565 gives similar interference results to the Non-Inhibitory model. It is not quite as large because the between-language distractors make it slightly more difficult for lexical selection to take place as V increases.

Results. Averaged simulation results by trial of the Inhibitory and Non-Inhibitory Models can be found in Fig. 1 (left and middle panels). To analyze within-language semantic effects, regression analyses with trial number (one through three) as the independent variable were used to predict mean reaction times. For the Inhibitory Model, each subsequent trial significantly increased reaction times ($\beta = 24.95$, $p < .01$), indicating that semantic relatedness on the previous trial interfered with naming on the current trial. Similar results were found for the Non-Inhibitory Model, ($\beta = 21.68$, $p < .01$).

The models make different predictions after a language switch. In order to determine if spreading activation interference effects were

[6] Although Declerck and Philipp (2018) have found evidence that balanced bilinguals also use inhibition

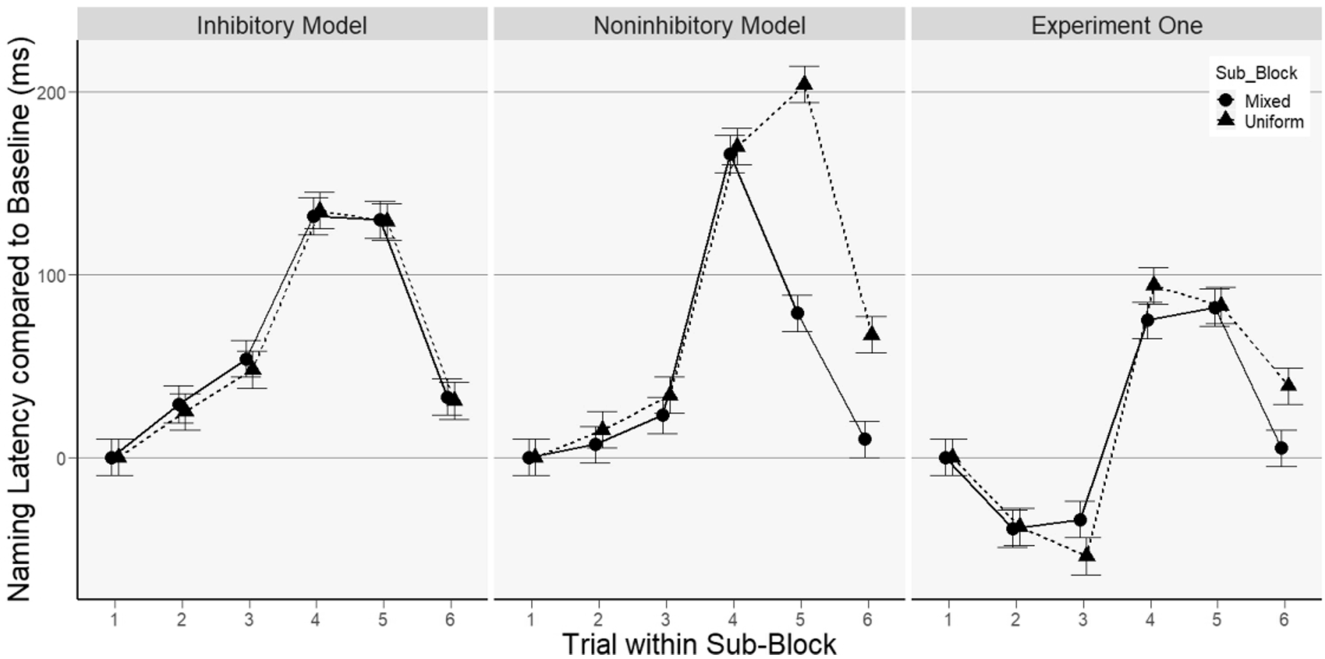


Fig. 1. Naming Latency *a priori* predictions for the Inhibitory Model (left), Non-Inhibitory Model (middle) and empirical data from Experiment 1 (right). Trials 1-6 were normalized by subtracting the naming latency from Trial 1 from each trial. In panel c, the only significant difference between mixed and uniform sub-blocks occurred on Trial 6.

abolished after a language switch, naming latencies of mixed and uniform sub-blocks were compared on trials five and six. For the Inhibitory Model on the fifth trial, (i.e., second switch trial), naming latencies in uniform sub-blocks ($M = 1059$, $SD = 53.20$) were 2 ms faster than naming latencies in mixed sub-blocks ($M = 1061$, $SD = 54.31$). In roughly 52% of the simulations on trial five, uniform sub-blocks were faster than mixed sub-blocks, $\chi^2_{1,N=200} = 0.16, p = 0.69$, suggesting that there is truly an underlying null effect because approximately half the time the difference is in the opposite direction. On the sixth trials (i.e. a stay following a switch), naming latencies in uniform sub-blocks ($M = 966, SD = 56.85$) were 9.51 ms faster than naming latencies in mixed sub-blocks ($M = 975, SD = 58.02$). In 52% of the simulations or trial six, uniform sub-blocks were slower than mixed sub-blocks, $\chi^2_{1,N=200} = 0.16$, $p = 0.69$. The results indicate that the Inhibitory Model predicts very little, if any, difference between uniform or mixed sub-blocks for trials five and six.

For the Non-Inhibitory Model on trial five, naming latencies in uniform blocks ($M = 1358, SD = 63.74$) were 134 ms slower than in mixed sub-blocks ($M = 1223, SD = 66.12$). In 92.5% of the simulations on trial five, uniform sub-blocks were slower than mixed sub-blocks, $\chi^2_{1,N=200} = 88.17, p < .001$. On trial six, naming latencies in uniform blocks ($M = 1220, SD = 62.42$) were 68 ms slower than in mixed sub-blocks ($M = 1152, SD = 62.42$). In roughly 93% of the simulations on trial six, uniform sub-blocks were slower than mixed sub-blocks, $\chi^2_{1,N=200} = 36.73$, $p < .001$. The non-inhibitory computational model predicts naming latencies to be longer on trials five and six for uniform sub-blocks than for mixed sub-blocks, and that there is a large effect size. In other words, spreading activation effects were not abolished after a language switch.

Discussion. Computational modeling was used to make predictions based on two verbal theories of bilingual language control. As expected, language switching abolished semantic effects for the Inhibitory Model, but did not abolish semantic effects in the Non-Inhibitory Model. These predictions are useful for estimating effect sizes and interpreting experimental results. This, in turn, can help adjudicate between the two verbal theories.

It should be noted that both computational models (and the verbal theories that they are based on) also predict within-language competitive effects (i.e., that naming latencies should increase from trial one to two and two to three). However, it is not at all guaranteed that this is the case. Next, we report empirical tests of these predictions.

Experiment 1

To test the models' predictions, we designed an experiment that tries to be as faithful as possible to the conditions imposed on the simulations. To date, we know of no studies that have tested how language switching affects spreading activation of semantically-related stimuli from one trial to the next. There have been at least three studies that examined naming latencies of semantically related trials separated by filler trials (Hong & Macwhinney, 2011; Lee & Williams, 2001; Runnqvist et al., 2012). Because this experiment does not have filler trials (and eliminates the possibility of learning effects), it not only tests the models' predictions, but also has broader theoretical implications for competition within language. For example, Navarrete et al. (2014) found facilitation when semantically related pictures were named one after another, but incremental learning effects once filler trials were introduced. We thought this was a possibility, but nonetheless we predicted within-language interference on stay trials mainly because (1) that is what the models predict when there is within-language competition and (2) it is common in many bilingual theories of speech production (e.g., the so-called Lexical Selection Mechanism; the Inhibitory Control Model, the Revised Hierarchical Model; Costa & Caramazza, 1999; Green, 1998; Kroll, Van Hell, Tokowicz, & Green, 2010). For example, Costa and Caramazza (1999, p. 232) assumed that "the ease with which a lexical node is selected depends on the activation level of competing lexical nodes." If there was facilitation on stay trials, then we predicted that inhibition would eliminate those effects as well.

Method

Participants. 45 English-Spanish speaking bilingual participants (71% female; 73% rated English as their L1) were recruited through the Psychology Department participant pool. One participant was removed due to failure to meet the requirements of the study. Additionally, three participants failed to complete all experimental blocks due to computer error or time constraints. In line with the literature on bilingual language production and comprehension (Caramazza, 1997; Linck et al., 2012; Meuter & Allport, 1999; Moreno, Federmeier, & Kutas, 2002), subjective questionnaires regarding their age of acquisition as well as self-ratings of their reading, writing and speaking ability of their languages were assessed using Likert scales. In addition, participants were given a more objective vocabulary measure, the Multilingual Naming Test (MINT; Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012), first in English and then in Spanish. The participants' MINT scores were used to determine L1 and L2 in the statistical analyses. If there was a tie, then their self-ratings were used. Twenty-six participants scored higher on the English portion of MINT, twelve scored higher on the Spanish portion, and 7 scored equally well on both portions. See Table 1 for information on participants' self-ratings of language ability and results of the Multilingual Naming Test (i.e., means and standard deviations).

Stimuli. 600 × 600 pixel color photographs from eight semantic categories were used as stimuli (6 pictures per category). The categories were *birds, body parts, clothes, fruits, furniture, musical instruments, vehicles* and *weapons*. Each stimulus was associated with a word to be named in the experiment. Between languages, words were controlled for in terms of word frequency, familiarity and prototypicality. Word frequency information for the picture names was taken from the Corpus of Contemporary English (COCA; Davies, 2017b) and Corpus del Español (Davies, 2017a). Familiarity and prototypicality ratings were taken from Schwanenflugel and Rey (1986). Care was taken to make sure none of the pictures had associative links (e.g., Lion-Tiger) by using the USF association norms (Nelson, McEvoy, & Schreiber, 2004).

Apparatus. Stimuli were presented using OpenSesame software (Mathôt, Schreij, & Theeuwes, 2012) on lab computers (Dell Optiplex 760). A microphone recorded participants' responses in order to evaluate naming latencies. Naming latencies for each trial were measured by a virtual voice-key, and verified in Praat (Boersma, 2006) and R (R Core Team, 2014).

