A Longitudinal Study on Lexical Accessibility and Executive Control in Adult Bilinguals

Julia Heimann

Doctor of Philosophy

University of York

Language and Linguistic Science

September 2023

Abstract

In recent decades, much research has explored the relationship between the bilingual experience, language switching, executive control mechanisms, and various external factors. The current study contributes to this literature by providing a comprehensive account of different factors that may impact language switching and executive control performance. Specifically, it aims to gain an understanding of the changes to language control and executive control changes in the first stages of immersion in an L2 environment and the interrelationship between language control and executive control while controlling for influences of language use.

The performance of 30 young adult Chinese-English bilinguals in a dual-language picture naming task, two executive control tasks (Simon, Flankers) and responses on an activity log questionnaire were investigated over six months upon initial immersion in the English L2 environment. The performance of 20 functionally monolingual English speakers with no L2 immersion background on the three reaction-time-based tasks were used as a proxy baseline measure against which to investigate the bilinguals' response times.

This approach allowed us to comprehensively portray lexical accessibility and the development of both language control and executive control in the initial stages of L2 immersion to understand better the interplay of the different factors upon each other across the course of development.

The synthesis of the results painted a convoluted picture, suggesting that the development of lexical access and executive control is not as straightforward as previously assumed: for instance, our results did not support the bilingual advantage hypothesis, as bilinguals surpassed the functionally monolingual speakers only in measures of the Flanker task performance. Interestingly, participants who were more balanced across languages performed better in the executive control tasks as opposed to participants who were more dominant in Chinese. Several factors of language use influenced both lexical access and executive control, including using L2 in academic settings and social contexts.

Acknowledgements

I wish to express my gratitude to my supervisors Prof. Monika Schmid and Prof. Leah Roberts for their continued support and guidance during my Ph.D., as well as their constructive feedback on my work.

I would like to thank my examiners Dr. Angela de Bruin and Professor Patrick Rebuschat for their comprehensive comments on my Ph.D. thesis, which have enriched the quality and depth of my research.

I would like to extend this gratitude to all academics, who - be it at the University of York, the University of Essex, the University of Edinburgh, the University of Zurich, or the University of Konstanz – have been part of my academic journey so far. Specifically, I would like to thank Dr. Angela De Bruin, Prof. Simone Pfenninger, Prof. Nathalie Giroud, Dr. Elisa Pellegrino, Prof. Volker Dellwo, Prof. Lena Jäger, Dr. Shayne Sloggett, Dr. Claire Childs, Prof. Danijela Trenkic, Dr. Laurel Lawyer, Dr. Karen Roehr-Brackin, Prof. Antonella Sorace, Prof. Vicky Chondrogianni and Dr. Barbora Skarabela for their invaluable suggestions and insightful discussions.

I am thankful for the unwavering support of my future husband Dr. Samuele Ambrosini, who has been supporting me every day by believing in me, listening to me, caring for me, sustaining me with coffee, and gifting me alpaca plushies.

I am deeply grateful for my papa Gunnar Heimann, who encouraged me every step of the way and helped me achieve the goals I set for myself. I could not have asked for a more supportive papa.

I would like to thank Dr. Catherine Baxter, my high school Italian teacher, for helping me discover my strengths and inspiring me to pursue my passion for languages and linguistics.

I also thank my friends Joseph Zivny, Jasmin Nohl, Antje Behler, Pelin Veit, Pori Saikia, and Dr. Moritz Janas, for their continued friendship and support during my Ph.D. journey.

Authorship Declaration

I declare that his thesis is a presentation of original work, and I am the sole author. This work has not previously been presented for a degree or the qualification at this institution or elsewhere. All sources are acknowledged as references.

The raw data utilised as part of his thesis has been provided by Prof. Monika Schmid.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACH	Adaptive Control Hypothesis
AoA	Age of Acquisition
ATH	Activation Threshold Hypothesis
CPM	Cognitive Process Model
EC	Executive Control
EM	Entrenchment Model
ICM	Inhibitory Control Model
L1	First Language
L2	Second Language
LC	Language Control
lmer	Linear Mixed Effects Regression
LoR	Length of Residence
RHM	Revised Hierarchical Model
RT	Reaction Time, Response Time
SD	Standard Deviation

1. Introduction

Research into bilingual lexical access has opened a window onto the discussion on the interplay of language control and executive control, which has dominated the research output of psycholinguists in the field of bilingualism for the past two decades. Lexical access is defined as "[...] the process by which an individual produces a specific word from their mental lexicon or recognizes it when used by others" (American Psychological Association, n.d.). The general notion is that bilingualism creates competition in lexical access – a competition the speaker must manage in order to enable the swift and accurate language switches we observe in their production. This competition necessitates an additional cognitive demand, which in turn engenders minor adjustments to the way bilinguals process language in moment-by-moment interactions. These adjustments are thought to gradually accrue over time and eventually lead to measurable neuroplastic changes.

Some researchers (e.g., Bialystok, 2009) argue that these neuroplastic changes facilitate inference management in the linguistic domain and generally transfer to executive control faculties, referring to this transfer as "bilingual advantage". Evidence for a bilingual advantage is currently contested, as, according to a meta-analysis conducted by Noort et al. (2009), over half of all studies on the topic lean toward its existence, but the remaining studies returning either no (17%) or mixed evidence (28%). Synthesizing data from 152 studies involving 891 effect sizes on adults, including unpublished data and various study-related factors, Lehtonen et al. (2018) initially find a slight advantage for bilinguals in inhibition, shifting, and working memory(effect size: $g= +0.06$ [0.00, +0.13]), which, however, disappeared after correcting for publication bias ($g= -0.07$ [−0.17, +0.04]).

Discrepancies in the findings for a bilingual advantage have several origins. For one, the selection of bilingual populations in the studies varies greatly, and individual differences, such as L2 immersion duration, L2 proficiency, pre-existing neurocognitive capacities, or sociocultural factors, further complicate interpretations. Methodological differences may also account for discrepancies, as researchers adopt non-standardised experimental designs or fail to control for individual variation, for instance, by neglecting the potential of multifactorial statistical analysis.

As such, a systematic understanding of how bilingualism contributes to a bilingual advantage is still lacking. Of particular interest are two questions: First, when precisely do the small-scale, real-time processing differences accrue to an extent at which differences between monolingual and bilingual lexical access and executive advantage are discernible? Second, which other factors (e.g., language use) affect lexical access and executive control performance changes?

This study is dedicated to exploring changes to language control on the one hand and executive control on the other. To address the challenge posed by the heterogeneity of the bilingual population, we select a relatively homogeneous group of Chinese-English bilingual university students, all of whom have just arrived in the United Kingdom at the time of first testing. We investigate their linguistic and cognitive performance and responses on an activity log questionnaire. We investigate the group in three sessions spread across six months. The first session is conducted shortly after the participants' first arrival, the second after three months, and the third after six months after their arrival. We analysed the development of bilingual lexical and changes to their executive control performance access across these sessions. We compared performance with speakers of English with no L2 immersion background by employing linear mixed models.

The overarching aim of this thesis is to explore the interplay of language control and executive control with a focus on the development in the first months of immersion in an L2- dominant environment while controlling for measures of language use. To do justice to the variety of aspects of this topic, this thesis is presented as a collection of three articles, each contributing to the overarching aim in different ways. In the following, this thesis introduction mirrors the articles' content division by providing three introductory subchapters to guide the reader through the individual subchapters in a structured manner.

The article chapters are arranged so that the preceding articles serve as the basis for the subsequent articles' (Figure 1). The preceding method chapter provides details on analysis choices.

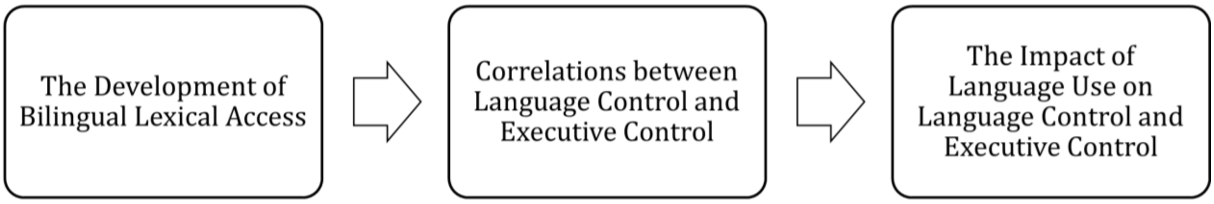


Figure 1: The Structure of the Dissertation

The first article– here presented as Chapter 3 – focuses on the development of lexical access times in both the participants' L1 and L2 as a function of immersion duration. For this, the article explores the data provided by the dual-language Picture Naming Task.

The second article–presented as Chapter 4 – concentrates on the interplay between language control and executive control. A longitudinal executive control data analysis explores whether bilinguals develop a bilingual advantage over time, and in comparison with English speakers with no L2 immersion background[1]. To this effect, the article utilises data from both the dual-language Picture Naming Task and the cognitive Simon Task and Flanker Task.

The third article– presented as Chapter 5 – considers both the interplay of language control and executive control and examines it in the light of derived measures of language use. This article draws from both the dual-language Picture Naming Task responses, the cognitive Simon Task and Flanker Task, and responses to a Language Activity Log Questionnaire.

In the following, this introduction will highlight the individual contributions of each article by providing a brief insight into the current state of the research and elaborating on the aims and purposes underpinning the current studies.

[1] In this thesis, these English speakers are sometimes referred to as "(functional) monolinguals" to facilitate discussions around the differences between this group and the group of Chinese students. Further details on this group are presented in Chapter 2.1. Participants.

1.1. Chapter 3: The Development of Bilingual Lexical Access

When native speakers of one language become immersed in an environment where another language is dominant, their utilisation of their first languages (L1) tends to decrease, resulting in subtle alterations in their language use. At the same time, their use of and proficiency in the second language(s)[2] (L2) typically increase as they adapt to the linguistic demands of the new environment. Chapter 3 – the article "The Development of Bilingual Lexical Access" – investigates these developments. In doing so, it touches on three main aspects. First, bilingualism and the nature of the two languages in the mind; second, lexical access – with a focus on bilingual lexical access; third, the development of this bilingual lexical access from "day one". This introduction will contextualise these main aspects to set the stage for Chapter 3.

The L1 and L2 exhibit contrasting characteristics: The L1 is typically acquired from birth; its acquisition is implicit and benefits from critical period advantages (Lenneberg, 1967; Johnson & Newport, 1989). In contrast, the L2 is typically acquired at a later age and its acquisition often explicit. The languages influence each other cross-linguistically, which means that all previously acquired languages affect the acquisition of subsequent languages and subsequent languages affect previously acquired languages.

For these reasons, the L2 lexicon is acquired differently than the L1 lexicon (e.g., Singleton, 1999). The term "lexicon" – or "mental lexicon" refers to the theoretical concept of a mental "word-storage" (Aitchison, 2012). A prominent model to account for the acquisition of the L2 – and the development of the L1 as a result of the acquisition of the L2 – is the Revised Hierarchical Model (henceforth: RHM; Kroll & Stewart, 1994; Kroll & Ma, 2017). This model assumes that there are links between the lexical sphere, i.e., the mental lexicon, and the conceptual sphere, i.e., the space where concepts or ideas are stored without their lexical form (see also: Levelt, 1989) and that these links are separate for either language. Figure 2 illustrates the RHM.

[2] The term L2 is used to mean any additional language(s) participants have acquired after their native language(s), as is the case in sequential bilinguals.

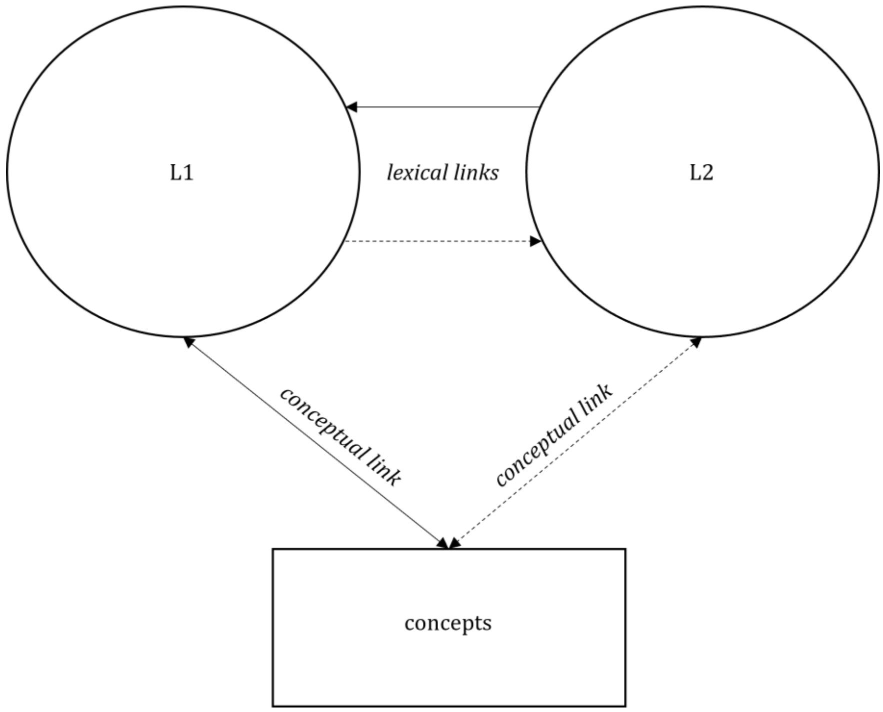


Figure 2: The Revised Hierarchical Model of bilingual lexical access.

The RHM proposes that no established links exist connecting the conceptual system with the second language (L2) lexicon during the early phases of bilingualism. As a result, speakers depend heavily on translation equivalents from their first language (L1) when retrieving words in the L2. As exposure to the L2 intensifies, a direct link between the second language (L2) and the conceptual store is established (Kroll & Stewart, 1994). At this stage, this L2 link is considered weaker and different from the L1 link (e.g., Meuter & Allport, 1999), as a result of different acquisition routes. As exposure and use of the L2 continues, the L2 link is thought to strengthen. The relative strength of either language's conceptual link is closely related to a concept here referred to as "language balance". Language balance reflects the ease with which each language can be accessed relative to the other.

It has been shown that the two language systems become co-activated during language production and comprehension (e.g., Kroll et al., 2006; Misra et al., 2012; Green, 2011). However, only one language can be produced at any one time. Thus, some sort of selection must occur. This selection must occur fast during language switching (when changing from one language to another), and language mixing (when using multiple languages at the same time). To ensure the selection of the target language, Green's (1998) Inhibitory Control Model (henceforth: ICM) proposes that it is necessary for a speaker to inhibit any non-target language representations actively. Active inhibition means employing mental resources to prevent the activation of any non-target representations. Active inhibition becomes increasingly necessary as the L2 matures. The two languages are said to "compete" against each other for selection (e.g., Green, 1998).

This competition has implications for both the development of the L2 – as well as the L1. The field of first language attrition studies the changes to the L1 due to this competition and the relative disuse of the L1. Schmid and Köpke (2009) find that the decline in the ability to promptly retrieve and recognize words in the first language is the most immediate effect of second language immersion.

Yet, the L1 is thought to be heavily "entrenched" in the mind (HJ Schmid, 2017) – both due to it being the first one to be acquired and it being used most frequently. As entrenched system, it is thought to be more difficult for this (L1) communicative routine to change. For this reason, L1 entrenchment is seen as a risk factor to L2 acquisition in the Unified Competition Model (UCM), as it impedes the establishment of the new system in the mind (MacWhinney, 2005).

Overall, the RHM, the ICM and the Unified Competition Model are useful theoretical frameworks to understand the developmental changes to L1 and L2 lexical access, respectively. However, current frameworks lack specific timeframes for specifically when and under which exact circumstances developmental changes might occur. For instance, in a longitudinal study, Baus et al. (2013) explore naming latencies and verbal fluency in a group of German university students during a semester abroad in Spain. The experiments were run upon the students' arrival to Spain and at their departures. Baus et al. (2013) find that participants became significantly slower at naming non-cognate items in their native language at the end of the immersion period but found no differences between the two testing points in the verbal fluency task. In another study, the verbal fluency task is found to be sensitive to the degree of immersion: Linck et al. (2009) find that as immersed L2 learners were significantly more impaired in producing L1 items in a verbal fluency task compared to their peers, who only learned L2 in the classroom in their L1-dominant environment. Most studies on bilingual lexical access – specifically those which combine aspects of executive control advantages – mainly involve bilinguals who have already spent several years in an L2 environment: for instance, 3.7 years in Bonfieni et al.'s study (2019); at least 6 years in Ooi et al.'s study (2018); and 3.8 years in Han et al.'s study (2022).

Overall, few studies have explored bilinguals' lexical accessibility in their early months of immersion within an L2-dominant environment using a bilingual longitudinal approach. Our present study aims to precisely address this gap by exploring their experiences in the early stages of L2 immersion in a comprehensive, longitudinal manner: At the beginning of the study, participants had recently commenced their university studies in the UK. The subsequent sessions – occurring at three and six months – track their developmental progress. We address the following research question:

How do naming latency, switch costs, and mix cost develop over time in the L1 and the L2, respectively?

Chapter 3 proposes the following hypotheses for the investigation of the research question:

The first hypothesis posits that during the first weeks of being in an English-speaking environment, bilingual participants will exhibit quicker naming latencies, as measured through response times, for items in their Chinese L1 compared to their English L2. This disparity arises because the L1 remains the more utilized and active language while naming in the L2 is likely to be slower due to the ongoing development of its conceptual associations.

The second hypothesis postulates that in subsequent sessions, there will be an increase in naming latencies for Chinese and switch costs associated with switching into Chinese. This is because connections to the Chinese lexicon are intentionally inhibited to aid L2 naming, resulting in the emergence of asymmetric switch costs. Concurrently, naming latencies in English and switch costs when switching into English will decrease, as bilinguals will develop stronger associations with their L2. These effects are expected to be most significant in the third session, aligning with our hypothesis that the duration of immersion correlates with proficiency levels.

The third hypothesis proposes that single-language blocks will consistently exhibit shorter naming latencies compared to mixed-language blocks. This is attributed to the potential for globally inhibiting the competing language in single-language blocks, which likely represents an efficient strategy. Additionally, as immersion duration increases, we contend that mix costs will diminish, reflecting bilinguals' improved capability to suppress the non-target language locally.

Finally, the last hypothesis serves as an "assumption check" by positing that English speakers with no L2 immersion background will consistently surpass bilinguals in performance due to the extra cognitive load faced by bilinguals, stemming from competition effects (e.g., Green, 1998).

We ran three linear mixed-effects models to test our hypotheses. All models had Response Time as the outcome variable. The data set was split in three different ways to test our hypothesis efficiently. If we had entered all data in a single model, we would have been less able to interpret certain interaction effects. Table 1 describes and illustrates – by way of greying out excluded trials – the ways in which the original dataset (Figure 3) was split to look at the differences between mono- and bilinguals in the single-language task, and the bilinguals' development of switch costs and mix cost, respectively.

