ABSTRACT

Title of Dissertation: THE WHAT AND WHERE OF CONTROL IN BILINGUAL LANGUAGE SWITCHING

Alison Ruppel Shell, Doctor of Philosophy, 2018

Dissertation directed by: Dr. L. Robert Slevc
Department of Psychology

Speakers of multiple languages must somehow express intended concepts using the appropriate lexical item in the intended language while not producing lexical items from another language that could equally well express the intended concepts. Thus bilingual speakers must presumably manage competition from these items active in their multiple languages in order to successfully communicate. However, it remains unclear where in the process of language production the competition exists, and what mechanisms are used to resolve the competition and successfully produce language. This dissertation set out to more robustly examine the implications of the prominent idea that domain general inhibitory control is used to inhibit the non-target language. To begin, I re-analyzed existing results from studies correlating measures of language switching and inhibitory control using a Bayesian approach. This reanalysis found that much of the previous literature either provides evidence *against* a relationship between a domain general inhibitory control task and language switching, or finds

little to no evidence for such a relationship. Across two experiments, I then assess the role of domain-general inhibitory control in bilingual lexical access using a dual-task design–combining a language switching task with a concurrent task taxing domain-general cognitive control–as well as an individual differences component in the relatively well-powered and pre-registered Experiment 2. In these experiments, I break down the complex process of inhibitory language control into possibly dissociable *levels* of control (control at the language level and control at the item level) and assess potentially dissociable *types* of control (proactive control used to bias and monitor for conflict more broadly, and reactive control used for dynamically selecting between languages at a trial by trial level). There was evidence against a role of reactive control in switching between languages at both the language and item level. There was some evidence, however, suggesting a potential role for proactive control or monitoring in a language switching context. Correlations between language switching costs and domain-general measures of inhibitory control suggest that language proficiency and flexibility of control may modulate the ability to reactively control language in a language switching context, however the specificity of these findings demonstrate the complexity of this relationship, in line with the mixed findings in the current literature.

THE WHAT AND WHERE OF CONTROL IN BILINGUAL LANGUAGE SWITCHING

by

Alison Ruppel Shell

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Dr. L. Robert Slevc, Chair
Dr. D.J. Bolger
Dr. Michael R. Dougherty
Dr. Jared Linck
Dr. Jared M. Novick

Acknowledgements

This work has been an incredible learning experience- not only learning about the complexities of the research- but also the challenges that come with this type of work. Though the aim of a dissertation is an independent project, there is no way I would have been able to do it alone and I am forever grateful for the many people who helped it come to fruition. To the Language Music and Cognition Lab and to all of the RA’s who helped to collect data, an essential component to the work! To my friends, colleagues, and running buddies- who have helped keep me stay sane through the process. To my committee, and in particular Jared Linck, who, along with my advisor Bob Slevc, got me started on this surprisingly complex and exhausting project seven years ago and has been supportive of the pursuit for answers ever since. To Bob- who has been with me from day one and never doubted I’d get it done- helping me keep spirits up, talking me through complex ideas, and working with me through my crazy technique of last-minute writing. To my amazing family, including my parents who have always inspired me and pushed me to persist when things get challenging, to Avery- who added some *extra* challenge to the process but also added joy to the days, and finally to my amazing, smart, patient and strong partner Jake -for modeling tenacity, and for his encouragement, and loyal blind faith through it all.

Table of Contents

Acknowledgements ii
Table of Contents iii
List of Tables v
List of Figures viii
Chapter 1: Introduction 1
Monolingual Language Control 4
Bilingual Language Control 6
The "Bilingual Advantage" 8
Language Switching 11
Relationships Between Individual Differences in Language and Domain General Control 13
Dual-task Assessment of Overlap Between Language and Domain-General Control 15
Re-evaluating Previous Work 17
Proactive vs. Reactive Inhibitory Control 21
Dual Levels of Control 24
Chapter 2: Experiment 1: Predictable Switching 31
Method 31
Participants 33
Materials and Procedures 33
Analysis and Results 35
Response Time Analysis 37
Accuracy Analysis 40
Chapter 3: Experiment 2: Taxing Control at the Task Level: Method 42
Participants 43
Procedure 43
Practice Counting 44
Language Switching 44
Simon Arrow Task 45
Single Language + Arrow Task 45
Combined Task 45
Single Language + Arrow Task (Part 2) 46
Individual Differences Assessments 46
Simon Arrow Task 46
AX-CPT 46
Language Proficiency 49
Chapter 4: Experiment 2: Analysis and Results 50
Switch Costs 50
Mixed Effects Analysis 51
Bayes Analysis 53
Mixing Costs 56
Mixed Effects Analysis 57

Bayes Analysis 59
Drift Diffusion Analysis 62
Individual Differences 65
Individual Differences Measures 66
Language Switching and Simon Task 67
Dual-task Language Switching and Simon Task 68
Language Mixing and Simon Task 69
Language Switching and AX-CPT 70
Language Mixing and AX-CPT 71
Language Proficiency 72
Chapter 5: Discussion 75
Appendices 86
References 89

List of Tables

Table 1. Bayes Factors for Reanalyzed Experiments 1-3 (Shell, 2015)

Table 2. Bayes factors for correlation between inhibitory control task (IC Task) and language switching (switch BF), language mixing (mix BF), and switch costs per language (L1, L2, L3).

Table 3. Experiment 1: Averaged Self-rated L2 Proficiency (Scale 0-10) and Percent Time Exposed to the L2 (from the modified LEAP-Q; Marian et al., 2007).

Table 4. Classification for Bayes Factors (adapted from Wagenmakers,Wetzels, Borsboom, & van der Maas, 2011; cf. Jeffreys, 1961).

Table 5. Experiment 1 Response Time (RT) in Milliseconds for Switch, Stay, and Cost (Switch-Stay RT) by Language and Congruency

Table 6. Experiment 1 Switch Cost Mixed Effects

Table 7. Experiment 1 Switch RTs Top 6 Best Fit Bayes Models

Table 8. Experiment 1 Mean Accuracy Rate for Stay Trials, Switch Trials and Switch Costs (Switch-stay Accuracy) by Language and Arrow Congruency

Table 9. Experiment 1 Switch Accuracy Top 6 Best Fit Bayes Models

Table 10. Experiment 2: Averaged Self-rated L2 Proficiency (Scale 0-10) and Percent Time Exposed to the L2 (from the modified LEAP-Q; Marian et al., 2007).

Table 11.Experiment 2 Response Time (RT) in Milliseconds for Switch, Stay, and Cost (switch-stay RT) by Language and Congruency

Table 12. Experiment 2 Language Switch Mixed Effects Results

Table 13. Experiment 2 Language Switching Mixed Effects for L1(top) and L2 (bottom)

Table 14. Experiment 2 Language Switching (RT) Top 6 Best Fit Bayes Models

Table 15. Experiment 2 Language Switching (RT) Top 3 Bayes Models: L1(top), L2(bottom)

Table 16. Experiment 2 Language Switching Accuracy: Top 6 Bayes Models

Table 17. Experiment 2 Mean Accuracy Rate for Stay Trials, Switch Trials and Switch Costs (Switch-Stay) by Language and Arrow Congruency

Table 18. Experiment 2 Response Time (RT) in Milliseconds for Single Language, Mixed Language, and Cost (mixed-single language RT) by Language and Congruency

Table 19. Experiment 2 Mix Cost Mixed Effects Results

Table 20. Experiment 2 Language Mix Cost Results by Language L1 (top) L2 (bottom)

Table 21. Experiment 2 Language Mixing RT: Top 6 Best Fit Bayes Models

Table 22. Experiment 2 Language Mixing RT: Top 3 Bayes Models in the L1 (top) and L2 (bottom)

Table 23. Experiment 2 Language Mixing Accuracy: Top 6 Bayes Models

Table 24. Experiment 2 Accuracy Rate for Single Language, Mixed Language, and Cost (mixed-single language) by Language and Congruency

Table 25. Experiment 2 Top 3 Bayes Models for Switch Costs using Drift Diffusion Parameters Ter, a, v

Table 26. Experiment 2: Top 3 Bayes Models for Mix Costs using Drift Diffusion Parameters Ter, a, v

Table 27. Internal Reliability Measures and Descriptive Statistics for Switch and Mix Costs, Simon Task Congruency, and AXCPT measures (RT in ms)

Table A.1. Experiment 1 (Untrimmed) Top 6 Best Fit Models of Language Switch from Bayes ANOVA

Table A.2. Experiment 2 (Untrimmed) Top 6 Best Fit Models of Language Switch from Bayes ANOVA

Table A.3. Experiment 2 (Untrimmed) Top 6 Best Fit Models of Language Mixing from Bayes ANOVA

Table B.1. Bayes Factors for correlations between Language Switch and Mix Costs and AX-CPT log RTs

Table B.2. Bayes Factors for correlations between Language Switch and Mix Costs and AX-CPT accuracy scores

Table C.1. Correlation matrix (Pearson's r) for all individual differences (log RT) measures

List of Figures

Figure 1. a) Model of monolingual speech production process. b) Model of bilingual speech production process. (from Shell, 2015; adapted from Levelt & Meyer, 1999; Kroll, Bobb, & Wodniecka, 2006).

Figure 2. Measures of domain general inhibitory control. Control is typically assessed as the difference in response time or accuracy (cost) on incongruent trial types (bottom row) compared to congruent trial types (top row).

Figure 3. Schematic of Green's (1998) Inhibitory Control Model (adapted from Declerck, Koch, & Philipp, 2015).

Figure 4. Schematic of trial from Experiment 1 (Shell, 2015), combining predictable language switching and Simon arrow task.

Figure 5. Schematic of a series of four trials in the predictable combined language switching and Simon arrow block from Experiment 1.

Figure 6. Experiment 1: Naming reaction time (log ms) by arrow congruency condition, switch condition and language.

Figure 7. Schematic of a single AX-CPT trial (A-X trial type).

Figure 8. Experiment 2 Naming reaction time (log ms) by arrow congruency condition, switch condition and language.

Figure 9. Experiment 2 Naming reaction time (log ms) by arrow congruency condition, block, and language.

Figure 10. Plotted relationship between self-rated L2 proficiency (scale 1-10) and L1 and L2 switch costs (log ms).

Chapter 1: Introduction

People talk a lot: on average, a typical speaker produces two to three words per second (Levelt & Meyer, 1999) out of the approximately 42,000 words that they know (Brysbaert, Stevens, Mandera, & Keuleers, 2016), and this all happens with apparent ease. On the back end of the process, language production models generally claim that speakers choose a certain word to produce by means of a process of spreading activation. More specifically, most models of word production (e.g., Levelt & Meyer, 1999) propose that activation from a concept to be expressed spreads to an array of relevant lexical items, *lemmas*, which carry semantic and syntactic information but not word-form information. This activation spreads both to the target lemma and also those lemmas sharing semantic properties with the target (see Figure 1a). Because this process results in multiple activated lexical items, lexical selection is generally assumed to be a competitive process.

Indeed, evidence of competition between semantically similar items has been demonstrated with tasks requiring naming items in the context of semantically similar competitors. The blocked naming paradigm, for example shows an increase in difficulty (slower naming times) when naming items among a set of semantically related items (e.g., Belke, 2008; Damian, Vigliocco, & Levelt, 2001). A similar paradigm used to investigate this semantic competition is the picture word interference (PWI) task, where a distractor word is presented along with the picture to be named. Typically in this task picture naming is slower when distractor words (presented either aurally and/or superimposed in print) are semantically related to the pictured item, suggesting interference (Damian & Bowers, 2003; Glaser, & Düngelhoff, 1984; Schriefers, Meyer, & Levelt, 1990). The degree of lexical competition is arguably even greater in bilingual speakers, who not only need to overcome this within-language competition from related items, but presumably also the competition from related items in their other language, assuming both languages are active during selection (see Figure 1b). Indeed, most evidence suggests that multiple languages are in fact active for bilinguals at the point of lexical selection, suggesting that this additional competition does exist in bilingual production (e.g. Green, 1998; Meuter & Allport, 1999; Philipp & Koch, 2009), and thus bilinguals must overcome both within- and between-language competition (e.g., Gollan & Silverberg, 2001).

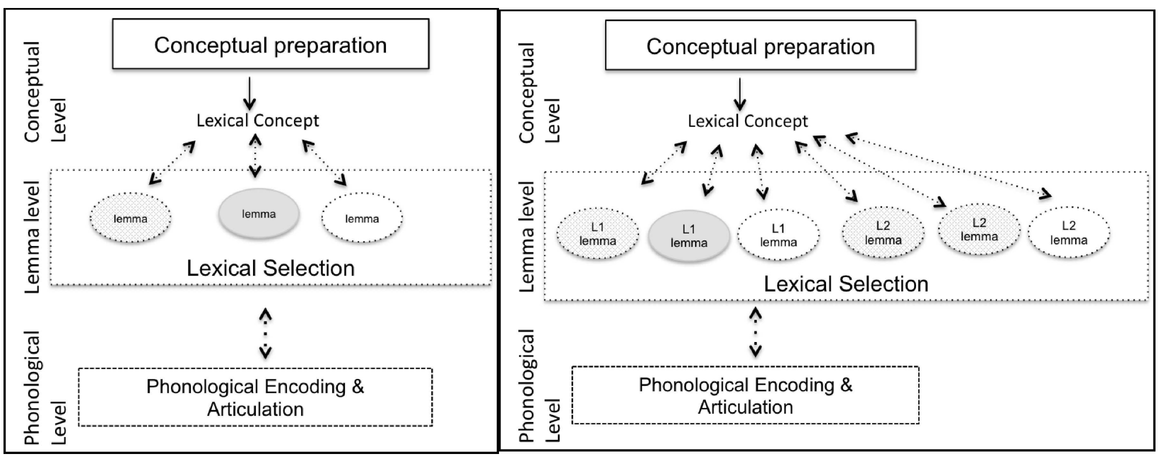

Figure 1.
a) (Left): Model of monolingual speech production process. Arrows represent activation between levels while shading in nodes represents amount of activation.
b) (Right): Model of bilingual speech production process. Note that the location of the lemma nodes is representative and does not specify that dominant language (L1) and second language (L2) lemmas are in distinct regions (Shell, 2015; adapted from Levelt & Meyer, 1999; Kroll, Bobb, & Wodniecka, 2006).

How bilinguals address and resolve this between-language competition is very much an ongoing debate. While some evidence supports the idea that cross-language interference is managed by language-specific mechanisms (i.e. the language specific model, Bloem & LaHeij, 2003), there is also considerable evidence suggesting that a domain-general system, such as the system used in general task switching, is involved in both monolingual and bilingual language production. This domain general control- that is, the processes used to manage conflict generally rather than one specific to a particular domain such as language- is likely multi-faceted and has been conceptualized in a number of ways including reactive inhibitory control (e.g. Green, 1998), proactive conflict monitoring (Botvinick, Braver, Barch, Carter, & Cohen, 2001), as well as cognitive control used to flexibly shift between use of proactive and reactive control (i.e., dual mechanisms of control: Braver, 2012).

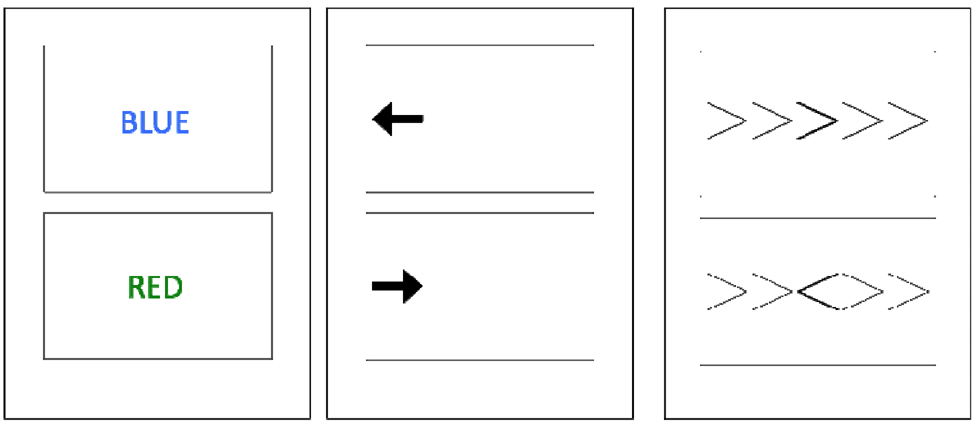

Figure 2.
Measures of domain general inhibitory control. Control is typically assessed as the difference in response time or accuracy (cost) on incongruent trial types (bottom row) compared to congruent trial types (top row)
(Left) Stroop task (Stroop, 1935). Participants must name the color the word is written in and ignore the color word itself.
(Middle) Simon arrow task (modified from Simon & Rudell, 1967). Participants must press an arrow key corresponding to the direction of the arrow while ignoring the arrow location on the screen.
(Right): Flanker. Participants must press an arrow key corresponding to the direction that the central arrow is pointing while ignoring the direction of the flanking arrows.

There are a few types of evidence for the involvement of some form of domain general control in language production, including studies investigating their shared neural processes (e.g. De Bruin, Roelofs, Dijkstra, & FitzPatrick, 2014; Guo, Liu, Misra, & Kroll, 2011), work showing interference between domain general control tasks and language selection tasks (e.g. Ferreira & Pashler, 1994; Shell, 2015), and work showing the relationship between domain general task performance and language control abilities (e.g. Kan et al., 2013; Linck, Schwieter, & Sunderman, 2011).

Monolingual Language Control

One type of evidence for a domain-general system in monolingual lexical selection comes from Ferreira and Pashler (2002), who found that pairing a word production task with a concurrent secondary (non-linguistic) task hindered secondary task performance. More specifically, they demonstrated that lemma selection, but not phoneme selection, interacted with performance on a concurrent tone-discrimination task. They argue that their findings support shared processing between language production and domain general processes at the response selection (i.e. lemma selection) but not the response execution (phoneme selection) level. Other evidence for a role of more general control in lexical selection has been presented with case studies of patients whose language deficits following a brain lesion could be tied to impairments in cognitive control more generally (e.g. Hamilton & Martin, 2005, Novick, Trueswell, & Thomson-Schill, 2010; Novick, Kan, & Thompson-Schill, 2013). Novick and colleagues (2013), for example, present a patient with damage to the left inferior frontal gyrus (LIFG: an area of the brain that is typically involved in dealing with conflict) who showed interesting patterns of task difficulty in both domain general interference resolution and particular language tasks. Specifically, the patient showed deficits on a recent probes task, a non-linguistic (though admittedly verbal in nature) task of inhibitory control where participants must respond ‘yes’ to items from the current (immediately preceding) list while still responding ‘no’ to recent/familiar items that occurred in the previous set. Similar deficits were observed in language production tasks that presumably require linguistic interference resolution. For example, the patient had difficulty naming pictures associated with multiple possible names (e.g., couch, sofa, loveseat), which requires selecting between these multiple possible responses, and difficulty naming multiple items from a large set with many possible responses (e.g. animals) compared to naming items from smaller (e.g. farm animals) response sets. In these tasks, conflict in response selection is inherently greater given a larger set of options to select from, as each item in the set can ostensibly interfere with selection of a single item.

Another patient (Hamilton & Martin, 2005) with left frontal and parietal damage also had impaired performance on both domain general and linguistic control tasks. The patient showed deficits on the recent probes task and the verbal Stroop task, in addition to impaired semantic short term memory- that is- increased interference from recently named items in a serial recall task. These deficits generally suggest difficulty with language control and selection under conflict or interference. Because these deficits were paired with cognitive deficits in the conflict resolution tasks involving similar processing demands, the authors suggest that the components of language production that are involved in dealing with linguistic competition may rely on the same general system used for resolving conflict more broadly. It is important to note however that the impaired performance on the domain general tasks for both of these patients was found only on tasks that were verbal in nature (in fact Martin and Hamilton did not find impaired performance on the nonverbal version of the Stroop), and so while these findings do suggest that verbal conflict resolution more broadly may be related to lexical selection, they do not support a role for non-linguistic conflict processes in lexical selection.