Procedure. Stimuli were presented in the center of a 15 in. 1600 × 900 pixel Dell computer screen. Participants were seated roughly 60 cm from the screen, with stimuli subtending a visual angle of roughly 10 degrees. After participants were familiarized with the pictures and their corresponding names, they completed a practice session. During the practice session, participants named each of the stimuli twice on the computer screen: once in their L1 and once in L2. They then started the experiment. They were asked to name the pictures as quickly and accurately as possible.

Language was cued based on the background color of the picture (*grey* or *light blue*), and background color was counterbalanced across participants. A single trial consisted of a fixation point, presentation of a stimulus and inter-stimulus interval. To make sure participants did not become accustomed to the timing of the pictures and consistent with other naming studies (e.g., Navarrete et al., 2014), the fixation point's duration varied between 250 and 700 ms across trials based on a uniform distribution. The mean of the uniform distribution was 500 ms. Stimuli were presented on a screen until a participant responded or until 3000 ms passed, whichever was shorter. A recording of the response started at the onset of the stimulus, and naming latencies were measured in milliseconds from the onset of the stimulus until the participant responded. Naming latencies were determined by a virtual voice-key. The inter-stimulus interval lasted 1500 ms after the participant responded. If a participant failed to respond within 3000 ms (i.e., a timeout), the program proceeded to the next trial. See Fig. 3 for a representation of a single trial. See Fig. 2a for a representation of how Sub-Blocks differed.

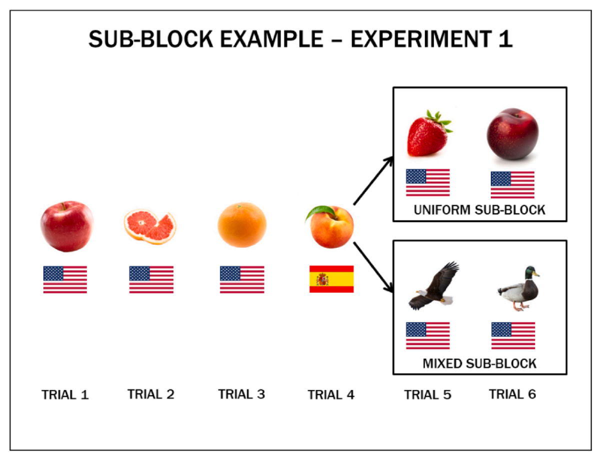
Participants named 768 trials in eight blocks. Each block contained 96 trials: 48 trials were named in English, and 48 were named in Spanish. Within each block, pictures were grouped into sub-blocks. Each sub-block consisted of 6 trials. Sub-blocks were divided into two types: uniform and mixed. Both types of sub-blocks cued trials according to the same language pattern (i.e., *stay, stay, stay, switch, switch, stay*).

In uniform sub-blocks, all trials came from the same semantic category. In mixed sub-blocks, semantic category changed on trial five. Each sub-block had a major semantic category, from which its words were quasi-randomly ordered. Mixed sub-blocks had a major semantic category (associated with the first 4 trials) and a minor semantic category (associated with the last two trials), with words from each being quasi-randomly selected. Picture stimuli were not repeated within a block

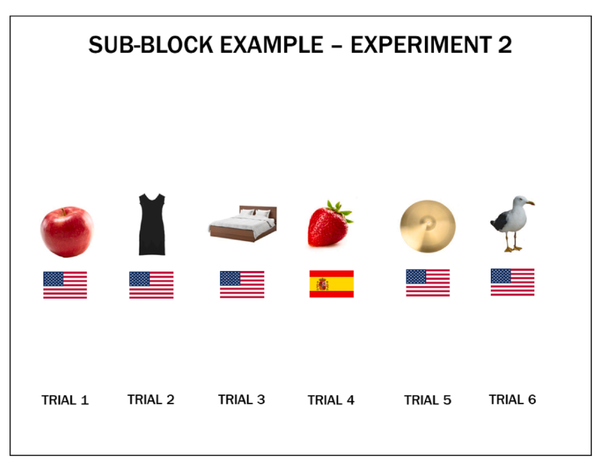
Table 1
Participants' language proficiency in Exp 1.

Measure	Language L1	L2
Self-Ratings		
Speaking (out of 7)	6.48 (0.7)	6.02 (0.8)
Reading (out of 7)	6.67 (1.0)	5.5 (0.9)
Writing (out of 7)	6.24 (1.2)	6.14 (1.0)
Age of Acquisition	2.23 (3.3)	4.48 (6.1)
MINT (% correct)	90 (10)	74 (13)

Note: MINT scores were used to determine bilinguals' L1 and L2 in statistical analyses. The L1 determined by the MINT scores may not necessarily correspond with participants' self ratings.



(a) Two presentations of a sub-block in Experiment 1, showing the difference between mixed and uniform sub-blocks



(b) Stimuli from semantic categories in Experiment 2 Stimuli were pseudorandomized so that neighbors were not presented one after another.

Fig. 2. Examples for How Semantically-Related Stimuli Were Presented in Exp. 1 and 2.

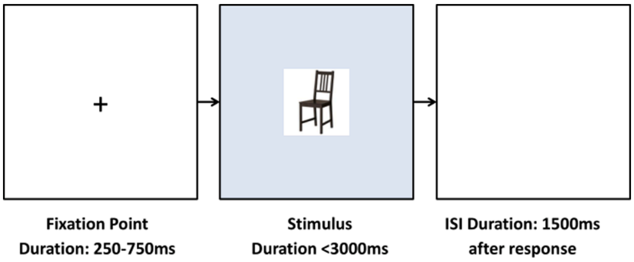


Fig. 3. A representation of a single trial. The color of the background cued participants to speak in either L1 or L2 and was counterbalanced across participants.

until all pictures from that category had been named. Additionally, each trial number across the sub-blocks was controlled for in terms of prototypicality, familiarity and word frequency.

In total, each participant was presented with 8 blocks of trials. Each block consisted of 16 sub-blocks: eight uniform sub-blocks and eight mixed sub-blocks. Within each block, words were presented once in English and once in Spanish, and this was counterbalanced. Type of sub-block alternated. Additionally, type of sub-block presented first within a block alternated. The order of the blocks was presented to participants according to a Balanced Latin Square design.

Results

Descriptives. Descriptive statistics are given in each table for each statistical analysis[7]. However, a summary of the reaction time data as a function of sub-block type and trial within sub-block can be found in Fig. 1.

Naming Latency Analyses on Trials 1-3. Naming latencies were analyzed using a Bayesian Hierarchical Model (BHM) using rjags[8] (Plummer, 2013). Trials 1-3 were analyzed separately from Trials 5-6. No analyses were conducted on Trial 4. The BHM assumes that reaction times come from an ex-Gaussian distribution, and provide posterior distribution estimates of each independent variable's effects on naming latencies in milliseconds. If 95% of the highest density interval (HDI) of the posterior distribution estimate does not include zero, then the effect is considered credible. The model can be used to estimate deflections (i.e., how much a condition is different from the grand mean in milliseconds) and mean differences. Deflection estimates are given in the tables. Mean differences are reported in the text.

To analyze Trials 1-3, trial number (one, two, three) and language (L1, L2) were input as independent variables. Participant, language of the stimulus (English, Spanish) and stimulus were controlled for. Only trials that were less than 500 ms were removed. 1237 trials (7.8%) were excluded from the analysis due to participant error. Of these, 654 (4.1%) were due to timeouts, 151 (0.9%) were intrusion errors, 368 (2.3%) were incorrect but semantically related/correct language responses, and 64 (<0.5%) were other errors (e.g., non-semantically related words, non-words, coughs etc.). A credible main effect of language was found. L1 trials were 13.21 ms slower than L2 trials, 95% HDI [6.25, 20.35]. Additionally, a credible main effect of trial type was found. The first trials in a sub-block were 33.11 ms slower than the second trials, 95% HDI [15.55, 47.67], and 38.45 ms slower than the third trials, 95% HDI [24.08, 55.65]. However, the second and third trials were not credibly different, 95% HDI [-18.93, 10.53]. Results are summarized in Table 2 and Fig. 4A.

As indicated in Fig. 4A, these main effects are qualified by a language by trial order interaction. For L1 trials, there is roughly equal facilitation from the first to second trials in a sub-block (24.49 ms, 95% HDI [−48.77, −2.89]) compared to the second and third trials (−23.90, 95% HDI [−45.51, −2.57]). However, for L2 trials, there is a relatively large facilitation effect from the first to second trials in a sub-block (−41.73 ms, 95% HDI [−63.47, −19.30]), but there is no credible difference between the second and third trials in a sub-block (13.22, 95% HDI [−7.52, 13.50]).

Error Analyses Trials 1-3. For error analyses, each trial was coded either 1 or 0 (correct, incorrect) based on the participant's response, and the data were input into a Bayesian model using rjags (Plummer, 2013). The model assumed each response is taken from a Bernoulli distribution, and a logistic link function was applied to the predictors to create a general linear model. The same predictors used to analyze trials 1-3 were input into this analysis. Overall, only a main effect of language was found: pictures named in L2 produced 1.4% more errors than pictures named in L1, 95% HDI [0.11, 4.7].

Naming Latency Analyses on Trial 5. Recall that the Inhibitory Model predicts no difference between sub-blocks on Trial 5, whereas the Non-Inhibitory Model predicts uniform sub-blocks to have longer naming latencies than mixed sub-blocks. In order to test this (i.e., whether spreading activation is eliminated after a language switch), each participant's naming latency data on trial five of each sub-block were input into the RT BHM as the dependent variable. Language (*L1*,*L2*) and sub-block type (mixed, uniform) were input as independent variables. Participant, stimulus and language of the stimulus (Spanish, English) were also input into the model to account for error associated with each source. 527 trials (9.9%) were excluded from the analysis due to participant error. Of these, 235 (4.5%) were due to timeouts, 134 (2.5%) were intrusion errors (i.e., wrong language), 130 (2.4%) were incorrect but semantically related/correct language responses, and 28 (<1%) were other errors (e.g., non-semantically related words, non-words, coughs etc.).

Table 2
Naming Latency Results on Trials 1-3 based on the Bayesian Model.