Providing a preview of our findings, our results revealed that the L1 remains predominantly active and readily accessible, whereas L2 retrieval is linked to extended naming latencies. Compared to the functionally monolingual English speaker's L1 retrieval, we observed that the L1 retrieval of bilinguals already incurs notably prolonged naming latencies, underscoring an increased processing demand experienced by bilinguals. In line with Meuter & Allport (1999), our findings revealed a marked reduction in L2 switch costs alongside an increase in L1 switch costs. Contrary to our expectations, we found that the mixed-language condition facilitated naming – particularly in the L2. Finally, we found that both the switch cost and mix cost asymmetry between L1 and L2 reduced as the bilinguals languages became more balanced.

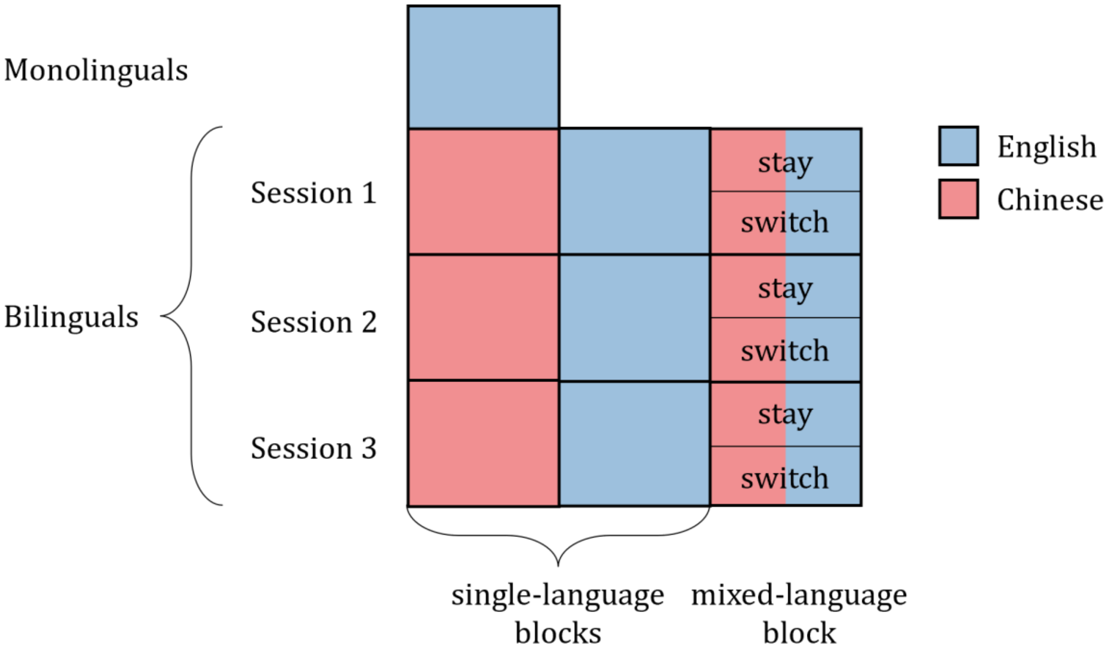
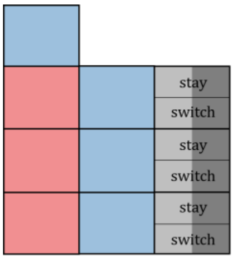


Figure 3: The original dual-language Picture Naming Task Dataset

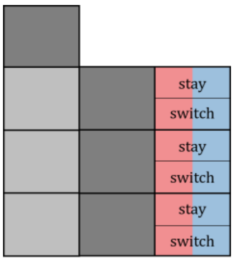
Table 1: Description of PNT dataset divisions for models in Chapter 3.

Model 1: Functional Monolinguals vs. Bilinguals



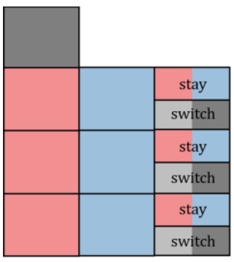
For model 1, we explored functional monolingual and bilingual "pure" response time, i.e., the response time without any additional task demands. To this effect, we excluded all mixed-language blocks. We differentiated between the bilinguals' languages and sessions.

Model 2: Bilingual Switch Cost



In model 2, we analysed bilingual switch costs by focusing on mixed-language blocks, excluding all single-language blocks. This helped us avoid a nested contrast of single vs. mixed (switch, non-switch), which aided the interpretation of the effects of switch cost.

Model 3: Bilingual Mix Cost



In model 3, we investigated bilingual mix cost by only selecting data of single-language blocks and "stay" ("non-switch") trials of the mixed-language block. Again, this meant we could avoid the nested contrast and could interpret the effects of mix cost more easily.

1.2. Chapter 4: Language Control Metrics Predict Executive Control Performance

The preceding introduction to the first article focused on how bilinguals selectively retrieve words from their languages. It explained the need for bilinguals to actively inhibit non-target language activation for accurate access to the target language, due to an ongoing co-activation of their languages (e.g., Green, 1998; Poulisse & Bongaerts, 1994; de Bot & Schreuder, 1993; Kroll & Sunderman, 2003; Kroll et al., 2006). Despite this additional requirement, bilinguals are known to swiftly transition between language systems with remarkable precision and notable ease (e.g., Heredia & Altarriba, 2001).

It is this observation that has allowed for an examination of the question of whether bilinguals develop efficient mechanisms to deal with this conflict– mechanisms, that may translate to non-linguistic skillsets, too. Chapter 4 – the article "Language Control Measures Predict Executive Control Performance" – investigates this question. In doing so, it touches on three main topics. First, the nature of executive control; second, neuroplasticity; third, the impact of language control on domain-general executive control. This introduction will contextualise these main topics to set the stage for Chapter 4.

The term "executive control" is a comprehensive label for numerous cognitive processes entailed in oversight and regulation of various higher-order functions, including decision-making, attention, problem-solving, planning, inhibition, and goal-setting. In experimental psychological research, various tasks are employed to assess the executive control of individuals. For instance, in the Simon Task (Simon, 1969), participants respond based on a specific stimulus feature (e.g., colour) while inhibiting information of its spatial location. This task examines the interference between task-relevant and -irrelevant features. In the Flanker Task (Eriksen and Eriksen, 1974), participants are shown a central target stimulus surrounded by other stimuli that are either matching or incongruent mismatching with the target. To be successful at the task, they must focus on the target and inhibit the influence of the surrounding distractors. In both these tasks, executive control is strongly associated with attention control[3].

[3] Other fields – such as neuroeconomics – have a slightly different understanding of executive control. For instance, in the Delayed Discounting Task (Kirby and Maraković, 1996), participants choose between smaller immediate rewards and larger delayed rewards, assessing their ability to delay gratification.

1. Introduction

Executive control performance varies between individuals. For instance, within groups characterised by lower attention levels, like children with ADHD, research has revealed that individuals face difficulties in inhibiting automatic responses when required, there is a reduced ability to delay gratification or resist temptations, and there are challenges in halting ongoing responses upon receiving signals and struggles in modifying response patterns despite feedback on errors (Barkley, 1999).

Specific experiences can influence performance on tests of executive control by leveraging the brain's ability to undergo changes, known as neuroplasticity. Neuroplasticity involves a reorganisation of connections and functions of the brain (Mateos-Aparicio & Rodríguez-Moreno, 2019). These adaptations enable the brain to adapt to novel circumstances and enhance its performance in familiar tasks. Examples of such specific experiences include enriched environments (Lövdén et al., 2005; Stine-Morrow et al., 2014; Zuelsdorff et al., 2019), healthy diets (Klimova et al., 2017), physical activity (Bamidis et al., 2014), musical activities (Mansens et al., 2018; Moreno et al., 2011), gaming (Strobach et al., 2012) and meditation (Teper & Inzlicht, 2013).

Bilingualism, too, has been proposed to be a specific experience that trains the brain in a way that could benefit certain tasks (e.g., Bialystok, 2009). Specifically, it is the aspect of language control within the bilingual experience – previously discussed in 1.1. Chapter 3: The Development of Bilingual Lexical Access– that is said to be like tasks involving attention control. Through similarity, supporters of the "bilingual advantage" argue that bilinguals, who regularly exercise language control, benefit from this experience beyond language, such as in executive control assessments.

This proposition gains support from multiple studies, such as those employing the Simon Task (e.g., Bialystok et al. 2008; Martin-Rhee & Bialystok, 2008) and Flanker-type tasks (e.g., Poarch & Van Hell, 2012; Yang et al., 2011), where bilinguals outperform monolinguals. Nevertheless, the findings of other studies do not corroborate the bilingual advantage (e.g., Gathercole et al., 2014; Morton & Harper, 2007; Antón et al., 2014; Paap & Greenberg, 2013). In a comprehensive meta-analysis on the bilingual advantage, Noort et al. (2019) found that slightly more than half of the studies evidenced positive bilingualism effects on cognitive control tasks not involving language, while around a third presented mixed and every sixth study presented conflicting results. Synthesizing data from 152 studies involving 891 effect sizes on adults, including unpublished data and various study-related factors, Lehtonen et al. (2018)

initially find a slight advantage for bilinguals in inhibition, shifting, and working memory (effect size: g= +0.06 [0.00, +0.13]), which, however, disappeared after correcting for publication bias (g= −0.07 [−0.17, +0.04]).

The inconclusive findings on the existence of a bilingual advantage highlight the need to further explore the intricate interplay of linguistic and executive control. Other more recent studies have explored linguistic and executive control metrics, such as Han et al. (2022), Bonfieni et al. (2019) and Ooi et al. (2018). Current frameworks lack specific details under which circumstances the bilinguals' executive control performance might change in a way that differs from that of monolinguals. Most studies on bilingual lexical access mainly involve bilinguals who have already spent several years in an L2 environment (e.g., 3.7 years in Bonfieni et al.'s (2019) study) and are not longitudinal in nature. Other studies, which do use a longitudinal design, investigate learners' in a home environment. For instance, Ramos et al. (2017) explore the relationship between learning a second language and executive control functions in elderly participants across one academic year. Further, Bak et al. (2016) explore the long-term effects on executive control functions following an intensive week of learning Scottish Gaelic, and separately, test the effects of repeated classroom exposure. To our knowledge, no previous research has examined bilinguals in their early immersion within an L2-dominant environment using a longitudinal approach.

Our present study aims to precisely address this gap by exploring the relationship between linguistic control and executive control at the initial stages of L2 immersion in a comprehensive, longitudinal manner: At the beginning of the study, participants had recently commenced their university studies in the UK. The subsequent sessions track their developmental progress at three and six months.

Further, our analysis includes an exploration of intra-individual variation (IAV). IAV concerns behavioural differences within individual bilinguals across contexts and/or at different moments of development, as well as nonlinear developmental changes (Hickmann et al., 2018). The consistency of attention control is an important cognitive trait related to a number of cognitive abilities, such as L2 learning (Unsworth, 2015) and may as such provide further insights into the relationship between the linguistic and the cognitive domains.

We address the following research questions:

RQ1: Is there a bilingual advantage in that bilinguals with L2 immersion experience outperform English speakers with no L2 immersion background in executive control tasks?

RQ2: If so, at which stage of the bilingual experience does it set in?

RQ3: Are bilingual individuals, who are better at switching between languages, also better at domain-general executive control tasks?

RQ4: Is there a correlation between measures of intra-individual variation (IAV) in linguistic control and executive control tasks?

Chapter 4 proposes the following hypotheses for the investigation of the research question:

The first hypothesis posits that bilinguals will consistently outperform English speakers with no L2 immersion experience. If the bilingual experience enhances executive control functions and so are more efficient at conflict management, as Bialystok (2009) proposes, then bilinguals will outperform functional monolinguals in tasks relating to inference suppression.

The second hypothesis postulates that there will be a significant effect of immersion duration on interference scores and mix cost. If the bilingual experience enhances functions of executive control through increased immersion duration, then bilinguals will become increasingly faster and better at inhibiting interfering information.

The third hypothesis proposes that there is a correlation between individual performance on the general-cognition executive control tasks and linguistic control tasks, meaning that those who perform better at linguistic control also perform better at the cognitive control tasks.

Following Pfenninger & Kliesch (2023), the final hypothesis proposes that there is a correlation between measures of IAV on general-cognition executive control tasks and on the picture naming task, which means that those who vary in one aspect of IAV in a task, then they will also vary in this aspect in another task. This would further support the claim for a bilingual advantage.

To test these hypotheses, the analysis draws from data of the dual-language Picture Naming Task – previously introduced in section 1.1. Chapter 3: The Development of Bilingual Lexical Access, as well as a Flanker, and a Simon Task. It employs several previously introduced metrics. Namely, Response Time, Switch Cost (+Asymmetry), Mix Cost (+Asymmetry), and Language Balance.

Further to these metrics, we introduce inference scores to assess participants' capacity to inhibit conflicting information in the Flanker Task and the Simon Task. Inference scores reflect inhibitory control, indicating increased difficulty in completing incongruent trials compared to congruent trials, typically calculated as the difference between the two. Table 2 summarises these metrics, their factor names and how they were calculated.

Table 2: Summary of Variables in Experimental Tasks

Factor	Marked (subtrahend)	Unmarked (minuend)	Measurement (difference)
Congruency	incongruent	congruent	interference scores
Task Type	mixed congruency	single congruency	mix cost
Trial Type	switch	non-switch	switch cost
Language	L2 (English)	L1 (Chinese)	language balance

Table 3 shows which metrics were used for language control and executive control.

Table 3: Summary of language control and executive control measures

	Language Control	Executive Control
Overall RT	X	X
Interference Scores		X
Switch Costs	X	
Mix Costs	X	X
Language Effects	X	

Further to the above, this analysis explores IAV metrics to understand whether there is a correlation of IAV metrics in language control measures and executive control measures. All metrics were represented Lowie and Verspoor's (2019) coefficient of variation (CV), which is calculated by dividing the standard deviation by the mean of the given subset. These metrics are presented in Chapters 2 and 4 to avoid redundancy.

We ran a series of linear mixed-effect regression models to explore the different aspects of the given research questions. For both the Flanker Task and the Simon Task we constructed two models. The first models compare the bilinguals' performance in the three sessions in either task against the functional monolinguals' performance regarding inference scores and mix costs. These models aim to investigate whether there is a bilingual advantage (RQ1) and, if so, when this advantage crystalises (RQ2). The second models exclude the functional monolinguals' performance and focus on the bilinguals' performance. In doing so, we can enrich the model with insights from the dual-language PNT and in this way explore whether bilingual individuals, who are better at switching between languages, are also better at domain-general executive control tasks (RQ3). Finally, a correlational Spearman analysis investigates whether there is a correlation between IAV measures in linguistic control and executive control tasks (RQ4).

Providing a preview of our findings, our results revealed that bilinguals exhibited better performance than functionally monolingual speakers across various executive control measures (RQ1), in line with studies supporting bilingual advantage. Longer immersion duration did not notably influence these outcomes (RQ2). Instead, language balance was pivotal, significantly impacting both language control and executive control performance (RQ3), highlighting its significance in the interplay between language control and cognitive abilities. Unlike related studies, linguistic and cognitive IAV measures did not correlate significantly (RQ4), which may be attributed to insufficient sample size.

1.3. Chapter 5: Effects of Language Use on Language Control and Executive Control

In section 1.2. Chapter 4: Language Control Metrics Predict Executive Control Performance above, we delved into whether bilinguals have an advantage in tasks relating to executive control due to their bilingual experience requiring some sort of language control mechanism (Bialystok, 2009). The core premise explored was that, given their need for an efficient language control mechanism which selects target representations by actively inhibiting non-target representations, this mechanism would transfer beyond linguistic benefits. The previous chapter concluded by providing mixed answers to the question of bilingual advantage, with several assessment metrics significantly improving, while others did not.

These mixed results motivated further investigation of the interplay between language control and executive control mechanisms. Specifically, Chapter 5 – the article "Effects of Language Use on Language Control and Executive Control" – investigates the role of language use habits on this relationship to explore the impact of different habits on the efficacy of training in the realm of language control to transfer to other cognitive domains. In doing so, it touches on three main topics. First, the nature of language control; second, the impact of different language use habits on language control mechanisms; and finally, the way in which these different habits transfer from language control to executive control. This section will contextualise these main aspects to set the stage for Chapter 5.

One consideration that emerged from the observation that some studies on the bilingual advantage have yielded positive results, while others have not (e.g., Noort et al. 2019) is that language control could have different levels of control. Different levels of control could be needed by different language use behaviours. For instance, between a professional translator, who uses two languages within the same context (a dual-language context bilingual), and someone who typically speaks English at work, but German with their family at home and thus only uses one language per context (a single-language context bilingual). Further, the type of switching may also differ, as switches may occur between sentences or within sentences ("intra-sentential switching"; Green & Abutalebi, 2013; Hartano & Yang, 2016).

1. Introduction

The Adaptive Control Hypothesis (ACH; Green & Abutalebi, 2013) and the Cognitive Process Model (CPM; Green & Li, 2014) consider the possibility that different switching behaviours impose demands on different parts of the control mechanism. As a result, these models posit that different control mechanisms may be trained depending on the individual experience of a bilingual. Regarding the question of the bilingual advantage, this distinction could further explain why some bilinguals seem to have an advantage in certain assessment metrics while others do not. Specifically, the ACH predicts that dual-context bilinguals require a higher taxing level of language control as opposed to single-context bilinguals and those who routinely mix between languages within utterances (Green & Abutalebi, 2013).

Overall, extensive research supports both the ACH and the CPM. For instance, Hartanto and Yang (2020) conducted a large sample size study, which revealed that those with an extended background in dual-language contexts needed less effort to switch between tasks than their counterparts who primarily spoke on a single-language basis. Furthermore, those who were regularly exposed to dense language-switching contexts were significantly better in goal maintenance and inhibitory control. Nevertheless, other studies did not find evidence for the predictions put forth by either ACH or CPM. De Bruin et al. (2015) compared active and inactive older adult bilinguals with monolinguals on executive control, considering various factors such as socio-economic status, education, IQ, gender, and age and found no difference in overall RT or the Simon effect. Consequently, the current evidence remains inconclusive.

For this reason, we suggest further focusing on external factors that impact the interaction between language use and linguistic and executive control. This study centres on the relative usage frequency of either language and the situational contexts in which it is used as modulating factors of language control on the one hand, and executive control on the other.

We address the following research question:

How does language use affect linguistic control and executive control?

Chapter 5 proposes the following hypotheses for the investigation of the research question:

The first hypothesis posits that those participants who regularly utilise their two languages within the same contexts will perform better in switch trials and in trials in the mixed-language condition as opposed to those participants who do not: Dual-context bilinguals have honed their switching skills, leading to more streamlined processes for managing conflicting linguistic representations, and, as a result, they are likely to encounter fewer challenges when switching between and mixing languages.