Other research has failed to find a relationship between non-verbal domain general and language control tasks in a typical monolingual population (Alario, Ziegler, Massol, & Cara, 2012). Alario and colleagues (2012) assessed the relationship between three tasks involving different forms of response selection among interference, including linguistic (blocked semantic naming: naming pictured items drawn from the same semantic category) and non-linguistic (Simon task and hue discrimination task) forms of discrimination. While all tasks showed a main effect of increased difficulty on the high interference trials, there was no evidence for a relationship between performance on non-linguistic and linguistic tasks, as would be predicted if they involved shared mechanisms. Evidence for shared mechanisms is often demonstrated in overlapping regions of activation in the brain during different tasks- typically taken to suggest that the tasks involved shared neural underpinnings or mechanisms. Piai, Roelofs, Achenson and Takashima (2013), however looked for these regions of neural overlap in participants during performance on three different tasks with varying need for language vs. domain general control: the PWI, Stroop, and Simon tasks. They found that while there were some overlapping regions of activation during these different tasks, there was also evidence for task-specific activity during the PWI task, which involves lexical competition among semantic competitors. These findings suggest that domain general control mechanisms may be shared between non-linguistic tasks and tasks of language control, but also that mechanisms specific to language control may be additionally involved. Therefore, in studying language production and control, both domain general and language-specific control processes should be considered viable mechanisms and may in fact work in concert to help control the complex process of language production.

Bilingual Language Control

One prominent theory of bilingual language control that proposes inhibition as its underlying mechanism is the Inhibitory Control Model (Green, 1998). Rooted in classic cognitive models of control, which suggest multi-level systems of controlling routine vs. non-routine behavior, the inhibitory model proposes that speakers use domain general inhibitory control mechanisms during language switching to suppress items in the non-target language (Green, 1998). More specifically, the model proposes that a language task schema- a top down mental device used for goal selection- is used to specify the language task, while at the lemma, or item level, language is specified by a language 'tag' (see Figure 3 below). After all items linked to active concepts have been activated, it is ensured that the correct item is chosen by reactively inhibiting the non-target tagged items (the amount of inhibition required is relative to the level of activation of the active items). As such, switching between languages requires control at both the schema level (to *bias selection* for the correct language task) and at the lemma level (to *reactively inhibit* all non-target tagged items).

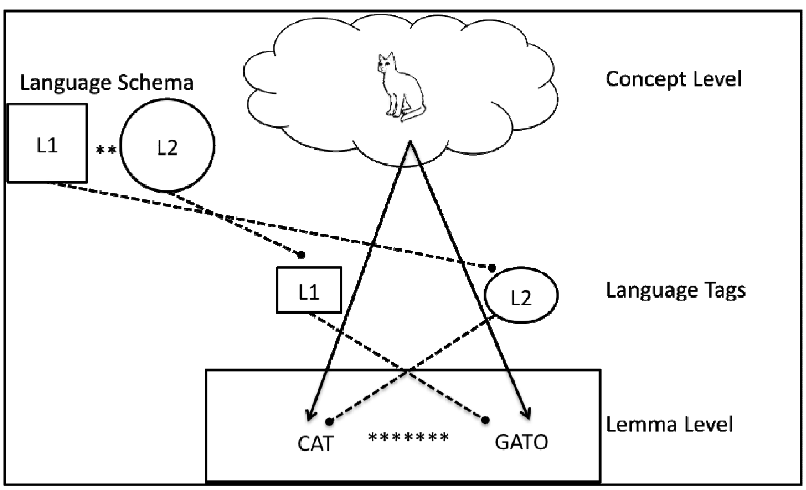

Figure 3. Schematic of Green's (1998) Inhibitory Control Model (adapted from Declerck, Koch, & Philipp, 2015). Arrowheads represent activation, dotted capped lines represent inhibition, and asterisks represent interference control.

This model helps to clarify the role of inhibitory mechanisms involved in language switching, and supports much of the experimental work suggesting this involvement. The evidence directly tying inhibitory control to bilingual language control, however, is certainly not conclusive. Much of the work assessing the mechanisms of lexical access in bilinguals has focused on a number of general questions, including potential advantages in non-linguistic cognitive tasks such as the Stroop, Simon, or flanker tasks (see Figure 2), as well as performance in language switching tasks, where bilinguals' ability to flexibly shift between multiple languages is typically measured. These various forms of evidence are often used to support shared mechanisms between these tasks and general bilingual language production and use, which will be outlined below.

The "Bilingual Advantage"

Much of the support for a prominent role of domain general inhibitory control in bilingual research comes from studies comparing task performance of monolinguals to bilinguals on non-linguistic inhibitory control tasks, often showing better performance in the bilingual group (e.g., Abutalebi et al., 2012; Bialystok, 1999). These findings are predicted if a bilingual's lifetime of using inhibitory control for dealing with multiple languages transfers to improved performance in other types of domain-general attention and inhibitory control tasks. This is often termed the "bilingual advantage" (Bialystok, 2009) and is an area of research that has become both quite popular and quite controversial (see below). Interestingly, the bilingual advantage in domain general inhibitory control tasks is typically not only limited to enhanced performance on those trials specifically requiring inhibitory control (e.g., incongruent trials), but instead often emerges as a more global benefit, for example, as faster responses on both congruent and incongruent trials (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004; Hilchey & Klein, 2011).

This global benefit in response times has been argued to be evidence *against* an advantage specific to dealing with conflict and so questions the claims of the bilingual advantage in inhibitory control (e.g., Hilchey & Klein, 2011). However, it may be that this global benefit disappears when conflict, or need for conflict monitoring is reduced. For example, Costa, Hernández, Costa-Faidella, and Núria Sebastián-Gallés (2009) tested bilingual and monolingual participants on versions of the flanker task with either mostly congruent or incongruent trials (low monitoring context) or a more even mix of incongruent and congruent trials (high monitoring context). While bilingual and monolinguals had similar performance in the low monitoring version, bilinguals were faster to respond to both congruent and incongruent trial types in the high monitoring version. When participants are unable to prepare for the trial type (target congruent or incongruent with the flankers) they must be constantly monitoring to respond, while monitoring needs are reduced when trial type is more consistent. As such, some have argued that the bilingual advantage stems not from conflict resolution or inhibitory abilities per se, but rather from an ability to flexibly deal with interference, sometimes called *conflict monitoring* (e.g., Abutalebi et al., 2012; Botvnick et al., 2001; Teubner-Rhodes, Bolger, & Novick, 2017). This ability for flexible control more recently has been considered a central component of the bilingual advantage (Abutalebi et al., 2012; Morales, Gomez-Ariza, & Bajo, 2013). That is, maintenance of two languages might require enhanced monitoring of the language environment to be *prepared* to deal with conflict, which could then impart enhanced domain general monitoring and adaptation abilities. In support of this, Wu and Thierry (2013) found that when bilingual participants performed a domain general control task (flanker task) that had been interleaved with distractor written words which they were told to ignore, their flanker performance was better when the words were presented in two languages, rather than just one. The authors suggest that for bilingual speakers, simply being in a bilingual context encourages enhanced monitoring generally, and thereby can improve performance on the non-linguistic flanker task. Similarly, research has found that bilinguals typically show an advantage on performance of the AX-CPT task - a task requiring participants to attend to a series of letter stimuli and to target “AX” trials; that is, when a valid cue (A) is followed by a valid probe (X), while non-target trials include either a valid cue followed by an invalid probe ("AY"trials) or an invalid cue followed by either a valid or invalid probe ("BX" or "BY" trials). Trials following the A cue presumably increase expectancy for a target response while maintenance of the B-cue context (i.e. any non- A cue trial) is necessary to override a tendency to respond to an X probe (Braver, 2012). Proficiency on this task demonstrates effective adjustment of control (Morales, et al., 2013) and so supports the theory that the bilingual advantage reflects underlying advantages in monitoring for conflict and adaptively engaging domain general inhibitory control, rather than inhibitory control ability itself (Hilchey & Klein, 2011).

Though there is much support for a relationship between bilingual language use and more general cognitive functioning, the field has yet to consistently agree upon the specifics of this relationship, and the advantage is not consistently found across studies (Paap & Greenberg, 2013). These inconsistencies may stem from variable definitions of the bilingual population (e.g. Luk & Bialystok, 2013; Yang, Hartano, & Yang, 2016), as well as variable measures of the corresponding ‘advantage’ (e.g. inhibitory control) which may be defined by different outcome measures (e.g. Flanker task, Simon task, Stroop task). Moreover, because comparisons of bilinguals and monolinguals are necessarily between-subject designs, it is easy to confound group differences in the outcome measure with the many other possible differences between groups (e.g. education level, socioeconomic status, etc.). And so, if group differences are found, it is unclear which of the variables may be driving the advantage. In addition, researchers have proposed a role of publication bias in further confusing the field, suggesting a biased tendency to publish results supporting the advantage while underrepresenting the findings which do not (De Bruin Treccani, & Della Salla, 2014, but see also Bialystok, Kroll, Green, MacWhinney & Craik, 2015). As such, while group difference findings can be a good starting point for understanding the underlying mechanisms used in bilingual control, it is not clear 1) that these findings should be taken as direct evidence that domain general inhibitory control is involved in bilingual language use or 2) if the evidence that exists is representative of the full body of research. The primary experiment here (Experiment 2) was therefore pre-registered and used Bayesian methods of analysis to better study, assess, and present the existing evidence for and against this hypothesized link.

Language Switching

Some of these interpretive limitations of the potential "bilingual advantage" in domain general cognitive control tasks are less problematic in work taking an experimental approach to lexical selection. Perhaps the most commonly used experimental paradigms to investigate this bilingual lexical competition and domain general control processes are language switching tasks. Note that while language switching tasks are most certainly not representative of a bilingual speaker's typical language experience and so might not reflect the nuances of language control in context, theses tasks can be used estimate language control in a more controlled setting and serve an efficient way to test predictions. In the typical language switching paradigm, speakers must name items in a given language and switch between two or more languages typically according to some sort of language cue. As is the case with basic task switching paradigms, there is usually a cost of switching, as measured by increased latency and/or reduced accuracy on trials where the task switches compared to trials where the task remains the same as the previous trial. Importantly, switch costs are usually larger when switching from a harder, or less practiced, to an easier, or more practiced, task. This switch cost *asymmetry* has been explained as being due to the need to more heavily inhibit interference from the more active dominant task in order to perform the non-dominant, target task. As the dominant task has been inhibited to a greater extent, there is a larger cost to overcome this inhibition when switching back into it (Wylie & Allport, 2000; cf. Monsell, Yeung, & Azuma, 2000). In terms of language switching, as the dominant language is generally more readily accessed, it competes more for selection, and must be inhibited to a greater extent when switching. And so, items in the dominant language require more activation to retrieve out of its inhibited state on the next trial. As such, in switching between languages, a speaker incurs a language-switching cost, and, in unbalanced bilinguals, an analogous switch cost asymmetry with a larger cost when switching into their dominant language (typically their first language, or L1) than when switching into their less dominant, or second language (L2) (e.g. Meuter & Allport, 1999; Guo, Liu, Misra, & Kroll, 2011). This asymmetry is often taken as evidence for use of inhibition in language switching, based on the logic that the asymmetry is due to an overcoming of previous inhibition. However this cost asymmetry is not always found, and switch costs can vary based on a number of task variables such as the timing of stimulus presentation, and switch cost scoring methods (e.g. Verhoef, Roelofs, & Chwilla, 2009; Hughes, Linck, Bowles, Koeth, & Bunting, 2014). Switch cost asymmetries could also be due to underlying mechanisms other than inhibition, for instance interference from persisting *activation* of the weaker task (Phillip, Gade & Koch, 2007). Moreover, if inhibition is being measured by these asymmetries, it is most likely only measuring local, or reactive inhibition (cf. Bobb & Wodniecka, 2013), and not the more global, proactive inhibition that I propose may be involved in language control. As such, relying on switch cost asymmetries to index inhibitory control during switching might not be a particularly effective measure.

Relationships Between Individual Differences in Language and Domain General Control

There are a number of studies in the literature investigating the relationship between individual differences in bilingual language control, as measured by language switch costs, and domain general task switching abilities. For example, Calabria and colleagues (2012) assessed both bilingual language switching and color-shape task switching in the same participants to see if performance in one task predicted performance in the other. In the language task, a representative flag presented above each picture cued the speaker to name the item either their L1 or L2 which required either staying in the same language as the previous trial (stay trials) or switching from one language to another (switch trials). In the color-shape switch task the same participants were cued to switch between categorizing shapes based on either their shape or color. In both cases, switch costs were measured as the difference between response latency and accuracy on switch vs. stay trials and the relationship between these costs was taken to represent the degree that these types of tasks involve shared processes. Though also prone to many of the pitfalls of non-experimental designs, correlational studies that assess task performance *within* participants can help to mitigate some of the major issues encountered when studying group differences. Importantly, correlational studies do not typically rely on a user-defined grouping to place participants into their respective group (e.g. monolingual vs. bilingual, high proficiency vs. low proficiency), which is often arbitrarily defined. In this case, Calabria and colleagues (2012) did not find a relationship between language and task switching costs, and this is in fact in line with the majority of studies measuring this relationship (Branzi, Calabria, Boscarino, & Costa, 2016; Calabria, Hernandez, Branzi & Costa. 2012; Calabria Branzi, Marne, Herandez & Costa, 2013; Jylkkä, Lehtonen, Lindholm, Kuusakoski, & Laine, 2017; Klecha, 2013; Prior & Gollan, 2012). More recently however, Declerck, Grangier, Koch, and Philipp (2017) were able to find this relationship under very specific circumstances. They used a very controlled paradigm to ensure that the tasks and stimuli for both linguistic and non-linguistic tasks were as similar as possible. Through a series of experiments they modified the cues, stimuli and response modality to be the same in both tasks, finding that the more overlap there was between tasks the more highly correlated they were. As such, they suggest that the previous failure to find a relationship between task and language switching may be due to methodological differences rather than a lack of shared processes. On the other hand, they address the fact that task switching and language switching did differ when using more distinct tasks which indicates that these types of control do not entirely overlap, but rather that their findings support that domain general mechanisms can be used to *support* language processing. These findings bring to light some of the challenges that plague these kinds of correlational studies- where the goal is to find evidence for transfer of skills across linguistic and non-linguistic domains while maintaining sufficient similarity at the level of the proposed shared underlying processes or mechanism. That is, in order to make generalizable claims about underlying shared *processes* between two tasks it is helpful to keep differences between the superficial components of the tasks to best pinpoint any relationship to the desired source (i.e. domain general control) rather than, e.g. speed of categorization.

Other correlational studies have assessed the relationship between language switching and more general measures of domain general control including the flanker (Gollan, Sandoval, & Salmon, 2011) and Simon tasks (de Bruin et al., 2014; Linck, et al., 2011) with mixed results, likely due to differences in task, population, and/or underpowered experiments. Based on the current research, it is not clear that these correlational findings lend much support to the theoretical overlap between domain general and language switching control. To obtain a better picture of the findings, the results from these studies will be compiled and reevaluated in a systematic manner as the first step of the current study- the process results will be reviewed in the following table (see Table 2 for a summary of the findings).

Dual-task Assessment of Overlap Between Language and Domain-General Control

As has been discussed, despite a fair amount of research assuming a role of domain general inhibitory control in language switching, support for this theory is mainly based on switch cost asymmetries, correlations between tasks, and group differences between monolingual and bilingual speakers, and not more experimental methods. Freund, Nozari, and Gordan (2016) point out moreover, that although advocates of domain general language control posit that conflict can be monitored both in language specific and in domain general systems, these claims do not typically commit to a *single* system responsible for dealing with this conflict. That is, despite a general consensus in the literature that language involves a *similar* type of conflict as we see in non-linguistic events, there is not compelling evidence that the processes are the *same*. Recently, a series of experiments using dual-task methods to tax domain general inhibitory control during language switching aimed to more directly test if the processes are in fact shared between linguistic and non-linguistic control. None of the results showed the expected interaction (based on the Inhibitory Control model) between linguistic and non-linguistic control tasks (Shell, 2015). More specifically, in these tasks participants were required to respond to a Simon arrow task (responding to the arrowhead direction while ignoring arrow location) while naming pictured items in the cued language in a language switching task. It was predicted that if switching required the same type of control as the arrow task, performing a language switch while dealing with interference from the Simon task would result in an over-additive interaction between tasks. That is switching would be harder (reflected by larger switch costs) when inhibitory control was taxed via incongruent arrow trials. The results, however did not show support for this interaction and so did not lend evidence to the Inhibitory Control Model.

This failure to find an over-additive interaction between language switching and demands on non-linguistic inhibition is not limited to the Simon arrow task, but has also emerged when secondary tasks manipulated semantic competition. Shell, Linck and Slevc (2015) found no interaction between interference in picture naming induced by semantically related distractors (i.e., the picture word interference (PWI) paradigm) and interference induced by switching languages. Similarly, Runnqvist, Strijkers, Alario, and Costa (2012) did not find evidence for an overlap between semantic- level inhibition and that used for language switching. They used a blocked semantic naming paradigm with language switches, and found that semantic competition was maintained across switch trials. As blocked semantic naming tasks are thought to create competition by building up interference through semantically related trials, if language switching does require inhibition of items in the non-target language, a language switch should resolve this competition. Runnqvist and colleagues (2012) however, found that semantic interference was maintained across language switches, again failing to support the inhibitory control theory.

Most of the research directly assessing inhibitory control thus far has failed to find evidence to support the Inhibitory Control model. However, most studies thus far have also not distinguished between the *levels* of control (as proposed by Green, 1998): a task-schema level responsible for biasing activation at the language level, and an item-level responsible for inhibiting incorrect lexical items. While it is yet undefined exactly what *kind* of control is relevant in language switching, or exactly how domain general control could deal with something like lexical representations, recent dual-task research aiming to assess if domain general control is involved may have overlooked this distinction. That is, if in fact domain general control is only relevant at one of the two proposed levels of language control, these dual-task designs aimed to specifically tax language control, may have targeted the wrong level, leading to the null findings.

The goal of the current study, therefore, is to reassess past findings to motivate a distinction between two types of control in bilingual selection (task-level and item-level control), and then to test some predictions of this distinction. The following section reports a systematic re-analysis of relevant past work (both my own data and from the literature) using Bayesian methods to assess the strength of evidence *for or against* a relationship between domain-general inhibitory control and language switching. That is, Bayesian analyses were used to determine the evidence for or against an interaction between these tasks in dual task studies and to assess correlational data that has looked for a relationship between domain general control and bilingual language switching. In addition to evaluating the strength of past evidence for (or against) this relationship, I aim to better locate when and where this relationship is found, in order to draw broader inferences about the levels of control involved in language switching.

Re-evaluating Previous Work

To directly assess the role of domain-general control mechanisms in bilingual language use, Shell (2015) used a dual task paradigm to tax domain general inhibitory control during a language switching task using a 2x2x2 design crossing language (L1/L2), switch (switch/stay), and domain-general conflict (high/low) (see Figure 4 for a schematic of a single trial). Through a series of three experiments, the expected over-additive interaction between switch and conflict, which would suggest shared resources between the two tasks, was not found. In fact, in some cases, analyses showed evidence for an effect in the opposite direction (i.e., a *reduced* language switching cost on incongruent arrow trials). As such, these null findings showed no evidence for the predicted role of domain general inhibitory control in language switching.