Source	Level	Mean (SE)	BHM Est.	Deflection	95% HDI	
					Lower	Upper
Grand Mean		1195 (3.44)	1198	NA	NA	NA
Language	L1	1201 (4.79)	1204	6.23*	0.12	12.64
	L2	1189 (4.94)	1192	−6.23*	−12.64	−0.12
Trial	One	1222 (6.35)	1222	23.89*	14.07	33.73
	Two	1185 (5.79)	1189	−9.18*	−17.93	−0.67
	Three	1181 (5.77)	1183	−14.65*	−23.43	−6.02
Trial by Language	L1 One	1228 (8.79)	1229	0.58	−8.55	9.41
	L1 Two	1203 (8.11)	1204	9.14	−0.11	18.09
	L1 Three	1175 (8.01)	1180	−9.38*	−18.39	−0.86
	L2 One	1203 (8.11)	1215	−0.58	−9.41	8.55
	L2 Two	1169 (8.26)	1173	−9.14	−18.09	0.11
	L2 Three	1186 (8.31)	1186	9.38*	0.86	18.39

*Indicates a credible deflection was found.

[7] The experiment controlled for repetition effects by not repeating a stimulus until all other stimuli had been named. However, to ensure that repetition effects were not confounding the results, we tested whether the independent variables interacted with repetition of the stimuli. No credible interactions were found in any of the analyses. The results are presented without repetition effects in the models. However, analyses with repetition effects can be found in the supplementary materials

[8] For all RT analyses, frequentist statistics were used to analyze the data and showed similar results to the Bayesian Model. Those analyses can be found in the supplementary materials

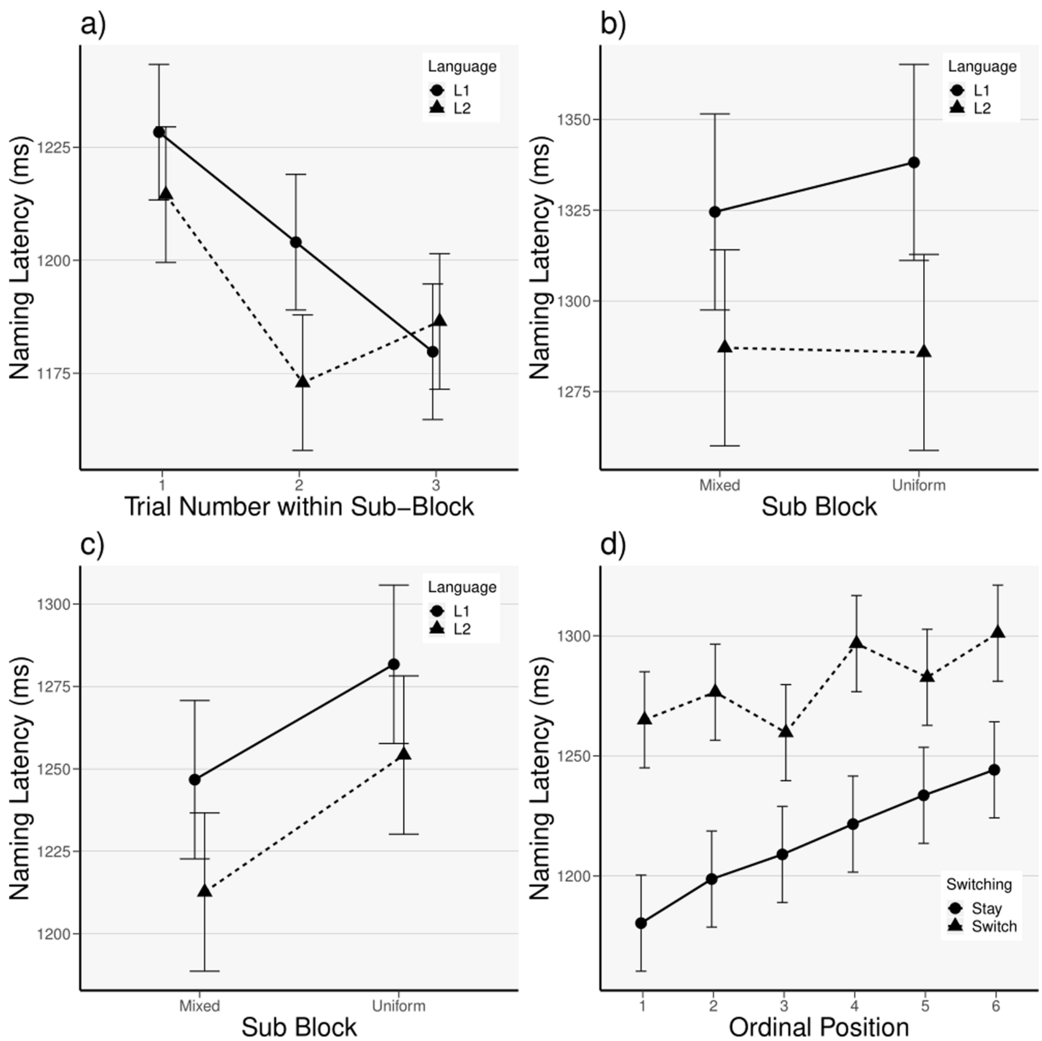


Fig. 4. Naming latency results for Experiments 1-2. Panel A: Trials 1-3 within a sub-block in Experiment 1. Panel B: Results for Trial 5 in Experiment 1. Panel C: Results for Trial 6 in Experiment 1. Panel D: Effects of ordinal position within a block in Experiment 2. Error bars represent 95% HDI estimates of the means, equivalent to roughly 2 standard errors.

Table 3
Naming Latency Results on Trials 5 based on the Bayesian Model.

Source	Level	Mean (SE)	BHM Est.	Deflection	95% HDI	
					Lower	Upper
Grand Mean		1306 (6.44)	1309	NA	1295.46	1322.26
Language	L1	1329 (9.03)	1331	22.29*	9.29	36.41
	L2	1283 (9.17)	1287	−22.29*	−36.41	−9.29
Block Type	Mixed	1303 (9.02)	1306	−3.18	−16.59	9.91
	Uniform	1309 (9.21)	1312	3.18	−9.91	16.59
Language by Block	L1 Mixed	1321 (12.55)	1324	−3.76	−17.09	10.31
	L1 Uni.	1337 (13.01)	1338	3.76	−10.31	17.09
	L2 Mixed	1284 (12.94)	1287	3.76	−10.31	17.09
	L2 Uni.	1281 (12.99)	1286	−3.76	−17.09	10.31

*Indicates a credible deflection was found.

Results of the analysis of Trial 5 are in shown in Table 3; the data are also depicted in Fig. 4B. Trial 5 was a switch trial, and consistent with previous findings, there was a main effect of language. L1 trials were 44.58 ms slower than L2 trials, 95% HDI [19.4, 46.8]. Type of sub-block (mixed, uniform) did not credibly affect reaction times. Neither was there an interaction between sub-block type and language.

Because a null effect was found, the mixed and uniform sub-block means on Trial 5 were calculated for each participant and input into JASP to calculate a Bayes Factor using a Bayesian paired-samples *t*-test. The results indicate that $BF_{10} = 0.167$, giving substantial evidence for the null hypothesis over the alternative. The facilitation on Trials 1-3 does seem to be eliminated after a language switch, supporting the predictions of the Inhibitory Model.

Error Analyses on Trial 5. The same predictors used to analyze naming latencies on Trial 5 were input into the error analysis. There were no main effects or interactions on error rates on Trial 5.

Naming Latency Analyses on Trial 6. The same Bayesian model used for Trial 5 data was used to analyze naming latencies for Trial 6 data. 449 trials (9.9%) were excluded from the analysis due to participant error. Of these, 232 (4.3%) were due to timeouts, 72 (1.3%) were intrusion errors, 120 (2.3%) were incorrect but semantically related/correct language responses, and 25 (<1%) were other errors (e.g., non-semantically related words, non-words, coughs etc.).

Consistent with the pattern in Fig. 4C, there was a credible main effect of language ($L1,L2$). L2 stimuli were named 30.8 ms faster than L1 stimuli, 95% HDI [-54.83, −5.58]. There was a credible main effect for type of block (mixed, uniform). Uniform blocks were 38.30 ms slower than mixed blocks, 95% HDI [14.7, 61.7]. There was no credible interaction between block type and language.

Error Analyses on Trial 6. The same predictors used to analyze naming latencies on Trial 6 were input into the error analysis. Overall, there was a main effect of Language where L2 trials produced 2.0% more errors than L1 trials, 95% [0.01, 2.9]. There was no main effect of Sub-Block or interaction between sub-block and Language.

Discussion. The results of Experiment 1 generally support the predictions made by the Inhibitory Model. For example, the inhibitory control model (Green, 1998) predicts that language switching should “lead to the abolition of both cross-language and within-language competitor priming.” On Trial 5, there was no credible difference in naming latencies between mixed sub-blocks and uniform sub-blocks, and a Bayes Factor gave substantial evidence for the null effect. Additionally, there was no difference in accuracy between mixed and uniform sub-blocks. Whatever effect spreading activation had on the previous trials was eliminated on trial five. There was a difference between mixed and uniform sub-blocks on Trial 6, but based on the facilitation on Trials 1-3, uniform blocks should have been faster than mixed sub-blocks.

In general, both models need to account for three things: (1) the facilitation found on Trials 1-3, (2) the null result of Trial 5 between Mixed and Uniform Sub-Blocks and (3) the interference found on Trial 6. The results of Trial 5 were more consistent with the *a priori* predictions of the Inhibitory computational model than with those of the Non-Inhibitory model. However, based on *a priori* modeling, neither model can explain the facilitation and interference found on the other trials. Having conducted a predictive test, we now turn to a descriptive test, i.e. the models’ abilities to fit the empirical data.

Fitting the models to the data of Exp. 1

In *Simulations*, *a priori* predictions were made for an Inhibitory model and a Non-Inhibitory model of bilingual naming. The results of Experiment 1 showed that both models fared poorly. The inhibitory model could predict similar naming latencies on Trial 5 between mixed and uniform sub-blocks. It did not predict the facilitation on Trials 1-3. It also did not predict naming to take longer on Trial 6 in uniform blocks compared to mixed blocks. The Non-Inhibitory Model could predict the interference on Trial 6 between mixed and uniform sub-blocks, but that result is inconsistent with the facilitation found on Trials 1-3 and the null result on Trial 5. If naming semantic neighbors within a language (i.e., on Trials 1-3) produces facilitation, then the Non-Inhibitory Model should predict mixed block stimuli to be named slower on Trials 5 and 6.