The second hypothesis proposes that indices of language use which predict a change in performance in the linguistic control measures also do so for executive control measures: If language control shares cognitive resources with general executive control, then enhancing language control through substantial language switching should also lead to improvements in participants' domain-general executive control measures.

To investigate the impact of different language use on the interplay of language control and executive control, this analysis considers several measures previously introduced in the introductions to chapters 3 and 4. Namely, Response Time, Switch Cost (+Asymmetry), Mix Cost (+Asymmetry), Language Balance, and inference scores. *Table 4* summarises the variables manipulated in the experiments, as well as their marked and unmarked categories and the measurements they enable.

Table 4: Summary of Variables in Experimental Tasks

Factor	Marked (subtrahend)	Unmarked (minuend)	Metric (difference)
Congruency	incongruent	congruent	interference scores
Task Type	mixed congruency	single congruency	mix cost
Trial Type	switch	non-switch	switch cost
Language	L2 (English)	L1 (Chinese)	language balance

Further to these concepts, this analysis utilized several measures of language use. These were derived measures from an Activity Log Questionnaire and operationalised as components using a principal component analysis. Based on these components, we created seven variables and computed factor loadings for each participant by session using a regression. The principal components include dual-language leisure activities (C1), dual-language academic preparation (C2), dual-language code-switching experience (C3), L1-focused academic support activities (C4), L2-focused social contacts (C5), L1-focused social contacts (C6), and L2-focused core academic activities (C7).

We used the activity components as factors in linear mixed effects regressions to investigate which aspects of language use could predict different performance metrics in language control and executive control tasks. We constructed one model for each of the three reaction-time-based experiments. The components were used to predict different performance metrics, as shown in Table 5.

Table 5: Summary of language control and executive control measures

	Language Control	Executive Control
Overall RT	X	X
Interference Scores		X
Switch Costs	X	
Mix Costs	X	X
Language Effects	X	

In a brief preview of our findings, our results partially support both the Adaptive Control Hypothesis (Green and Abutalebi 2013) and the Cognitive Process Model (Green and Li, 2014). Increased use of L2 in social and academic contexts was linked to better performance on multiple measures of language and executive control.

Overall, our studies' results painted a mixed picture on the bilingual advantage, suggesting that the development of lexical access and executive control is not as straightforward as previously assumed: for instance, our results did not support the anticipated cost of mixing languages, as trials in the mixed language block elicited significantly faster reaction times than trials in the single language block. Furthermore, bilinguals surpassed functionally monolingual speakers only in measures of the Flanker task performance. Interestingly, participants who were more balanced across languages performed better in the executive control tasks as opposed to participants who were more dominant in Chinese. Several factors of language use influenced both lexical access and executive control, including using L2 in academic settings and social contexts.

2. Methodological Details

This chapter outlines the overall structure and design of the study. Section 2.1. Participants offers detailed information on the participants, while Section 2.2. Experiments and Questionnaires presents information on the battery of tests employed and 2.3. Procedure details how the experiments were delivered. 2.4. Metrics presents the metrics we used to operationalise performance measures in language control and executive control and 2.5. Data Analysis presents the inferential methods employed. Details on the individual models are provided in Appendix 1: Model Summaries.

Please observe that every separate empirical analysis employs its own methodology section, outlining participant details and the materials employed for data collection. Consequently, there will be instances of duplication. To minimize redundancy and repetition, readers are directed to relevant chapters whenever deemed appropriate. Within this section, particular emphasis has been placed on aspects that were omitted from individual studies due to spatial limitations or an excess of intricate particulars.

2.1. Participants

The studies in this thesis each employed the same sample of 30 Chinese-English bilingual adult participants, aged between 19 and 32 (mean = 23.5 years). Two thirds identified as female (N=20) and one third identified as male (N=10). All reported having spoken some form of Chinese (26 Mandarin, 3 Cantonese, 1 Taiwanese) as their first language. Most had recently arrived in the United Kingdom to pursue university degrees (LOR = 0.4 years), with six individuals taking part in pre-sessional English courses and sixteen enrolled in an English language class at the time of initial testing. The age of first exposure to English was at 8.5 years of age on average (SD = 3 years). Of those reporting knowledge of third languages, third languages featured include French (7), Japanese (5), Cantonese (5), Spanish (4), Taiwanese (2), Korean (2), German (2). Table 6: Details on the bilingual speakers' backgrounds

describes the English speakers' backgrounds in detail.

Throughout the study duration, the attrition of the bilingual participants was such that of the 30 who took part in the first session, 27 completed the executive control tasks in Session 2 while 21 completed the PNT, and session 3 saw 27 participants complete the full battery of tasks. Data from all available participants were considered for analysis.

Table 6: Details on the bilingual speakers' backgrounds

Part.	Age	Gender	AOA EN (y)	IELTS[4]	C-T I[5]	C-T III[6]	Other L2s
201	26	Male	12	7.5	92	118	
202	24	Male	10	6.5	71	68	
203	26	Male	12	6.5	55	92	
204	21	Female	N.A.	N.A.	85	71	
205	24	Female	6	6.5	32	64	Japanese, French
206	22	Male	10	6.5	50	95	German
207	22	Female	N.A.	N.A.	71	43	
208	23	Female	7	7.0	51	N.A.	Spanish
209	23	Male	11	6.5	68	85	
210	32	Female	N.A.	N.A.	75	59	
211	25	Female	9	7.5	106	68	Taiwanese, Spanish
212	25	Female	10	7.5	105	88	Korean, French
213	23	Male	9	7.5	115	118	Cantonese, French
214	24	Female	7	7.0	75	77	Cantonese, French
215	24	Female	9	7.0	64	88	Japanese
216	22	Female	9	6.5	81	55	French
217	24	Female	10	6.5	99	100	Korean, French
218	22	Female	9	7.0	68	95	Japanese, French
219	22	Male	6	6.0	71	N.A.	
220	24	Female	10	5.5	43	N.A.	Cantonese
221	24	Female	15	7.0	64	78	
222	23	Female	11	5.5	81	63	French
223	21	Female	7	6.0	43	41	Japanese
224	21	Female	N.A.	N.A.	73	105	
225	29	Female	10	7.5	95	99	Taiwanese, Spanish, German
226	18	Male	4	7.5	114	114	Cantonese, French
227	21	Female	12	7.0	41	85	French
228	25	Female	5	6.5	92	83	Japanese, Spanish
229	25	Male	4	7.5	95	105	French
230	20	Male	6	7.0	59	92	Cantonese

[4] Self-reported
[5] Results of English language C-test administered during the first session.
[6] Results of English language C-test administered during the third session.

The functionally monolingual group consisted of 20 English adults from the UK aged 19 to 53 (mean = 26.6 years; SD = 12). There is no significant age difference between the bilingual group and the functionally monolingual group. Thirteen were female and seven male, all native speakers of English with 11 self-reporting prior formal language instruction which they had not used since school days. Table 7 describes the English speakers' backgrounds in detail.

Table 7: Details on the functionally monolingual speakers' backgrounds

Part.	**Age**	**Gender**	**L2**	**AOA L2**	**C-Test**[7]
301	49	Female	French, Italian	11, 18	119
302	46	Male	French	11	99
303	53	Female			137
304	21	Female	French, German	6, 13	109
305	21	Female	French	12	91
306	19	Male			105
307	19	Female	French, Spanish	10, 14	112
308	52	Male	French, German	NA, NA	142
309	21	Female			96
310	21	Female	French, German	12, 15	102
311	21	Male	Welsh	5	122
312	19	Female	French, German	6,11	114
313	21	Female	French	11	119
314	21	Female			122
315	22	Female			77
316	22	Female			102
317	22	Female			122
318	20	Male			114
319	20	Male	English, French	4, 11	102
320	22	Male			114

[7] Results of English language C-test administered during the first session.

The (functionally) monolingual group of English speakers was included to provide a brief exploratory insight into the differences between them and the bilingual Chinese-English speakers. Resource limitations meant that it was not possible to recruit a better matching, and more homogeneously monolingual, participant group. Through the heterogeneity introduced by variables such as the previous L2 exposure reported by just over half of the English native speakers, we are unable to exclude potential confounds in our analysis. This implies that the meaningfulness of any found association is inevitably reduced (Rothmann et al., 2023). Further, as native English speakers only completed the executive control tasks in the first session in this preliminary exploration, we are not able to cleanly interpret any effects of immersion duration for bilinguals, as these may be explained by practise effects. Resource limitations meant the functionally monolingual group only repeated the Picture Naming Task once – in the third session. Ideally, we would have liked to recruit a control group from China. This ideal group would differ from the experimental group only in L2 immersion duration. Age, gender, sociolinguistic background as well as other potential confounds (mentioned in later chapters) would ideally have been matched between the two groups. On top, we would have liked to measure performances of the within the same number of sessions to be able to account for any practise effects.

To estimate the sample size needed we used G*Power (3.1.9.7; Faul et al., 2007, 2009) and data from our previous experiments and literature. Estimating a size effect of 0.85, a fixed $\alpha=0.05$ and applying a 0.67 ratio allocation of participants (bilinguals/monolinguals), 20 and 30 participants would yield $>$ 80% power (82.26%) to detect a significant difference be the experimental groups.

However, given that a larger number of bilingual participants did not complete all three sessions – as described above – we acknowledge that the actual power of the study falls short of our initial expectation and that thus a higher number of participants would have been desirable.

2.2. Experiments and Questionnaires

The present study utilised a battery of three timed tasks. A dual-language picture naming task measured the participants' lexical accessibility and language switch efficiency by evaluating their response times to pictures in both English and Chinese. In addition, two executive control tasks were employed to test the ability to selectively attend to target information and inhibit conflicting information during visuospatial processing. The Simon Task and the Flanker Task have different conceptualizations within the Dimensional Overlap Model (Kornblum et al., 1990): for the Flanker Task, the Dimensional Overlap Model posits that it is an example of a task involving a Stimulus-Stimulus conflict. In such tasks, there is a conflict between items of the same stimulus (i.e., arrows) which originate and are resolved at the perceptual level. In contrast, for the Simon Task, the model posits that the task involves a stimulus-response conflict. In such instances, the conflict arises due to the stimulus dimension (here: stimulus location) being irrelevant to the response rule and the response dimension (Kornblum, 1994; Paap, 2019; Blumenfeld & Marian, 2014). Using these two different kinds of tasks allows us to better understand which kind of conflict monitoring mechanisms are employed in linguistic control.

Participants also completed an Activity Log to explore how they used their languages in everyday life. The research assistant conducting the experiment was a functionally monolingual speaker of English and all instructions were given in English. The experimental battery was created by the research group and drew from different previous studies cited below.

The thesis is organized to feature the battery of tasks in the following manner: the individual articles are dedicated to each exploring distinct facets of this experimental battery. Each subsequent chapter introduces an additional set of tasks. This approach enabled us to initially concentrate on the data derived from the tasks individually and subsequently explore their interactions with data derived from other tasks. Table 8 illustrates the employment of tasks across the three article chapters.

Table 8: Employment of Experimental Battery Across the three Chapters

Experiment	Article 1	Article 2	Article 3
PNT	X	X	X
Simon/Flanker		X	X
Activity Log			X
Background Questionnaire	X	X	X

2.2.1. (Dual-Language) Picture Naming Task

In the picture naming task, participants were asked to name the displayed pictures as fast as possible while coloured frames around the picture indicated whether they should name the item in Chinese (red) or English (blue)[8]. Monolinguals only completed the task in English without any visual cues. To avoid repetition, details on the task are provided in 3.4.3.1. Dual-Language Picture Naming Task.

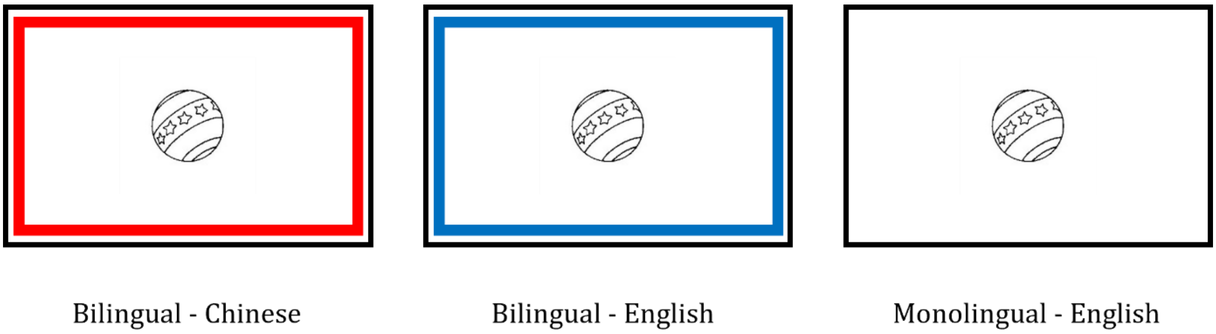


Figure 4: Bilingual and monolingual PNT prompts

The pictures used to elicit naming, their target names, and their corresponding "ID" are provided in Appendix 3: Pictures Used in the PNT. The pictures have been drawn from Bates et al. (2003). Different lists were used for the different blocks (Chinese, English, mixed) and sessions. The order of presentation of pictures was randomised within the lists. The nine lists – one for each of the three blocks and three sessions – are provided in Appendix 4: Lists Used in the PNT By Block And By Session.

[8] Reviewers have noted that the employment of only one colour cue per language results in a confound, whereby, during a language switch, participants have the additional task requirement of processing the change in colour on top of the actual language switch (e.g., Heikoop et al., 2016). A better design would include two or three colours by language, such that a change in colour would also accompany non-switch trials and so lead to an improved comparability between switch and non-switch trials.

2.2.2. Flanker Task

In the Flanker task (Eriksen & Eriksen, 1974), participants are presented with five arrows on the screen with one target arrow at the centre and two non-target arrows on either side. The direction of the non-target items can either correspond to that of the target item (congruent Flankers) or they can show the opposite direction of the target item (incongruent Flankers). Participants in this study are asked to indicate the direction of the target stimulus by respectively pressing either the “3” key or “9” key on the keyboard for left and right directionality, respectively.

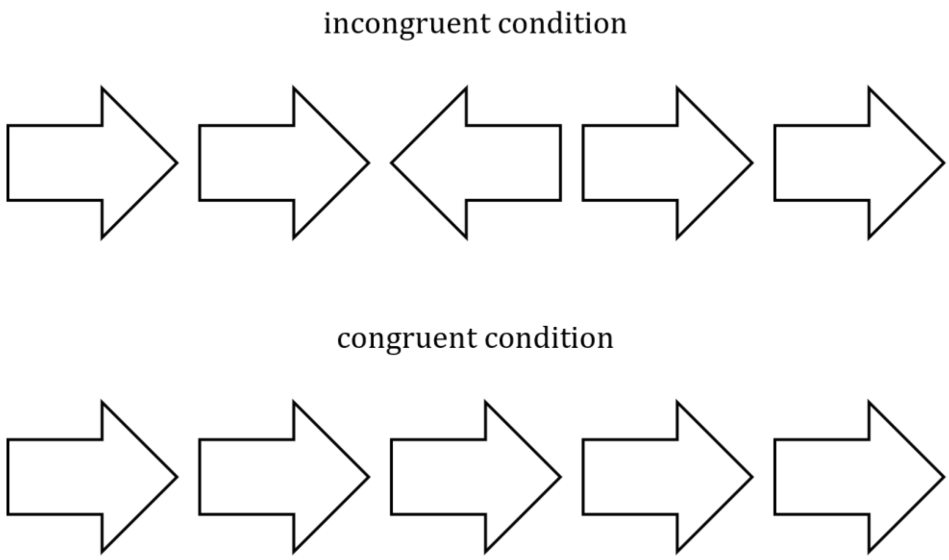


Figure 5: Incongruent and congruent conditions in the Flanker Task

2.2.3. Simon Task

In the Simon task (Simon & Rudell, 1967), participants are required to suppress spatial information and prioritize colour information. They are directed to press the left arrow key upon sighting a red square and the right arrow key upon sighting a green square, regardless of the stimulus side on the screen. Trials are "congruent" when the colour cue appears on the screen side corresponding to the relative position of the response key (e.g., a green square shown on the right side of the screen).

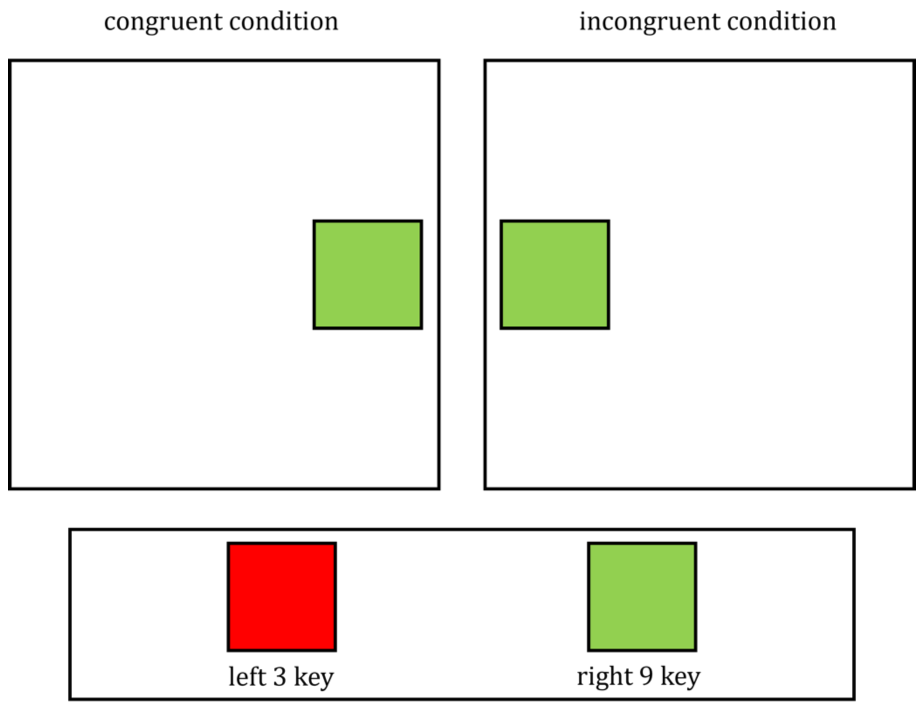


Figure 6: Congruent and incongruent conditions in the Simon Task

2.2.4. Design of timed experiments

All three timed tasks had four components to them (Figure 7). All tasks were preceded by eight mock trials to accustom the participant to the task demands. The second two components of each task were presented as single-task conditions. In the single-task conditions, a block of 20 congruent trials was followed by a block of 40 incongruent trials in the executive control task. In the language control task, 50 Chinese trials were followed by 50 English trials. The final component of each task was a "mixed condition" block, where the previous blocks' trials feature in a semi-alternating manner. The semi-random alternation is such that half the trials were English, and half were Chinese and of those, about half the trials were preceded by a trial of the other language ("switch"), and the other half was preceded by a trial of the same language ("stay"). For the functionally monolingual group, there was no distinction between blocks in the picture naming task, as they were only tested in English.