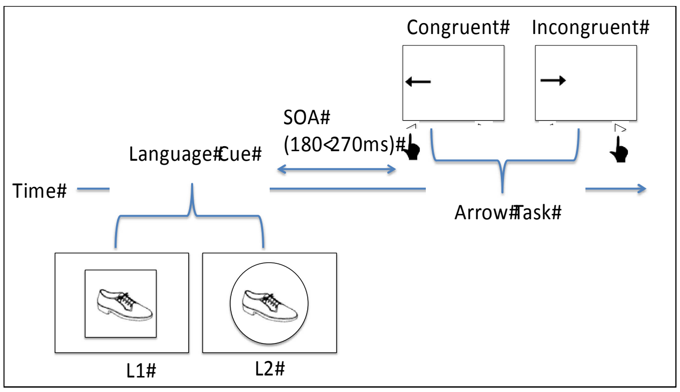

Figure 4. Schematic of trial from Experiment 1 (Shell, 2015), combining predictable language switching and Simon arrow task.

However, while the analyses for the above experiments were unable to find support for the predicted over-additive interaction, the null hypothesis significance testing methods used for statistical analysis cannot provide evidence for the alternative hypothesis that there was *not* an interaction between tasks. Bayesian approaches, unlike null hypothesis significance testing, however are able to quantify the *degree* of evidence supporting the hypotheses (including the null hypothesis) given the data and so can specify the likelihood of different possible models to represent the data. As such, re-evaluation of past work using Bayesian analyses can help determine the strength of evidence for or against a model that includes the interaction.

All Bayesian analyses were conducted using the BayesFactor package (version 0.9.12-2; Morey, Rouder, & Jamil, 2015) for the statistical software R (version 3.3.3; R Core Team, 2017). Re-analysis of reaction time data from Shell (2015) Experiments 1-3 was conducted to evaluate the fit of the interaction model (Switch * Conflict condition) using mixed factor Bayesian ANOVA: Switch condition (switch/stay), Conflict condition (high/low) and Language (L1/L2) were included as fixed effects, and Subject and Item were entered as random effects to assess the model fit.[1] In addition, in order to assess evidence supporting the directional hypothesis of an over-additive interaction, (predicted increased switch costs during incongruent arrow trials) directional Bayesian t-tests were performed on the per-subject switch costs (mean RT stay-mean RT switch). In these analyses the Bayes Factor (BF) refers to the odds of the hypothesis model relative to the odds of the null model (which includes the random effects). A Bayes factor greater than one indicates the odds of the tested model is greater than the odds of the null, while a Bayes Factor less than one is in favor the null, with odds close to one being weak, or anecdotal evidence. Wagenmakers, Wetzels, Borsboom, and van der Maas, (2011) define a BF of greater than 3 as indicating moderate evidence for the tested model while a BF of less than .33 indicates moderate evidence for the null hypothesis.

As shown in Table 1, Bayesian analyses do not support any interaction- either finding evidence for no interaction whatsoever or failing to have enough evidence to either support or deny an interaction (i.e. anecdotal evidence). Importantly, directional t-tests, which tested if switch costs were larger in the high conflict, compared to the low conflict condition, show substantial support for the null model- that switch costs were *not* larger during high conflict trials than low (see Table 1 below for Bayes Factors). To summarize, the analyses suggest that the failure to find support for an interaction between domain general conflict and language switching costs in previous work can be interpreted as evidence *against* such interactions, and thus do not

[1] Sample R syntax for the mixed effects Bayesian ANOVA with Subject and Item included as random effects: anovaBF(logTime~SwitchCond*ConflictCond*Language+Subject+Item, data = data, whichRandom = c("Subject", "Item").

support the prediction for a relationship between domain general control tasks and language switching in general.

Table 1

Bayes Factors for Reanalyzed Experiments 1-3 (Shell, 2015)

Experiment	**Interaction model**	**Directional t-test**
Exp1 (Pics)	0.404 ()	0.002 (--)
Exp2 DCA (Digits)	0.0003 (--)	0.083 (-)
Exp3 (PWI)	0.61 ()	0.073 (-)

Note: Signs in parentheses represent strength of support for the model: () = anecdotal, substantial (-) strong (--) evidence in favor of the null.

To evaluate the evidence for or against a relationship between language switching and domain general control from correlational findings in the current literature, Bayesian correlational analyses were conducted on all published studies that included measures of both language switching and a non-linguistic inhibitory control tasks for each participant and provided sufficient data for this analysis. When raw data was available, Bayesian analyses were conducted on switch costs and domain general control effects (e.g. incongruent - congruent response time) using the regressionBF function. When raw data were not available, estimated Bayes factors were calculated using the linearReg.R2stat function from the BayesFactor package (Morey Rouder, & Jamil, 2015) which calculates an estimated Bayes factor based on the reported r-statistic and number of participants (N) for all research that reported these values (see Table 2 for results).

Table 2

Bayes factors for correlation between inhibitory control task (IC Task) and language switching (switch BF), language mixing (mix BF), and switch costs per language (L1, L2, L3).

Study Citation	N	IC Task	Switch BF	L1 Switch	L2 Switch	L3 Switch	Mix BF
Branzi et al., 2016	62	Switch	0.098 (-)				
Calabria et al., 2012	14	Switch	0.262 ( )	0.286	0.283 ( )		
Calabria et al., 2013	60	Switch	0.1 (-)				
DeClerk et al., 2017 Exp 1	62	Switch	0.941				
DeClerk et al., 2017 Exp 2	62	Switch	4.065				
DeClerk et al., 2017 Exp 3	62	Switch	11.857				
Klecha 2013	22	Switch	0.525 ( )		1.377 ( )		1.87 ( )
Prior & Gollan, 2012	104	Switch	0.063 (--)				2214.97 (+++)
deBruin 2014	24	Simon		1.526 ( )	6.139 (+)	0.811 ( )	
Linck et al., 2011	56	Simon	0.787 ( )	3.46 (+)	0.294(-)	0.376 ( )	
Shell, 2015, Exp. 1	37	Simon	0.42( )	0.33( )	0.39( )		
Shell, 2015, Exp. 2	37	Simon	0.39( )	0.32( )	0.31( )		

*Note: * Indicates estimated BF. Symbols in parentheses represent support for the model: Anecdotal (), Substantial (-), Strong (--) evidence for null. Substantial (+), Strong (++) evidence for alternative*

The findings from the above analyses do not appear to be consistent or lend support to a clear conclusion. Most of the support is anecdotal, and the limited findings that provide more robust evidence show evidence both for and against a relationship. The strongest evidence for any relationship is found not in language *switch* costs, but rather in language mix costs (Jyttkla, et al., 2017; Prior & Gollan, 2012; Soveri Rodriguez-Fornells, & Laine, 2011). Language mix costs measure the cost of being in a mixed language context (i.e. language switching) compared to a single language context. This distinction between switch and mix costs will be further examined here as a means of measuring differing types of control in language switching.

Proactive vs. Reactive Inhibitory Control

Additionally, while none of the correlational studies above found evidence for a relationship between overall switch costs and domain general control, there was some evidence for this relationship when switching costs were measured separately for each language (de Bruin et al., 2014; Linck et al., 2011). The results from these studies, however do not correspond;

while Linck et al., (2011) found evidence for a positive relationship between Simon costs and language switch costs when switching into the dominant L1 and a negative relationship when switching into the L2, de Bruin et al., (2014) did not find a relationship for switches into the L1 but found a positive relationship when switching into the L2. In both cases, these findings are argued to reflect reliance on inhibitory control; either to suppress the L1 when switching into a less dominant language (de Bruin et al., 2014) or to re-engage L1 after being temporarily deactivated (Linck et al., 2011). Though the results both suggest domain general control in language switching, they clearly contradict each other in the specifics of the relationship. Of course, as is often the case in the literature, key differences between study design could drive discrepancies in the results. For instance, one major difference was the proportion of switch trials during the switch tasks- with 66% of trials being switches in the de Bruin and colleague's (2014) paradigm but only 33% in the study by Linck and colleagues (2011). This switching context is likely to have an influence on a participant's expectations and *preparation* for a switch trial. That is, in a high switch context, a participant may increase inhibitory control in advance or more globally rather than rely on a more localized inhibition. Accordingly, while the original Inhibitory Control Model (Green, 1998) suggested that the control required for language switching is *reactive,* whereby words in both languages are activated and then reactive control is used to inhibit the non-target words, a more recently revised Inhibitory Control Model (Green & Abutalebi, 2013) suggests that *proactive* control may also be used adaptively to bias the activation levels of the competing languages depending on the language context.

A similar theory has been suggested outside of the language switching domain. Braver's (2012) *dual mechanisms of control* framework suggests two levels of control: 1) proactive control which biases resources and prevents potential interference before it occurs and 2) reactive control which resolves interference after its detected (see also Rogers & Monsell, 1995 for a related theory). Accordingly, the differences in switch context (ratio of switch to stay trials) might bias reliance on proactive or reactive control, with proactive control being advantageous when switch trials are more frequent. Along these lines, it has been suggested that unbalanced bilinguals likely exert a general level of proactive control on their dominant L1 when in a mixed language context in preparation to produce the less dominant L2 (e.g., Yang, Ye, Wang, Zhou, & Wu, 2018). As such these different forms of control may be more or less relevant in different contexts and so may relate differentially to different measures of control.

While the focus of most of the work discussed above has been on experimentally testing predictions of the reactive Inhibitory Control Model (Green, 1998), a more recent version of that model based on the adaptive control hypothesis suggests multiple layers of control involved in language switching (e.g., Abutalebi & Green, 2013; cf. Braver, 2012) which should be considered when evaluating language control. It may be possible to dissociate between proactive and reactive control in language switching tasks by contrasting two different measures of language control in switching paradigms: *switch costs,* measured as the increased latency to respond on a switch trial compared to a stay trial in the same block, and *mix costs*, measured by comparing stay trials in a mixed language block to trials in a pure language block. Mix costs presumably reflect the need to monitor the uncertain environment, which can be considered a form of proactive control, while switch costs reflect the cost of dynamically responding to a cue; that is, reactive control. It may be that previous evidence for a relationship between mixing costs and domain-general inhibitory control costs (Klecha, 2013; Prior & Gollan, 2012) reflects shared proactive, rather than reactive control mechanisms. It could therefore be the case that a relationship between domain general control and language control is only found in contexts or using measures of proactive control (e.g., high conflict contexts, or mixing costs).

Language switching studies often find that switch and mix costs are differentially affected by manipulations such as cue-stimulus timing (e.g., Ma, Li, & Guo, 2016), and are differentially correlated with individual difference measures (e.g., Prior & MacWhinney, 2010; Prior & Gollan, 2011). These cost measures may therefore be capturing distinct components of control and might speak to the debate about a general processing advantage for bilinguals in in high conflict contexts (Costa, 2009; Hilchey & Klein, 2011). Evaluating the effect of taxing domain general inhibitory control on both switch and mix costs in the current study will allow for better specification of the type of control required in language switching. Additionally, assessing mixing costs will be important in correlational analyses between domain general control tasks and language switching to capture the potential shared role of proactive control in these tasks.

Dual Levels of Control

As discussed above, the original Inhibitory Control Model (Green, 1998) proposed a two-level system of language control, with control both at the task schema level (i.e., the language) and at the lemma, or item level (i.e., the specific lexical items) to inhibit all non-target items. In many language switching studies, paradigms are not designed to properly distinguish behavior at these two levels. That is to say, most switching paradigms require participants to switch languages according to a language cue which is presented simultaneously with the item, allowing both language and item responses to be prepared at the same time. Because these responses can be prepared, and executed, simultaneously in these designs it is difficult to discriminate between control requirements at each level of interference, if there are differences.

A few studies have tried to distinguish between levels of language control by separating the language cue from the stimulus, either by manipulating the timing of the stimulus and language cue, or manipulating the predictability of the task, which allows for planning the response in advance of a cue (e.g., Chang, Xie, Li, Wang, & Liu, 2016; Declerck & Philipp, 2015; Verhoef Roelofs & Chwilla, 2010; Liu et al., 2018). Based on these findings, there is some evidence that processes relevant to task switching may occur at the item selection stage. For example, Chang and colleagues (2016) used a cued language switching design where the stimulus and cue were presented separately, including both trials where the cue was presented before the stimulus, and trials with the stimulus presented before the cue. Participants performed a language switching task -naming digits in either their L1 or L2 based on the language cue and behavioral and neural responses were recorded. The amplitude of neural activity in response to events was measured using event related potentials (ERPs), which were time-locked to either the cue or the stimulus to determine when switch-related activity occurred. If participants use proactive control at the task-schema (language) level, switching-related activity would be predicted to occur at the language cue onset, while if participants control language at the item level, switching activity would be predicted at the presentation of the stimulus. ERPs time locked to the stimuli (and so ostensibly measuring the neural response when selecting the language at the item level of control), but not the language cues, were different for switch than for stay trials. That is, there was evidence for switch costs at the item level, but not at the language level. These findings begin to add some preliminary support for Green's (1998) theoretical dual-level model, with evidence for a distinction between these levels of control, and further may suggest that switching activity mainly occurs at the item level not the language level. On the other hand there is also evidence that a language switch can be prepared in advance of the stimulus, which supports a role of control for language switching at the language level (e.g. Verhoef et al., 2010). In line with this, Wu and Thierry (2017) used ERP measures to look at the neural activation that occurs in preparation for language switching and found a difference in activity following the cue for L2 compared to L1 (greater activity following the L2 cue). Because it is generally thought that different amounts of control are needed for dealing with the L1 vs. the L2, the difference in activity at the point of the language cue suggests proactive language control may occur before stimulus onset. While these findings have different conclusions, they all support a distinction between control at the language level and the item level, however there is not yet sufficient research to clarify if the type of control at each level is a domain general control or one specific to language. As such one of the goals of Experiment 2 is to assess the effects of domain general control at each level during a language switch by manipulating participants' ability to prepare for the language in advance of the stimulus.

In the previous dual-task experiments described above (Shell, 2015), language switches were predictable (presented using the "alternating runs" method), where the language to be used on the upcoming trial was always known in advance. This predictability actually removed the need for a task (language) cue at all (other than as a reminder) and participants could ostensibly prepare for each task in advance of the task cue. Thus, even if domain-general inhibitory control *is* used to switch between task sets or languages, in this paradigm it could be that this control was initiated before the stimulus (and secondary task) appeared, thus decoupling the dual-tasks intended to coincide to assess their interactive effect. That a language switch can be prepared in advance of the stimulus has been shown in some of the EEG work described above as well as in studies finding that switch costs still occur in trials following a switch *cue* even when there is no actual need to respond (e.g. Lenartowicz, Yeung, & Cohen, 2011; Wu & Thierry, 2017). This suggests that switch costs may in part reflect proactive adjustments at the task level that can occur in advance- or even in the absence of- the stimulus. Specifically, previous dual task investigations of language switching (Shell, 2015) have timed the secondary task to appear at stimulus onset (or shortly thereafter) in order to tax control during the language switch. However, if a language switch can be prepared in *advance* of the stimulus, then the secondary task may not have an opportunity to interfere with the switch as intended. The timing of the secondary task in these studies, however, did still occur at the point when an individual lexical item had to be selected. Thus, in these previous designs, the dual task may *not* have interfered directly at the language task level, which could explain the failure to find an interaction with the language switch, but presumably still could have interfered with lexical selection at the item level. As there was no interaction between the two tasks, one might interpret the null findings to mean that domain general control is *not* involved at the item selection stage (see, e.g. Alario et al., 2012; Freund et al., 2016), yet it still might be involved at the language selection stage.

Selection of the appropriate language in a language switch task can be likened to the processes of task selection in the classic task switching paradigm. That is, language selection can be considered a higher order goal, or task (e.g. ‘speak in English’ or ‘report the color of the item’) , with selection of a particular item within that language below it (e.g. ‘cat’ vs. ‘red’). As such, it is feasible that domain general processes that manage task selection, (e.g., maintenance and biasing of goals) may be involved in this broad level goal shift. Models of cognitive control (e.g., Miller & Cohen, 2001) suggest a domain general process for helping to represent and maintain goals and the means to achieve them, as well as for providing important signals that can bias response execution, particularly in the face of competition. This process is seen as domain general in that it is responsible for general goal maintenance and task selection across a range of tasks, including for example, a Simon arrow task and a language selection task. On the other hand, while there is some evidence that lemma, or item selection may require domain general cognitive control, for example item selection under high conflict (Novick, et al., 2013) it is not clear that the same process used to bias attention to allow for selection at a higher level is also used to inhibit specific competitors at the item level. That is, domain general cognitive control may be useful to bias activation and allow for eventual item selection without acting directly on activation or inhibition at the item level.

To investigate this possibility, the current study manipulated demands on task level (language) versus item level control during language switching in production. By the account described here, if domain general control mechanisms apply to language-level processes, but not to item level processes, interactions with simultaneous demands on domain general inhibitory control (via a dual task) are not expected when task level control demands are relatively low (Exp 1; cf. Shell, 2015), but are expected when task-level demands are high (Exp 2). These experiments also included separate measures of language switch and language mix costs, which likely reflect similar but distinct components of control that may be employed differentially at the different levels of language control. Additionally, in order to better capture the process of the language switch task, which I have suggested might involve several stages, I include an additional method for analyzing task performance, drift diffusion, that may better reflect these stages (c.f. Schmitz & Voss, 2012). Drift diffusion modeling takes both reaction time and accuracy of task performance to decompose responses into separate measures of non-decision time (encoding), drift rate, or decision time (planning and response selection) and response criterion (a threshold for making a response based on risk or confidence). Schmitz and Voss (2012) used drift diffusion modeling to decompose task switching into preparation and response selection stages. By manipulating different aspects of the task switching task, they found that the nondecision parameter increased in trials where task preparation was not possible, suggesting it reflects task set preparation. Drift rate, reflecting speed of evidence accumulation, was shown to be faster in predictable and stay trials, as well as for participants with higher levels of fluid intelligence. Lastly, the response criterion, reflecting a response caution threshold, was adjusted to be more stringent in task-switch vs. task-pure blocks. These measures might differentially show the effect of the paired arrow task on language switching. For example, response criterion should be reflected in measures that compare single and dual-language contexts (i.e. mix costs).

To determine how individual differences in experience with language control might drive differences in its relationship to domain general control, I assessed how individual differences in measures of language switching (and language mixing) relate to individual differences in language use and proficiency, inhibitory control abilities (assessed with a non-linguistic inhibitory control task), and specific measures from the AX-CPT task (Braver, 2012) that distinguish reliance on proactive and reactive control. In addition to measuring switching ability using the classic latency analysis of switch costs, rate residualized switch costs, which incorporate both latency and accuracy as the *rate* of accurate responses, were calculated (cf., Hughes et al., 2014). By including both response time and accuracy in the same measurement this index of switch costs can help account for the speed-accuracy trade off that often is ignored in behavioral response time data, and therefore may provide a more valid assessment of an individual's switching ability. By including a range of tasks, incorporating accuracy into response measures, and testing a larger sample than in most previous work, the goal is to determine more specifically which meaningful relationships do exist between switch and mixing costs and measures of non-linguistic control.

In order to increase openness, reproducibility, and valid interpretations of the data (cf. Munafò et al., 2017; Wagenmakers, Wetzels, Borsboom, Van der Mass, & Kievit, 2012), the details of the methods and analysis for Experiment 2 in this study were pre-registered (before the start of data collection) with AsPredicted (https://aspredicted.org/): an online repository where details of the study, hypotheses, and analyses are stored and made available publicly to ensure conclusions are based on a priori decisions and hypotheses. A non-editable registration form is publically available at https://aspredicted.org/sd4pu.pdf. All decisions made in the methods and analysis were made in accordance with the pre-registration except where noted.