Because the facilitation on Trials 1-3 suggests within-language competition is absent, we removed the assumption of within language competition to create two new models, and we fit the models to the individual and average data. We refer to these models as the Non-Competitive Within-Language (NCWL) Inhibitory and Non-Competitive Within-Language Non-Inhibitory Models. Details of the fits to individual subjects and methods used to fit the models can be found in the *Appendix B*. Removing the within language constraint should demonstrate that (1) within-language competition is absent on Trials 1-3 and (2) the facilitation found on Trials 1-3 is the result of a different mechanism than the interference found on Trial 6 (i.e., uniform sub-blocks being named slower than mixed sub-blocks). We expect that the NCWL models will be able to account for the facilitation observed and improve overall model fit. See the *Appendix A* for information on how we removed the within-language competition assumption.

We fit the models to the participants’ data individually, and to the average data. Individual results are found in *Appendix B*. For the averaged data, the models were fit using random search (Bergstra & Bengio, 2012). Parameters were allowed to vary randomly through parameter space to find the best values that could account for the data. To avoid overfitting, only three parameters were chosen for each model. For the Inhibitory Model, the competition parameter (V), the language strength parameter on stay trials (L) and inhibition parameter (h) were allowed to vary. The Non-Inhibitory Model does not have an inhibition parameter. Thus, the language strength parameter on switch trials was allowed to vary (L_{switch}), along with the competition parameter (V), and the language strength parameter on stay trials (L_{stay}). It should be noted that the inhibition parameter helps determine the language strength parameter on switch trials for the Inhibitory Model, and thus analogous parameters were chosen for both models. Additionally, a simplifying assumption was made by setting L2 parameters equal to the corresponding L1 free parameters. The remaining parameters, listed in Table 4, we kept equal to the values used in pre-experimental predictions (see *Simulations*).

We averaged the data across participants for both mixed and uniform sub-blocks. We fit the models and calculated RMSE values across all four models (the Inhibitory Model, Non-Inhibitory, NCWL Inhibitory, NCWL Non-Inhibitory) using 5000 iterations.

Fit of the models to the average data

All participant data were aggregated and averaged by sub-block and trial number. The original and new models were fit to the averages. See Fig. 5 for a comparison of average fits of the models to the data of Experiment 1. Because there are only five observations per sub-block (not including Trial 1 - the baseline), no statistical analyses were used. However, RMSE values were found for each model. Averaging across sub-block type, the NCWL Inhibitory Model had the lowest error for both mixed (RMSE = 16.78 ms) and Uniform (RMSE = 32.70) sub-blocks. Interestingly, the NCWL Non-Inhibitory Model performed worst (mixed RMSE = 25.47, uniform RMSE = 36.48). The original Inhibitory Model performed similarly (Mixed RMSE = 23.54, Uniform RMSE = 35.84) to the original Non-Inhibitory Model (Mixed RMSE = 24.41, Uniform RMSE = 34.42).

Considering the best-fitting parameter values of the original models in Table 4, an interesting pattern emerges. Specifically, within-language competition is virtually absent according to the original models ($V \approx .5$; i.e., the first three trials all have similar RTs). For the original Inhibitory Model, the fits thereby suggest that the bulk of the competition comes from between-language distractors. However, to create switch costs in the Non-Inhibitory model, there has to be at least some within-language competition (i.e., between-language distractors have no effect on RTs). Turning to the L parameters, it is clear that the original Non-Inhibitory model solves this by assuming activation spreads to the lexicon

Table 4
Best fitting parameters for the Inhibitory and Non-Inhibitory Models of All Participant Data. The Original Models assume within language competition. The New Models assume no within-language competition.

	Inhibitory		Non-Inhibitory	
Parameters	Original Model	NCWL Model	Original Model	NCWL Model
$L_{stay,L1}$	6.3	1.08	6.1	8.7
$L_{stay,L2}$	6.3	1.08	6.1	8.7
h_1	1.2	1.88	NA	NA
h_2	1.2	1.88	NA	NA
$L_{switch,L1}$	NA	NA	118	139
$L_{switch,L2}$	NA	NA	118	139
c	0.01	0.01	0.01	0.01
Y_2	37.5	3.24	NA	NA
Y_2	37.5	3.24	NA	NA
R_1	3	3	3	3
R_2	3	3	3	3
p_1	0.8	0.40	0.75	0.40
V	0.49	0.53	0.51	0.53

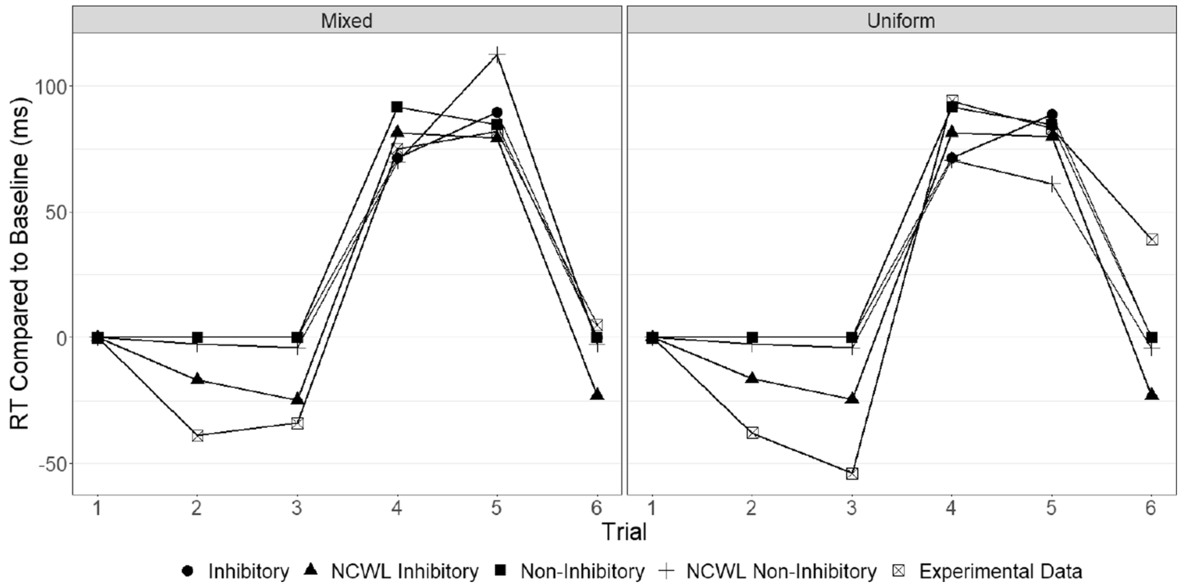


Fig. 5. Average fits for the non-inhibitory, inhibitory, NCWL Non-Inhibitory and NCWL Inhibitory models.

extremely slowly on switch trials. Both models assume spreading activation has little effect on RTs.

On the other hand, the new models also showed interesting results. The NCWL Inhibitory Model performed best. The V parameter is now above the theoretical 0.5 threshhold for competition. The increase in the V parameter made it necessary for the target to receive activation more quickly (i.e., the L parameter is smaller compared to the original model), which also increased the inhibition parameter h. This allowed for the facilitation to be abolished after Trial 4 of a sub-block, allowing it to fit the RTs on Trial 5. In contrast, the NCWL Non-Inhibitory Model took the same tactic as the original: Assume spreading activation had a small effect, and account for switch costs by having activation spread extremely slowly to the lexicon. However, it was punished on Trial 5 as mixed blocks were disproportionately affected by the slow activation parameter and spreading activation. Namely, the lemma on Trial 5 in mixed sub-blocks had to be activated from its resting level. Without the benefit from spreading activation of previous trials and with the slow activation parameter, RTs increased drastically. If the NCWL Non-Inhibitory model allows for even modest facilitation on the first three trials, it cannot fit Trials 5 and 6 at all. This result demonstrates that any interference found on Trials 5 or 6 cannot be the result of the same mechanism that produced the facilitation on Trials 1-3.

Despite some individual differences, our conclusions also generally hold when examining fits to individual data, as reported in *Appendix B*. Upon observing the individual iterations of the New Models, it is apparent that both models could account for relatively large facilitation effects on Trials 1-3. However, this would make them fit poorly on Trial 6. The RTs on Trial 6 would be too fast. This increased the RMSE. These results suggest that the facilitation on the first three trials is the result of a different mechanism than the mechanism causing interference on Trial 6.

Discussion

The model fits produced two relatively clear conclusions. (1) The NCWL Inhibitory Model qualitatively fit the average data best: it could account for the facilitation found in the first three trials while still showing equivalent RTs between mixed and Uniform sub-blocks on Trial 5. (2) None of the models could account for facilitation on Trials 1-3 and still show interference on Trial 6 (i.e., Uniform Trial 6 RT > Mixed Trial 6 RT).

Although unlikely, the facilitation on Trials 1-3 might suggest that the stimuli within a semantic category were not sufficiently related to produce interference. It is possible that we chose semantic neighbors that were not in fact semantic neighbors (i.e., and hence no spreading activation actually occurred). This can be tested using the continuous naming paradigm, in which semantically related stimuli are named with filler trials. Prior investigations have used this paradigm to demonstrate that interference occurs each time a semantic neighbor is presented (Navarrete et al., 2014). Instead of presenting semantically related stimuli serially, we separated them in Experiment 2 by filler trials. By so doing, we aim to test whether (1) our stimuli were sufficiently related within a category and whether, as we suspect, (2) incremental learning effects have been mistaken as spreading activation effects in previous bilingual naming studies (Hong & Macwhinney, 2011; Lee & Williams, 2001; Runnqvist et al., 2012).