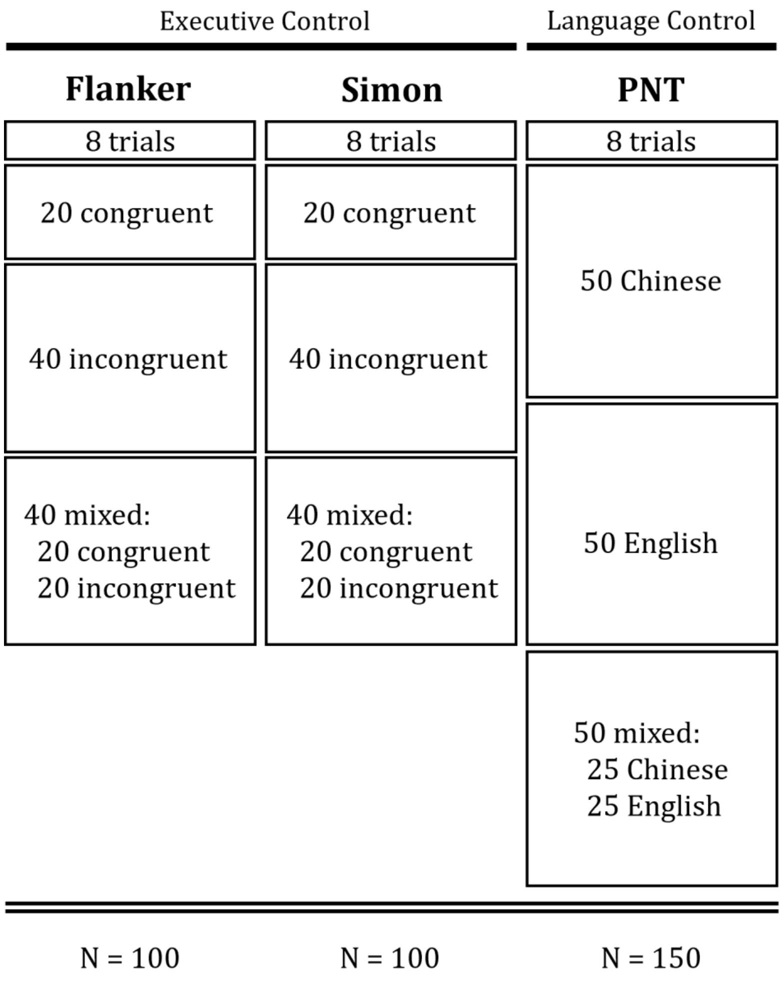


Figure 7: Study Design. N refers to the number of experimental trials.

The order in which the three blocks was presented was fixed throughout the study. Doing so, we aimed to eliminate effects varying block orders may have had. Varying block orders could have introduced a further confounding effect: for instance, if half the participants had completed the cognitively more demanding mixed condition blocks first, then it could have been the case that they performed significantly slower on the the subsequent single condition blocks, as they might have found these blocks comparatively less engaging. The same logic can be applied to the order of presentation of single condition blocks, i.e., alternating between single condition Chinese and English blocks, and congruent and incongruent blocks, within their respective tasks' starting positions; and the order of task presentation (Flanker, Simon, PNT).

We acknowledge, however, that this block order constant may engender further confounding effects, such as practise effects, whereby participants become faster at completing individual trials because of repeated exposures. The individual items were randomised within each block as we do not expect the order of individual items to significantly affect average reaction time performances.

Within the mixed block, there are only 12-13 trials per condition (50 trials divided by 4 conditions: L1 switch, L2 switch, L1 non-switch, L2 non-switch). We acknowledge that this limited number of trials may have implications for the statistical power of our analyses, potentially compromising our ability to detect small or moderate effects reliably (e.g., Brysbaert & Stevens, 2018). The consequences of low statistical power include an increased risk of Type II errors (false negatives), where true effects may go undetected due to insufficient statistical sensitivity. Additionally, low power undermines the precision and reliability of effect estimates, potentially leading to inflated Type I error rates (false positives) and diminished confidence in the validity of our findings. What is more, potential trial exclusions (e.g., due to exceedingly high reaction times, or null responses) may further jeopardize the power of our study[9].

[9] In a foreshadowing to the results, our study did see a relatively high number of trial exclusions (Appendix 2.A: Data Exclusion Rates). We recommend a higher number of trials for future studies.

2.2.5. Activity Log

During each of the three sessions, all bilingual participants completed a written questionnaire inquiring about their daily language usage. The survey encompassed inquiries regarding the number of hours allocated to various activities (e.g., attending classes, watching movies, socializing), along with the percentage of time spent engaged in English or Chinese interactions. The complete list of questions from the activity log questionnaire can be found in Appendix 5: Full catalogue of questions contained in the Activity Log.

2.2.6. Language Background Questionnaire

As part of the initial session, all participants completed a sociolinguistic background questionnaire. The questions covered a range of topics, including date of birth, gender, knowledge of any second language, order of language acquisition, date of arrival in the United Kingdom, language test results, and participation in pre-sessional English classes.

2.2.7. C-Test

The full experimental battery also included English C-tests (Keijzer, 2007) to obtain a measure of participants' English proficiency. The results are presented in 2.1. Participants.

2.3. Procedure

The participants completed the tasks in three separate sessions; the initial session occurred in October 2016, followed by a second one from late January to early February 2017 and culminating with a third one in April 2017. All tasks were conducted in person in the Reaction Time lab of the Department of Language and Linguistics at the University of Essex. The experiment was carried out by a research assistant. The Picture Naming Task was conducted as part of all three sessions for bilinguals, whereas the monolingual controls only undertook the task during the first and last sessions. The monolinguals participated in the executive control tasks only as part of the first session. The language background questionnaire was distributed to both groups at the start of the study. The experiments were presented to the participants on a Windows computer using E-Prime. PNT responses were recorded on a Tascam recorder. The participants completed the Flanker Task first, the Simon Task second, the dual-language Picture Naming third, the C-Test fourth, and the Activity Log Questionnaire, on paper, last. Participants did not receive compensation for their participation. Participants provided informed consent. The study obtained ethical approval from the University of Essex.

2.4. Metrics

Throughout the dissertation, we employ a range of derived metrics. In the following, we provide an overview of these metrics and briefly outline how they have been calculated.

Overall Response Time (RT) represents the time required to process a single trial. It represents how the dependent variable was measured throughout the study. In interaction effects with factors “trial-type”, “condition”, and “congruency”, it serves as the basis for the calculation of switch and mix costs, and inference scores, respectively.

Switch costs reflect the relative difficulty with which new, corresponding mental task sets are adopted in a mixed-task block. They are measured as the difference in RT between similar (non-switch) and different (switch) trails.[10]

Switch Cost Asymmetry shows the relative difficulty with which either language is re-selected following a preceding inhibition. It is calculated by subtracting the average switch costs from one language (in our study, L1), from the other (L2).

Mix costs measure the ability to monitor conflict between tasks and keep two task sets partially activated (Segal et al.; 2021). They are usually measured as the difference between non-switch trials in the mixed block vs. single trials in a single-task block. In our study, we operationalise mix costs for investigating inhibition and competition when the two languages are used interchangeably (mixed block) vs. in more separate contexts (single-language block).

We have introduced the term **mix cost asymmetry** in our study to reflect the discrepancy between mix costs in the L1 and the L2. We have calculated this measure by subtracting the average mix costs in the L1 from the average mix cost in the L2.

Language Balance portrays the relative language access ease, revealing how the bilinguals' languages interact. Following Birdsong (2016), we computed a "between-language subtractive differentials" measure for language balance by subtracting

10 We do not measure **global switch costs** (difference between single-language blocks in two different scenarios: one where the L1 block follows the L2 block, and the other where the order is reversed) and instead test the L1 single-language block at the beginning of each experiment to avoid inflated L1 and allow for a clean comparison of language balance.

performance scores between languages, typically computed through response time averages. This measure in our study serves as a proxy for L2 use relative to L1 and access ease.

Interference scores are a measure of inhibitory control which reveal the relative, added difficulty with which incongruent trials are completed as opposed to congruent trials. They are typically measured as the difference between congruent and incongruent trials. In Chapters 4 and 5, we investigate interference scores in the Flanker and the Simon tasks to evaluate the participants' ability to inhibit conflicting information.

Table 9 summarises the variables manipulated in the experiments, as well as their marked and unmarked categories and the measurements they enable.

Table 9: Summary of Variables in Experimental Tasks

Factor	Marked (subtrahend)	Unmarked (minuend)	Measurement (difference)
Congruency	incongruent	congruent	interference scores
Task Type	mixed congruency	single congruency	mix cost
Trial Type	switch	non-switch	switch cost
Language	L2 (English)	L1 (Chinese)	language balance

Further to the above, Article 2 (Chapter 4) – Language Control Measures Predict Executive Control Performance –includes derived IAV metrics. IAV metrics are useful because they provide a secondary aspect to the data, as opposed to only relying on mean scores (for a discussion of this, see Birdsong 2023; Pfenninger & Kliesch, 2023). For all measures of IAV, we employed Lowie and Verspoor's (2019) coefficient of variation (CV), which is calculated by dividing the standard deviation by the mean of the given sample.

Inconsistency - reflecting the variation within a particular task demand within a single assessment – was calculated using means and SDs of Chinese and English trials in single-language condition (PNT) and congruent trials in the single-congruency condition (Simon, Flanker) – by participant for each session. In so doing, we aim to capture inconsistency within the the least complex trials, i.e., those without any additional task demands (e.g., switching, mixing) to control for any effects thereof.

Variability – reflecting developmental changes – was calculated using **across-session** means and SDs of Chinese and English trials in single-language condition (PNT) and congruent trials in the single-congruency condition (Simon, Flanker) – by participant.

Dispersion - relating to the influence of condition and task demands – was calculated using means and SDs across different task demands (Chinese/English; single/mixed; switch/non-switch; congruent/incongruent) within a single assessment – by participant by session. In so doing, we aim to capture the overall variation across task demands within a task.

Table 10 summarises in which articles the individual metrics were used and whether they were used to capture performance in the PNT – thus language control (LC); or the Simon or Flanker Task – thus executive control (EC).

Table 10: Employment of metrics across articles and tasks.

Metric	Article 1	Article 2	Article 3
RT	LC	LC & EC	LC & EC
Switch Cost	LC	LC	LC
Mix Costs	LC	LC & EC	LC & EC
Switch Cost Asym.	LC	LC	LC
Mix Cost Asym.	LC	LC & EC	LC & EC
Language Balance	LC	LC	LC
Inference Scores	-	EC	EC
IAV metrics	-	LC & EC	-

Further to the above, Article 3 (Chapter 5) – Effects of Language Use on Language Control and Executive Control – includes seven activity components. The principal components include dual-language leisure activities (C1), dual-language academic preparation (C2), dual-language code-switching experience (C3), L1-focused academic support activities (C4), L2-focused social contacts (C5), L1-focused social contacts (C6), and L2-focused core academic activities (C7). The details of how these components were calculated are provided in 5.4.4.2. Activity Log Questionnaire.

2.5. Data Analysis

Details on how the recordings of the PNT were operationalised are provided in 3.4.4. Data Analysis; the pre-processing of Simon and Flanker Tasks are detailed in 4.4.4.2. Flanker and Simon Tasks. 5.4.4.2. Activity Log Questionnaire details the pre-processing steps of the answers obtained from the Activity Log Questionnaire and on the Principal Components Analysis. Throughout the dissertation, we employ a range of derived metrics, which are described above (2.4. Metrics).

Further to the analysis provided in the articles, Appendix 2: PNT Data Audit provides additional analyses on the data validity of the PNT. The results of these analyses are discussed in section of the thesis' discussion.

All three studies employed linear mixed-effects models to examine our hypotheses by using Jamovi and R (The Jamovi Project, 2022; R Core Team, 2021; Gallucci, 2019). The choice of analysis fell on linear mixed-effects models, as they are a popular analysis of choice in the social sciences, as they can be used to examine the effects of both fixed and random factors on a dependent variable. Random factors allow for the consideration of individual variability and captures the hierarchical structure of the data. The formula for a linear mixed-effects model can be represented as:

$$Y = X\beta + Z\gamma + \epsilon,$$

where Y represents the dependent variable, X is the design matrix for fixed effects, containing predictors associated with the fixed effects coefficients β, Z is the design matrix for random effects, corresponding to the random effects coefficients γ, and ϵ denotes the error term, representing unexplained variability (Singer and Willett, 2003).

To construct an optimal model, we incrementally added random and fixed factors while keeping those that significantly enhanced the more complex model. For models using Flanker Task or Simon Task data, we employed the random intercept of "participant" to allow participants to have differing intercepts. For the models employing PNT data, we employed the random slopes "Session|participant" and "Language|item". The former allows the participants to not only have different intercepts, but also different slopes, while the latter allows different items to differ between two languages (Winter, 2013).

We used REML for model comparison. The variables were coded using simple contrasts. Simple coding closely resembles dummy coding, as each level is compared to a reference level. The intercept differs in that in dummy coding, the intercept represents the cell mean of the reference group, whereas in simple coding, it corresponds to the mean of cell means (Statistical Consulting Group, n.d.). In Chapter 5, we coded Congruency and Condition using Helmert contrasts. This meant that the effect size was inverted from negative to positive to facilitate the interpretation of certain interaction effects.

Furthermore, we utilised the Akaike Information Criterion (AIC) to estimate prediction error and performed an ANOVA test to assess the models' fitting relative to one another. We also generated subsets of the complete data sets along with appropriate fixed and random factors to enable us to test our predictions accurately.

Appendix 1: Model Summaries provides a centralised overview of every analysis' models alongside specifics concerning how the models' factors were coded, and the coefficients scaled.

3. The Development of Bilingual Lexical Access

Naming Latencies, Switch Costs and Mixed Costs Within the First Six Months of L2 Immersion

Abstract

Bilinguals have access to two language systems simultaneously and can switch between them with great ease and speed. The present study investigates the changes in bilinguals' language-switching abilities as a function of immersion duration. Thirty Chinese-English bilinguals completed a dual-language picture-naming task at three testing times during their first year in the United Kingdom. Linear mixed-effect regression models provided mixed evidence for improving bilingual naming latency as a function of immersion duration. While overall naming latency decreases, the decrease is more salient in the bilinguals' L1 than in their L2. Language balance had a significant impact on several measures of language control, but it did not significantly change over the testing period.

3.1. Introduction

When native speakers of one language immerse themselves in the linguistic environment of another language, the input and use of their first language (L1) decreases and small changes in their language use are sometimes observed. For instance, speakers may find it harder to access lexical items, become more disfluent in their speech, or experience cross-linguistic influence from the new language (L2). This development is referred to as "first language attrition"; it affects different aspects of language to varying extents, in varying order and rates (Sorace, 2011, Schmid, 2011; Schmid and Köpke, 2017; Paradis, 2007) and is governed by external factors, such as length of residence, age of emigration and exposure to either language (Schmid et al., 2022).

It is a common assumption that the lexicon and the speakers' access to it are particularly sensitive to language attrition, changing first, fast and to greater extents as opposed to areas such as phonetics or morphosyntax. For instance, Schmid and Köpke (2009) find that the decline in the ability to promptly retrieve and recognize words in the first language is the most immediate effect of second language immersion in both experimental settings and in free speech. Paradis (2007) further argues that the decline in lexical retrieval ability is directly linked to L1 disuse.

To date, however, there are no longitudinal studies that map the attrition process of the lexicon in its initial stages. As such, a systematic understanding of how exactly immersion duration contributes to lexical retrieval difficulties is still lacking. Of particular interest are two questions: first, **when exactly** the small-scale, real-time processing differences accrue to an extent at which differences between monolingual and bilingual lexical access are discernible. Second, **how gradual or linear the changes in lexical access are**. As such, investigating the impact of immersion duration on lexical attrition is crucial in understanding which changes occur and when.

The current study investigates patterns in the development of bilinguals' lexical access and switching ability as a function of immersion duration. We examine the effect of L2 immersion on lexical access and code-switching by investigating performance in a cohort of adult Chinese-English bilinguals in their first year of residence in the United Kingdom. Using a dual-language picture naming task, we measure overall response times (RT), switch and mix costs in their Chinese L1 and English L2 across three sessions spanning nine months in total. The first session is conducted at the time of the participants' first arrival to the UK, the second after three months, and the third after six months. We control for several linguistic and extralinguistic factors and conduct a range of linear mixed-effect models.

3.2. Literature Review

Several differences in learners' language skills have have previously been reported following increased exposure to a L2, such as in a study abroad or other L2 immersion context: Oral production skills (e.g., Davidson, 2010; Freed, 1995) as well as the acquisition of vocabulary (Dewey, 2008; Foster, 2009) are believed to be the skills that benefit the most from an immersion experience. In contrast, in other areas such as morphosyntax (e.g., Collentine, 2004; DeKeyser, 1991) and phonology (e.g., Díaz-Campos, 2004; Mora, 2008), it has been reported that L2 learners do not experience significant gains after a period abroad. Positive outcomes as a result of an immersive L2 experience have been documented in the area of pragmatics and sociolinguistics (e.g., DuFon & Churchill, 2006; Félix-Brasdefer, 2004).

In the past, most studies have focussed exclusively on second language learners. These studies investigate, for instance, the cross-linguistic influence of the L1 lexicon on the developing L2 lexicon (e.g., Pavlenko, 2009, Ringbom & Jarvis, 2009). On the other hand, fewer studies have considered the impact of the second language on the first. In the past, the first language lexicon has been considered as mostly stable linguistic knowledge, which is not prone to significant changes. More recently, however, psycholinguistic and neurolinguistic studies have revealed measurable differences in lexical access to the L1 in bilinguals, supporting a notion that the L1 lexicon and its access are permeable to external influences (e.g., Steinhauser & Kasparian, 2020; Pierce et al., 2014; Misra, Guo, Bobb, & Kroll, 2012; see Köpke & Keijzer for an overview).

This chapter provides a brief introduction into various theories on bilingual lexical access, namely, the Inhibitory Control Model (Green 1998), the Revised Hierarchical Model (Kroll and Stewart 1994), the Activation Threshold Hypothesis (Paradis, 2004; 2007) and finally, the Entrenchment Model (HJ Schmid, 2017). I will discuss implications for first language attriters and discuss the development of bilingual lexical access from its starting point, the monolingual lexicon.