Chapter 2: Experiment 1: Predictable Switching

Method

Experiment 1 served as a conceptual replication to the dual-task switching paradigm from Shell (2015), using a modified participant group (speakers with a broader range of language experience) and stimulus set (Arabic digits) but keeping the underlying design and method consistent. As such, the paradigm required speakers to simultaneously perform a language switching task and the arrow version of the Simon task- a secondary task thought to involve non-linguistic inhibitory control. As in Shell (2015), the purpose of pairing these tasks was to increase demand on domain general inhibitory control during an incongruent arrow trial in the Simon task to determine if the mechanisms involved in this control are shared with those involved in language switching (see Figure 5 for a schematic).

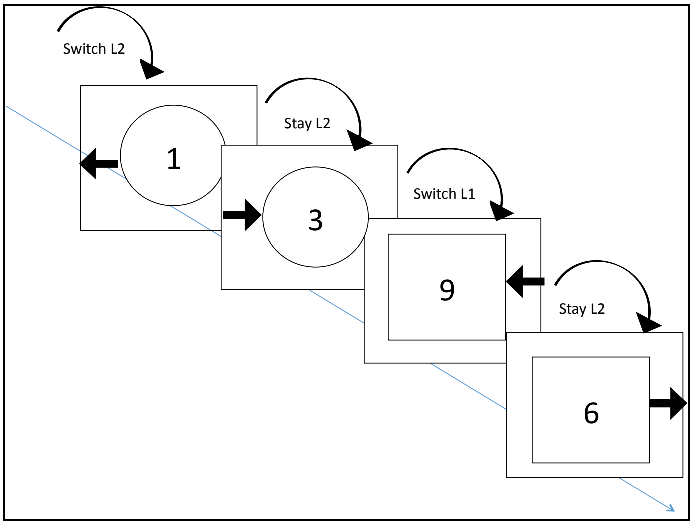

Figure 5. Schematic of a series of four trials in the predictable combined language switching and Simon arrow block from Experiment 1.

To allow for a broad range of bilingual experience (providing a better ability to correlate differences in language experience and performance on our tasks) as well as to increase the size of our participant pool, the following experiments recruited individuals with various degrees of L2 proficiency from the general University subject pool, rather than specially recruiting language students with a restricted range of L2 proficiency. Accordingly, a stimulus set of single-digit numbers was selected with the assumption that even low proficiency speakers are comfortable naming numbers, and so would be able to participate in the task. While there is some research to suggest that digit naming may differ somewhat from item naming (Declerck, Koch, & Philipp, 2012), the general findings from the original paradigm (Shell, 2015) were predicted to remain. That is, I predicted to replicate the *lack* of an over-additive interaction between language switch and arrow congruency (i.e., while there is a predicted main effect of both language switching and of arrow congruency in the Simon arrow task, there is no predicted interaction between the two tasks). As discussed above, this result may indicate that the predictability of the language switch allows speakers to perform a language switch (at the ‘task’ level) in advance of the arrow task.

That is, this task (like those tasks in Shell, 2015), yokes non-linguistic demands on inhibitory control only with *item* selection, and not with *language/task* selection. This experiment extends those in Shell (2015) by replicating the general pattern of results and extending this pattern to a broader range of speakers. In addition, this experiment serves as a well-matched comparison for Experiment 2, which uses the same stimuli and participant population.

Participants

Forty-two participants were recruited from the Psychology subject pool with the prerequisite that participants must be native English speakers and be able to count to 10 in their second language; either Spanish (31), French (3), Mandarin (3), or Hebrew (5) (see Table 3 for self-rated language proficiency summary). All participants were compensated with extra credit for their Psychology course.

Table 3
Experiment 1: Averaged Self-rated L2 Proficiency (Scale 0-10) and Percent Time Exposed to the L2 (from the modified LEAP-Q; Marian et al., 2007).

Proficiency L2	Speaking	Understanding	Writing	% Exposed
Self-rated (0-10)	5.73	6.13	5.90	9.72

Materials and Procedures

After brief instruction and consent, participants completed a series of blocks of trials, starting off with counting and digit naming practice, followed by language switching, the Simon arrow task and then the combined language switch and arrow block. All sessions ended with participants filling out an abbreviated language experience (LEAP-Q: Marian, Blumfield, & Kaushanskaya, 2007) survey; a self report measure which included questions about languages known, order of dominance, order of acquisition, percent time exposed to each, and proficiency (scale 0-10).

Digit practice. Participants practiced naming numbers 0-9 in English and then in their L2. L2 practice began with naming digits in counting order to ensure comfort with the items before then naming in a random order.

Switching Practice. Participants practiced switching between naming items in their L1 and L2 in a predictable alternating runs design (L1, L1, L2, L2). A language cue appeared with the onset of the digit, each digit appearing in the center of a cue (square for L1, circle for L2). There were 80 trials in the switching block, items counterbalanced by switch condition (switch, stay) and language (L1, L2). Trials were presented in a pseudo randomized fixed order to minimize repetition.

Simon Arrow Task. Participants then performed the arrow variant of the Simon task (Simon & Rudell, 1967). Following a centrally located fixation cross, which appeared for 750 ms, an arrow symbol pointing to the left or right appeared on either the left or right side of the computer screen. Participants were told to respond with a button press to the direction the arrow was pointing (m for right pointing and z for left pointing) while ignoring the location of the arrow. The arrow remained on the screen until a key response. Auditory feedback (beep "basso") followed an incorrect response to encourage accuracy. There were a total of 80 trials with an equal number of left and right arrows appearing equally on the left and right side of the screen, resulting in 1/2 of trials being congruent (arrow direction matched location) and 1/2 being incongruent (arrow direction did not match location).

Language Switching and Arrow Task. Finally, participants were instructed to perform the language switching task at the same time as the arrow task – with the arrow appearing 200ms after the onset of the cue and stimulus onset, chosen so that the arrow would likely appear during lemma selection, based on estimates of the timing of word production (Indefrey & Levelt, 2004). Thus, the conflict from the arrow was timed to tax inhibitory resources at the predicted time point of lexical conflict in the language switching task. They were told to not prioritize either task and to perform each one as quickly and accurately as possible. There were four blocks of 80 trials each, all items appearing in a pseudo-randomized order to avoid repetition of trial type, with a self-timed break between blocks. All items were counterbalanced by language, switch type and arrow congruency (See Figure 5 for a schematic of a series of trials).

Analysis and Results

Analysis for all experimental effects was conducted both using frequentist ANOVA and Bayesian analysis.[2] Voice key response times (RTs) were log transformed and trimmed per subject (see below for details) and analyzed using both mixed effects models with the lme4 package (version 1.1.13; Bates, Maechler, Bolker & Walker, 2015) and mixed factor Bayesian ANOVAs using the BayesFactor package (version 9.12.2; Morey, et al., 2015) for the statistical software R (version 3.3.3; R Core Team, 2017). For all mixed effects analyses of response time data reported in the following two experiments, the fully specified model would not converge, therefore all random effect terms were included but were uncorrelated. P-values were calculated using the lmerTest package (lmerTest version 1.0; Kuznetsova, Brockhoff & Christensen, 2017), which estimates p-values and degrees of freedom from the mixed effects model t-test based on Satterthwaite approximation for denominator degrees of freedom.

[2] Pre-registration for Experiment 2 specified use of Bayes analysis and so all conclusions in the following studies are based on Bayesian analysis results. Frequentist results are reported here for comparison with previous studies.

Bayesian analyses included all fixed effects and interactions as well as random effects. Bayes Factors were calculated using the default prior distribution of 0.5, considered a medium wide distribution, which specifies that 50% of the effect sizes fall within the range -.707 and .707 (square root of 0.5), selected to be impartial about the expected effect size. The Bayes Factor for model fit is adjusted for the complexity of the model and model fit is assessed by model comparison, comparing each model to the null model (which include only the random effects Subject and Item). For each analysis the top six best fit models are reported and model comparisons were conducted between the models with and without the primary effects of theoretical interest to determine the degree of evidence for those effects (because these models share the same null denominator, they were simply input as a ratio to compare evidence for one over the other). The results of these comparisons are used to determine if adding the predictor to the model adds sufficient explanatory power to overcome the Bayesian penalties for increasing model complexity. While Bayes factors qualify evidence for or against a model on a continuous scale, in order to clarify the results for reporting and interpretation of model fit, I provide a well-accepted method of Bayes Factor classification in Table 4 (Wagenmakers,Wetzels, Borsboom, & van der Maas, 2011; cf. Jeffreys, 1961).

Table 4
Classification for Bayes Factors (adapted from Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011; cf. Jeffreys, 1961).

Bayes Factor (BF)	Interpretation
>100	Extreme evidence for model
30-100	Very strong evidence for model
10-30	Strong evidence for model
3-10	Substantial evidence for model
1-3	Anecdotal evidence for model
1	No evidence for model
1/3-1	Anecdotal evidence against model
1/10-1	Substantial evidence against model
1/30-1/10	Strong against model
1/100-1/30	Very strong evidence against model
<1/100	Extreme evidence against model

Response Time Analysis

For the latency analyses, all trials with errors in the verbal response (342, 2.6%) or the arrow response (309, 2.3%) and all voice key detection errors (i.e. when a trial was skipped or not appropriately detected due to the microphone calibration or extraneous noises; 183, 1.4%) were removed from the reaction time data analysis[3]. Following this, RTs greater than three standard deviations from each subject's mean (198, 1.6%) were removed from analysis, however analyses were conducted on both trimmed and untrimmed data to ensure that the trimming did not change results (see Appendix A for untrimmed Bayesian results). In total these trimming criteria led to the removal of 815 (6.1% of all) trials.

As can be seen in Figure 6, participants showed robust switch costs, but these costs were not larger in the incongruent than congruent condition as predicted by the hypothesis of shared inhibitory resources used for language switching and the incongruent arrow trials (see Table 5

[3] Note that this differs slightly from the exclusion criteria in Experiment 2, where I also removed all trials following a naming response in the wrong language.

for switch cost descriptives). These observations were supported both by frequentist and Bayesian analyses.

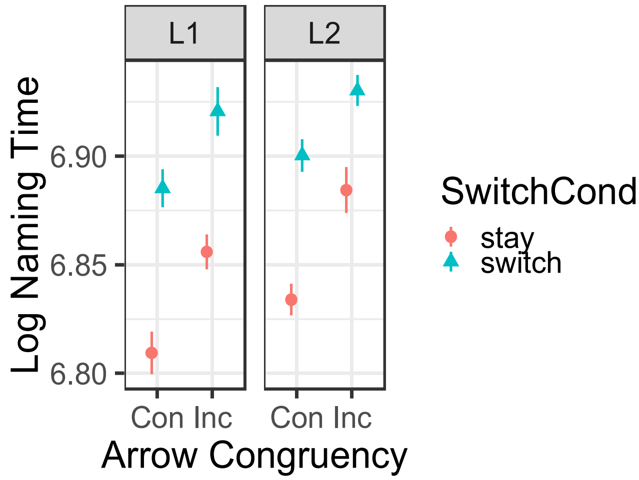

Figure 6. Experiment 1: Naming reaction time (log ms) by arrow congruency condition, switch condition and language. Plotted data are means of subject means (note that analyses were conducted over non-averaged log-transformed data). Error bars indicate standard error of the mean.

Table 5
Experiment 1 Response Time (RT) in Milliseconds for Switch, Stay, and Cost (Switch-Stay RT) by Language and Congruency

Language	Congruency	Stay RT	Switch RT	Switch Cost
L1	Congruent	949.81	1027.70	77.89
L1	Incongruent	990.30	1060.17	69.87
L2	Congruent	977.36	1051.93	74.57
L2	Incongruent	1036.50	1079.63	43.14

Response times were analyzed using mixed effects modeling with Switch Condition, Congruency, and Language as fixed effects and Subject and Item as random effects (see Table 6 for model syntax and analysis results). There was a significant main effect of Switch Condition, with longer RTs for switch than stay trials, and of the Congruency Condition, with longer RTs during incongruent than congruent arrow trials (see Table 4 for summarized RT results). There was no main effect for language or any interactions between conditions.

Table 6
Experiment 1 Switch Cost Mixed Effects Results

Naming Time Fixed Effects	Estimate	Std Error	df	t value	Pr(>|t|)
Intercept	6.8774	0.0323	45.27	212.78	<.001
Switch Cond	0.0632	0.0095	15.27	6.627	<.001
Congruency Cond	0.0407	0.0058	15.35	7.034	<.001
Lang	0.0190	0.0151	28.73	1.254	0.220
SwitchCond *Congruency	-0.0169	0.0105	9.04	-1.606	0.143
Switch Cond *Lang	-0.0124	0.0147	12.96	-0.846	0.413
Congruency *Lang	0.0001	0.0142	10.29	0.010	0.992
SwitchCond*Congruency*Lang	-0.0067	0.0321	9.020	-0.208	0.840

Note. Model syntax: lmer(logVKRT ~ switchCond*ConCond*Lang + (switchCond*ConCond*Lang||Subject) + (switchCond*ConCond*Lang||Item), data=acc.trimmed)

For the Bayes latency analysis, Switch condition, Congruency and Language were included as fixed effects, and Subject and Item were entered as random effects to assess model fit. The best fitting model included all main effects- that is, Congruency, Switch Condition, and Language, as well as an interaction of Language*Congruency and Language*Switch (See Table 7 for the top six models that fit the data). The best fitting model did not include the hypothesized Switch*Congruency interaction, however the next best fit model did include this interaction. To determine the relative evidence for the best fit model over the model with the added hypothesized Switch*Congruency interaction, I divided the Bayes factors from the each model for model comparison and found anecdotal to substantial evidence against the model with the interaction (BF = .280). These data therefore provide no support for a language by congruency interaction, but also give no convincing evidence against an effect of the arrow task on language switching.

Table 7

Experiment 1 Switch RTs Top 6 Best Fit Bayes Models

Model Fit	Bayes Factor +/- error
Congruency + SwitchCond + Language + Subject + Item	3.899e+74 ±3.58%
Congruency + SwitchCond + Congruency*SwitchCond + Language + Subject + Item	1.078e+74 ±4.53%
Congruency + SwitchCond + Language + SwitchCond*Language + Subject + Item	3.646e+73 ±2.29%
Congruency + SwitchCond + Congruency*SwitchCond + Language + SwitchCond*Language + Subject + Item	1.046e+73 ±4.89%
Congruency + SwitchCond + Language + Congruency*Language + Subject + Item	5.332e+72 ±27.36%
Congruency+ SwitchCond + Congruency*SwitchCond + Language + Congruency*Language + Subject + Item	3.396e+72 ±9.22%

Note: Model syntax: anovaBF(logVKRT ~ switchCond + ConCond + Lang+ switchCond*ConCond*Lang + Subject+Item, data = acc.trimmed, whichRandom = c("Subject", "Item"))

These findings are in line with predictions based on Shell (2015), with main effects of each of the factors, but no evidence for an interaction between language switching and arrow congruency in the domain-general Simon arrow task (see Figure 6 for plot of RT by condition).

Accuracy Analysis

Naming accuracy was assessed as response accuracy (both correct language and item name) from all naming trials where the corresponding arrow response was correct (thus excluding 309, 2.3% of naming trials). For the Bayesian accuracy analysis, accuracy was calculated as the arcsine-transformed proportion correct per subject per condition. Naming accuracy was relatively high across all conditions (see Table 8). The best fitting model included only switch condition (BF = 7.89 E+06). That is, there is no evidence that language[4] or arrow congruency had an effect on naming response accuracy (see Table 9).

[4] The 2nd best fit model did include an effect of Language, however model comparison shows only anecdotal evidence that a model including a main effect of language is less likely than one without (BF = .681).

Table 8

Experiment 1 Mean Accuracy Rate for Stay Trials, Switch Trials and Switch Costs (Switch-stay Accuracy) by Language and Arrow Congruency

Language	Congruency	Stay Acc	Switch Acc	Switch Cost
L1	Congruent	0.98	0.95	0.03
L1	Incongruent	0.98	0.96	0.02
L2	Congruent	0.98	0.98	0.01
L2	Incongruent	0.99	0.97	0.02

Table 9

Experiment 1 Switch Accuracy Top 6 Best Fit Bayes Models

Model Fit	Bayes Factor +/- error
SwitchCond + Subject	7.89E+06 ±1.31%
SwitchCond + Lang + Subject	5.37E+06 ±6.41%
SwitchCond + Lang + SwitchCond:Lang + Subject	1.89E+06 ±2.89%
SwitchCond + Congruency + Subject	1.02E+06 ±2.01%
SwitchCond + Congruency + Lang + Subject	6.66E+05 ±2.77%
SwitchCond + Congruency + Lang + SwitchCond:Lang + Subject	2.56E+05 ±4.53%

Note. model syntax: anovaBF(arc.acc ~ switchCond + ConCond + Lang+ switchCond*ConCond*Lang + Subject, data = bysubs.acc, whichRandom = c("Subject"))

The findings from Experiment 1, which paired a predictable language switch task with a Simon arrow task to tax control at the item level, lend further support to previous findings of no interaction between tasks (Shell, 2015). These results suggest that the type of control required for response inhibition in the Simon task is not involved at the item selection level in the language switching task, even when using a smaller set of items to select from (digits) in a sample with more variable language experience. The language switches in Experiment 1 were predictable, where the language to be used on the upcoming trial was always known in advance and so participants could prepare for each language in advance of the task cue / trial onset. As such, despite pairing a secondary control task with the language switch task, it is quite possible that the switch, and any form of control involved, was initiated before the secondary task and therefore the two paired tasks may have actually occurred separately in time. These findings serve as a comparison study for Experiment 2, described below.

Chapter 3: Experiment 2: Taxing Control at the Task Level: Method

In order to better evaluate the role of a secondary task on the language switch, Experiment 2 used an unpredictable, cued language switching task. With an unpredictable language switch, the target language should not be able to be (reliably) prepared prior to stimulus (and arrow) onset. This was expected to constrain both ‘task’ and ‘item’ level selection to begin at the point of stimulus onset; contrasting with the procedure in Experiment 1, where task-level control could be engaged in advance to prepare for an upcoming predictable language switch. To further reduce any advance preparation, the language cue was presented simultaneously with, rather than before, the to-be-named item.

In addition, to maximize the level of conflict on switch trials, as well as to be more comparable to previous unpredictable switch paradigms (e.g. Linck et al., 2011; Meuter & Allport, 1999), the ratio of stay to switch trials was increased so that two thirds of trials were repeat trials and one-third switch trials, thereby making the switch trials less likely. This context of reduced switching (relative to Experiment 1) may also bias the speaker’s control state towards goal maintenance, or proactive control and reduce cognitive flexibility (Fröber & Dreisbach, 2017), and thus increase the cost of switching. Also in contrast to Experiment 1, Experiment 2 included additional blocks of naming in one language at a time. As such, one of the major aims of Experiment 2 was to evaluate the effect of taxing domain general inhibitory control on both switch (stay-switch trial types within mixed block) and mix costs (single language trials-mixed language stay trials), thereby allowing for better specification of the type of control required in language switching.

Participants

Participants were recruited and compensated as in Experiment 1. A total of 145 participants completed the study; 90 participants spoke Spanish as their L2, 27 French, 14 Mandarin, 8 Hebrew, 2 Russian, 2 Italian, 1 German, and 1 Hindi. Participants' language backgrounds can be seen in Table 10.

Procedure

After brief instruction and consent, all participants completed the language switch task, followed by the secondary measure of inhibitory control, the AX-CPT (see Figure 7 and below for details), which requires attending to a continuous series of letter stimuli to monitor for paired items among distractors (see below for details), and ended by filling out questions from the abbreviated language experience (LEAP-Q) survey. The switching task consisted of one block of digit naming practice and five experimental blocks with instructions and self-timed breaks between each. Participants began with counting and digit naming practice in English and in their second language, followed by practice switching between languages. Following the language switching practice they performed the Simon arrow task which acted to both provide practice on that part of the task and also as a separate measure of nonlinguistic inhibitory control ability. Next participants performed two blocks of combined single-language and arrow trials, one each for L1 and L2 (order was counterbalanced between groups). Finally participants completed the combined language switch and arrow task, followed by two more sets of single-language and arrow trials, one for each L1 and L2, but in the opposite order as the first two single-language blocks.