Experiment 2

Experiment 1 found that priming effects were eliminated after a language switch. This supports the Inhibitory over the Non-Inhibitory model. However, Runnqvist et al. (2012) found that cumulative semantic interference was not reduced in the continuous naming paradigm. They took this as evidence against inhibition. But, those results may simply indicate that language switching does not affect incremental learning. In order to test this alternative explanation, we instructed participants to name semantically-related neighbors, this time separated by filler trials, (i.e., the continuous naming paradigm). If cumulative semantic interference is the result of a learning mechanism that is independent of semantic priming, then a specific prediction can be made: When introducing filler trials between related stimuli, cumulative semantic interference should occur on both stay and switch trials and should should show the same pattern regardless of the number of filler trials. With each presentation of a semantic neighbor within a block, naming latency should increase 10–30 ms (Howard et al., 2006; Navarrete et al., 2012; Navarrete et al., 2010). By the same token, such a result would also indicate that the stimuli within the semantic categories in Experiment 1 were sufficiently related.

Method

Participants. 45 English-Spanish speaking bilingual participants (66% female) were recruited through the Psychology Department participant pool. The same questionnaire used in Experiment 1 was given to participants in Experiment 2, as well as the Multilingual Naming Test. 30 participants scored higher on the English portion of MINT than on the Spanish portion, nine scored higher on Spanish, and five scored equally well on both portions. One participant was excluded because they could not name more than 10% of the pictures in L2 of MINT. Like in Experiment 1, the participants' MINT scores were used to determine L1 and L2 in the statistical analyses. If there was a tie, then their self-ratings were used. See Table 5 for a summary of participants' self-ratings of language proficiency and the results of the Multilingual Naming Test.

Stimuli and Apparatus. The same stimuli and apparatus used in Experiment 1 were used in Experiment 2.

Procedure. The procedure in Experiment 2 was similar to that of Experiment 1 with one major exception: Stimuli within a block were pseudorandomized so that at least one intervening trial separated semantically related stimuli. Because of this, the mixed/uniform sub-block distinction no longer applies. Participants still named pictures in the same language order (*Stay, Stay, Stay, Switch, Switch, Stay*). The intervening/filler trials should introduce incremental learning effects. The number of filler trials was calculated by how many stimuli from another category were named between two semantically related stimuli. For example, if participants name *apple, dress, bed, fresa, tambourine, seagull*, then *dress* and *bed* would create a lag of two (i.e., two fillers) for *fresa*. 50% of trials had five or fewer fillers, while 50% of trials had five or more fillers. Like in Experiment 1, no word was repeated until all other stimuli were presented and language (Spanish, English) was counterbalanced. See Fig. 2b for a representation of a sub-block.

Results

Naming Latency Analysis. A Bayesian model that is similar to the models used for analyzing naming latencies in Experiment 1 was used to assess incremental learning effects in this experiment. Participant, stimulus, language (*L1, L2*) and language of the stimulus (*English, Spanish*) were controlled for. Trial type (*stay, switch*) and ordinal presentation of a semantically related stimulus (one through six) were input as independent variables. Consistent with previous literature, a picture from one semantic category served as a filler trial for pictures from another semantic category. 3116 trials (10.2%) were removed due to participant error. Of these, 1333 (4.4%) were due to timeouts, 627 (2.1%) were intrusion errors, 983 (3.2%) were incorrect but semantically related/correct language responses, and 173 (< 0.6%) were other errors (e.g., non-semantically related words, non-words, coughs etc.).

As is visually apparent in Fig. 4D, there was a main effect of ordinal presentation of a semantically related stimulus. On average, naming latency credibly increased by 9.98 ms for each presentation of a semantic neighbor, 95% HDI [5.46, 14.21]. Additionally, there was a main effect of trial type. Stay trials were named 65.60 ms faster than switch trials, 95% HDI [-77.20, −55.01]. There was no interaction between language and presentation order. The results indicate that language switching had little to no effect on CSI effects.

We also checked whether number of filler trials influenced the CSI effect. Controlling for the same variables as the previous analysis, ordinal position (one through six) and number of filler trials (greater than or equal to five and less than five) were input as independent variables. The average slope with fewer than five intervening trials is 12.85 ms per ordinal presentation, 95% HDI [4.88, 19.6], and it is nearly identical to the average slope of five or more trials, 11.35 ms per ordinal presentation, 95% HDI [6.22, 16.48]. Critically, the difference between the slopes is only 1.40 ms per presentation and is not credible, 95% HDI [-7.72, 9.89]. The results indicate that increasing the number of filler trials between semantically related stimuli does not decrease CSI, and it supports the idea that incremental learning created the semantic interference found in this experiment.

Error Analyses. A Bayesian model that is similar to the models used for analyzing accuracy in Experiment 1 was used to assess incremental learning effects in Experiment 2. Participant, stimulus, language (*L1, L2*) and language of the stimulus (*English, Spanish*) were controlled for. Trial type (stay, switch) and ordinal presentation of a semantically related stimulus (one through six) were input as independent variables. There was a main effect of type of trial. Switch trials produced 1.88% more errors than stay trials, and the difference was credible, 95% HDI [0.26, 4.6]. There was also a main effect of ordinal presentation. Naming accuracy decreased by roughly 0.64% with each presentation of a semantic stimulus, 95% HDI [-1.66, −0.09]. There was no credible interaction.

Table 5
Participants' language proficiency in Exp 2.

Measure	Language L1	L2
Self-Ratings		
Speaking (out of 7)	6.61 (0.5)	6.09 (0.9)
Reading (out of 7)	6.48 (0.7)	6.18 (1.3)
Writing (out of 7)	6.30 (0.7)	6.02 (1.0)
Age of Acquisition	2.65 (3.3)	4.38 (5.7)
MINT (% correct)	89 (10)	73 (12)

Note: MINT scores were used to determine bilinguals' L1 and L2 in statistical analyses. The L1 determined by the MINT scores may not necessarily correspond with participants' self ratings.

Discussion

Experiment 2 demonstrates that language switching does not abolish cumulative semantic interference in the continuous naming paradigm. With each presentation of a semantic neighbor, naming latencies increased by a constant amount. This was true for both stay and switch trials. Contrasting these results with the results of Experiment 1, we see evidence that cumulative semantic interference is not the result of spreading activation. Rather, it is the result of another mechanism. A likely candidate is incremental learning, as suggested by several researchers (Damian & Als, 2005; Howard et al., 2006; Navarrete et al., 2012; Navarrete et al., 2010). In addition, the results of Experiment 2 reinforce the conclusions of Experiment 1. Namely, the stimuli were sufficiently related. Thus, switching languages abolished facilitation effects in Experiment 1, supporting an inhibitory model of bilingual language production.

General discussion

In this study, we detailed two computational models to test verbal theories of bilingual language control in speech production: an Inhibitory model and a Non-Inhibitory model. We used those models to make predictions for how naming latencies would change when participants named semantically-related stimuli under different conditions (i.e., uniform sub-blocks and mixed sub-blocks). The Inhibitory Model predicted spreading activation effects (i.e., interference) would be abolished following a language switch. The Non-Inhibitory Model predicted spreading activation effects would continue after a language switch. The results of Experiment 1 indicate that spreading activation effects were indeed eliminated. However, facilitation was found when naming pictures within a language, which is something neither model predicted. We then fit the original models to the individual data, and the Inhibitory Model fit the individual data better than the Non-Inhibitory Model (see *Appendix B*). We then reran the models, assuming no within-language competition, and qualitatively, the NCWL Inhibitory Model was the only model that could explain both the facilitation of Trials 1-3 of a sub-block and the equivalent RTs on Trial 5 between Mixed and Uniform sub-blocks.

Experiment 2 suggests that the continuous naming paradigm may not be a valid way of testing whether switching languages abolishes residual, spreading activation, and calls into question some of the conclusions made from previous research (Hong & Macwhinney, 2011; Lee &

Williams, 2001; Runnqvist et al., 2012). Results indicated that it takes longer to name each additional presentation of a semantically related stimulus, regardless of language switching. Additionally, the number of filler trials had no effect on the interference effects found. If activation of semantic neighbors creates interference, then more filler trials should decrease the interference effect. However, this did not happen.

Within-language competition in bilingual models of production

An influential assumption in monolingual and bilingual models of speech production is that words compete for selection within a lexicon. For example, Costa and Caramazza (1999) proposed a non-inhibitory model of bilingual production, but they still assumed a competitive process within a language: "the degree of activation of [same language] non-target nodes affects the ease with which the target word will be selected" (p. 232). Similarly, Green (1998), when arguing for inhibition in the inhibitory control model, states "individuals have difficulty regulating the competition amongst lemmas via the semantic route" (p. 73). Monolingual models also make this assumption (Harley, 1993; Levelt et al., 1999; Roelofs, 1992).

Why do the production models make this assumption? Although the answer is beyond the scope of this paper, it is likely based on earlier studies that elicited speech errors (see Levelt, 1999) or incremental learning effects through the continuous naming paradigm or from picture-word interference tasks (Schriefers, Meyer, & Levelt, 1990). We have already argued that interference under the continuous naming paradigm is most likely due to incremental learning and not competition/activation. We will not belabor this point. On the other hand, picture-word interference tasks may increase the activation of a semantic neighbor in a way that slows naming, but such slowing may not be attributable to competition within the lexicon (e.g., Mahon et al., 2007). One might argue that as a participant is about to say the target word, the distractor gets chosen mistakenly due to its visual similarity to the target. The distractor gets temporarily selected, but its production is blocked by an internal monitoring mechanism, thereby preventing articulation (Hartsuiker & Kolk, 2001). This could produce the interference found in such studies.

In Experiment 1, naming semantic neighbors one after another induced facilitation within a language. This result is consistent with recent monolingual studies, (Navarrete et al., 2010; Navarrete et al., 2012; Navarrete et al., 2014). In particular, it is consistent with a non-competitive model of speech production within one language. It is not consistent with either of the original computational models designed in this study. This was evidenced by the fact that fitting the original models to participant data failed when those data showed facilitation. Experiment 1's results suggest that bilingual models of speech production need to be updated to account for the within-language facilitation found. More research is needed to replicate these findings.

Finally, It seems unlikely based on Experiment 1 that activation within the semantic network creates "very transitory" facilitation while long lags create inhibition (Wheeldon & Monsell, 1994, p. 345). If this were the case, then by Trial 3 of a sub-block in Experiment 1, the interference in the lexicon should have offset the facilitation and a return to baseline should occur. However, Trial 3 still showed facilitation. This result is consistent with Navarrete et al. (2014), who found similar facilitatory effects even after five presentations of semantically related stimuli. It is also consistent with the NCWL Inhibitory Model that assumed no within-language competition.