The traditional models of lexical access comprise four phases: conception, grammatical encoding (lemma-level), phonological encoding (lexeme-level), and articulation (Levelt, 1989). While Levelt (1989) argues for a modular top-down process, Dell (1986) supports an interactive and cascaded process.

Levelt and Dell's models introduce useful concepts for considerations on bilingual lexical access. For instance, the distinction into different levels of processing leave space for the following questions: is there a level at which language is selected, at which point only target-language lexemes are activated? Or do the two language systems share activation at the lexical level?

Numerous studies show that two language systems become co-activated during production and comprehension (e.g., Kroll et al., 2006, Misra, Guo, Bobb, & Kroll, 2012, Green, 2011), highlighting the need for an inhibitory mechanism at a "local" level (de Groot & Christoffels, 2006). However, previous literature has also shown costs associated with up- and downregulating language systems "globally" (de Groot & Christoffels, 2006), suggesting that there is an overarching mechanism for language schema (e.g., Meuter & Allport 1999; Costa & Santesteban, 2004; Campbell, 2005).

The kind of inhibition applied is thought to depend upon different factors, such as the type of linguistic task at hand and the speaker's proficiency in either language. For instance, an individual who would speak English at their workplace and German with their family could globally upregulate German and downregulate English when returning home. By contrast, a simultaneous translator, who needs to frequently switch between languages, cannot globally downregulate either language and would need to locally inhibit non-target representations instead.

The theoretical framework supporting the notion of inhibition in bilingual access is Green's (1998) Inhibitory Control Model. He proposes that language production is akin to non-linguistic activities which are either routine (e.g., the native L1) or non-routine (e.g., the newly acquired L2). There is competition for selection between them, as the non-target lexicon becomes co-activated during processing, which, in turn, necessitates its active inhibition. (Green, 1998; Kroll et al. 2006).

The status of the two lexica is generally considered unequal, with the L1 lexicon sporting comparatively stronger conceptual links, the assumption is that the L1 will require a larger inhibitory effort to be actively suppressed to facilitate L2 production. Support for increased L1 inhibitory requirements is evidenced by asymmetric inhibition costs. Meuter and Allport (1999) found that bilinguals were significantly slower naming L1 items when they followed L2 items as opposed to the other way around. The authors concluded that, in this scenario, the bilinguals recruited more inhibitory resources to actively suppress the L1 during the L2 speaking block, leading to a measurable lower performance when prompted to select it in the following block.

Kroll and Stewart's Revised Hierarchical Model (Kroll & Stewart, 1994; Kroll & Ma, 2017) incorporates the notion of competition between the two linguistic systems at the lexical level. However, they propose that this competition emerges during bilingual development. They propose that at the beginning of bilingualism, there are no connections linking the conceptual system to the L2 lexicon, and speakers depend heavily on L1 translation equivalents when processing words in their L2. Linking this idea back to Dell's model, single L2 lexemes would link to their L1 translation equivalents in a similar manner to synonyms, which have specific tags and would be selected depending on, for instance, the register of the conversation.

Without a direct conceptual link, L2 access would be delayed. However, with no direct competitor at the level of lexical selection, there is no need for the inhibition of the L1 (Kroll & Ma, 2017). As such, the L1 lexicon is still overwhelmingly active and readily available to the speaker, while the L2 faces a temporal delay in access. As the exposure to the new linguistic environment and use of the L2 lexicon increase, Kroll and Stewart (1994) suggest that a direct conceptual link is developed, albeit weaker than that of the L1.[11] With the development of a second conceptual link, the need for selection between the two lexica arises.

As discussed before, the introduction of competition between the systems of unequal standing would incur asymmetric switch costs. However, with increased L2 proficiency, studies on cued language switching find that asymmetric inhibition costs reduce (Costa & Santesteban, 2004; Costa et al. 2006; Meuter & Allport, 1999). This could be either i) because the active inhibition of the L1 is only a temporary measure of language control until L2 representations become more stabilized, or ii) because the L2 too needs to be

[11] At the same time, a weak lexical link is established from the L2 to the L1, indicating that some L2 translation equivalents are now used to inform concepts in the L1.

inhibited more if it's more dominant, leading to more comparable L1 and L2 inhibition levels, and thus symmetrical switching costs

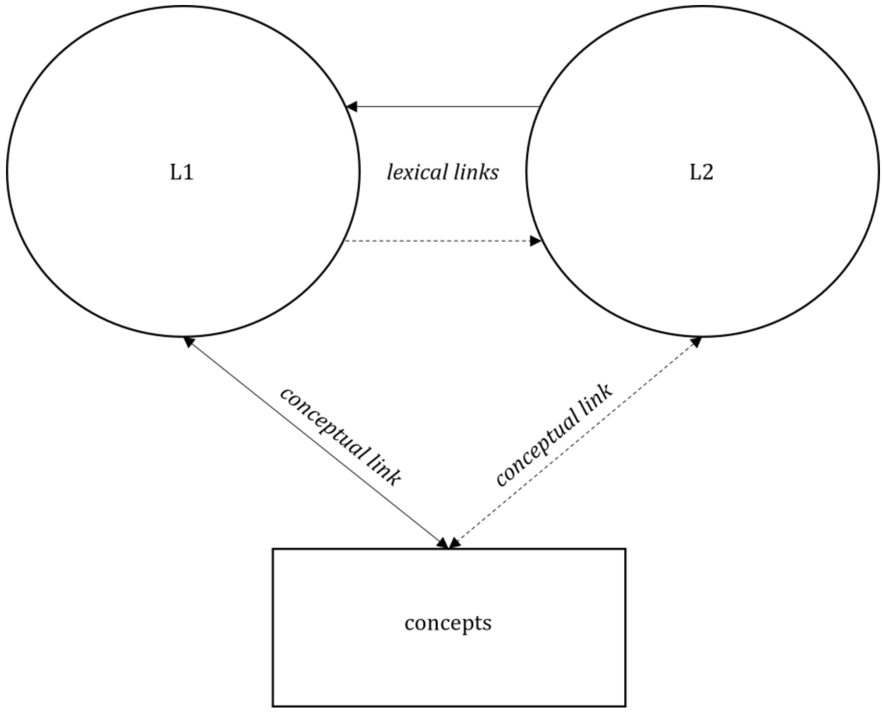


Figure 8: Revised Hierarchical Model of bilingual lexical access.

While the Revised Hierarchical Model can account for changes in bilingual proficiency over time, it falls short on making a statement on the exact process by which the L2 lexicon is separated from its L1 counterpart and, thus far, empirical evidence to support such a transition is lacking. Further, the model lacks details on how the two language systems differ from a monolingual language system.

Bilinguals have repeatedly been shown to perform certain tasks more slowly than their monolingual counterparts, e.g., Kilborn (1989) in the processing of sentences, Mack (1983) in naming and lexical decision and Mägiste (1979, 1985) in naming, recall and matching. It is as such largely acknowledged that bilinguals are not "just" the sum of two monolinguals.

In balanced bilinguals, longer response times may be indicative of non-selective lexical access (e.g., De Bruijn et al., 2001; Green, 2003), or of difficulty in managing two linguistic systems in real time (Green, 1998). Longer response times in unbalanced individuals are often encountered in the weaker language and are therefore often used as a proxy for language balance. In late bilinguals, L2 response times are often longer than L1 ones. However, many studies have found that L1 response times may eventually slow down while those of the L2 become faster, suggesting a reversal of this pattern (Schmid & Jarvis, 2014).

A recurring point of discussion in recent research on the attrition of L1 lexical access is the question whether changes to L1 lexical access are due to competition effects, decreasing exposure/use frequency, or a combination thereof. In this chapter we will discuss the Activation Threshold Hypothesis (henceforth: ATH; Paradis, 2004; 2007) and the Entrenchment Model (HJ Schmid, 2017) and their implications on this debate.

The ATH predicts that language disuse causes its attrition, facilitating access to the most frequently used items of the L2 and making it harder to activate their lesser used L1 equivalents. The mechanism which regulates upkeep and loss does so through lowering and raising activation thresholds – the amount of neural energy necessary to access a piece of information - which are associated with all items across the speakers' languages.

Low activation thresholds facilitate retrieval, while high activation thresholds impede it. When any item is used, the activation threshold associated with it is lowered, meaning a subsequent use of the item will entail less neural effort. With disuse, activation thresholds will increase again over time. As such, frequently and more recently used items will have lower activation thresholds, whereas disused items will have higher thresholds and will be more difficult to retrieve, as their retrieval entails a larger neural effort. In this view, a bilingual speaker immersed in an L2 environment will gradually face more effortful access routes to their L1 vocabulary as a result of higher activation thresholds of these items. Studies have shown that immersion in an L2 environment exerts significant pressure on L1 retrieval ability, as even short periods of L2 immersion led to a measurable change in L1 access times (Schmid & Yılmaz, 2021; Yılmaz & Schmid, 2012; Schmid & Jarvis, 2014).

Extending the ATH, it could be argued that bilinguals would eventually lose all access to L1, as long-term immersion in an L2 environment would critically raise activation thresholds to an extent which would make L1 retrieval an increasingly difficult task. In a computational psycholinguistic approach, Meara (2004) simulated lexical attrition using a random autonomous Boolean network model. In various scenarios mimicking attrition events, all models predict an eventual catastrophic cascade, where the deactivation of some words inevitably leads to the dramatic decrease of access to most words. However, even long-term immigrants tend to retain a good command of their L1, pointing toward a somewhat special status of L1 lexical knowledge.

The Entrenchment Model (HJ Schmid, 2017) argues that through frequent use, linguistic knowledge can become entrenched, which renders it less permeable to change and allows it to be represented more "holistically" and processed more automatically as opposed to less frequently used linguistic knowledge (Steinkrauss & Schmid, 2017). In other words, entrenchment means that such knowledge is "anchored more deeply" in the mental framework, similarly to how certain muscle groups receive some sort of muscle memory upon being trained sufficiently (e.g., Krakauer & Shadmehr, 2006): the resulting memory trace allows the trained action to be repeated with less conscious effort – even after longer periods of disuse. In monolinguals, the L1 is thought to be heavily entrenched in the mind, both due to it being the first one to be acquired, and it being used most frequently. As such, the L1 enjoys a special status. In the Unified Competition Model (MacWhinney, 2005), L1 entrenchment is seen as a risk factor to L2 acquisition, as it impedes the establishment of the new system in the mind.

With decreasing frequency of L1 use, however, it is argued that elements may face disentrenchment. Disentrenchment is when previously entrenched knowledge gradually loses its special status and is processed less automatically as a result (Steinkrauss & Schmid, 2017). The effect of this can be understood as language attrition, which Köpke & Schmid (2004, p.5) define as "the non-pathological decrease in a language that had previously been acquired by an individual" To date, it is unclear how this deterioration proceeds and whether this process can be solely attributed to disuse.[13]

For instance, in a longitudinal study, Baus et al. (2013) explore naming latencies and verbal fluency in a group of German university students during a semester abroad in Spain. The experiments were run upon the students' arrival to Spain and at their departures. Baus et al. (2013) find that participants became significantly slower at naming non-cognate items in their native language at the end of the immersion period but found no differences between the two testing points in the verbal fluency task. In another study, the verbal fluency task is found to be sensitive to the degree of immersion: Linck et al. (2009) find that as immersed L2 learners were significantly

[13] For a discussion on competition versus frequency effects in language attrition, refer to Schmid & Yılmaz (2021).

more impaired in producing L1 items in a verbal fluency task compared to their peers, who only learned L2 in the classroom in their L1-dominant environment.

From the above review of literature, four to five distinct stages in the development of bilingual lexical access can be identified, which each have different implications for response times in either language (Table 11). The acronyms in brackets reference the theories discussed above.

Following the classification of the development of the bilingual lexicon into different stages, several questions arise:

- When can we expect each stage to take place?
- Do stages overlap?
- Do stages follow in a linear manner?
- When do competition effects take place?
- How do global vs. local inhibition compare over time?

Longitudinal designs allow for the examination of changes to lexical accessibility through different stages of the bilingual experience. However, few studies provide information on the length of residence, age, or age of acquisition of bilingual participants, or employ repeated measures design to compare the individual development of bilingualism, lexical accessibility, and inhibitory capacity. As a result, it is difficult to draw any conclusions on the time course of changes in language control.

Table 11: The timeline of bilingual lexical development.

Stage	RT L1	RT L2	Description
0	low	-	One lexicon means there is no competition between language systems. Lexical items are readily available.
1	low	high	(RHM:) The first L2 items are learned “attach to” their L1 translation equivalents. (IC, EM, ATH:) The competition starts here, but the L1 is still overwhelmingly active, causing delays in L2 retrieval.
2	high	medium	(RHM) The L2 lexicon has formed its own conceptual links. There is now competition at the level of selection. The routine (ICM), heavily entrenched (EM) L1 with still relatively low activation threshold (ATH) must be actively suppressed to allow for the selection of the non-routine L2, resulting in higher L1 RTs in switch conditions.
3	medium	medium	The L2 lexicon has established itself as a routine (ICM) language, but competition between the language systems remains, so RTs will be slower than monolingual RT. Asymmetric switch costs may arise a result of the L1 having entrenched structures, which cause the competition to be unequal.
4	high	medium	Disuse of the L1 will eventually lead to the decoupling (EM) and raising of activation thresholds (ATH) of L1 items, which will become more difficult to access, resulting, once again, in higher response times. Having become the default language (ICM), it is now the L2 which needs to be inhibited to facilitate L1 production, leading to a reversed switch cost asymmetry.

3.3. The Current Study

The different theoretical frameworks explored above provide testable hypotheses for different stages in the development of the bilingual lexicon. Currently, however, these frameworks do not specify specific timeframes in which we may expect such changes to take place. Other studies on bilingual lexical access tend to take bilinguals into consideration, who have lived several years in the L2 environment already (e.g., 3.7 years in Bonfieni et al. (2019)'s study). To our knowledge, however, no previous literature has so far considered bilinguals in their first months upon immersion in a L2-dominant environment in a longitudinal approach.

The present study investigates the development of L1 vs. L2 lexical accessibility as a function of immersion duration in the L2. Similarly to recent previous studies (e.g., Bonfieni et al., 2019), we explore overall response times as well as switch costs and mix costs as proxies for relative inhibitory control effort on a dual-language picture naming task in a group of Chinese-English bilinguals as part of a longitudinal study spanning six months. At the beginning, participants had just started their university studies in the UK. The second and third sessions capture the development after three and six months, respectively. We use various measures to control for extralinguistic factors, such as age and language proficiency, and to account for further individual variation. Monolingual performance proxies are provided by matched controls.

In the dual-language picture naming task participants are instructed to name pictures in each language separately in single-task blocks and to name pictures in one of the two languages in a pseudo-randomized manner in the mixed task block.

In this way, it is possible to gather different measures of language accessibility and switch efficiency for either language:

- **(Overall) response time** (RT) reflects the overall average latency needed to process trials in various conditions. It is a useful measure in itself and is used in the calculations in the following measures.
- **Switch costs** reflect the relative difficulty with which new, corresponding mental task sets are adopted in a mixed-task block. They are measured as the difference in RT between similar (non-switch) and different (switch) trails[14].
- **Mix costs**[15] are a measure of the ability to monitor conflict between tasks and keep two task sets partially activated (Segal et al.; 2021). They are usually measured as the difference between non-switch trials in the mixed block vs. single trials in a single-task block. In our study, we operationalize mix costs for the investigation of inhibition and competition at local level (mixed block) vs. at global level (single-language block).
- **Language Balance** describes the relative ease of access to one language as opposed to another and aids in understanding how the bilinguals' two languages interact with each other. It is usually calculated by subtracting the average response time for one language from the average response time of the other.

[14] We do not measure **asymmetric global switch costs** (difference between single-language blocks in two different scenarios: one where the L1 block follows the L2 block, and the other where the order is reversed) and instead test the L1 single-language block at the beginning of each trial to avoid inflated L1 and allow for a clean comparison of language balance.

[15] Segal et al. (2021) found that mix costs are typically higher than switch costs and, as such, create more opportunity for variability. This increased variability allows for a better comparison of mix costs between sessions.

3.3.2. Hypotheses

We address the following research questions and hypotheses:

How do naming latency, switch costs, and mix cost develop over time in the L1 and the L2, respectively?

We hypothesize that i) within the first weeks after arrival in an English-speaking environment, bilingual participants will be faster to name items in their Chinese L1 as opposed to their English L2, as the L1 is still the most used and active. L2 naming will be slower as L2 conceptual links are only being developed.

Further, we hypothesize that ii) in the following sessions, naming latencies in Chinese and switch costs into Chinese will increase, as links to the Chinese lexicon are actively suppressed to facilitate L2 naming, leading to asymmetric switch costs. At the same time, naming latencies in English and switch costs into English will decrease, as bilinguals will establish increasingly stronger links with the L2. The effects will be most pronounced in the third session, as we hypothesize immersion duration to be a function of proficiency.

As for mix costs, we hypothesize that iii) single-language blocks will consistently incur smaller naming latencies than mixed-language blocks, as single-language blocks allow for the global inhibition of the competing language and is likely to be employed as efficient strategy. As immersion duration increases, we argue that mix cost decreases as bilinguals will become better at locally inhibiting the intrusive language. We hypothesize that when mixing languages, the more dominant L1 will be more inhibited than the L2, leading to longer L1 naming latencies. As immersion duration progresses, we hypothesize that this asymmetry decreases.

Finally, we propose that vi) bilinguals will be consistently outperformed by monolinguals, as bilinguals confront an additional processing burden posed by competition effects (e.g., Green, 1998).

3.4. Methods

3.4.1. Apparatus

A dual-language picture naming task was presented to the participants on a Windows computer using E-Prime with a Serial Response Box.

3.4.2. Participants

A total of 30 Chinese-English bilingual adults participated in this study, ranging from 19 to 32 years of age (mean = 23.5 years). All Chinese-English bilinguals reported speaking some form of Chinese (26 Mandarin, 3 Cantonese, 1 Taiwanese) as their first language. Twenty participants identified as female, and ten as male. Most had only just arrived in the United Kingdom in the month prior to the first testing session to pursue university degrees. Six individuals took part in pre-sessional English courses, and sixteen stated that they were enrolled in an English language class at the point of first testing. C-tests revealed that participants had comparable proficiency levels.

Throughout the study duration, the attrition of the bilingual participants was such that of the 30 who took part in the first session, 27 completed the executive control tasks in Session 2 while 21 completed the PNT, and session 3 saw 27 participants complete the full battery of tasks. Data from all available participants were considered for analysis.

The control group comprised 20 English functionally monolingual adults between the ages of 19 and 53 years (mean = 26.6 years) from the United Kingdom, of which thirteen identified as female and seven identified as male. All control participants have English as their first language. Eleven stated they had had some formal language education at school that they had not made use of since.

3.4.3. Procedure

All participants completed a picture-naming task and a language background questionnaire as part of a study encompassing a battery of different tasks, including a Simon task and a Flanker task and an Activity Log.