Table 10
Experiment 2: Averaged Self-rated L2 Proficiency (Scale 0-10) and Percent Time Exposed to the L2 (from the modified LEAP-Q; Marian et al., 2007).

Proficiency L2	Speaking	Understanding	Writing	% Exposed
Self-rated (0-10)	5.52	6.28	5.04	10.86

Practice Counting

Participants practiced counting digits 1-9 in L1 and L2 in serial order to ensure comfort with the task and calibration of the voice key trigger. The digits were each proceeded by a fixation cue (plus sign) and remained on the screen until the voice key was detected. Following this, and the experimenter's corrections or reminders on any erroneous trials, participants practiced naming the digits 1-9 presented in a randomized order in their L2 twice to ensure comfort with the L2 names. During naming practice the items were presented along with their language cue: each digit centered in either a box (L1) or a circle (L2). In total there were 18 serially ordered practice trials, nine in each L1 and L2, and 18 randomized practice trials in L2 for a total of 36 practice digit naming trials.

Language Switching

Once comfortable with the digit names, participants practiced the language switching task (this was also included to measure of switching ability). The language switch block included 108 trials with 36 (1/3) being switch trials. Each digit appeared equally in each language for both switch and stay trials. Trials were presented in a fixed pseudorandom order to have no more than four stay trials in a row and to minimize digit repetition by allowing for a digit to repeat itself a maximum of one time on consecutive trials. Each item was presented simultaneously with its respective language cue (box or circle) and remained on the screen until the voice key was detected. A fixation sign in the center of the screen appeared between each trial for 500ms. Accuracy was coded by an experimenter in the room.

Simon Arrow Task

Participants performed the Simon arrow task, following the same protocol and procedures as Experiment 1, crossing arrow location and arrow direction, thus yielding a set of 80 arrow trials fully balanced by location, direction, and arrow congruency (see above for details).

Single Language + Arrow Task

Each single language block contained 72 trials balanced by item and arrow condition, with half of trials paired with incongruent and half paired with congruent arrows. Trials were presented in a fixed pseudorandom order, again allowing a maximum of two consecutive repeated digits while additionally constraining the arrow direction, location, and congruency conditions each to a maximum of three consecutive repetitions. Each digit was presented simultaneously with language cue (held consistent throughout block) and arrow onset. A fixation sign appeared in the center of the screen after each trial for 100 milliseconds. There were two single language blocks (one each per language, order counterbalanced by participant).

Combined Task

The mixed block included 216 total trials with 72 (1/3) switch trials and 144 stay trials. Within switch and stay trial types, trials were fully balanced by arrow congruency and language, and each digit was presented equally in each condition. Digits 1-9 were presented in a fixed, pseudorandomized order to minimize item repetition (as in the other blocks) with a maximum of two consecutive repeated digits, and three consecutive arrow direction, location, or congruency type repetitions. Digits were again presented simultaneously with a language cue and arrow to maximize the overlap between language switch and arrow task. A fixation sign appeared in the center of the screen between each trial for 500ms. Four self-timed breaks were evenly spaced throughout the task (every 54 trials).

Single Language + Arrow Task (Part 2)

The final block was the same design as the fourth block, including single language digit naming and the arrow task, with the order of language reversed. Two sets of single language trials one in each language (L1, L2) were pseudorandomized and balanced in the same manner as the previous block.

Individual Differences Assessments

Simon Arrow Task

The arrow version of the Simon task (used as the secondary task in Experiments 1 and 2), when performed independently from the language switch task, served as an individual difference measure of nonlinguistic inhibitory control (e.g., deBruin et al., 2014; Linck et al., 2011).

AX-CPT

In addition to the Simon arrow task, following completion of the language switching tasks, participants performed an additional measure of cognitive control ability: the AX-CPT task. This task requires participants to monitor for a specific pattern of letters and differentiates proactive monitoring from reactive control (Braver, 2012). The AX-CPT was implemented using E-prime software (Psychology Software Tools, Pittsburgh, PA). Each trial included five letters presented on a black screen individually. Each trial began with a cue (any letter except X, K, or Y) presented in red, followed by three distractor stimuli used to prevent rehearsal (any letter except, A, X, B, Y, or K) presented in white, and then a probe letter (any letter except A, B, or K) in red. Each cue, distractor and probe was presented for 300 ms followed by an inter-trial interval of 1000ms (see Figure 7 for trial schematic).

Participants were told to respond to each trial with either a “yes” or “no” response with the appropriate button on the button box. The cue and following three distractors were always to be responded to with a “no” response while the probe stimulus would only receive a “yes” response if it was an X and the cue had been an A (AX trial type). Other trial types included AY (A followed by a letter other than X), BX (any non-A letter followed by an X) or BY (any non-A letter followed by any non-X letter). Responses were recorded using a button box, responding with one hand on the “yes” key and the other for the “no” key. Yes and no key locations were counterbalanced across participants. Seventy percent of the trials consisted of an A followed by an X (AX trials) to encourage preparation for a "yes" response. Each of the other response pair trial types (i.e. AY, BY, BX) each occurred on ten percent of the trials.

After reading through instructions, each participant started with practice block of 10 trials with mean reaction time and accuracy provided as feedback to encourage speed and accuracy. Following completion of the practice block, participants began the experimental block which included 100 trials and had a self-timed break at the halfway point.

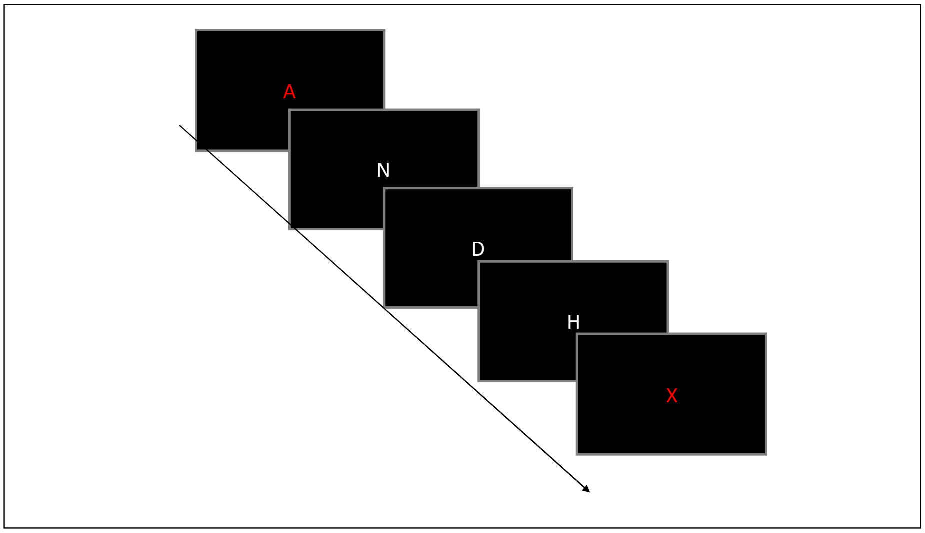

Figure 7. Schematic of a single AX-CPT trial (A-X trial type). Participants should attend to the red letters, watching for a red “A” followed by a red “X” and ignore the distractor white trials. Participants must still respond with a “NO” button response to the first four trials and only respond “YES” to the final trial if the A-X pattern conditions are met. That is, for this trial the correct set of responses would be “NO”, “NO”, “NO”, “NO”, “YES”.

While both reactive and proactive control can be used in this task, there is likely a tradeoff between reliance on each, wherein increased reliance on proactive control will benefit BX trials (by proactively engaging control to inhibit a ‘yes’ response to a X probe) but cause more false alarms on AY trials (where proactively preparing for a ‘yes’ response given the A cue could induce false alarms for a non-X probe) (Braver, 2012). Mean RT and accuracy were scored for target trials (AX), lure trials (AY), probe trials (BX), and control trials (BY). The main outcome of interest here is the relative reliance on proactive control versus reactive control (proactive index), calculated by (AY-BX)/(AY+BX) for both RT and accuracy (Braver et al., 2001). A positive value here indicates higher engagement of proactive control, marked by higher AY interference, while a negative value indicates more reliance on reactive control, marked by higher BX interference (Braver, Paxton, Locke, & Barch, 2009).

Language Proficiency

As in Experiment 1, an abbreviated version of the Language Experience and Proficiency Questionnaire (LEAP-Q: Marian, et al., 2007) was used to measure individual differences in language use and proficiency. Questions included the number of languages spoken, order of acquisition and dominance, self-rated proficiency in speaking, reading, and writing the two languages used during the study, and an additional question asking how if participants ever switched between languages within a sentence or conversation (code switching) (see Table 10).

Chapter 4: Experiment 2: Analysis and Results

Switch Costs

Data cleaning and analysis for Experiment 2 were conducted according to the same protocol as in Experiment 1, but with additional trials removed according to the preregistered decision to remove naming trials following a naming trial where the wrong language was used. These trials were removed because a trial's "switch" or "stay" status is dependent on the language used in previous trial and therefore use of the wrong language on the previous trial could impact the switch or stay status of the current trial. As in Experiment 1, I used both mixed effects and Bayesian ANOVA modeling on RT and accuracy outcomes. In addition, I used drift diffusion modeling methods (Wagenmakers van der Maas, & Grasman, 2007) which incorporate both RT and accuracy to break down responses into three behavioral dependent variables: non-decision time (T_{er}), decision time or drift rate (v), and response caution, or boundary separation (a). Each of these three variables was analyzed separately.

As can be seen in Figure 8, participants showed robust switch costs, but these costs were not larger in the incongruent than congruent condition (in fact, in the L2 they appear to be *smaller* in the incongruent condition) as predicted by the hypothesis of shared inhibitory resources used for language switching and responding on incongruent arrow trials (see Table 11 for switch cost descriptives). These observations were supported both by frequentist and Bayesian analyses.

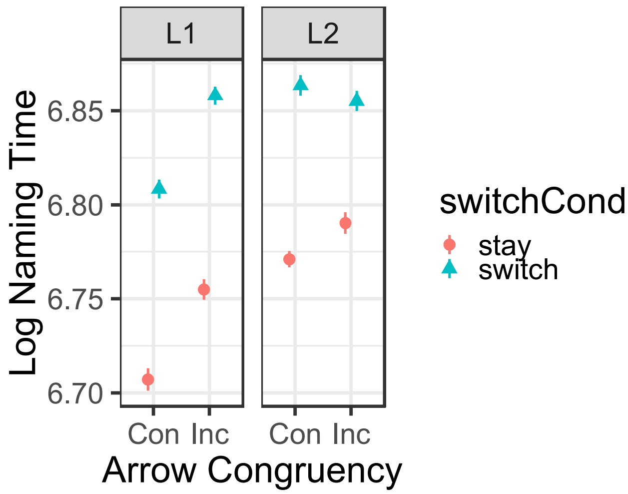

Figure 8. Experiment 2 Naming reaction time (log ms) by arrow congruency condition, switch condition and language. Plotted data are means of subject means (note that analyses were conducted over non-averaged log-transformed data). Error bars indicate standard error of the mean.

Table 11
Experiment 2 Response Time (RT) in Milliseconds for Switch, Stay, and Cost (switch-stay RT) by Language and Congruency

Language	Congruency	Stay RT	Switch RT	Switch Cost
L1	Congruent	857.23	946.81	89.59
L1	Incongruent	902.35	996.62	94.27
L2	Congruent	918.63	1000.85	82.22
L2	Incongruent	941.90	997.84	55.94

Mixed Effects Analysis

All trials with errors in the verbal naming response (1180, 3.8%) or the arrow key response (720, 2.3% of trials) and all voice key detection errors (i.e. when a trial was skipped or not appropriately detected due to the microphone calibration or extraneous noises, 224, 0.7% of trials) were removed from the reaction time data analysis. Following this all trials which following a trial with a language error (were removed. Finally, RTs greater than 3 standard deviations from each subject’s mean (472, 1.7% of trials) were removed from analysis. In total these criteria led to the removal of 3444 (16.5%) of all trials. Analyses were computed for both trimmed and untrimmed data and found that trimming did not significantly influence results (see Appendix A for untrimmed results).

The results of the mixed effects analysis on the model with uncorrelated random slopes included all main effects: a main effect of Switching, with longer RT's for switch than stay trials, a main effect of arrow Congruency, with longer arrow RT's on incongruent compared with congruent trials, and Language, with longer RTs for the L2 than the L1 (see Table 11 for by-subject means). In addition, all interactions were significant: Language by Congruency, Language by Switch Condition, Congruency by Switch Condition, and the three way interaction between Language, Congruency, and Switch Condition (see Table 12).

Table 12
Experiment 2 Language Switch Mixed Effects Results

Naming Time Fixed Effects	Estimate	Std Error	t value	Pr(>|t|)
Intercept	6.804	0.019	361.592	<.001
Switch	0.089	0.007	13.434	<.001
Congruency	0.026	0.006	4.244	<.001
Language	0.037	0.012	3.088	<.001
Switch*Congruency	-0.012	0.005	-2.249	0.026
Switch*Language	-0.017	0.007	-2.478	0.014
Cong*Language	-0.044	0.005	-8.416	<.001
Switch*Congruency*Language	-0.028	0.011	-2.539	0.012

The three-way interaction with language suggests that the two way interaction between Congruency and Switch condition varies by language. To better understand this difference, I ran separate exploratory analyses for just the L1 and just the L2 conditions and found evidence for an (underadditive) interaction between Congruency and Switch in the L2 but not in the L1 (see Table 13 for statistical results and Figure 8 for a visual representation of these interactions). These separate analyses reveal a significant Switch*Congruency interaction in the L2 but not in the L1 suggesting an effect of arrow congruency on L2 switch costs with smaller switch costs in L2 but not in L1.

Table 13

Experiment 2 Language Switching Mixed Effects for L1(top) and L2 (bottom)

L1 Naming Fixed Effects	Estimate	Std Error	t value	Pr(>|t|)
Intercept	6.78E+00	0.018	380.23	<.001
Switch	1.01E-01	0.012	8.425	<.001
Congruency	4.84E-02	0.01	4.8	0.001
Switch*Congruency	4.86E-04	0.027	0.018	0.986
L2 Naming Fixed Effects	**Estimate**	**Std Error**	**t value**	**Pr(>|t|)**
Intercept	6.82E+00	0.021	318.875	<.001
Switch	7.98E-02	0.009	9.355	<.001
Congruency	4.49E-03	0.01	0.445	0.667
Switch*Congruency	-2.72E-02	0.012	-2.316	0.045

Bayes Analysis

As in Experiment 1, Bayesian ANOVA was used to determine the best fit model given the switching data, while providing the relative evidence for each model. The best fitting model included main effects of Switching, Congruency, and Language as well as an interaction between Switch and Language and an interaction between Congruency and Language but not the hypothesized interaction between Congruency and Switch Condition (see Table 14 BF for top 6 best fitting models).

Table 14

Experiment 2 Language Switching (RT) Top 6 Best Fit Bayes Models

Model Fit	Bayes Factor +/- error
SwitchCond + Congruency + Lang + SwitchCond*Lang + Congruency*Lang + Subject + Item	8.725e+333 ±2.81%
SwitchCond + Congruency + SwitchCond*Congruency + Lang + SwitchCond*Lang + Congruency*Lang + Subject + Item	4.931e+333 ±10.1%
SwitchCond + Congruency + SwitchCond*Congruency + Lang + SwitchCond*Lang + Congruency*Lang + SwitchCond*Congruency*Lang + Subject + Item	3.461e+333 ±8.17%
SwitchCond + Congruency + Lang + Congruency*Lang + Subject + Item	7.941e+331 ±6%
SwitchCond + Congruency + SwitchCond*Congruency + Lang + Congruency*Lang + Subject + Item	4.292e+331 ±4.8%
SwitchCond + Congruency + Lang + SwitchCond*Lang + Subject + Item	2.981e+303 ±3.45%

The second best fit model however did include the Switch by Congruency interaction, and so model comparison of these two models was used to determine the evidence for the interaction. The results of this comparison showed substantial evidence in favor of the model without the hypothesized interaction (BF = 5.06), thus the data support the lack of an interaction between language switching and arrow congruency. Analysis on untrimmed data showed a similar pattern (see Appendix A for untrimmed results).

To test for the inclusion of the three-way interaction between Language, Congruency, and Block, I ran a model comparison for the model with the three-way interaction (and all other main effects and interactions) and its counterpart with all two-way interactions and main effects but without the three-way interaction. Although there was anecdotal evidence against the inclusion of the three-way interaction in the model (BF =. 702) I nonetheless performed separate tests for the Switch*Congruency interaction for each language for comparison with the analysis above. As in the mixed effects findings, the switch-by-congruency interaction was supported in the L2 (BF = 3.765 E+ 95; model comparison shows strong evidence for this model over the reduced model: BF= 26.328) but the lack of interaction was supported in the L1 (BF = 1.184; model comparison shows strong evidence for the model without the interaction compared to the model with it: BF = 32.906), although again note that these effects by language were not clearly supported by the three-way interaction in the omnibus test (see Table 15 for by-language results).

Table 15

Experiment 2 Language Switching (RT) Top 3 Bayes Models: L1(top), L2(bottom)

L1 Model Fit	Bayes Factor +/- error
Switch Condition + Congruency + Subject + Item	1.184e+183 ±2.27%
Switch Condition + Congruency + Switch Condition*Congruen	3.416e+181 ±2.1%
Switch Condition + Subject + Item	1.546e+143 ±1.85%
Congruency + Subject + Item	4.281e+37 ±1.17%
L2 Model Fit	**Bayes Factor +/- error**
Switch Condition + Congruency + Switch Condition*Congruen	3.775e+95 ±3.14%
Switch Condition + Subject + Item	1.480e+95 ±1.13%
Switch Condition + Congruency + Subject + Item	4.1202e+94 ±2.01%
Congruency + Subject + Item	0.306 ±1.12%

As in Experiment 1, naming accuracy was calculated using the arcsine-transformed proportion correct per subject per condition. The best fitting model included only switch condition (BF = 2.115 e+34). To determine the evidence for the Switch*Congruency interaction I compared the model which included main effects of Switch and Congruency alone and the model with both the main effects and the interaction and found anecdotal to substantial evidence against the model with the interaction (BF = 0.336) (see Table 16 for analysis and Table 17 for mean accuracy rates by condition).

Table 16

Experiment 2 Language Switching Accuracy: Top 6 Bayes Models

Model Fit	Bayes Factor +/- error
switchCond +Subject	2.115E+34 ±26.84%
switchCond +Lang + Subject	2.441E+33 ±2.41%
switchCond +ConCond + Subject	1.764E+33 ±1.69%
switchCond +ConCond + Lang + ConCond*Lang + Subject	2.865E+32 ±6.06%
switchCond +ConCond + switchCond*ConCond + Subject	2.782E+32 ±5.02%
switchCond +ConCond + Lang + Subject	2.688E+32 ±1.85%

Table 17
Experiment 2 Mean Accuracy Rate for Stay Trials, Switch Trials and Switch Costs (Switch-stay) by Language and Arrow Congruency

Language	Congruency	Stay Acc	Switch Acc	Switch Cost
L1	Congruent	0.981	0.940	0.042
L1	Incongruent	0.974	0.929	0.046
L2	Congruent	0.981	0.935	0.046
L2	Incongruent	0.981	0.944	0.037

Mixing Costs

Mixing costs were calculated and analyses were conducted to assess the effect of being in a mixed language block compared to a single language block on trimmed and log transformed RTs, as well as how this interacts with the Simon arrow task. To determine mix costs, RT and accuracy data from the pure language blocks (naming items in a single language while performing the Simon arrow task) were compared to only the stay trials in the mixed language blocks. As can be seen in Figure 9, participants' RTs showed robust mixing costs in with a pattern of larger mix costs in the L1 than the L2 and on incongruent than congruent Simon arrow trials (see Table 18 for switch cost descriptives). Frequentist and Bayesian analyses generally confirm these observations; though see below for details (untrimmed RTs show the same pattern of results, see Appendix A).