Reexamining between-language competition in bilingual speech production

In Experiment 1, facilitation was found when naming within a language. This calls into question whether competition between words exists within a lexicon during speech production. A pressing question then arises: If no within-language competition exists, does between-language competition exist? One of the reasons inhibition was proposed as a controlling mechanism is the so-called "hard problem" of bilingual lexical selection where multiple lexical representations become active across both languages (Finkbeiner, Gollan, & Caramazza, 2006). It is unclear how the correct word, much less the correct language, is chosen if competition exists at every level. If competition does not occur within a language, then it is hard to imagine why it would exist between languages, tempting the conclusion by extension that there is no between-language competition either. This line of reasoning would seem to eliminate the "hard problem" entirely.

At this point, we argue against making the leap that between-language lexical selection is non-competitive. Let us consider more closely the results of Experiment 1. Those results showed that after participants named a series of semantically related stimuli, a language switch eliminated the facilitation that was previously observed. Conceptually, this suggests that activation flowed from the semantic network to the target lexicon, priming semantic neighbors in that language (i.e., Trials 1-3 of a sub-block). After the language switch, that activation was eliminated to such a degree that naming another semantically-related neighbor had the same naming latency as an unrelated word (i.e., uniform Trial 5 RT = mixed Trial 5 RT). This is not a trivial result, and a Bayes Factor analysis indicated strong evidence for the null finding. Thus, results suggest inhibition is used to modulate activity of the two languages.

One wrinkle in the results of Experiment 1 is that there was a difference between mixed and uniform sub-blocks on Trial 6. One might argue that Trial 5 of a sub-block is a switch trial, and that comparing mixed and uniform sub-blocks on Trial 6 is a better way to determine whether inhibition occurred. To this, we offer three responses.

First, if there is no inhibition, then one should expect the same pattern of results on Trials 1-3 as on Trials 5-6. In other words, if there is facilitation found on Trials 1-3, then naming latencies in uniform blocks should be faster on Trials 5-6 than in mixed blocks. If interference is found on Trials 1-3, then naming latencies in uniform blocks should be slower on Trials 5-6 than in mixed blocks. Facilitation was found on Trials 1-3, but uniform blocks were either the same as mixed blocks (i.e., Trial 5), or slower than mixed blocks (i.e., Trial 6). Fitting the computational models to the participant data further confirmed this (see *Appendix B*). Both the Inhibitory and Non-Inhibitory models performed similarly in predicting participants' uniform-block latencies, but the Inhibitory model outperformed the Non-Inhibitory model in predicting participants' mixed-block latencies. It seems more likely that the mixed/uniform difference found on Trial 6 reflects another mechanism like incremental learning than that it reflects spreading activation. Second, if there is no inhibition, then there is no *a priori* reason to believe that a switch trial should behave differently from stay trials in terms of priming effects. Experiment 2, for example, showed similar incremental learning effects on both stay and switch trials. Third, the NCWL Non-Inhibitory Model failed at modeling the average data because if it allowed for the needed facilitation on Trials 1-3, it would not be able to fit Trials 5-6. These results suggest inhibition is used to suppress the non-target language, which in turn implies between language competition.

Hybrid models

At first glance, it might seem strange to propose a hybrid model that assumes competition between languages but no competition within a language. Yet, previous theories have suggested there is within-language competition but no between-language competition (Costa & Caramazza, 1999; Costa & Santesteban, 2004; Costa et al., 2006). We see no reason why the reverse should not be considered, and we argue that this possibility should be taken seriously. Follow up research is needed to verify our findings.

Consider a monolingual model that assumes the most active word/concept is selected during production, and that the activation of its semantic neighbors does not create interference (e.g., Caramazza, 1997; Dell, 1986). Such models could work well for monolinguals (except in the relatively infrequent cases of memory search failures, e.g. tip-of-the tongue phenomena, in which retrieval is blocked repeatedly by some prepotent but erroneous response). However, what would happen if a monolingual started learning a second language? If both lexicons become active, and L1 is stronger than L2, then L1 words would be selected most, if not all, of the time. Some form of language control would still be needed as the language learner's stronger L1 nodes would be de facto competitors to their L2 nodes.

Now consider a more balanced bilingual within the context of non-competitive models. The situation is similar to an unbalanced bilingual. During production, both languages would (at least initially) be highly active. In many cases, two words (i.e., the target and its translation) would almost completely overlap in their semantic representations (e.g., *perro/dog*). What determines which word is chosen in such a model? Without some sort of control, the most active word would "win" and output would look random, even when a bilingual is talking to a monolingual. In this way, bilinguals must constantly deal with "bivalent responses" (see Finkbeiner et al., 2006).

Thus, a model that assumes no within-language competition would still need inhibition to control between language translations. The argument is bolstered by the fact that out of the four models tested here, the NCWL Non-inhibitory model (i.e., a non-hybrid model) provided the poorest fit to the averaged data.

Activation vs. incremental learning

Results from Experiment 2 appear to show incremental learning effects. When separated by filler trials, naming latencies increased each time a semantically-related neighbor was presented. In Experiment 1, when semantic neighbors were presented without fillers, facilitation was observed. In Experiment 2 with filler trials, naming latencies increased with each presentation of semantic neighbor. This was true for stay and switch trials. The effect was similar when there were few fillers (≤ 5) or many fillers (>5). The contrasting results from the two experiments, and the results of the new models (i.e., the more facilitation the models assume on Trials 1-3, the less accurate they are fitting Trial 6) indicate that two different mechanisms are being used. We suggest one possibility is that spreading activation created the priming effects in Experiment 1 on Trials 1-3, and incremental learning created the interference in Trial 6 in and in Experiment 2. We argue that inhibition only affects activation due to spreading activation. Therefore, any conclusions about inhibition based on the continuous naming paradigm (i.e., separating semantic neighbors with filler trials) should be treated with caution.

But this paper does little to undermine previous bilingual research that used variants of the continuous naming paradigm. Rather, it clarifies the meticulous work already done, and suggests a new avenue of research. Our results lead us to think that bilingual speech production should be examined through the lens of activation/inhibition as well as through the lens of learning. There exists an extensive body of research in the memory domain that to date has not, but can be, leveraged to clarify and inform issues of bilingual language production. The research tradition we refer to has classically dealt with very similar issues and constructs of competition, interference, inhibition, and executive control of output, often with the use of verbal stimuli (see e.g., Delprato, 1971; Delprato, 2005; McGeoch & Irion, 1952; Slamecka, 1967; Underwood, 1983). We expect that pursuit of this unexplored connection between the literatures will lead to important advances in our understanding of bilingual language control. We also argue that previous bilingual control studies that look at repetition priming (e.g., Kleinman & Gollan, 2018), and/or semantic interference (e.g., Lee & Williams, 2001; Hong & Macwhinney, 2011) can be reexamined under the lens of incremental learning effects.

How a non-inhibitory model could still work

We have presented two main models in this study: an Inhibitory Model and a Non-Inhibitory Model. We tested each model three times: (1) once to make predictions before gathering data (2) fitting the models to the data from Experiment 1 assuming within-language competition and (3) fitting the models to the data from Experiment 1 assuming no within-language competition. In so doing, we made assumptions, which we feel are reasonable and grounded in past experiments. However, as a thought experiment, we would like to discuss how the Non-Inhibitory Model could still work.

The major assumption we made is that spreading activation affects both languages. This idea is consistent with experimental results and current theories that show both languages become active during speech production (e.g., Costa et al., 1999; Dijkstra et al., 2019; Hermans et al., 1998; Kheder & Kaan, 2019; Santesteban & Schwieter, 2019). However, if spreading activation only affects one language, then a non-inhibitory model could still be viable, but with some caveats. Consider a bilingual naming six stimuli in a uniform sub-block. On trials 1-3, they name *apple, grapefruit, orange* in their L1. No spreading activation occurs in their L2 on the fourth trial as they name *peach*. Activation in the first language decays during this time. When switching back into their L1 to name *strawberry*, all the residual spreading activation has decayed as well as any facilitation. We think this is an unlikely explanation. However, it should be testable with the right experimental paradigm (e.g., using a modified blocked naming/continuous naming paradigm like those used in the study).

Additionally, a non-inhibitory model may still work for specific individuals. Although the NCWL Non-Inhibitory Model fit the average data worse than the NCWL Inhibitory Model did, it sometimes performed similarly when fitting to participants individually (see *Appendix B*). This might suggest that inhibition is used by the majority of participants, but a few may be relying on a non-inhibitory mechanism. However, the individual fits are clearly noisy, which is typical in comparing individual data to group averages. This could make the participant fits appear more similar between the models than they are. Regardless, a study specifically designed to examine individual differences may be able to determine whether some bilinguals use a non-inhibitory lexical mechanism.

Reverse dominance effects: proactive or reactive inhibition?

We now consider whether the reverse dominance found in Experiment 1 reflects proactive inhibition, reactive inhibition or both. The inhibitory models we presented assume a type of global inhibition is applied to both languages, and that the non-target language remains inhibited until a language switch. However, we did not specifically assume *proactive inhibition*. This type of inhibition is meant to try to "resolve anticipated language interference" before it happens (Declerck et al., 2020). *Reactive inhibition*, on the other hand, applies inhibition that is proportional to how active the distractors are. Let us compare what proactive and reactive inhibition would predict trial by trial in Experiment 1.

During an experimental sub-block, participants were presented with the following trial types: *stay, stay, stay, switch, switch, stay*. This pattern repeated throughout the experiment. After switch trials (e.g., Trial 5 of Experiment 1), theories of proactive and reactive inhibition would predict similar naming latencies: longer switch RTs in L1 than in L2. The strength of L1 would create strong reactive inhibition on L2 trials, which would increase RTs when participants switch back into L1. Proactive inhibition would create inhibiton on each L1 trial to facilitate naming in L2, also increasing L1 RTs. It is difficult to determine which type of inhibition is used based only on results from Trial 5 of Experiment 1.