Participants completed these tasks during three temporally distinct sessions. The first session took place at the start of the academic year in October 2016, the second one in late January and early February 2017, and the third and final one took place in April 2017. Bilinguals performed the picture naming task at all three sessions. Monolingual controls completed the picture naming task in sessions one and three. The language background questionnaire was administered to both groups at the time of first testing.

3.4.3.1. Dual-Language Picture Naming Task

In the dual-language picture naming task, bilingual participants were shown a picture and asked to name the item in one of two languages as fast as they could. Responses were given orally and recorded as a sound file. The task was presented to the participants on a Windows computer using E-Prime with a Serial Response Box and voice trigger. A coloured frame around the picture indicated whether the item should be named in Chinese (red) or English (blue). The experiment was split into four consecutive blocks: a training block consisting of 8 trials (4 Chinese, 4 English), a Chinese-target block of 50 trials, an English-target block of 50 trials, and a mixed block with both Chinese and English targets alternating semi-randomly across 50 trials. The semi-random alternation is such that there are 25 trials of either language in an order whereby about half the trials were preceded by a trial of the other language ("switch"), and the other half was preceded by a trial of the same language ("stay"). The order in which the three blocks was presented was fixed throughout the study. Doing so, we aimed to eliminate effects varying block orders may have had (see 2.2.4. Design of timed experiments). For monolinguals, there was no such distinction between blocks, as they were only tested in English, and no visual cue to response language was included. Each trial begins with an acoustic signal of 100ms, with the onset of the presentation of the visual stimulus occurring 500ms after the onset of the sound cue. The self-paced trial ends with the participant pressing the space bar.

The pictures for the task were deployed randomly across the experimental blocks and were selected from a battery composed of a total of 512 pictures. The battery is based on Bates et al.'s (2003) comparison of timed picture naming in seven languages, which provides information about differences in naming latencies and word frequencies cross-linguistically. Further, the word frequencies from this task were used to control for word frequency effects in our experiment for both English and Chinese.

3.4.3.2. Language Background Questionnaire

All participants completed a language background questionnaire as part of the first session. Questions included date of birth, gender, and knowledge of any second language. For the bilingual group, questions also included the date of arrival in the United Kingdom, order of acquisition of languages, any language test results (IELTS, TOEFL, university-internal language examinations), and whether they had taken part in any pre-sessional English classes (and if so, how many hours per week, and for how long), attendance of language classes (and if so, how many hours per week, for how long, and what kind of exercises are involved).

3.4.4. Data Analysis

Using the voice recordings collected during the experiment, we manually measured reaction times and labelled responses in Praat (Boersma & Weenik, 2018). Reaction times were measured from the onset of the acoustic signal to the onset of the participant response. 500ms were subtracted from the measured time to reflect the difference in time between signal onset and visual stimulus onset.

Responses were coded as follows: response language was coded "ENG" for English responses, and "CH" for Chinese responses. Item responses were coded "0" for any missed trials, i.e., where no response was attempted by the participant, "1" when the response exactly matched the target item, "2" when the response was deemed an appropriate synonym for the target (e.g., "kitten" for "cat"), "3" for responses which were semantically further removed from the target, for instance, distantly related responses (e.g. "dog" for "cat") or hypernyms (e.g., “animal" for "cat") and "4" for any responses that did not resemble the target in any way. Language accuracy was coded as "1" where the response language fit the target language, and "0" where it did not. Item accuracy was coded "1" where the item response was previously coded as either "1" or "2", and else "0". Trials were coded as "non-switch" when the previous trial aimed to elicit a response in the same language as the current trial, and "switch" where the previous trial target language differed from the current one.

The reaction times was normally distributed. We levelled reaction times higher than the mean plus two standard deviations (2770ms) and removed data from the first trial in each block as well as trials with reaction times below 250ms. We retained inaccurate and missed trials (i.e., those where language or item accuracy was coded "0"), as well as trials with RTs beyond 4000ms for the analysis of excluded items but removed trials of this description from all other analyses. Trials succeeding an excluded item retained their previously coded trial type to maximise on available data. A comparatively large number of trials were excluded from analysis: 24.1% of all bilingual trials and 10.4% of all monolingual trials (Appendix 2.A: Data Exclusion Rates). Comparable studies only excluded around 5% of their data (e.g., Bonfieni, 2019).

We calculated switch cost for bilinguals for each session and language by subtracting the average response times of the switch trials of the mixed block from the average reaction time of the non-stay trials (see Declerck and Philipp, 2015).

We calculated bilinguals' mix costs for each session and language by subtracting the average response times in the non-switch trials in the mixed-task task type and trials in the single-task task type of the corresponding language.

Following Birdsong (2016), we calculated a measure for language balance by calculating "between-language subtractive differentials", which entails subtracting performance scores in either language. In this study, we calculated bilinguals' overall response times average in single-task blocks for each session and in each language and subtracting the average reaction time for the Chinese single-task block from the average reaction time for the English single-task block in each session - for all participants individually. A perfectly balanced bilingual would average a score of zero. A positive deviation from 0 indicates Chinese dominance, as Chinese items are named faster as opposed to English items, while a negative deviation from 0 indicates English dominance, as Chinese items are named slower as opposed to English items. Dominance does not reflect proficiency.

We ran a series of linear mixed effects models to test our hypotheses using Jamovi and R (The Jamovi Project, 2022; Gallucci, 2019; R Core Team, 2021). In addition, we ran a generalized mixed effect model for the analysis of excluded trials. For each analysis, we constructed an optimal model by adding random and fixed factors on a step-by-step basis, retaining factors only when the more complex model is significantly superior to the reduced model. Subsets of the full data sets were created, and fixed and random factors were adopted to specifically test the predictions we have set out. We used the Akaike Information Criterion (AIC) as an estimator of prediction error and conduct an ANOVA test for the models' goodness-of-fit relative to one another.

We split the dataset in different ways to construct interpretable models to provide evidence for our hypotheses. First, to test how monolinguals compare to bilinguals in both their Chinese L1 and their English L2, we created a subset that only includes data on the single-task task types. This allows us to compare "pure" vocabulary retrieval times for two groups without the added difficulties brought about by language switches in the mixed-language task. For this analysis, we created the variable "GroupLanguage", which divides the bilingual group's trials further into "Bilingual-Chinese trials" and

"Bilingual-English trials" and facilitates interpretation between the bilingual's naming times in either language and the monolingual control's naming times. As monolinguals did not complete the picture naming task in session 2, we exclude this session for this analysis. Factors tested in this analysis included session, Group Language, frequency[16], Percentage of Excluded Trials and interaction effects.

Second, to explore the cost of switching between the languages, we created a subset that includes only the bilingual mixed-task results and no single-task results. Doing so permits us to focus on the effects affecting overall mixed-task RT, and specifically, switch costs (the effect of trial type (switch vs. non-switch)) and factors affecting the latter. For this analysis, we entered language, language balance, frequency, session, and trial type and interaction effects into the model to optimize it.

Third, we investigate mix costs and create a subset of bilingual data, which excludes all switch trials and only retains non-switch trials in both mixed-task and single-task task types. Doing so allows for clean testing of the effect of task type (mix cost: mixed-task vs. single-task) without the added difficulty posed by switch trials in the mixed-task task type. This model was optimized by factors including Tasktype, Language, Language Balance, Frequency, Percentage of Excluded Trials and interaction effects. The Percentage of Excluded trails was added as a proxy for task involvement.

For all the above analyses we employed *Participant|Session* as a random factor, which allows for the participants' performance to vary over session, and *Item|Language*, which lets the random effect of the item vary across languages. Both random factors significantly contribute to the model fit in either analysis.

Fourth, to illustrate the interplay of switch cost asymmetry, mix cost asymmetry and language balance, we included analysis which takes language balance as dependent variable and session, switch- and mix cost asymmetry as predictors, and participant ID as a random factor.

[16] The frequency measure was based on Bates et al. (2003).

3.5. Results

3.5.1. Single task/ Overall RT

Analysing performance in the single-language conditions allows for meaningful comparisons to be made between bilingual and monolingual response times and provides an overview of the overall bilingual development. Thus, we first report and analyse the single-task data, i.e., excluding any switch trials. Figure 9 shows the bilinguals' performance of response times with respect to the monolinguals' overall performance.

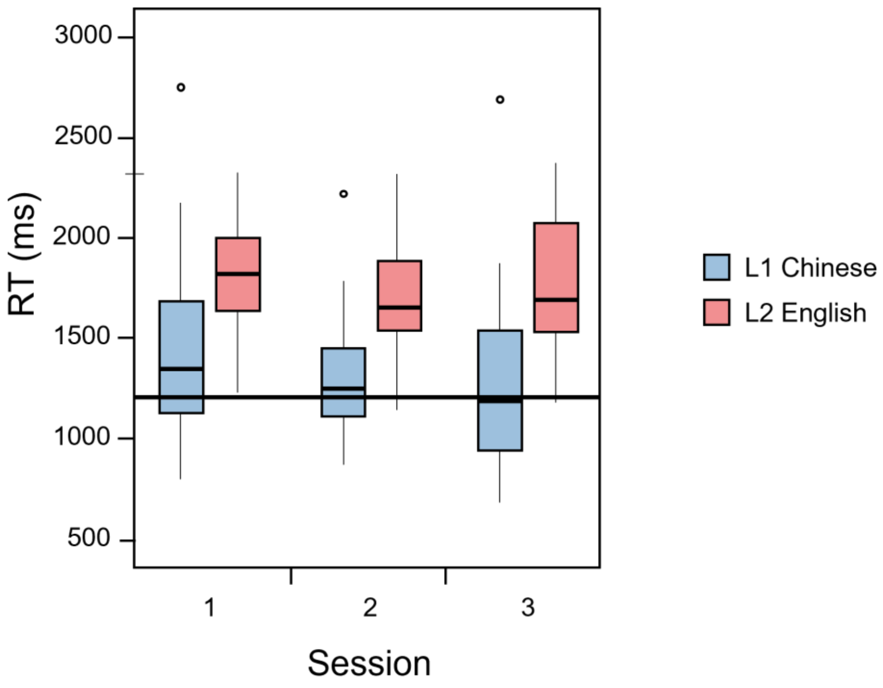


Figure 9: Average RT in the single-language PNT blocks by language and Session[17]. Whiskers reflect extreme values that are no outliers (within 1.5x IQR).

[17] The line at RT=1186ms represents the L1 English speakers response time average.

As L1 English speakers did not partake in a picture naming task in Session 2, we exclude all Session 2 data from the current analysis to facilitate modelling differences between the functionally monolingual speakers and bilinguals.

Formula: (lm, REML) RT ~ 1+ GroupLanguage*Session + frequency + percentage of excluded trials + (1+Session|participant) + (1+GroupLanguage|item), data = subset of non-switch trials, excluding Session 2

Table 12: Model for bilingual and monolingual single-task naming latencies.

Fixed effects	Estimate	Std. error	P-Value
(Intercept)	1435.1	36.5	< .001
Session 3	-87.8	28.7	<.004
GroupLanguage L2 English	508.4	75.4	< .001
GroupLanguage L1 Chinese	26.0	75.6	0.73
frequency	-25.7	11.7	<.03
Percentage of excluded trials	1171.7	176.6	< .001

The analysis of overall naming latencies revealed a significant main effect of Session ($p=0.01$), revealing that naming latencies became shorter between Sessions 1 and 3 ($\beta=-87.8$, $SE=28.7$, $p=0.04$), as well as a main effect of GroupLanguage ($p<0.001$), as bilinguals had longer naming latencies than monolinguals in their English L2 ($\beta=508.4$, $SE=75.4$, $p<0.001$) but not their Chinese L1 ($\beta=26$, $SE=75.6$, $p=0.73$), and a significant main effect of frequency ($p<0.05$), which indicates that more frequent words were named faster. The inclusion of %excluded trials significantly improved the model ($p<0.001$) and shows that those participants who had the most difficulty with the task also recorded larger naming latencies ($\beta=1171.7$, $SE=176.6$, $p=<.001$).

3.5.2. Switch Cost

Switch cost is calculated by subtracting the average response time in switch trials from the average response time in non-switch (or "stay") trials in the mixed-task condition. Figure 10 shows the development of average RT of switch and non-switch trials in Chinese and English.

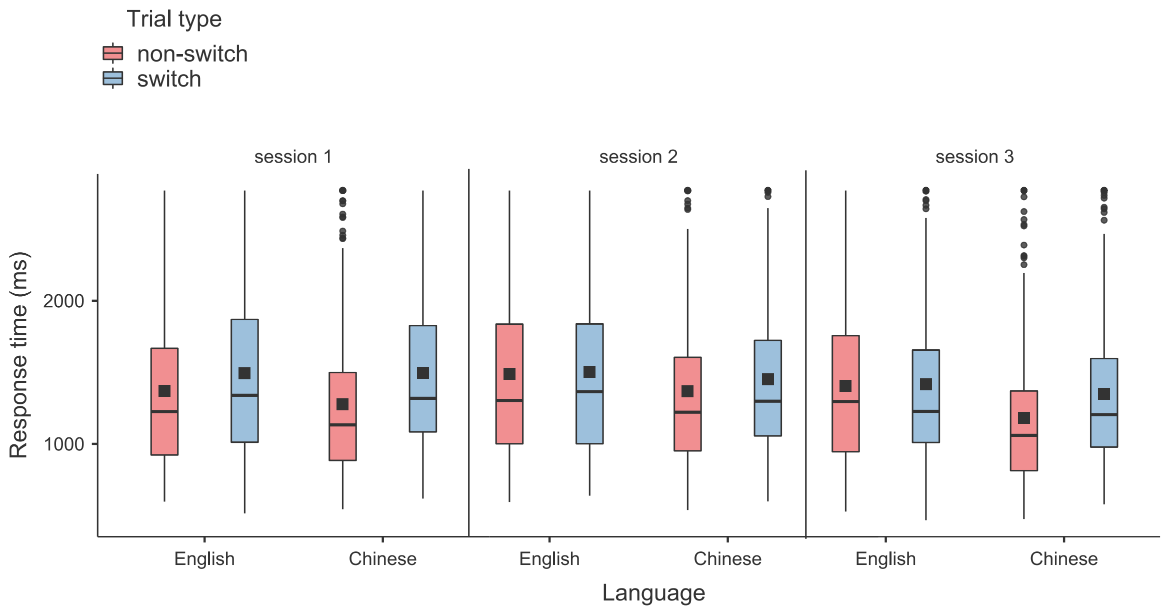


Figure 10: Average RT of switch and non-switch trials by Session and language. Whiskers reflect extreme values that are no outliers (within 1.5x IQR). Black boxes represent averages.

Figure 11 shows the magnitude of the bilinguals' switch costs into Chinese and English.

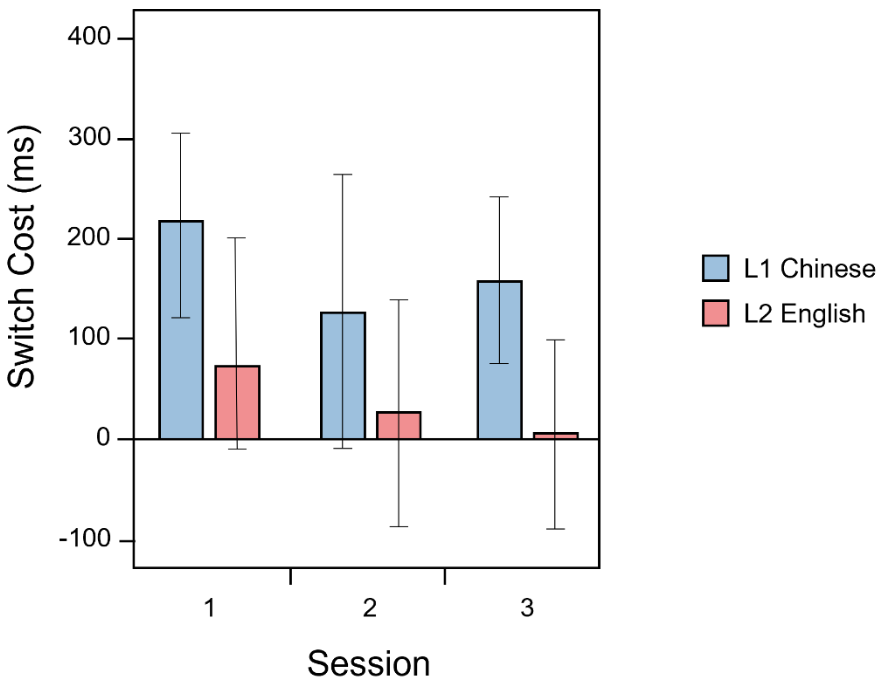


Figure 11: Switch cost into Chinese and English by Session.
Error bars reflect 95% Confidence Intervals.

The analysis of switch cost (the difference in RT between non-switch and switch trials in the mixed-task task type) revealed a significant main effect of trial type ($p<0.001$), echoing the finding that switch trials were slower than non-switch trials ($\beta=100.4$, $SE=17.2$, $p<0.001$), as well as a main effect of language ($p <0.001$), as Chinese items were named faster than English items ($\beta=-129.7$, $SE=37.2$, $p<0.001$), and a significant main effect of frequency ($p<0.001$), which indicates that more frequent words were named faster.

Formula: (lm, REML) RT ~ trial_type*language*language_balance + trial_type*Session frequency + (1+language|item) + (1+Session|participant), data = subset of bilinguals' mixed-task trials

Table 13: Model for bilinguals' switch costs in Chinese and English[18].

	Fixed effects	Estimate	Std. error	P-Value
	(Intercept)	1423.7	48.9	<.001
overall effects	**language Chinese**	-129.7	37.3	<.001
	language balance	0.001	0.07	0.93
	frequency	-60.9	15.1	<.001
	Session 2	-1.6	32.9	0.96
	Session 3	-49.7	34.9	0.17
	language Chinese: language_balance	-0.5	0.06	<.001
sw. cost effects	**Trial-type switch**	100.4	17.2	<.001
	Trial-type switch: language Chinese	138.5	33.5	<.001
	Trial-type switch: language_balance	0.08	0.06	0.16
	Trial-type switch: Session 2	-79.2	43.1	0.07
	Trial-type switch: Session 3	-84.1	39.8	<.05
	Trial-type switch: language Chinese: language balance	0.3	0.1	0.006

The addition of a fixed main effect for Session did not lead to an improvement in model fit compared to the model without (p = 0.38), reflecting no significant change in overall reaction time with increased immersion duration (β=-1.6, SE=32.9, p=0.96 for Session 2, and β=-49.7, SE=34.9, p=0.17 for Session 3).

Similarly, the addition of the interaction between trial type and Session marginally failed to significantly improve the model (p=0.06), reflecting a non-significant change in switch cost between Session 1 and Session 2 (β=-79.2, SE=43.1, p=0.07). However, switch costs did significantly decrease between Session 1 and Session 3 (β=-79.2, SE=43.1, p=0.04). We have retained this interaction to reflect this reliable decline.