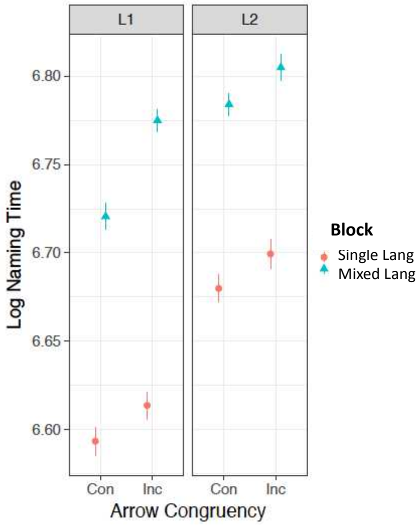

Figure 9. Experiment 2 Naming reaction time (log ms) by arrow congruency condition, block, and language. Plotted data are means of subject means (note that analyses were conducted over non-averaged log-transformed data). Error bars indicate standard error of the mean.

Table 18

Experiment 2 Response Time (RT) in Milliseconds for Single Language, Mixed Language, and Cost (mixed-single language RT) by Language and Congruency

Language	Congruency	Single Lang RT	Mixed (Stay) RT	Mix Cost
L1	Congruent	751.176	854.389	103.213
L1	Incongruent	770.668	895.360	124.693
L2	Congruent	834.540	912.464	77.924
L2	Incongruent	847.483	931.766	84.284

Mixed Effects Analysis

Block (pure/mixed), Arrow Congruency, and Language were included as fixed effects, and Subject and Item were entered as random effects to assess the model fit for the mixed effects model specified with uncorrelated random slopes. There were significant main effects for all factors. That is, there was a main effect of block, with longer naming RTs in the mixed than the pure language block, a main effect of Arrow Congruency, with longer naming RTs on incongruent trials than congruent trials, and a main effect of language, with longer naming RTs on L2 than L1 trials. There was also an interaction between block and language, with larger mix costs in the L1 than the L2 (see Table 18 for means). To better understand this interaction, I additionally assessed mix costs (Block) separately in the L1 and in the L2 and found evidence for mix costs and congruency effects in both the L1 and the L2, but found evidence for a Block*Congruency interaction only in the L1 (See Table 19). That is, there was evidence for an interactive effect of arrow congruency on language mix costs in the L1, but not in the L2 (although note that the three-way Block*Congruency*Language interaction was not significant in the omnibus analysis).

Table 19

Experiment 2 Mix Cost Mixed Effects Results

Naming Time Fixed Effects	Estimate	Std Error	t value	Pr(>|t|)
Intercept	6.697	0.017	397.663	<.001
Switch	0.025	0.005	5.118	<.001
Block	0.115	0.008	15.354	<.001
Language	0.064	0.012	5.244	<.001
Congruency*Block	0.013	0.006	2.161	0.062
Congruency*Language	-0.02	0.009	-2.225	0.057
Block*Language	-0.04	0.01	-4.172	<.001
Congruency*Block*Language	-0.025	0.02	-1.255	0.243

Table 20

Experiment 2 Language Mix Cost Results by Language L1 (top) L2 (bottom)

L1 Naming Fixed Effects	Estimate	Std Error	t value	Pr(>|t|)
Intercept	6.669	0.016	424.528	<.001
Congruency	0.035	0.007	5.250	<.001
Block	0.136	0.009	15.734	<.001
Congruency*Block	0.025	0.001	2.338	0.044
L2 Naming Fixed Effects	**Estimate**	**Std Error**	**t value**	**Pr(>|t|)**
Intercept	6.734	0.020	334.660	<.001
Congruency	0.016	0.006	2.456	0.036
Block	0.096	0.009	10.312	<.001
Congruency*Block	0.002	0.012	0.137	0.894

Bayes Analysis

The best fit model in the Bayes analysis of mix costs included all three main effects (Block, Language, Arrow Congruency) and all interactions: Language by Congruency, Language by Block, and the hypothesized Block by Congruency interaction, as well as the three-way interaction between Block, Congruency, and Language (see Table 21 for results). The top three models all included the hypothesized interaction between Block and Congruency- the hypothesized interaction with the top fit model additionally including the three-way interaction between Block, Congruency, and Language. To better understand this interaction, I performed two follow-up model comparisons. To assess the evidence for the three-way interaction, I compared the top fit model (full model including all interactions and the three way interaction) to its counterpart with all two-way but without the three way interaction, finding strong evidence for the full model with the three way interaction (BF = 15.538). In order to more directly test for the inclusion of the hypothesized Block*Congruency interaction, I also compared a model with all three two-way interactions (but no three-way interaction) to its reduced counterpart without the Block*Congruency interaction, finding strong evidence for the model including the block*congruency interaction (BF = 5.311). As such, these data support the hypothesis for an interaction between Block and Congruency, suggesting larger mix costs during incongruent arrow trial types.

Table 21
Experiment 2 Language Mixing RT: Top 6 Best Fit Bayes Models

Model Fit	Bayes Factor +/- error
Block + Congruency + Block*Congruency + Lang + Block*Lang + Congruency*Lang + Block*Congruency*Lang + Subject + Item	1.818e+680 ±5.21%
Block + Congruency + Block*Congruency + Lang + Block*Lang + Subject + Item	1.304e+679 ±6.08%
Block + Congruency + Block*Congruency + Lang + Block*Lang + Congruency*Lang + Subject + Item	1.170e+679 ±8.67%
Block + Congruency + Lang + Block*Lang + Subject + Item	2.523e+678 ±5.15%
Block + Congruency + Lang + Block*Lang + Congruency*Lang + Subject + Item	2.202e+678 ±9.32%
Block + Congruency + Block*Congruency + Lang + Subject + Item	1.386e+670 ±2.41%

Because of the three way interaction with language, I conducted separate analyses for the L1 and the L2 to determine if there was evidence for an interaction between congruency and mix costs in each language. The top fit model from the analysis of the L1 included main effects and an interaction between Block and Congruency. To test the relative evidence for this model over the main effects only model, these models were compared, finding extreme evidence for the model with the interaction (BF =19.229E+4). On the other hand, in the L2 analysis, the top fit model did not include Block*Congruency interaction, and model comparison between the main effects only model and model with the interaction found strong evidence against the interaction model (BF =0.0192). That is, there was evidence for an effect of arrow congruency on mixing costs in the L1 but not in the L2 (see Table 22 for by-language results).

Table 22

Experiment 2 Language Mixing RT: Top 3 Bayes Models in the L1 (top) and L2 (bottom)

L1 Model Fit	Bayes Factor +/- error
Block + Congruency + Block*Congruency + Subject + Item	6.321e+476 ±3.79%
Block + Congruency + Subject + Item	3.287e+471 ±3.75%
Block + Subject + Item	1.112e+448 ±2.33%
Congruency + Subject + Item	6.454e+22 ±1.27%

L2 Model Fit	Bayes Factor +/- error
Block + Congruency + Subject + Item	4.192e+150 ±2.91%
Block + Congruency + Block*Congruency + Subject + Item	8.082e+148 ±2.63%
Block + Subject + Item	1.692e+147 ±1.11%
Congruency + Subject + Item	3195.599 ±1.19%

As in analysis of switching costs, naming accuracy was calculated for mixing costs using the arcsine-transformed proportion correct per subject per condition. The best fitting model included only Block (BF =2.743e+58). To determine the evidence for the Block*Congruency interaction I compared the model which included main effects of Block and Congruency alone and the model with both the main effects and the interaction and found anecdotal evidence against the model with the interaction (BF = 0.815). (See Table 23 for analysis and Table 24 for mean accuracy rates by condition).

Table 23

Experiment 2 Language Mixing Accuracy: Top 6 Bayes Models

Model Fit	Bayes Factor +/- error
Block + Subject	2.743e+58±2.59%
Block + Lang + Block*Lang + Subject	1.476e+58±1.6%
Block + Congruency + Subject	9.278e+57±1.23%
Block + Congruency + Block*Congruency + Subject	7.558e+57±6.48%
Block + Congruency + Lang + Block*Lang + Subject	5.471e+57±3.91%
Block + Congruency + Block*Congruency + Lang + Block*Lang + Subject	4.153e+57±3.08%

Table 24

Experiment 2 Accuracy Rate for Single Language, Mixed Language, and Cost (mixed-single language) by Language and Congruency

Language	Congruency	Pure Acc	Mix Acc	Mix Cost
L1	Congruent	1.00	0.98	-0.02
L1	Incongruent	1.00	0.97	-0.03
L2	Congruent	1.00	0.98	-0.02
L2	Incongruent	1.00	0.98	-0.02

Drift Diffusion Analysis

Drift diffusion parameters on naming responses (incorporating both latency and accuracy) were calculated to better capture the decision process involved in naming during a language switching task and the effects of the dual task paradigm. Specifically I looked at the parameters non-decision time (T_{er}), decision time or drift rate (v), and response caution, or boundary separation (a) as they might reflect the hypothesized interaction between the arrow task and language switch or mix costs.

After removing trials with incorrect arrow responses (720, 2.3%) and trimming RTs to three standard deviations of the participant's mean, RTs means and variance, as well as naming accuracy were calculated per subject by condition (Switch Condition, Language, Congruency) to be entered into the drift diffusion model using the EZ drift diffusion calculation (Wagenmakers, et al., 2007). Mean accuracy rates were generally high on this task (97%), which suggests high values for drift diffusion parameters, drift rate parameters (v), and boundary separation (a).

When the values for drift rate and boundary separation are high, this could result in low error or error free performance, that is 100% correct responses on certain trial types. The EZ diffusion calculation involves the logit function and so 100% accuracy results in an infinite solution. The solution used here, based on the solution from the EZ diffusion model (Wagenmakers, et al., 2007) is to apply an edge-correction method, replacing P (proportion correct) with a value that corresponds to one half of an error: P = 1- 1/2n (n being the number of trials included in the calculated P).

Drift diffusion analysis of switch costs from the combined block. A mixed effects Bayes analysis was used to estimate the best fit model with each of the drift diffusion parameters calculated including drift rate (*v*), boundary separation (*a*) and non-decision time (T_{er}) and model comparison compared each best fit model with the Block*Congruency interaction to that without. Fitting with the results using the other metrics above, there was evidence against an effect of arrow congruency on mixing costs for all of the drift diffusion parameters, with strong evidence against the interaction in T_{er} (BF = 0.082), substantial evidence against the interaction in *a* (BF = 0.155) and *v* (BF = 0.179). To summarize, there is evidence that the best fit models for switching costs do not include the interaction of interest (Switch*Congruency) for any of the drift diffusion parameters (See Table 25).

Table 25

Experiment 2 Top 3 Bayes Models for Switch Costs using Drift Diffusion Parameters Ter, a, v

Ter	Model	Bayes Factor +/- error
Model 1	Switch+Congruency+Lang+Switch*Lang+Subject	2.284E+12 ±5.63%
Model 2	Switch+Congruency+Lang+Switch*Lang+ConCond*Lang+Subject	8.088E+11 ±6.49%
Model 3	Switch+Congruency+Lang+Subject	2.862E+11 ±2.44%
a	Model	Bayes Factor +/- error
Model 1	Switch+Subject	5.449E+96 ±1.22%
Model 2	Switch+Lang+Subject	3.837E+95 ±3.79%
Model 3	Switch+Congruency+Subject	3.703E+95 ±3.22%
v	Model	Bayes Factor +/- error
Model 1	Switch+Congruency+Lang+Congruency*Lang+Subjsct	1.593E+109 ±2.57%
Model 2	Switch+Lang+Subject	9.258E+108 ±5.75%
Model 3	Switch+Subject	4.227E+108 ±2.93%

Drift diffusion analysis of mix costs from the combined block. As described above, pure language block trials and stay trials from the switch blocks were merged together to calculate mix costs for naming trials with accurate arrow responses. As in the switch cost analysis, mixed effects Bayes analysis was used to estimate the best fit model with each of the drift diffusion parameters calculated including drift rate (*v*), boundary separation (*a*) and non-decision time (T_{er}) and model comparison compared each best fit model with the Block*Congruency interaction to that without. Fitting with the results using the other metrics above, there was evidence against an effect of arrow congruency on mixing costs for all of the drift diffusion parameters, with anecdotal evidence against the interaction in T_{er} (BF = .623), substantial evidence against the interaction in *a* (BF = .109), and anecdotal evidence against the interaction in *v* (BF =.640). (See Table 26 for top 3 models for each)

Table 26

Experiment 2: Top 3 Bayes Models for Mix Costs using Drift Diffusion Parameters Ter, a, v

Ter	Model	Bayes Factor +/- error
Model 1	Block+Congruency+Lang+Subject	1.071E+35 ±4%
Model 2	Block+Congruency+Lang+Block*Lang+Subject	2.559E+34 ±11.58%
Model 3	Block+Congruency+Block*Congruency+Lang+Subject	1.594E+34 ±6.08%
a	Model	Bayes Factor +/- error
Model 1	Block+Lang+Block*Lang+Subject	5.340E+105 ±3.24%
Model 2	Block+Congruency+Lang+Block*Lang+Subject	3.304E+104 ±3.38%
Model 3	Block+Congruency+Lang+Block*Lang+Congruency*Lang +Subject	4.165E+103 ±2.52%
v	Model	Bayes Factor +/- error
Model 1	Block+Congruency+Lang+Block*Lang+Subject	1.375E+144 ±10.795%
Model 2	Block+Congruency+Lang+Block*Lang+Congruency*Lang+Subject	7.960E+143 ±2.75%
Model 3	Block+Lang+Block*Lang+Subject	3.224E+143 ±1.64%

<u>*Individual Differences*</u>

To assess whether individuals' ability to switch between languages correlates with nonlinguistic measures of inhibitory control, Bayesian correlations were conducted between measures of domain-general control abilities (separately for performance on the Simon arrow task and the AX-CPT) and both language switching and language mixing costs. Internal consistency of each of the cost measures (switch, mix, and Simon arrow costs) was calculated using split half reliability- correlating the effect from one half of the items (odd-numbered trials) with the other half (even numbered trials) corrected according to the Spearman–Brown prediction formula (Brown, 1910; Spearman, 1910). Internal consistency of the AX-CPT measures was calculated using Chronbach's coefficient alpha (Cronbach, 1951). Table 27 below shows descriptives statistics and internal consistency scores for each of the measures[5].

[5] Note that the AX-CPT reliability is difficult to measure because the AX trials are so much more frequent (7x) than the other trial types with only 10 of each the AY and BX trial types used to disentangle proactive from reactive control. Additionally to assess reliability of RTs, only accurate trials are included, reducing the number of valid trials per subject to a greater extent. As such, reliability for these measures is typically low- often below .60 for healthy young adults (cf. Cooper, Gonthier, Barch and Braver, 2017).

To more robustly measure an individual's' switching *ability*, rate residualized switch costs were calculated to incorporate both speed and accuracy - which help address concerns of a speed-accuracy tradeoff - into a single measure of switching that could be correlated with the individual differences measures (Hughes et al., 2014). The rate of correct responses (i.e. number of correct responses per second) for switch and stay trials were calculated and then the residualized difference between these rates were calculated per subject following Hughes and colleagues (2014). Rate residual switch costs were calculated per subject both across languages and separately for L1 and L2. The scores reflect each subject's residualized switch cost- that is, the difference between observed and predicted switch trial rates, where a larger switch cost equates to a more negative residual score.

Individual Differences Measures

Both RT and accuracy were recorded in the Simon arrow task and the AX-CPT. For the Simon task, effects were calculated as the difference between scores on incongruent arrow trials and congruent trials. For the AX-CPT, mean RT and accuracy were scored for target trials (AX), lure trials (AY), probe trials (BX), and control trials (BY). While each trial type was scored separately, the main outcome measure of interest for the AX-CPT was the relative reliance on proactive control versus reactive control, calculated by (AY-BX)/(AY+BX) for both RT and accuracy (Braver et al., 2001)

First, to best match past correlational analyses of language switching and inhibitory control (e.g., Calabria et al., 2012; 2013; deBruin et al., 2014; Linck et al., 2011), language switch costs were drawn from the switch only block (i.e., without a secondary task), while inhibitory control was assessed by the arrow congruency effect from the Simon arrow task (incongruent arrow-congruent arrow responses). If language switching relies on the same type of control required for the Simon arrow task, then we might expect a relationship between performance on the language switching and the arrow task, assuming shared mechanisms. Because the Inhibitory Control Model (Green, 1998) proposes that inhibitory control may be differentially applied when switching into the L2 and the L1, supported by different patterns for the relationship between switching into the dominant or less dominant language (e.g., de Bruin et al., 2014; Linck et al., 2011), correlations between Simon performance and language switching were conducted both across languages, as well as separately for switches into the L1 and into the L2. Data trimming and log transformation of the RT data for all tasks was performed here in the same manner as described for the switching task above. Mean switch costs and arrow congruency effects were calculated for each subject. Both Pearson's correlation and Bayesian correlation (using the regressionBF function) were conducted to find the relationship between language switch costs and arrow congruency effects.

Language Switching and Simon Task

Frequentist correlational analysis results found no support for a relationship between language switching and Simon arrow performance in latencies (r = 0.06, p = 0.47) or in terms of rate residualized switch costs (r = -0.04, p = .61), while Bayesian correlation results provide substantial evidence against this relationship both in terms of latencies (BF = 0.27) and rate residualized switch costs (BF = .20) (See Table 27 for descriptives of the switch cost measure in the switch only block).

Given higher level interactions with language found above and in the literature, I ran these analyses separately for switches into the L1 and into the L2 to see if the relationship might be specific to a single language. Results again found no evidence for a relationship between arrow costs and switches into the L1 in latencies (r =. 04, p =.64) or in terms of rate residualized switch costs (r =-.14, p =.10) while Bayesian correlation results show substantial evidence against this relationship in latencies (BF = .21) and anecdotal evidence in rate residualized switch costs (BF = .61). Similarly, for switches into the L2, frequentist correlational analyses did not find support for this relationship in latencies (r =. 07, p = .40) or rate residualized switch costs (r =-.06, p = .48) while Bayesian correlation results show substantial evidence against this relationship in latencies (BF= .27) and rate residualized switch costs (BF= .22). In sum, these data provide consistent evidence against a relationship between mechanisms of language switching and overcoming incongruent arrow direction/location pairings in the Simon task.

Dual-task Language Switching and Simon Task

In addition to calculating switch costs from the pure switching task, I also calculated switch costs from the combined arrow and language block, which contained more trials and so might better capture and represent individual differences in the ability to switch between languages. Switch costs were calculated by collapsing across arrow types and looking only at switch vs. stay trials for switch costs. Given that there was no evidence for an effect of arrow congruency on switch condition reported above, arrow condition likely did not influence switch costs overall. Again, there were no significant correlations between switch costs and arrow costs in latency (r = .03, p =.18) or in terms of rate residualized switch costs (r = -.04, p = .62) and Bayes analysis found substantial evidence against this relationship in both latency (BF = 0.19) and rate residualized switch costs (BF = .20) (See Table 27 for internal reliability of the switch cost measure in the dual-task block).

When assessed separately for switches into each language, there was not support for the relationship when switching into the L1 in terms of latency (r = 13. p = 0.11) or rate residualized switch costs (r = .01, p =.90), while Bayes analysis provides anecdotal evidence against the relationship in latency (BF = .41) and substantial evidence in terms of rate residualized costs (BF = .18).