However, the results of Trials 1-3 seem to slightly favor a non-proactive inhibitory account. If bilinguals anticipate interference from L1, then after they name a picture on Trial 1 of a sub-block, they should proactively re-inhibit L1. This would counteract the priming observed on Trials 2 and 3 in L1. One might argue that as bilinguals progressed in the experiment, they learned the trial order, and "proactively used proactive inhibition." We doubt this explanation however, because an analysis of repetition effects showed no two- or three-way interactions between repetition of the stimuli, language and sub-block trial number (see Supplementary Materials). Additionally, participants were not told that trial type would repeat, and informal interviews with several participants after the Experiment ended indicated they did not catch on to the trial order.

How, then, do the models account for reverse dominance, especially given the fact that it occurred in both experiments on stay and switch trials? We present two non-mutually exclusive ideas. First, proactive control may happen, but it is non-inhibitory in nature. Sspreading activation can still happen between semantic neighbors on stay trials and accounts for reverse dominance. Second, strong inhibition of the unintended language creates quick RTs, especially if within-language competition is absent. Consider *Eq. 4* of the Inhibitory Model. The faster distractors decrease in activation, the more quickly a word is selected. Because proficient bilinguals' L1 and L2 are similar in strength, slightly stronger L1 inhibition compared to L2 could cause L1 stay trials to be longer than L2 stay trials.[9] Thus, the model provides a mathematical explanation for the "overshoot" hypothesis proposed by Declerck et al. (2020) and Gollan and Ferreira (2009). Understanding reverse dominance was not a goal of this study, and future research is needed to better understand this question.

Conclusion

The modeling and experimental work we report suggest that language switching abolishes spreading activation effects, but cumulative semantic interference (created by a more gradual, incremental learning process) is unaffected by language switching. Experiment 1 provides evidence that bilinguals use inhibition in order to control language output, consistent with the Inhibitory Model. But it also demonstrates the need to update models of monolingual and bilingual lexical access to account for the facilitation that was found. Experiment 1 suggests that spreading activation does not create within-language competition among lexical entries. Experiment 2 suggests that models of bilingual language control should incorporate a mechanism of incremental learning. This study also gives researchers a new tool that can allow them to test for inhibition instead of switch costs. In answering the question, "How do bilinguals control their language output," the answer is by using global inhibition and by continual (and incremental) adaptation to their environment.

CRediT authorship contribution statement

Mark Lowry: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Chad Dubé:** Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review & editing, Resources, Supervision. **Elizabeth Schotter:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Resources, Supervision.

Appendix A. NCWL models

In order to remove the within-language competition assumption of the Non-Competitive Within-Language (NCWL) Inhibitory and Non-Competitive Within-Language Non-Inhibitory Models, we had to modify Eqs. 4 and 8 of the original models. In the original models, as the target activation increased, within-language distractor activation also increased. The increasing distractor activation affected how quickly a lemma was selected. To get rid of this effect, the within-language distractor activations were replaced by constants in those equations. To keep the new models as similar to the original models as possible, the sum (S) of the initial activation levels for the within-language competitors were used. Thus, Eq. 4 becomes

$$a_x \geqslant V \times (S + \sum a_b) \tag{10}$$

where $\sum a_b$ is the sum total of the between language distractors. Because the Non-Inhibitory Model has no between language distractors, Eq. 8 becomes

$$a_x \geqslant V \times (S) \tag{11}$$

It should be noted that activation still spreads to the non-target nodes. The only difference is that those activation levels do not contribute to the above equations.

Additionally, the models needed to be modified in order to create spreading activation effects. In the original models, the p_1 parameter was set to around 0.75-0.8. This allowed the target to receive more activation than the distractors. On the next semantically related trial, the target would become a distractor and increase naming latencies. However, because there is no longer any within-language competition between words, the p_1 parameter needed to be lower in order to give new targets sufficient priming on the next trial. There were other ways of accomplishing this goal, but they required changing multiple parameters. Modifying p_1 ensured that the new and original models were as similar as possible. One other change was made: the inhibition parameter (h) was no longer proportional to the reactivation parameter (Y). This is due to the fact that without within-language distractors, the switch cost was almost entirely determined by the inhibition parameter anyway (i.e., the faster the between-language words are inhibited, the faster the switch cost). Making h proportional to Y was redundant and unnecessary. However, the reactivation parameter was still proportional to the language strength parameter, just like the NCWL Non-Inhibitory Model.

Appendix B. Individual participant fits

We fit the original models to each participant's individual data. For each participant in Experiment 1, means were computed for each trial within both types of sub-blocks (12 total means for each participant: six uniform means and six mixed means). The first trial of a sub-block was then subtracted from each trial to normalize the data. The original models were fit to each participant's naming latency data using custom scripts programmed

[9] Some reactivation of L2 would also be needed between L1 stay trials.

in R. To simplify parameter estimation, the noise parameters were removed and the three remaining parameters for each model were allowed to vary randomly. The predictions were normalized as in the participants' data, and then compared to a participant's responses for both mixed and uniform sub-blocks. Root Mean Square Error (RMSE) was then calculated for the participant to estimate the goodness of fit of the model to that participant's data; higher RMSE values suggest a worse fit. This process was repeated 300 times for each participant. The iteration with the lowest average RMSE of both types of sub-blocks was then chosen, and the RMSE, parameter estimates, and fitted latencies were saved.

As an exploratory analysis, we also wanted to test the idea that less proficient bilinguals use inhibition, but more proficient bilinguals use a lexical selection mechanism (see Costa et al., 2006). If so, then individual participant fit (i.e., RMSE) for the Non-Inhibitory Models should be positively correlated with proficiency. We used the difference between L1 and L2 scores on the MINT as a measure of proficiency. Each participants' RMSE was used to determine model fit.

B.1. Fit of original models to individual participant data

For Uniform sub-blocks, the Inhibitory model had an average participant RMSE of 81.11 ($SD = 41.91$). Average RMSE of the Non-Inhibitory Model was equal to 78.58 ($SD = 40.05$). A Bayesian paired samples *t*-test showed that the difference was weak or anecdotal, $BF_{10} = 1.36$. In other words, there isn't much evidence that the non-inhibitory performed better on Uniform sub-blocks compared to the inhibitory model.

For Mixed sub-blocks, the Inhibitory model had an average participant RMSE of 70.09 ($SD = 39.63$). Average RMSE of the Non-Inhibitory Model was equal to 80.11 ($SD = 40.65$). A Bayesian Paired Samples *t*-test showed that the difference was strong, $BF_{10} = 35.51$. In other words, there is substantial evidence that the Inhibitory Model fit the data better than the Non-Inhibitory Model on mixed sub-blocks.

Fig. B.1 shows the three best(labeled 1-3) and worst (labeled 41–43) participant fits for the original models. The best fitting participants were those with the lowest RMSE. The worst fitting participants were those with the highest RMSE. Notice that the models fit the data well when there is little or no facilitation. When there is facilitation, the models fail.

B.2. Fit of new models to individual participant data

For Uniform sub-blocks, the NCWL Inhibitory model had an average participant RMSE of 84.57 ($SD = 40.43$). Average RMSE of the Non-Inhibitory Model was equal to 85.03 ($SD = 41.14$). A Bayesian paired samples *t*-test showed no credible difference, $BF_{10} = 0.16$. In other words, the null hypothesis is favored (i.e., both models performed equally well).

For Mixed sub-blocks, the Inhibitory model had an average participant RMSE of 68.56 ($SD = 35.87$). Average RMSE of the Non-Inhibitory Model was equal to 66.52 ($SD = 35.80$). and is an improvement over the original. A Bayesian Paired Samples *t*-test showed no credible difference between the NCWL Inhibitory and NCWL Non-Inhibitory, $BF_{10} = 0.21$. In other words, the null hypothesis is favored (i.e., both models performed equally well).

Fig. B.2 shows the three best (labeled 1-3) and worst (labeled 41–43) participant fits for each model. Notice that qualitatively, the NCWL Inhibitory model can now account for facilitation on the first three trials (e.g., 41).

To test whether the NCWL Non-Inhibitory Model fit more proficient bilinguals better than for less proficient bilinguals, we ran a Pearson's correlation using the participants' MINT scores (L1 vocabulary score - L2 vocabulary score) and individual RMSE. As participants got more proficient, their RMSE scores decreased (i.e., the fit was better), but the effect was small and not significant, $r(42)=0.09$, $p = 0.58$.

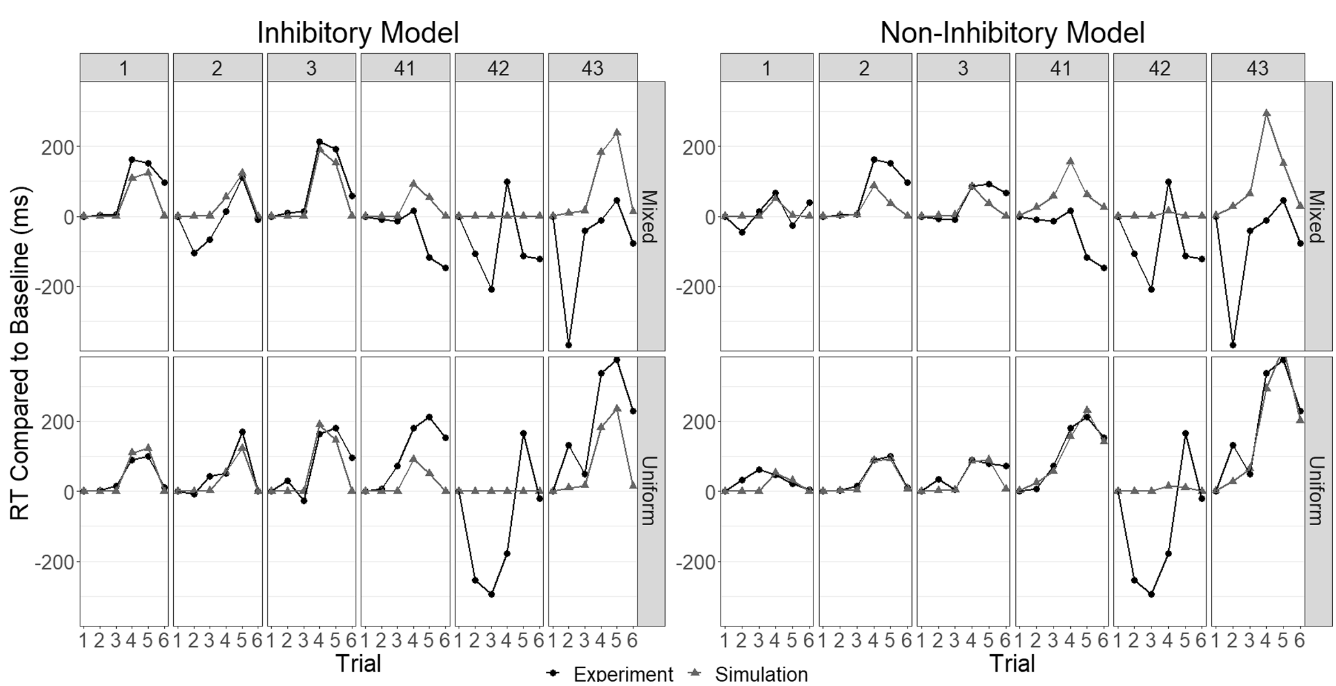


Fig. B.1. Fits of the original models to the data of six participants (3 best and 3 worst). The language strength parameter (L), the inhibition parameter (h) and competition parameter (V) were allowed to change. Fits for Inhibitory Model are shown on the left graphs, and Non-Inhibitory Model are shown on the right graphs. The top panels show fits for Uniform Sub-Blocks, and the bottom panels show fits for Mixed Sub-Blocks.