[18] Factors and interactions that contributed to an improvement in model fit are marked in **bold.** Factors that did not contribute to a better model fit were retained in the model where it was contained in a higher-order effect, i.e., an interaction effect, and significantly contributed to model fit in this constellation.

In line with previous literature on asymmetric switch costs, there was a reliable interaction between trial type and language ($p<0.001$), which revealed that switches into Chinese were more costly than switches into English ($\beta=138.5$, $SE=33.5$, $p<0.001$).

The addition to an interaction effect of trial type, Session and language failed to significantly improve the model, suggesting that switch cost into Chinese (and English) did not significantly change over the testing period.

The interaction between language and language balance significantly improved the model ($p<0.001$), indicating that participants who are less proficient in English named Chinese targets faster as opposed to English targets ($\beta=-0.5$, $SE=0.06$, $p=0.001$). Higher L2 proficiency led to smaller discrepancies between L2 and L1 naming latencies.

Furthermore, a significant interaction effect between trial type, language, and language balance ($p<0.001$) reflected that those bilinguals, who are more proficient in English (who had smaller RTs in the English single-language task as opposed to in the Chinese single-language task) had a smaller switch cost asymmetry with switches into Chinese becoming increasingly slower (and eventually slower than switches into English). Figure 12 illustrates switch cost asymmetry between Chinese and English at three different levels of language balance.[19]

[19] For ease of visual presentation, we divided participants into discrete proficiency levels for the purpose of this figure. However, proficiency was entered into all regression models as an interval variable.

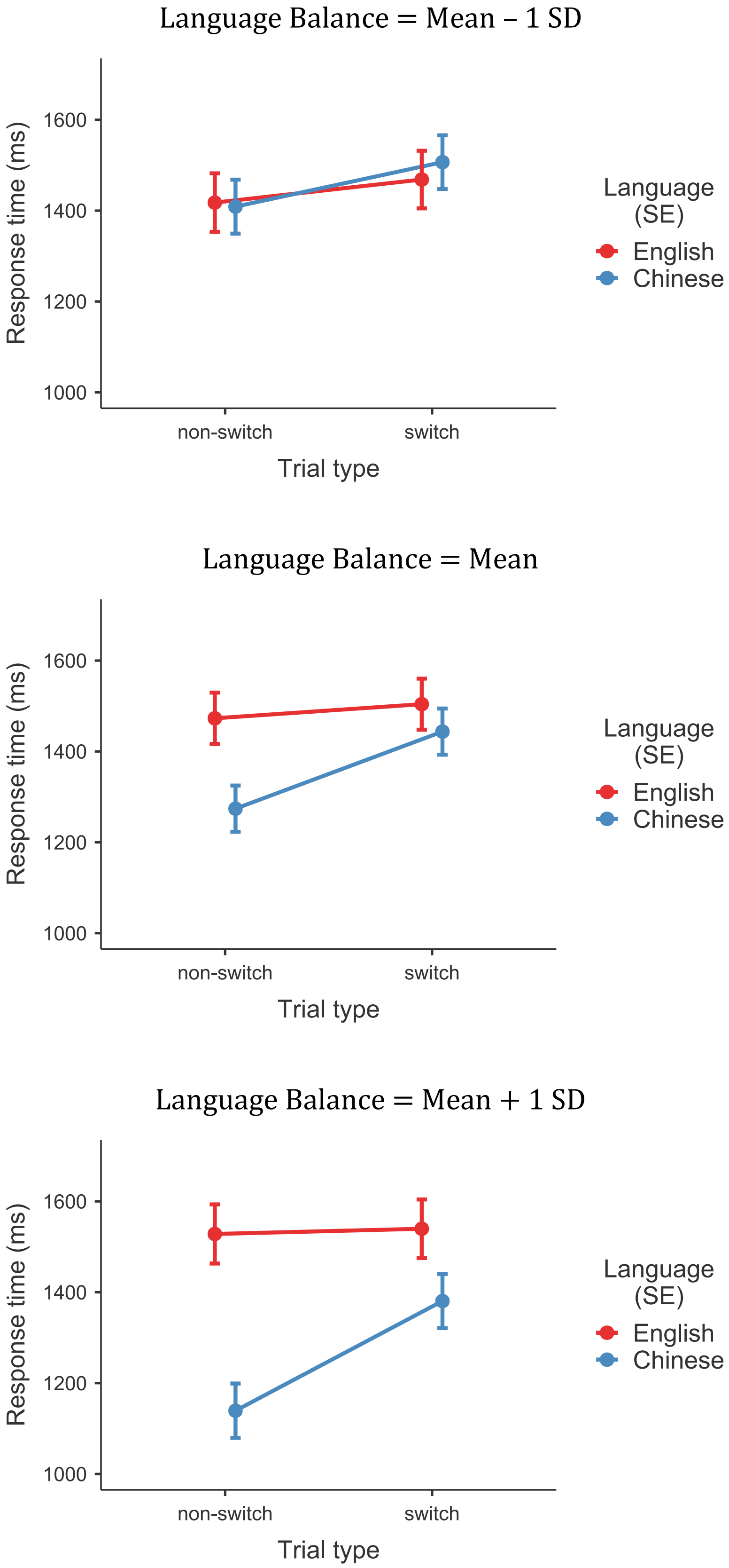


Figure 12: Switch cost asymmetry between Chinese and English by language balance. Error bars reflect Standard Errors.

3.5.3. Mix Cost

Mix cost is calculated by subtracting the average response time in non-switch (or "stay") trials in the mixed-task condition and average response times of trials in the single-language task. Figure 13 shows the development of average RT of switch and non-switch trials in Chinese and English.

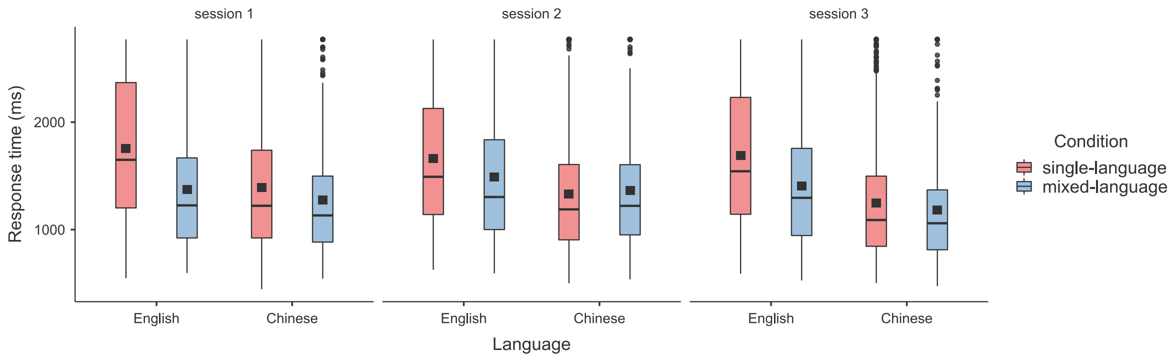


Figure 13: Average RT of mix and single trials by Session and language. Whiskers reflect extreme values that are no outliers (within 1.5x IQR). Black boxes represent averages.

Figure 14 shows the magnitude of the bilinguals' mix "costs" into Chinese and English.

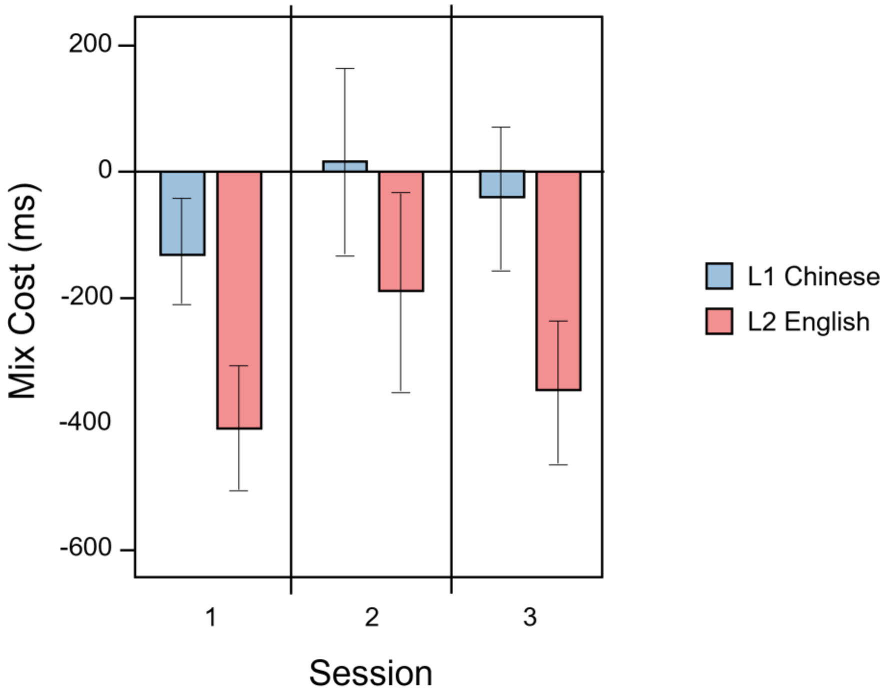


Figure 14: Mix cost of Chinese and English by Session.
Error bars reflect 95% Confidence Intervals.

The analysis of mix cost (RT in single-task trials and non-switch trials in the mixed condition) showed a significant main effect of task type ($p<0.001$), reflecting the finding that trials in the mixed condition were faster than trials in the single-task condition (β=-170.6, SE=27.5, $p<0.001$), as well as a main effect of Session ($p<0.02$), as overall RT significantly decreased between Session 1 and Session 2 (β=-113.7, SE=39.2, $p<0.001$), but not significantly so between Session 1 and Session 3 (β=-47.1, SE=32.5, p=0.2).

There was also a significant interaction effect of task type and Session ($p<0.001$), as mix cost was significantly higher in Session 2 (β=176.9, SE=38.4, $p<0.001$) and Session 3 (β85.6, SE=33.4, p=0.01) than in Session 1.

Formula: (lm, REML) RT ~ task_type*language*language_balance + task_type*Session + task_type*frequency + task_type:%excluded_trials + (1+language|item) + (1+Session|participant), data = subset of bilingual non-switch trials

Table 14: Model for bilinguals' mix costs in Chinese and English[20].

	Fixed effects	Estimate	Std. error	P-Value
	(Intercept)	1466.0	47.8	<.001
overall effects	**Language Chinese**	-333.3	27.0	<.001
	Frequency	-61.8	11.5	<.001
	Percentage of excluded-trials	944.0	209.1	<.001
	Language Balance	0.1	0.1	0.6
	Session 2	-113.7	39.2	0.01
	Session 3	-47.1	32.5	0.2
	Language Chinese: Language Balance	-0.8	0.05	<.001
mix cost effects	**Task type Mixed**	-170.6	27.5	<.001
	Task type Mixed: Language Chinese	286.0	53.9	<.001
	Task type Mixed: Session 2	176.9	38.4	<.001
	Task type Mixed: Session 3	85.6	33.4	0.01
	Task-type Mixed: language balance	-0.1	0.05	0.01
	Task type Mixed: Percentage of excluded-trials	-386.0	190.2	0.04
	Task type Mixed: Frequency	-50.6	21.8	0.02
	Task type Mixed: Language Chinese: Language Balance	0.4	0.1	<.001

There was a main effect of language ($p < 0.001$), as Chinese items were named faster than English items (β=-333.3, SE=27.0, $p<0.001$), and a reliable interaction effect between trial type and language, indicating that Chinese mix cost is greater than English mix cost, i.e., in the mixed task type, English trials were named much faster than Chinese trials (β=286.0, SE=53.9, $p<0.001$).

A significant main effect of frequency ($p<0.001$) revealed that more frequent words were named faster (β=-61.8, SE=11.5, $p<0.001$), and a considerable interaction effect of task type and frequency ($p<0.02$) reflects that higher frequency further facilitated

[20] Factors and interactions that contributed to an improvement in model fit are marked in **bold**. Factors that did not contribute to a better model fit were retained in the model where it was contained in a higher-order effect, i.e., an interaction effect, and significantly contributed to model fit in this constellation.

naming in the mixed condition to a larger extent as opposed to the single-task condition (β=-50.6, SE=21.8, $p<0.02$).

A significant main effect of overall trial exclusion rates ($p<0.001$) showed that those participants with a higher percentage of excluded trials (no or inaccurate response, RT > 4000ms) also performed worse in terms of response times (β=944.0, SE=209.1, $p<0.001$), and a significant interaction between task-type and exclusion trials revealed that mixing significantly facilitated faster naming for those with higher trial exclusion rates (β=-386.0, SE=190.2, $p<0.04$).

The interactions between language and language balance, and task-type and language balance significantly improved the model (both $p<0.001$), revealing that participants who are less proficient in English named Chinese targets faster as opposed to English targets (β=-0.5, SE=0.06, $p=0.001$), and that those less proficient had an additional facilitation effect in the mixed condition (β=-0.1, SE=0.05, $p=0.001$).

As was the case for switch costs, there was also a significant 3-way-interaction between task type, language, and language balance ($p<0.001$), which reflected that more balanced bilinguals had a smaller mix costs asymmetry. Figure 15 shows an overview of the mix cost asymmetry between Chinese and English at three different levels of language balance.

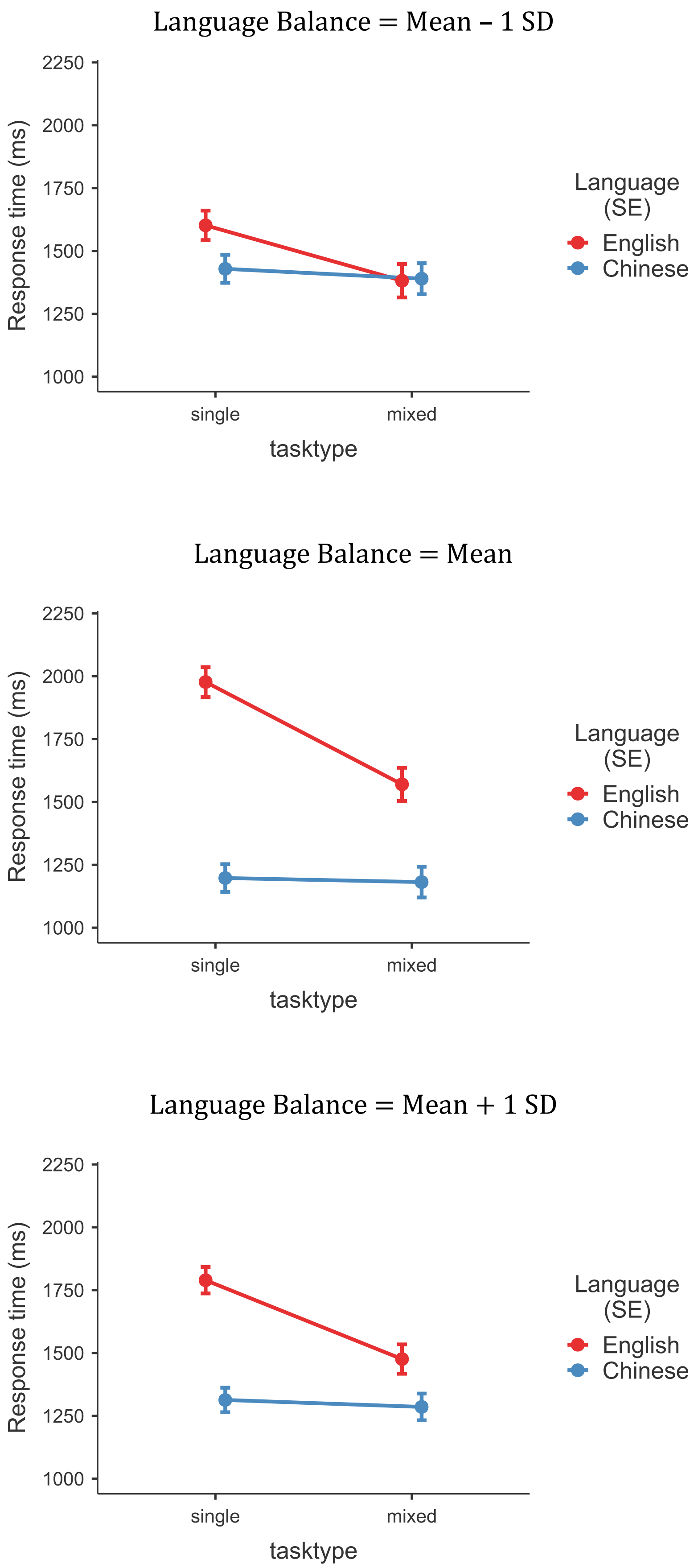


Figure 15: Mix cost asymmetry by language balance. Error bars reflect Standard Errors.

3.5.4. Language Balance

Formula: (lm, REML) LanguageBalance ~ Session+SC-ASYM+MC-ASYM data = summary of PNT outcomes per participant by Session

Table 15: Model for bilinguals' language balance.

Fixed effects	Estimate	Std. error	P-Value
(Intercept)	426.2	42.1	<0.001
Switch Cost Asymmetry	0.3	0.06	<0.001
Mix Cost Asymmetry	0.5	0.06	<0.001
Session 2	-8.3	32.6	0.801
Session 3	80.1	29.9	0.011

The linear mixed effects regression model (Table 15) shows that language balance significantly differed between Sessions 1 and 3 (β=-8.3, SE=32.6, $p<0.011$), but not between Sessions 1 and 2 ($p=0.8$) A post-hoc analysis with Bonferroni correction of Session showed that Language Balance also significantly differed between Sessions 2 and 3 ($p_{bonferroni} < 0.04$). The change was nonlinear, as language balance decreased between Session 1 and Session 2 but increased between Session 2 and Session 3. This means that overall, response times in Chinese as opposed to English were fastest in Session 3.

Adding PNT outcome variables Switch Cost Asymmetry and Mix Cost Asymmetry as co-variates further significantly improved model fit ($p<0.001$ in each instance). Both variables have a strong positive correlation with Language Balance.

As L2 balance increases (it becomes easier to access English as opposed to Chinese), switch cost asymmetry reduces (Chinese is no longer as actively inhibited) and even turns into higher switching cost into English for very proficient participants.

Similarly, as L2 proficiency increases, mix “cost” asymmetry reduces, as overall English response times decrease and so the effect of mixing no longer provides an increased facilitation effect. The facilitation effect of mixing at high naming speeds (achieved through high proficiency) plateaus.

3.6. Discussion

The aim of the present study was to investigate longitudinally the effect of immersion duration on bilingual lexical access and executive control in Chinese bilinguals and to evaluate the interconnectedness between language control and executive control. We employed dual-language picture naming tasks, as well as Flanker and Simon tasks at three different time points during their first year of stay in the United Kingdom.