Finally, there was no relationship between the Simon task and switches into the L2 in terms of latency (r = -.09, p = .28) or rate residualized switch costs (r =-.02, p =.82) while Bayesian analyses found anecdotal evidence against the relationship both in latency (BF = .33) and rate residualized cost (BF = .75). That is, there is no evidence for a relationship between switching costs and performance on the Simon arrow task, either in the single-task language switching blocks or in the dual-task language switching block, and Bayesian correlations suggest reasonable evidence against such a relationship. These data thus do not support the claim that individual differences in language switching relate to performance on domain general control tasks such as the Simon task.

Language Mixing and Simon Task

There has been some evidence for a relationship between inhibitory control measures and *mix* costs (rather than switch costs) in prior work (e.g. Prior & Gollan, 2012). While Experiment 2 found support for an interaction between mixing costs and the inhibitory control task, which suggests shared processes between the two, the correlational analyses here found no evidence for a relationship between individual differences in language mixing costs and arrow congruency effects in latency (r = -.05, p =.55) or rate residualized mix costs (r = -.04, p = .64) with Bayesian analysis showing substantial evidence *against* the relationship in latency (BF = .20) or rate residualized costs (BF = .20) (See Table 27 for descriptives of the mix cost measure).

When this relationship was tested separately for mix costs in the L1 and the L2, there was no evidence for the relationship in L1 latency (r = -.17, p = 0.05) or rate residualized cost (r = -.04, p = .62) while Bayes analyses found anecdotal evidence against the relation in latency (BF =

0.77), and substantial evidence against the relation in terms of rate residualized costs. There was also no evidence for this relationship with mix costs in the L2 in terms of latency (r = .07, p = .39) or rate residualized costs (r = -.04, p = .61) while Bayes analysis found substantial evidence against the relation in costs for both latency (BF = .26) and rate residualized mix costs (BF = .20). That is, there is no evidence for a relationship between mixing costs and performance on the Simon arrow task, with some evidence against this relationship.

Table 27
Internal Reliability Measures and Descriptive Statistics for Switch and Mix Costs, Simon Task Congruency, and AXCPT measures (RT in ms)

Measure	Reliability	Mean (ms)	SD (ms)	Range (ms)
Switch Costs (switch only)	0.515	127.907	76.769	-23.072, 490.132
Switch Costs (dual task)	0.615	84.006	59.309	-16.095, 392.310
Mix costs	0.947	92.467	93.951	-96.193, 748.919
Simon Cost	0.672	26.330	60.206	-90.249, 368.345
AX-CPT: AX *	0.97	299.488	85.702	101, 863
AX-CPT: AY *	0.76	396.998	99.994	122, 958
AX-CPT: BX *	N/A	221.157	85.451	101, 792
AX-CPT: BY *	0.87	248.046	110.969	100, 865
AX-CPT proactive index	N/A	-433.463	118.639	-100.286, -210.000

Note. *AX-CPT measures were calculated with Cronbach's coefficient alpha. All other measures listed were calculated with split half reliability on the difference score measures (see text).
[a] There were not enough valid trials to calculate internal reliability for the BX measure or the proactive index.

Language Switching and AX-CPT

Per subject means were calculated for reaction time and accuracy for each of the four trial types: AX, AY, BX, and BY. Only accurate responses were included in the reaction time analysis and reaction times were log transformed. Based on Braver and colleagues (Braver et al., 2009), a proactive behavioral index measure was calculated as difference score of performance between AY and BX trial types ([AY-BX]/[AY+BX]) to reflect reliance on proactive, compared to reactive control.

If language control relies on domain-general control mechanisms for task/language level control (versus item-selection level control), then participants who rely relatively more on proactive control than reactive would be expected to demonstrate smaller mix costs, which presumably reflect proactive control, but may show larger switch costs if these reflect reactive control. To assess this, switch costs and mix costs were separately correlated with the AX-CPT proactive control index. There was substantial evidence against this relationship in switch costs both across languages (BF = .23) and when switching into the L1 (BF = .29) or L2 (BF = .18).

Language Mixing and AX-CPT

There was also evidence against a relationship between mix costs across languages (BF = .20) and for this relationship specific to the L1 (BF = .29) and the L2 (BF = .18). While there was no evidence for a relationship with the general index of control, as an exploratory analysis, I separately correlated switch and mix costs with each of the outcome measures in the AXCPT task to see if there was anything of interest (see Appendix B for tables of all Bayes Factor correlation results). The only evidence for a relationship with AXCPT measures was between switching costs in the L2 and RTs for AX (BF = 2.94; anecdotal) and AY (BF= 7.37; strong) trial types. While AX trials response times might simply reflect a general speed of processing, AY trial types are of theoretical interest. AY trial types have been suggested to reflect reactive control needed to reactively inhibit the prepotent 'yes' response to the Y following the A cue (Cooper, Gonthier, Barch, & Braver, 2017) and also the efficient *adjustment* of proactive and reactive control (Morales, Yudes, Gómez-Ariza, & Bajo, 2015). That is, while an A cue engages monitoring for the target, the non-target Y probe trial requires adjustment to disengage. Participants who were faster to accurately respond on these trials, thus reflecting better adjustments, proactive and reactive control, also had smaller costs when switching into the L2 (see discussion below).

Language Proficiency

Bilingual second language proficiency has been shown to modulate performance both in language switching and domain general control tasks (e.g. Tao, Taft, & Gollan, 2015). Similarly, as proposed by the adaptive control hypothesis (Green & Abutalebi, 2013) and supported in a review of the literature (Yang et al., 2016), bilingual language use and experience (e.g. frequent switching) also modulates language control. Importantly, some research has found a bilingual advantage but only in those bilinguals who frequently switch languages, suggesting some sort of minimum threshold of switching experience or proficiency necessary to benefit domain general control. These observations suggest that language use and proficiency may modulate performance on language switching tasks as well as performance on tasks of domain general control and so it is worth considering that the current study's majority unbalanced bilinguals may not have sufficient language-switch experience to show a relationship between language and domain general control.

Bayesian regression analyses showed substantial evidence for a relationship between self-reported L2 speaking proficiency and the cost of switching into the L1 (BF = 9.43), but only anecdotal evidence for a relationship with switch costs into the L2 (BF = 1.61). Frequentist correlational analyses also show for support this relationship in the L1 (r = -0.25, p = 0.004) and in the L2 (r = -0.15, p = .09). On the other hand, when speaking proficiency was regressed on mix costs, there was anecdotal evidence *against* a relationship with mix costs in the L1 (BF =0.595) and somewhat stronger evidence against a relationship in the L2 (BF = 0.183). Frequentist correlational analyses similarly found no evidence for a relationship between L2 proficiency and mix costs in the L1 (r = -.14, p = .09) or the L2 (r = -.01. p = .95). That is, language proficiency appears to be related to language switching costs on switches into L1 but not into L2 (see Figure 10), while proficiency is not related to language mixing costs in L1 or L2. This supports a distinction between the effect of language proficiency on switching and mixing costs in the L1 and L2.

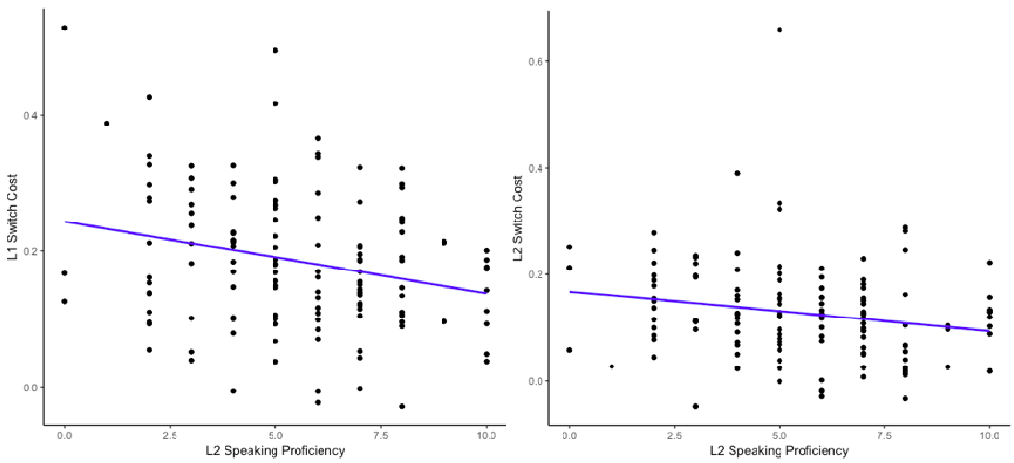

Figure 10. Plotted relationship between self-rated L2 proficiency (scale 1-10) and L1 (left) and L2 (right) switch costs (logRT).

In summary, Experiment 2, which paired the Simon arrow task with an unpredictable language switch task, aimed to tax inhibitory control at both the language and the item level, measured by an interaction between the arrow task and switching, or mixing costs. As in Experiment 1, there was no interaction between the arrow task and language switching. This interaction was however found in the measure of language mix costs (in particular in the L1). Mixing costs here index the control required in a mixed language block, suggesting that incongruent arrow trials interfered with this proactive control. The individual differences measures assessed the relationship between language switching and language mixing ability with second language proficiency, the Simon arrow and AX-CPT performance[6]. These findings from these correlations suggest a relationship between second language proficiency and L1 switch costs, with smaller switch costs for more proficient speakers. In addition there was a relationship between performance on the AX-CPT, in particular RTs on the AY trials, and switch costs in the L2 wherein participants with faster RTs on AY trials had smaller switch costs in the L2. (See Appendix C for a complete correlational matrix of all individual differences variables).

[6] Note that due to the generally low internal consistency of these measures, these correlations are likely under-estimated.

Chapter 5: Discussion

Though it is generally agreed upon that bilingual language production is a process that involves competition, the details of where in the process the competition exists, and what mechanisms are used in order to resolve the competition and successfully produce language, are still unclear. This dissertation set to more robustly examine the implications of the prominent theoretical Inhibitory Control Model (Green, 1998; Green & Abutalebi, 2013), which proposes that inhibitory control is used to inhibit the nontarget language. Although this account has been quite influential, the evidence to support this theory remains quite mixed. Much of the evidence for inhibitory control in language selection is built around inconsistent findings of the relationship between linguistic and nonlinguistic control; for example, with some studies finding this relationship only in specific contexts (e.g. Costa et al., 2009) or with specific stimuli (e.g. DeClerck et al., 2017). In addition, much of this past evidence is based on the interpretation that an asymmetry in switch costs, where switches to L1 show greater costs than switches to L2, yet this asymmetry is only inconsistently found and may not in fact be an effective index of inhibition (Bobb & Wodniecka, 2013). Moreover, the Inhibitory Control Model does not specify the details of where, how, or what type of inhibitory control is applied in the process of bilingual language production. As such, researchers are prone to introducing their own definition of what is meant, resulting in an extremely mixed bag of findings and little consensus in the field.

One contribution of this dissertation was to re-analyze existing results from studies correlating measures of language switching and inhibitory control using a Bayesian approach. I was able to assess the strength of evidence for or against correlations between individual differences in performance on domain-general inhibitory control tasks and language switching costs. This reanalysis found that many of the studies, including my previous experiments (Shell, 2015), either provide evidence *against* a relationship between a domain general inhibitory control task and language switching, or find little to no evidence for this relationship. Importantly, those studies that had data supporting a relationship between linguistic and domain general control found it only under specific circumstances, for instance with extremely controlled and similar stimuli for both tasks (DeClerck et al., 2017), when switching into (or out of) a dominant language (de Bruin 2014; Linck et al., 2011), or only in measurement of mix costs, but not switch costs (Prior & Gollan, 2012). It thus seems that much of the evidence suggesting that domain-general inhibitory control plays an important role in bilingual language use is based on relatively weak data. Accordingly, my reanalysis of the existing correlational data between language switching and non-linguistic control provides some reference for the status of the current findings yet is inconclusive at best in terms of conclusions about inhibitory control in language switching.

In this dissertation, I set out to more robustly assess the contribution of domain-general inhibitory control in resolving interference from a non-target language in bilingual language production, under the (well-supported) assumption that bilinguals do indeed deal with both within- and between-language competition (e.g., Green, 1998; Gollan & Silverberg, 2001; Meuter & Allport, 1999; Philipp & Koch, 2009; see Kroll, et al., 2006 for a review). Importantly, I aimed to break down a somewhat complex model of language control into possibly dissociable *levels* of control: control at the language level and control at the item level. That is, speakers likely need to both select the appropriate language to use (at a global language level), and also select a lexical item from a specific language (at an item level), with potential conflict existing at each level (i.e., between languages and between lexical items). In addition I aimed to investigate the different *types* of control that may be involved in language switching: proactive control used to bias and monitor for conflict more broadly, and reactive used for dynamically controlling at a trial by trial level. It may be that past studies are differentially measuring levels and types of language control, which may or may not be involved or related to language performance, leading to mixed findings.

Experiments 1 and 2 were used to disentangle control at the language and at the item level, two possible loci of control proposed by the Inhibitory Control model (Green 1998; Green & Abutalebi, 2013), which remain relatively ambiguous in the current research. To do so I relied on the language switching paradigm and assessed costs associated with the control needed to remain in or switch between languages.

Experiment 1 combined a predictable language switching task with a common domain-general task of inhibitory control – the Simon arrow task – to interfere with processes that occurred during lexical retrieval, but not necessarily at the language level. In this experiment, language switching was predictable, and so participants could have prepared for the appropriate language, but still presumably could not have prepared to produce a specific lexical item in that language, prior to the demands on domain-general inhibitory control associated with onset of the Simon arrow. If selecting a lexical item from a specific language (for a bilingual speaker) relies on the same cognitive mechanisms involved in overcoming interference in the Simon arrow task, then there should be a larger switch cost on switch trials paired with an incongruent Simon arrow, compared to a congruent arrow- that is-an overadditive interaction between Language Switch and Arrow Congruency. While the data did support main effects for both the switch task and congruency task, there was no support for an interaction between tasks (in fact the pattern of interaction was in the underadditive direction). Bayesian analyses supported this, with anecdotal evidence that the model that best fits these data does not include this interaction. These findings substantiate the results from similar dual task studies in my previous research (Shell, 2015) and suggest that the processes for dealing with conflict on the Simon task and lexical selection during a language switch are likely not shared. Instead, bilingual lexical selection may rely on some type of language specific control (Bloem & LaHeij, 2003) or not involve inhibition at all, but rather management of language use by language *activation* rather than *inhibition* (e.g., Phillip, et al., 2007).

In contrast to Experiment 1, which involved taxing item-level control, Experiment 2 aimed to primarily tax *language* level control. In addition, Experiment 2 was designed as a relatively well-powered, preregistered experiment in order to avoid biases in data collection or analysis and therefore to better draw meaningful conclusions from the data based on a priori hypotheses. Experiment 2 presented a modified version of the task which used an unpredictable switching paradigm so that the language switch could not be reliably planned in advance, with the aim that the Simon arrow task would more directly interfere with switching at the language level (as well as at the item level). Again, evidence for this interference would be an overadditive interaction between Switch and Congruency, and again, there was no evidence to support such an interaction. In line with Experiment 1, there was no evidence of an overadditive interaction, again, failing to support shared processes between the Simon task and language switching, even at the more global level of language control.

Thus, the above findings failed to find support for (and, in some cases, found evidence against) an interaction between language switching costs and domain general (Simon arrow) interference, even when distinguishing between possible *levels* of control used during language switching (and so do not find support for the Inhibitory Control model). A second aim of this study was to distinguish between *types* of control that might be used at either of these levels: proactive vs. reactive control. While language switching could presumably involve both reactive control- that is- the control used to deal with conflict when conflict is encountered- and proactive control, used for advanced preparation for encountering conflict- switch costs may only reflect reactive control. Therefore, in this study I assessed proactive control using mix costs: the difference between naming items in a single language block and naming items in a dual language block. This mixed language (switching) block generally requires being in a state of preparation for an upcoming language switch, and so was predicted to generally increase the engagement of monitoring or control on all trials, including the stay trials where there was no immediate need for control (cf. work on conflict monitoring in language switching; e.g., Green & Abutalebi, 2013; Jylkkä et al., 2017 Tuebner-Rhodes et al., 2017).

Mix costs from Experiment 2 were measured by comparing non-switch trials from the switch block and trials from the pure language blocks. Support for an effect of the Simon task on proactive control was thus measured as an overadditive interaction between Block and Arrow Congruency. In contrast to the switching cost data described above, these data in fact supported this interaction for mix costs. However, this interaction was qualified by a three-way interaction between Block, Congruency, and Language, indicating that this interaction differed by language. That is, interference from the incongruent arrow trials was only found during the mixed language context for L1 stay trials (compared to the baseline trials in an L1 only context). These findings suggest that the type of control required to successfully name items in the L1 in a mixed language context may be shared with the control used for responding to incongruent arrow trials. Although this conclusion may seem unlikely, other research supports a role for proactive control of the L1 in a mixed language context (e.g. Christoffels, Firk, & Schiller, 2007;). For example, Christoffels et al., (2007) assessed language production in pure and mixed language blocks and found that the L1 was significantly slowed in the mixed block while the L2 was not particularly impacted. In addition, they assessed the effect of cognates (items whose name in both languages are similar- e.g. “three” and “tres”) on naming latency, which implicates influence from the non-target language, and found a cognate facilitation effect that was larger for L1 than L2 in the mixed condition, and larger for L2 than L1 in the blocked condition. Together these results suggest that the L1 may be globally inhibited in a mixed language context and therefore less able to influence the L2. Thus, that there is a relationship between the arrow task and mix costs, but not switch costs, may be because mix costs are better able to capture the global, proactive control of the L1 involved on both switch and stay trials, which is not captured by switch cost measures. Though this specific effect is still relatively tentative as it stands among much other evidence against the interaction between the arrow and language switching tasks, it is nevertheless a worthwhile finding to pursue for future investigations.

A perhaps somewhat more obvious predictor of language control ability is experience and proficiency with the languages. Language control unsurprisingly, has been shown to relate to language experience. (e.g. frequent language switching) and proficiency (see Yang et al., 2016 for a review). Similarly, bilingual second language proficiency has been shown to modulate performance both in language switching and domain general control tasks (e.g. Tao et al., 2015). Together, these observations suggest that language use and proficiency may modulate performance on language switching tasks as well as performance on tasks of domain general control (cf. ‘the bilingual advantage’; Bialystok, 2009) which may be reflected in language switching or mixing costs. The Inhibitory Control Model (Green, 1998) predicts that the amount of inhibitory control required to switch into the L1 or L2 should be proportional to the relative imbalance in proficiency between these languages. That is, to use the weaker L2 requires stronger inhibition of L1 and thus larger cost of switching back into the L1. Indeed, this relationship was confirmed here: participants who were less proficient in their L2 had a larger cost of switching into their L1. On the other hand, there was no relationship between language proficiency and the cost of mixing. It is not surprising that the effects of language proficiency are only reflected in switch costs, as the relative inhibition of each language explained by the Inhibitory Control Model is presumably a reactive control measured by trial to trial effects of switching rather between languages, rather than proactive global inhibition of a language which would be reflected in mixing costs.

To more specifically target the role of proactive control in language switching in this study, I compared performance on the AX-CPT measures of proactive and reactive control to switching and mixing costs. While there were no relationships between language switching and the proactive control index used to capture relative reliance on proactive vs. reactive control, there was still strong evidence for a relationship between AY response times and L2 switching costs (with anecdotal evidence for this relationship in the AX trials). That is, participants who were more prepared to respond following the A cue (particularly with a "no" response to the Y probe) also show reduced switch costs in the L2. Studies investigating differences between bilingual and monolingual speakers on this task have found that bilinguals tend to perform better than monolinguals on the AY trials, suggesting more effective adjustments of proactive and reactive control (Morales, Gómez-Ariza, & Bajo, 2013; Morales et al., 2015). In the context of the AX-CPT, with a high number of AX trials, when a Y probe is encountered following an A cue which presumably engages monitoring, participants need to reactively inhibit their prepotent response. Similarly, according to the Inhibitory Control Model (Green, 1998), to switch into the

L2 in a mixed language context, speakers reactively inhibit items in their L1. Importantly, it is worth considering the proposal that the AY measure indexes *adjustments* of proactive and reactive control. If so, this relationship between AY response times and smaller switch costs in the L2 suggests that it may be this flexible ability to adjust use of proactive and reactive control, rather than simply the ability to inhibit, that benefits bilingual language switching. This flexible control is in line with the conflict monitoring account of bilingual language control (e.g., Abutalebi et al., 2012; Botvnick et al., 2001; Teubner-Rhodes, Bolger, & Novick, 2017).