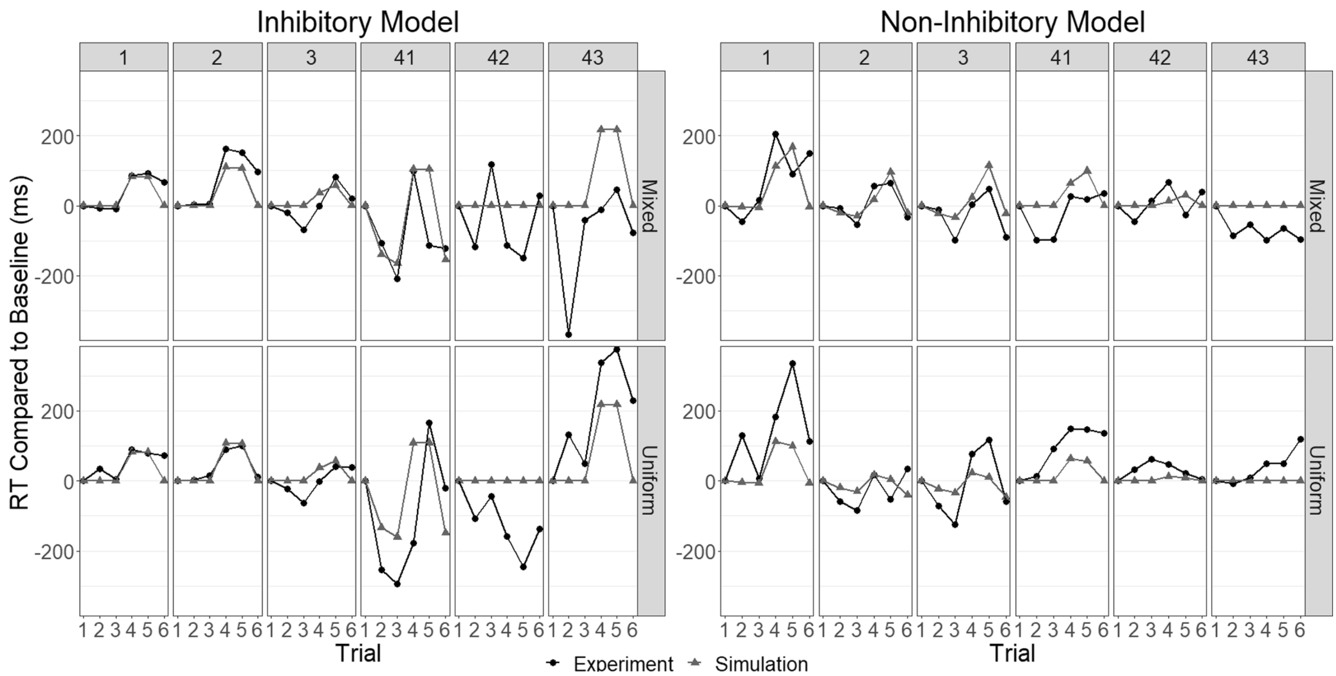


Fig. B.2. Fits of the new models (i.e., no within language competition) to the data of six participants (3 best and 3 worst). The language strength parameter (*L*), the inhibition parameter (*h*) and competition parameter (*V*) were allowed to change. Fits for NCWL Inhibitory Model are shown on the left, and NCWL Non-Inhibitory Model are shown on the right. The top panels show fits for Mixed Sub-Blocks, and the bottom panels show fits for Uniform Sub-Blocks.

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jml.2020.104195.

References

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305.

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

Boersma, P. (2006). Praat: Doing phonetics by computer. http://www.praat.org/.

Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2016). Language control in bilinguals: Monitoring and response selection. *Cerebral Cortex (New York, N.Y.: 1991), 26*, 2367–2380.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*(1), 177–208.

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

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256.

Costa, A. (2005). *Handbook of Bilingualism: Psycholinguistic Approaches., chapter Lexical Access in Bilingual Production* (pp. 308–325). Oxford University Press.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from SpanishSpanish bilinguals. *Bilingualism: Language and Cognition, 2*, 231–244.

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

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

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.

Costa, A., Pannunzi, M., Deco, G., & Pickering, M. J. (2017). Do bilinguals automatically activate their native language when they are not using it? *Cognitive Science, 41*(6), 1629–1644.

Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1372.

Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review, 10*(1), 111–117.

Damian, M. F., Vigliocco, G., & Levelt, W. J. (2001). Effects of semantic context in the naming of pictures and words. *Cognition, 81*(3), B77–B86.

Davies, M. (2002–2017a). Corpus del Español: 100 million words, 1200s–1900s. Available online at http://www.corpusdelespanol.org.

Davies, M. (2008–2017b). The Corpus of Contemporary American English (COCA): 520 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

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

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

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). Erp evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in Psychology, 1*, 177.

Delprato, D. J. (1971). Specific-pair interference on recall and associative-matching retention tests. *The American Journal of Psychology*, 185–193.

Delprato, D. J. (2005). Retroactive interference as a function of degree of interpolated study without overt retrieval practice. *Psychonomic Bulletin & Review, 12*(2), 345–349.

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition, 22*(4), 657–679.

Eriksen, B., & Eriksen, C. (1974). Effectsofnoiselettersuponidentificationofa targetletterinanonsearchtask.

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*(5), 1075.

Finkbeiner, M., Gollan, T. H., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism: Language and Cognition, 9*(2), 153–166.

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

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition, 15*(3), 594–615.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67–81.
Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530.
Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes, 8*(3), 291–309.
Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42*(2), 113–157.
Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition, 1*, 213–229.
Hong, L., & Macwhinney, B. (2011). Semantic competitor priming within and across languages: The interplay of vocabulary knowledge, learning experience and working memory capacity. *Bilingualism: Language and Cognition, 14*(4), 433–443.
Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100*(3), 464–482.
Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition, 4*, 169–178.
Jacobs, A., Fricke, M., & Kroll, J. F. (2016). Cross-language activation begins during speech planning and extends into second language speech. *Language Learning, 66*, 324–353.
Kheder, S., & Kaan, E. (2019). Lexical selection, cross-language interaction, and switch costs in habitually codeswitching bilinguals. *Bilingualism: Language and Cognition, 22*(3), 569–589.
Kleinman, D., & Gollan, T. H. (2018). Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition, 173*, 115–132.
Kroll, J., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174.
Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica, 128*(3), 416–430.
Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*, 373–381.
Lee, M.-W., & Williams, J. N. (2001). Lexical access in spoken word production by bilinguals: Evidence from the semantic competitor priming paradigm. *Bilingualism: Language and Cognition, 4*(3), 233–248.
Levelt. (1999). Models of word production. *Trends in cognitive sciences, 3*, 223–232.
Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38.
Linck, J.a., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition, 15*, 651–662.
Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization, Number 8*. Oxford University Press on Demand.
Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & 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*(3), 503.
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7.
McGeoch, J. A., & Irion, A. L. (1952). *Psychology of human learning* (revised ed.). New York: Longmans.
Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25–40.
Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language, 67*(1), 224–237.
Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (Words): An electrophysiological study of code switching. *Brain and Language, 80*(2), 188–207.
Navarrete, E., Mahon, B. Z., & Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica, 134*(3), 279–289.
Navarrete, E., Del Prato, P., & Mahon, B. Z. (2012). Factors determining semantic facilitation and interference in the cyclic naming paradigm. *Frontiers in Psychology, 3*, 38.
Navarrete, E., Prato, P. D., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language, 76*, 253–272.
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc, 36*, 402–407.
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*(2), 227–252.
Plummer, M. (2013). rjags: Bayesian graphical models using MCMC. R package version 4.6, 3.
R Core Team. (2014). *R: A language and environment for statistical computing* (p. 2013). Vienna, Austria: R Foundation for Statistical Computing.
Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*(1–3), 107–142.
Runnqvist, E., Strijkers, K., Alario, F.-X., & Costa, A. (2012). Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language, 66*(4), 850–869.
Santesteban, M., & Schwieter, J. W. (2019). 6 lexical selection and competition in bilinguals. *Bilingual Lexical Ambiguity Resolution*, 126.
Schriefers, H., Meyer, A., & Levelt, W. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*(1), 86–102.
Schwanenflugel, Paula J., & Rey, Mario (1986). The relationship between category typicality and concept familiarity: Evidence from Spanish-and English-speaking monolinguals. *Memory & Cognition, 14*, 150–163. https://doi.org/10.3758/BF03198375.
Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J.-M. (2016). Using brain potentials to functionally localise stroop-like effects in colour and picture naming: perceptual encoding versus word planning. *PloS One, 11*(9).
Slamecka, N. (Ed.), (1967). *Human memory and learning: Selected readings*. New York, NY: Oxford University Press.
Underwood, B. (1983). *Attributes of memory*. Glenview, IL: Scott, Foresman.
Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition, 110*(1), 84–99.
Wheeldon, L. R., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language, 33*(3), 332–356.
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).