To summarize, there is no clear relationship between language switching and inhibitory control as found by a reanalysis of the literature, and this dissertation found similarly mixed findings - with support for this relationship only using measures and in particular language contexts. There was no evidence for an interaction between language switching and the Simon arrow task, finding evidence against a role of reactive at both the language and item level. There was some evidence, however for an effect of the Simon arrow task on language mix costs in the L1 which better capture proactive control of the L1, in line with the idea that unbalanced bilinguals may proactively inhibit their L1 when in a mixed language context, and in line with accounts of general conflict monitoring in bilingual language control. Although this effect is tentative, it suggests a potential role for proactive control that is often overlooked in the language switching literature.

Individual difference measures on the other hand were shown to only correlate with language switching costs, with smaller switch costs into the L1 for more proficient speakers, and smaller switch costs into the L2 for participants who performed better on AY trials of the AX-CPT, reflecting adjustments of control. These findings suggest that language proficiency and flexibility of control may modulate the ability to reactively control language in a language switching context, however the specificity of these findings demonstrate the complexity of the relationship.

The mixed findings from this dissertation clearly suggests that there is not a straightforward relationship between domain general control as measured by the Simon arrow task or the AX-CPT and language control as measured by a cued language switching task. These results may be interpreted to mean that there is indeed no such relationship, and instead language control might rely on language-specific control mechanisms. For example it may be that the type of control that deals with resolving competition from lexical competitors in (e.g. in a language switching or picture word interference task) is not necessarily the same as the control used to deal with competing responses on a Simon arrow task (cf. e.g., Piai et al., 2013). This language-specific account fits with other claims of the relative modularity of language (e.g. Caplan & Waters, 1999). Another interpretation of these findings is that bilingual language switching does not rely on inhibition of the non-target item or language, but rather relies on activation of the target to bias selection (e.g. Phillip et al., 2007). In fact many of the arguments for inhibition fail to address a potential role of activation. For example, asymmetric switch costs are argued to reflect a greater degree of control needed for inhibiting a dominant language, which is likely inherently more active than the less dominant language, making it more difficult to eventually switch back into (Meuter & Allport, 1999). This argument suggests that switching *out* of the more active L1 should require more inhibitory control, yet it somewhat unclear how inhibition is involved when switching back *into* the L1 (perhaps activation?). That is, even if inhibition is involved in language switching, it is worth considering the role of activation in overcoming this previous inhibition.

Still, I did find some suggestive patterns for involvement of inhibitory control that seem to differ by language and the type of control assessed (i.e. proactive control in the mix cost vs. reactive control measured by the switch costs). These patterns raise another possibility, which is that bilingual language control does in fact rely, at least in part on domain general inhibition, however only in specific ways (as has been made clear by the inconsistencies found in the field currently). This highlights the fact that there has been remarkably little clarity about exactly how or when inhibition might be involved in language control. For instance, inhibition may be involved in proactively monitoring for conflict (and effectively adjusting control) in a mixed language context, or involved in globally inhibiting the dominant language. It could also be that inhibitory control is used to more rapidly inhibit the competing item or language, or to flexibly release previous inhibition. These distinctions may be the key to more consistently locating the potential role of inhibition in language control.

In light of the mixed findings, this dissertation sought to better understand the mechanisms that allow speakers to successfully speak- a process that is deceivingly complex and competitive -with the goal of understanding the mechanisms involved in resolving this competition. The predominant belief in the field is based on influential theories suggesting that domain general inhibitory control is involved in managing this conflict. These theories have also been used to substantiate the (relatively controversial) findings of bilingual advantages on tasks of inhibitory control- that is- that a benefit to domain general control from a bilingual's everyday language. The results of the reanalysis of the available language switching data and the experiments of this dissertation however do not support these theories. They instead support the need to better define and locate how, what, and where inhibitory control might be used in bilingual language control (as well as monolingual language control) and to additionally consider the other mechanisms that might contribute to success in navigating bilingual language production.

Appendices

Appendix A

Table A.1

Experiment 1 (Untrimmed) Top 6 Best Fit Models of Language Switch from Bayes ANOVA

Model Fit	Bayes Factor +/- error
Switch Cond + Congruency + Lang + SwitchCond *Lang + Congruency* Lang + Subject + Item	2.067e+258 ±4.07%
Switch Cond + Congruency + SwitchCond *Congruency + Lang + SwitchCond *Lang + Congruency* Lang + Subject + Item	1.319e+258 ±2.8%
Switch Cond + Congruency + Lang + Congruency* Lang + Subject + Item	7.615e+257 ±3.89%
Switch Cond + Congruency + SwitchCond *Congruency + Lang + Congruency* Lang + Subject + Item	5.480e+257 ±3%
Switch Cond + Congruency + SwitchCond *Congruency + Lang + SwitchCond *Lang + Congruency* Lang + SwitchCond *Congruency* Lang + Subject + Item	5.243e+257 ±5.01%
Switch Cond + Congruency + Lang + SwitchCond *Lang + Subject + Item	2.813e+245 ±2.97%

Table A.2

Experiment 2 (Untrimmed) Top 6 Best Fit Models of Language Switch from Bayes ANOVA

Model Fit	Bayes Factor +/- error
SwitchCond + Congruency + Language + SwitchCond*Language + Congruency*Language + Subject + Item	2.067e+258±4.07%
SwitchCond + Congruency + SwitchCond*Congruency + Language + SwitchCond*Language + Congruency*Language + Subject + Item	1.319e+258±2.8%
SwitchCond + Congruency + Language + Congruency*Language + Subject + Item	7.615e+257±3.89%
SwitchCond + Congruency + SwitchCond*Congruency + Language + Congruency*Language + Subject + Item	5.480e+257±3%
SwitchCond + Congruency + SwitchCond*Congruency + Language + SwitchCond*Language + Congruency*Language + SwitchCond*Congruency*Language + Subject + Item	5.243e+257±5.01%
SwitchCond + Congruency + Language + SwitchCond*Language + Subject + Item	2.813e+245±2.97%

Table A.3

Experiment 2 (Untrimmed) Top 6 Best Fit Models of Language Mixing from Bayes ANOVA

Model Fit	Bayes Factor +/- error
Congruency + Lang + Congruency*Lang + Block + Congruency*Block + Lang*Block + Congruency*Lang*Block + Subject + Item	7.710e+694 ±3.14%
Congruency + Lang + Congruency*Lang + Block + Congruency*Block + Lang*Block + Subject + Item	7.539e+693 ±3.73%
Congruency + Lang + Block + Congruency*Block + Lang*Block + Subject + Item	5.005e+693 ±3.53%
Congruency + Lang + Congruency*Lang + Block + Lang*Block + Subject + Item	2.08e+693 ±3.4%
Congruency + Lang + Block + Lang*Block + Subject + Item	1.816e+693 ±5.43%
Congruency + Lang + Congruency*Lang + Block + Congruency*Block + Subject + Item	8.55e+685 ±3.45%

Appendix B

Table B.1

Bayes Factors for correlations between Language Switch and Mix Costs and AX-CPT log RTs

AXCPT Measure	Overall Switch	L1 Switch	L2 Switch	Overall Mix	L1 Mix	L2 Mix
AX	1.75	0.30	2.94	0.39	0.24	0.53
AY	6.83	0.54	7.37	0.36	0.22	0.48
BX	0.23	0.29	0.18	0.2	0.29	0.21
BY	0.22	0.23	0.18	0.19	0.18	0.18
proactive RT	0.23	0.29	0.18	0.20	0.29	0.18

Note. proactive RT = proactive control index on RT (AY-BX)/(AY+BX)

Table B.2

Bayes Factors for correlations between Language Switch and Mix Costs and AX-CPT accuracy scores

AXCPT Measure	Overall Switch	L1 Switch	L2 Switch	Overall Mix	L1 Mix	L2 Mix
AX	0.32	0.27	0.23	0.22	0.21	0.22
AY	0.40	0.25	0.34	0.2	0.21	0.20
BX	0.26	0.18	0.37	0.22	0.19	0.24
BY	0.79	0.29	0.82	0.18	0.18	0.19
proactive acc	0.27	0.18	0.38	0.22	0.29	0.24

Note. proactive acc = proactive control index (AY-BX)/(AY+BX)

Appendix C

Table C.1

Correlation matrix (Pearson's r) for all individual differences (log RT) measures

ID Measure	Switch Cost	Switch Cost L1	Switch Cost L2	Mix Cost	Mix Cost L1	Mix Cost L2	Simon Effect	AX-CPT pro	AX-CPT AX	AX-CPT AY	AX-CPT BX	AX-CPT BY	L2 prof (LEAPQ)
Switch Cost All	-	**0.828**	**0.791**	0.006	-0.034	0.048	0.061	0.032	0.156	**0.177**	0.109	0.146	**0.247**
Switch Cost L1	-	-	**0.324**	0.077	0.069	0.0692	0.04	-0.008	0.027	0.057	0.055	0.031	**-0.231**
Switch Cost L2	-	-	-	0.066	-0.123	0.006	0.07	0.061	**0.213**	**0.227**	0.117	**0.213**	-0.146
Mix Cost All	-	-	-	-	**0.898**	**0.887**	-0.051	-0.099	0.073	0.054	0.135	0.019	0.086
Mix Cost L1	-	-	-	-	-	**0.594**	**-0.166**	-0.106	0.07	0.024	0.117	-0.019	-0.141
Mix Cost L2	-	-	-	-	-	-	0.074	0.073	0.059	0.073	0.124	0.057	0.006
Simon Effect	-	-	-	-	-	-	-	**-0.198**	**0.251**	0.156	**0.314**	**0.185**	0.123
AX-CPT pro	-	-	-	-	-	-	-	-	-0.06	**0.318**	**-0.694**	**-0.28**	0.009
AXCPT-AX	-	-	-	-	-	-	-	-	-	**0.728**	**0.618**	**0.582**	-0.049
AXCPT-AY	-	-	-	-	-	-	-	-	-	-	**0.461**	**0.473**	-0.07
AXCPT-BX	-	-	-	-	-	-	-	-	-	-	-	**0.624**	-0.06
AXCPT-BY	-	-	-	-	-	-	-	-	-	-	-	-	-0.05
L2 prof	-	-	-	-	-	-	-	-	-	-	-	-	-

Note. L2 prof = LEAP-Q self-rated Speaking proficiency; AXCPT_pro = proactive control index (AY-BX)/(AY+BX). Bolded items indicate significant $p<.05$.

References

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., … Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, *22*(9), 2076–86.

Alario, F., Ziegler, J. C., Massol, S., & de Cara, B. (2012). Probing the link between cognitive control and lexical selection in monolingual speakers. *Topics in Cognitive Psychology*, *112*, 545–559.

Bates, D.,Maechler, M., Bolker, B. Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin & Review*, *15*(2), 357–363.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, *70*(3), 636–644.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism*, *12*(1), 3–11.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*(2), 290–303.

Bialystok, E., Kroll, J. F., Green, D. W., MacWhinney, B., & Craik, F. I. M. (2015). Publication bias and the validity of evidence: What's the connection? *Psychological Science*, *26*, 944–946.

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language*, *48*(3), 468–488.

Braver, T. S., Barch, D. M., Kelley, W. M., Buckner, R. L., Cohen, N. J., Miezin, F. M., … Petersen, S. E. (2001). Direct comparison of prefrontal cortex regions engaged by working and long-term memory tasks. *NeuroImage*, *14*, 48–59.

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.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

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

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–13.

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(18), 7351–7356.

Brown, W. (1910). Some experimental results in the correlation of mental abilities 1. *British Journal of Psychology, 1904□1920*, *3*, 296-322.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*, 1–11.

Calabria, M., Branzi, F. M., Marne, P., Hernández, M., & Costa, A. (2013). Age-related effects over bilingual language control and executive control. *Bilingualism: Language and Cognition*, (November 2015), 1–14.

Calabria, M., Hernández, M., Branzi, F. M., & Costa, A. (2012). Qualitative differences between bilingual language control and executive control: Evidence from task-switching. *Frontiers in Psychology*, *2*(399), 1–10.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 77–126.

Chang, S., Xie, J., Li, L., Wang, R., & Liu, M. (2016). Switch costs occur at lemma stage when bilinguals name digits: evidence from language-switching and event-related potentials. *Frontiers in Psychology*, *7*, 1–9.

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

Cooper, S. R., Gonthier, C., Barch, D. M., & Braver, T. S. (2017). The role of psychometrics in individual differences research in cognition: A case study of the AX-CPT. *Frontiers in Psychology*, *8*(1482), 1–16.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*(2), 135–149.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297–334.

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. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, *81*, 77–86.

De Bruin, A., Roelofs, A., Dijkstra, T., & FitzPatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: FMRI evidence from trilingual speakers. *NeuroImage*, *90*, 348–359.

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*(1), 99–107.

Declerck, M., Koch, I., & Philipp, A. M. (2012). Digits vs. pictures: The influence of stimulus type on language switching. *Bilingualism: Language and Cognition*, *15*(4), 896–904.

Declerck, M., Koch, I., & Philipp, A. M. (2015). The minimum requirements of language control : Evidence from sequential predictability effects in language switching. *Journal of Experimental Psychology*, *41*(2), 377–394.

Declerck, M., Grainger, J., Koch, I., & Philipp, A. M. (2017). Is language control just a form of executive control? Evidence for overlapping processes in language switching and task switching. *Journal of Memory and Language*, *95*, 138–145.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology Learning Memory Cognition*, *28*(6), 1187–1199.

Freund, M., Nozari, N., & Gordon, B. (2016). Conflict-based regulation of control in language production. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1625–1630). Austin, TX: Cognitive Science Society, Inc.

Fröber, K., & Dreisbach, G. (2017). Keep flexible – Keep switching! The influence of forced task switching on voluntary task switching. *Cognition*, *162*, 48–53.

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology*, *10*(5), 640–654.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew English bilinguals. *Bilingualism: Language and Cognition*, *4*(1), 63–83.

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

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530.

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals. *NeuroImage*, *56*(4), 2300–9.

Hamilton, A. C., & Martin, R. C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 1–13.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, *18*, 625–658.

Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior Research Methods*, *46*, 702–721.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–44.

Jeffreys, H. (1961). The Theory of Probability (3rd ed.). Oxford. p. 432

Jylkkä, J., Lehtonen, M., Lindholm, F., Kuusakoski, A., & Laine, M. (2017). The relationship between general executive functions and bilingual switching and monitoring in language production. *Bilingualism: Language and Cognition*, *21*(3), 505-522.

Kan, I. P., Teubner-Rhodes, S., Drummey, A. B., Nutile, L., Krupa, L., & Novick, J. M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition*, *129*(3), 637–51.

Klecha, A. (2013). Language and task switching in Polish-English bilinguals. *Psychology of Language and Communications*, *17*(1), 17–36.

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

Kuznetsova, A, Brockhoff, P., &. Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26.

Lenartowicz, A., Yeung, N., & Cohen, J. D. (2011). No-go trials can modulate switch cost by interfering with effects of task preparation. *Psychological Research*, *75*, 66–76.

Levelt, W. J. M., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Linck, J. A., Schwieter, J. W., & Sunderman, G. (2011). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition*, *15*(3), 651–662.

Liu, Y., Chang, S., Li, L., Liu, W., Chen, D., Zhang, J., & Wang, R. (2018). Effects of naming language and switch predictability on switch costs in bilingual language production. *Frontiers in Psychology*, *9*(649), 1–13.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *24*(5), 605–621.

Ma, F., Li, S., & Guo, T. (2016). Reactive and proactive control in bilingual word production: An investigation of influential factors. *Journal of Memory and Language*, *86*, 35–59.

Marian, V., Blumfield, H., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967.

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

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202.

Monsell, S., Yeung, N., & Azuma, R. (2000). Reconfiguration of task-set: Is it easier to switch to the weaker task? *Psychological Research*, *63,* 250–264.

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology*, *25*(5), 531–546.

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology*, *25*(5), 531–546.

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). BayesFactor: Computation of Bayes factors for common designs. *R package version 0.9*, *8*.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie, N., … Wagenmakers, E. (2017). A manifesto for reproducible science. *Nature Human Behavior, 1*(21), 1–9.

Novick, J. M., Kan, I. P., & Thompson-Schill, S. L. (2013). A case for conflict across multiple domains : Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, *26*(6), 527–567.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, *4*(10), 906–924.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–244.

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

Philipp, A. M., & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching between languages in naming tasks? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(5), 1187–95.

Piai, V., Roelofs, A., Acheson, D. J., & Takashima, A. (2013). Attention for speaking: domain-general control from the anterior cingulate cortex in spoken word production. *Frontiers in Human Neuroscience*, *7*(832), 1-14.

Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(2), 253.

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

Prior, A., & Golan, T.H. (2012). The elusive link between language control and executive control: A case of limited transfer. *Journal of Cognitive Psychology*, *100*(2), 130–134.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from http://www.pstnet.com.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rogers, D., & Monsell, S. (1995). Cost of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology. General*, *124*(2), 207–231.

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.

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross language intrusion errors in aging billinguals reveal the link between executive control and language selectioon. *Psychological Sciences*, *22*(9), 1155–1164.

Schmitz, F., & Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 222–250.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the Time course of lexical access in language production : Picture-word interference studies. *Journal of Memory and Language*, *29*, 86–102.

Shell, A. R. (2015). The role of domain general cognitive mechanisms in bilingual language production.

Shell, A. R., Linck, J. A., & Slevc, L. R. (2015). Examining the role of inhibitory control in bilingual language switching. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, 2146–2151.

Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3) 300-304.

Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. *Frontiers in Psychology,* 2(183), 1–8.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 1904□1920*, *3*(3), 271-295.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.

Tao, L., Taft, M., & Gollan, T. H. (2015). The bilingual switching advantage: sometimes related to bilingual proficiency, sometimes not. *Journal of the International Neuropsychological Society*, *21*(7), 531–544.

Teubner-Rhodes, S., Bolger, D. J., & Novick, J. M. (2017). Conflict monitoring and detection in the bilingual brain. *Bilingualism: Language and Cognition*, 1–25.

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.

Verhoef, K. M. W., Roelofs, A., & Chwilla, D. J. (2010). Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *Journal of Cognitive Neuroscience*, *22*(8), 1832–1843.

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–32.

Wagenmakers, E. -J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, *7*(6), 632–638.

Wu, Y. J., & Thierry, G. (2013). Fast modulation of executive function by language context in bilinguals. *The Journal of Neuroscience The Official Journal of the Society for Neuroscience*, *33*, 13533–13537.

Wu, Y. J., & Thierry, G. (2017). Brain potentials predict language selection before speech onset in bilinguals. *Brain and Language*, *171*, 23–30.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of “switch costs”. *Psychological Research*, *63*(3–4), 212–33.

Yang, H., Hartanto, A., & Yang, S. (2016). The complex nature of bilinguals’ language usage modulates task-switching outcomes. *Frontiers in Psychology*, *7*, 1-8.

Yang, J., Ye, J., Wang, R., Zhou, K., & Wu, Y. J. (2018). Bilingual contexts modulate the inhibitory control network. *Frontiers in Psychology*, *9*(395), 1–14.