Monolinguals only surpass bilinguals' L2 and bilingual L1 RT significantly decreases over time while L2 RT does not. We predicted that monolinguals would consistently outperform bilinguals in the dual-picture naming task. To test this prediction, we compared monolingual single-task performance to bilingual single-task performance in their L1 Chinese and their L2 English. Yet, contrary to expectations, the analysis of naming latencies in the single-language task condition revealed that throughout the testing period monolinguals only significantly surpassed bilinguals in their English L2, but not significantly so in their L1.

We predicted that Chinese lexical access would gradually decelerate over time while English lexical access would accelerate, as English conceptual ties were strengthened while Chinese lexical items were increasingly disused and/or inhibited to facilitate English access. Contrary to expectations, Chinese naming latencies did not significantly decelerate over time and English naming latencies did not significantly improve within the testing period, either.

These findings are at odds with previous findings in various investigations of L1 lexical attrition, which report that delayed lexical access is one of the earliest and most prominent effects of L2 immersion (e.g., Schmid and Köpke, 2009). A possible explanation for this result could be that the bilinguals in our sample have not yet reached the point at which attrition effects become discernible. Specifically, it could be the case that they have not immersed themselves into the new environment as much as anticipated. Linck et al. (2009) find that as immersed L2 learners were significantly more impaired in producing L1 items in a verbal fluency task compared to their peers, who only learned L2 in the classroom in their L1-dominant environment. However, while Baus et al. (2013) find no statistically significant indicators of L1 attrition in their longitudinally employed verbal fluency task, they do find longer naming latencies for L1 in their picture naming task. These findings suggest that different comparability across the tasks and populations – and, therefore, interpretations on the depth of the bilinguals' immersion, is not straightforward. **Switch costs decrease over time, but switch cost asymmetry does not.** We have found that overall switch cost significantly decreased between Sessions 1 and 3, suggesting that bilinguals do become better at managing their two language systems over time. Due to the lack of a suitable control group this finding may be a result of repeated task exposures as opposed to an effect of L2 immersion duration. This decrease in switch cost was independent of language, as the interaction effect of trial type, Session and language failed to significantly improve model fit. Switch cost into English decreases between Session 1 and consecutive Sessions, but not significantly so, while switch cost into Chinese is lowest in Session 2, but again, not significantly lower than in Sessions 1 and 3 when controlling for confounding factors.

In line with observations by Meuter and Allport (1999), we did, however, find evidence for asymmetric switch costs, as switches into the Chinese L1 were significantly more costly than into the English L2. However, this asymmetry is independent of immersion duration.

Switch cost asymmetry depends on L2 proficiency. We have found a significant interaction effect of between trial type, language, and language balance, which suggests those bilinguals who are more proficient in the L2 have a smaller switch cost asymmetry. This finding is in line with other studies on cued language switching, which have found a link between L2 proficiency (where language balance is such that L2 naming is less costly than L1 naming).

The analysis of language balance supports the notion that switch cost asymmetry depends heavily on L2 proficiency. As we did not find a significant decrease in switch cost asymmetry, and neither did we find an increase in L2 proficiency between Sessions 1 and 3. Instead, the data suggest a development toward faster L1 response times. Curiously, when switch cost into Chinese was lowest in Session 2, language proficiency was at its lowest, too. We conclude that L2 proficiency strongly impacts switch cost asymmetry and since L2 proficiency did not significantly increase during the testing period, switch cost asymmetry did not recede, either.

We found no evidence for mix cost. Mixed-task task types are assumed to elicit larger elicitation times, as rapidly switching is thought to be more taxing: two mental task sets need to remain co-activated while the selection of the non-target must be selectively inhibited. Yet, the **mixed-task task type was found to significantly facilitate naming** with targets named 170ms faster on average.

Similar studies have previously found mix costs in cued language switching tasks. For instance, Bonfieni et. al (2019) investigated mix and switch costs in two groups of participants within an experimental design structurally similar to the one presented in the current study. The Italian-English group comprised Italian native speakers, who were late bilinguals and had lived an average of 3.7 years in the UK, while the Italian-Sardinian group had been regularly immersed in either language environment from a young age. Bonfieni et al. (2019) report that both groups were highly proficient in their respective L2s. In their study, they found L1 mix costs averaging 45msecs (43) and 40msecs (37) and L2 mix costs averaging 64msecs (39) and 50msecs (49), respectively.

De Bruin et al. (2018) investigated voluntary and cued language switches. Their results indicated that in tasks involving voluntary and cued language switching, response times were longer for switch trials compared to non-switch trials. Interestingly, while they found evidence for a mix cost in the cued task; they found a mix benefit in the voluntary task.

Ma et al. (2015) report finding mix cost in their structurally similar experiments involving groups of unbalanced Chinese-English bilinguals. In their study, they experimentally manipulated the Cue-Stimulus-Interval (CSI), which is the time between the cue (presentation of a red or blue dot representative for the target language) and the stimulus (a digit between 0 to 9). Their prediction was that an increased CSI allows for more preparatory time. Indeed, they found that mix costs were highest when the CSI interval was 0msecs. Considering our experiment had 0 ms CSI, this raises the question of why we did not find any mix cost.

Our data seems to suggest that the nature of the rapid switching facilitates naming: with two language systems/lexica co-activated, neither are globally inhibited. Perhaps global inhibition exerts a higher cognitive load than local inhibition. It would seem that this is a phenomenon of less proficient individuals, as increasing proficiency decreased the facilitation effect ($\beta=-0.1$, $SE=0.05$, $p=0.001$).

Linking this finding back to the Revised Hierarchical Model, this finding could be evidence for L2 lexical items being closely associated with their L1 translation equivalents, at a time when L2 conceptual links have yet to be established. In this view, it could be easier to have lexicons activated, and inhibit locally, as opposed to inhibit globally, where the target word may be closely linked to an item that is now less accessible. The finding that mixing facilitation is larger in English than in Chinese (β=286.0, SE=53.9, $p<0.001$) supports this idea: there are weak L2 conceptual links, so it would be easier to have access / co-activation of the L1.

We also find a **mix "cost" (or rather: "facilitation") asymmetry**, with mixing in the English L2 being significantly more facilitative than in the L1. Curiously, like the switch cost asymmetry, this facilitation strongly depends on L2 proficiency, but not Session. To our knowledge, no other study has previously found or commented on the relationship between language proficiency and mix cost. As with switch cost, the effect of mixing is lowest in Session 2, too. Again, we conclude that L2 proficiency strongly impacts mix cost asymmetry and since L2 proficiency did not significantly increase during the testing period, mix cost asymmetry did not recede, either.

Given the majority of studies found a mix cost, our finding of a "mix benefit" is likely a result of methodological differences. In our case, it is likely that the facilitation is due to a practise effect: the mixed block was always completed last, as the order of blocks were not counterbalanced. The finding that less proficient bilinguals experience a larger benefit supports this idea, as they may benefit more from repeated trails.

To avoid redundancy, an overview of the limitations of this analysis as well as of the study as a whole are presented in 6.3. Limitations. Specifically, the appropriateness of employing a functionally monolingual group as comparative measure as well as practise effects within the individual experimental sessions and between experimental sessions run are reflected upon in this subchapter.

3.7. Conclusion

The aim of the present study was to investigate longitudinally the effect of immersion duration on bilingual lexical access in Chinese bilinguals. We found that in early stages of bilingual development, the L1 is still overwhelmingly active and easy to access, while L2 retrieval is associated with longer naming latencies. However, we found evidence that the bilingual's L1 retrieval already elicits significantly larger naming latencies than monolingual L1 retrieval, indicating a larger processing burden in bilinguals. Furthermore, our results show a significant drop in L2 switch costs while L1 switch costs increase. Interestingly, we found that tasks in which the bilingual is asked to switch between two languages facilitate naming more than single-language tasks, which suggests that the local inhibition of the non-target language may be less costly at these early stages of L2 immersion. We found asymmetries between L1 and L2 in both switch cost and mix cost, and that these symmetries decline with increasing language balance/L2 proficiency.

4. Language Control Measures Predict Executive Control Performance

A Longitudinal Study Within the First Six Months of L2 Immersion

Abstract

Bilinguals switch between languages with great ease and speed. Executive control, the system enabling these switches, is thought to be trained during the bilingual experience, resulting in improved performance as opposed to monolinguals ("bilingual advantage"). The present study investigates the changes in bilinguals' language switching abilities and general executive control as a function of immersion duration. Thirty Chinese-English bilinguals completed a picture-naming task, a Flanker, and a Simon Task at three testing times during their first year in the United Kingdom. Linear mixed-effect regression models provide mixed evidence for improving bilingual naming latency and executive control as a function of exposure duration. While overall naming latency decreases, the decrease is more salient in the bilinguals' L1 than in their L2. Similarly, while some indices of executive control prove superior in later sessions, this finding is inconsistent: L2 use had a significant impact on several measures of language control and executive control, but it did not significantly improve over the testing period.

4.1. Introduction

In the past decade, there has been a growing interest in how bilinguals selectively access words across their languages, as proficient bilinguals switch swiftly between their language systems with great accuracy and relative ease (Heredia & Altarriba, 2001). Several studies have shown that bilinguals' languages are in a constant state of co-activation (e.g., Poulisse & Bongaerts 1994; de Bot & Schreuder, 1993; Kroll & Sunderman 2003; Kroll et al. 2006). Hence, it has been proposed that bilinguals need to actively inhibit the activation of the non-target language to allow error-free access to the target language (Green, 1998) in a process referred to here as language control.

A recurring question in the field of bilingualism and cognition is whether increased training in language control translates to domain-general executive control – and for whom. Previous research suggests a link between linguistic and non-linguistic executive control functions. For instance, bilinguals who had difficulties switching between naming numbers and adding numbers also displayed difficulties when switching between languages in a dual-language elicitation task (Gollan et al., 2014). Similarly, Prior & Gollan (2011) found that bilinguals who switch their languages often have smaller non-linguistic task-switching costs. Task-switching costs occur as mental sets need to be updated when tasks of a different nature immediately follow the completion of a previous task.

Bialystok (2009) concluded that the bilingual experience significantly improves executive function through the link between language control and executive control, giving rise to a "bilingual advantage". Supporting this hypothesis are several studies where bilinguals outperform monolinguals, including studies involving the Simon Task (e.g., Bialystok et al. 2008; Martin-Rhee & Bialystok, 2008) and Flanker-type tasks (e.g., Poarch & Van Hell, 2012; Yang et al., 2011).

Other studies, however, do not find evidence for a bilingual advantage (e.g., Gathercole et al., 2014; Morton & Harper, 2007; Antón et al., 2014; Paap & Greenberg, 2013). Overall, in a meta-analysis encompassing 46 original studies on the bilingual advantage, Noort et al. (2019) concluded that 54% of the studies did find beneficial effects of bilingualism on cognitive control tasks, while the rest returned either mixed (28%) or contradicting (17%) evidence. Synthesizing data from 152 studies involving 891 effect sizes on adults, including unpublished data and various study-related factors, Lehtonen et al. (2018) initially find a slight advantage for bilinguals in inhibition, shifting, and working memory (effect size: g= +0.06 [0.00, +0.13]), which, however, disappeared after correcting for publication bias (g= −0.07 [−0.17, +0.04]).

The discrepancies between studies stem from a variety of causes. Noort et al. (2009) highlighted methodological differences, such as variation in the selection of bilingual populations, the adoption of non-standardized tests, and the lack of longitudinal designs. Baum and Titone (2014) point out that studies tend to make relatively coarse comparisons between bilinguals and monolinguals, for instance, by neglecting bilinguals' inherent individual variability. Individual differences, such as pre-existing neurocognitive capacities or sociocultural factors, could further affect the types of communicative experiences encountered. So, variations within the bilingual experience influence the relationship between language and executive control. Both Noort et al. (2009) and Baum and Titone (2014) highlight the need for multifactorial statistical analyses that appreciate the multidimensionality of the issue.

Therefore, understanding what factors affect the language switching ability is essential to describe language control and to associate different dimensions of the bilingual experience with its cognitive effects.

To contribute to the general understanding of the reasons behind the discrepancies in findings, the current study investigates patterns in the development of bilinguals' executive control by exploring executive control in verbal and nonverbal tasks as a function of immersion duration. Our main hypothesis is that there is a significant effect of immersion duration on language control and executive control performance: If the bilingual experience gradually enhances functions of language control over time (i.e., second language (L2) immersion duration) – and there is a link between language control and executive control, then bilinguals will become increasingly faster at executive control tasks, too.

We examine the effect of prolonged L2 immersion on code-switching and executive control by investigating performance on different tasks in a cohort of young adult Chinese-English bilinguals in their first year of university in the United Kingdom. We measure overall response times (RT), switch and mix costs in their Chinese L1 and English L2 as part of a dual picture naming task, and congruency costs in single and mixed conditions in Flanker and Simon tasks across three sessions. The first session is conducted shortly after the participants' first arrival to the UK, the second after three months, and the third after six months. We control for individual variation by employing linear mixed-effect models.

4.2. Literature Review

Executive control can be formally characterized as having three different main functions (Miyake et al., 2000; Friedman & Miyake, 2004; Friedman et al., 2006): First, the inhibition of irrelevant information to enable focus on the target information, e.g., inhibiting the non-target-language lexeme in favour of the target-language lexeme. Second, the shifting of mental tasks and rules, i.e., switching between languages where appropriate. Third, the updating and maintaining of representations stored in working memory.

Executive control naturally varies between and within individuals. For instance, age has a large impact on executive control functions, as different executive skillsets are boosted and attenuated at different stages of life (Salthouse 1985; 2000): Executive function develops in the early years and continues to augment during childhood and adolescence (Best & Miller 2010). In early adulthood cognitive functions associated with fluid intelligence, such as working memory, long-term memory, processing speed, and spatial ability begin to decline. Crystallized functions, like verbal ability (e.g., vocabulary), the capacity to utilize previously acquired knowledge and experiences, personality, information processing, and comprehension, tend to exhibit growth throughout adulthood, but gradually decline after the age of 60 (Murman 2015; Schaie 1994; Rohwedder & Willis, 2010; Salthouse 2010; Verhaeghen & Cerella, 2002).

Executive control can also vary among individuals, and certain life experiences, stimulating leisure activities and interventions have been shown to enhance performance on measures of executive control and may mitigate the adverse effects of structural and functional decline, such as an enriched environment (Lövdén et al. 2005; Stine-Morrow et al. 2014; Zuelsdorff et al. 2019), healthy nutrition (Klimova et al. 2017), physical activity (Bamidis et al., 2014) or making music (Mansens et al., 2018; Moreno et al. 2011).

The bilingual advantage view proposes that bilingualism, too, influences executive control. Bialystok (2009) suggests that the neurological structures that enable moment-by-moment language processing differ between monolinguals and bilinguals and that these differences accrue. As a result, a substantial neuroplastic change in bilinguals develops over time. As outlined above, evidence for a bilingual advantage is currently mixed. The question remains whether the differences in linguistic experience lastingly modify brain structures concerned with executive functions. In the following, we identify and discuss the essential premises for a bilingual advantage to emerge:

Premise 1: "Executive control is a part of the bilingual experience." Bilinguals need to regulate their languages. Green's (1998) Inhibitory Control Model provides a theoretical framework for bilingual language control. Only the target language item is selected during lexical access, while the non-target language vocabulary selection is actively inhibited. Active suppression is necessary, as non-target vocabularies become co-activated during processing (Green, 1998; Poulisse & Bongaerts, 1994; de Bot & Schreuder, 1993; Kroll & Sunderman, 2003; Kroll et al., 2006).

Overall, bilingual vocabularies are in a constant state of co-activation in a compound/unitary system, where all languages employ the same neural structures. It is thus impossible to entirely forego the processing of the non-target language (Paradis, 2004). Neuroimaging studies support this notion by showing activation in neural regions associated with monitoring and interference inhibition during bilingual language processing (Luk et al., 2012).

Inhibition is the focus of research investigating the relationship between bilingualism and executive functions, as it is considered the primary ability required to resolve conflicts between competing languages. Updating and switching mental sets also play a role in the language control mechanism. When language switches occur, the regulatory mechanism needs to switch tasks and update to reflect the current language requirement.

Premise 2: "The bilingual experience induces changes in the brain, which accrue over time. Bilingualism affects neuroplasticity. Through training – the sustained, repeated execution of a particular task – processes can become more efficient, resulting in superior task performance, for instance, increasing speed and accuracy. "

Neuroplasticity is the underlying neurological ability enabling the human brain to actively rewire neural networks (e.g., Mateos-Aparicio & Rodríguez-Moreno, 2019). "Structural neuroplasticity" refers to the change in the number of synapses and dendrites on a neuron, while "functional neuroplasticity" refers to the transformation of the available properties of neurons, e.g., the amount of neurotransmitter released and the number of post-synaptic receptors (idem). Such modifications aim to adapt to changes in the environment, like improving performance in familiar tasks.

Several studies show structural changes in different occupational groups and communities and that individual experiences could be responsible for these changes. For instance, taxi drivers have enlarged regions of the hippocampus responsible for navigation (Maguire et al., 2000), musicians tend to have an increased cortical representation for the fingers they play with (Elbert et al., 1995) and bilinguals tend to have denser grey matter in the left inferior parietal cortex, which is associated with language (Mechelli et al., 2004).

Similarly, as different occupational groups face additional processing costs, bilinguals too face an additional processing burden: the active suppression of the non-target language during lexical retrieval. Active cross-language tasks necessitate more cognitive resources, which are reflected in larger overall lexical retrieval times when compared to monolinguals (Gollan et al., 2005; Gollan et al., 2008; Gollan et al., 2011; Ivanova & Costa, 2008; for review, Bialystok et al., 2009).

Greater inhibitory control is particularly relevant for the first language (Green, 1998). The first language is more readily available, "dominant" at the onset of bilingualism due to its early, prolonged, and competition-less usage. On the other side, the second language is still establishing its mental representation (e.g., Kroll & Stewart, 1994; Kroll & Ma, 2017 and slowly “entrenching” itself (HJ Schmid, 2017).

This unequal status of the languages means that increased executive control needs to be exercised on the activation of the L1 to facilitate L2 production. Meuter & Allport (1999) show that bilinguals take significantly longer to complete an L1 speaking trial, when it immediately follows an L2 speaking trial, as opposed to vice versa. These asymmetric switch costs can be interpreted as a training effect with the brain actively adjusting to facilitate task demands. There is a range of potential contributors – next to the structural and functional changes – including inhibitory neurons, which release chemical messages to supress and inhibit other neurons actively.

Premise 3: “Domain-specific executive control translates to domain-general executive control.” Bialystok (2009) suggests that training executive functions used in the linguistic domain, such as code-switching, translates to generally more efficient executive mechanisms, giving rise to a bilingual advantage. Indeed, similar inhibitory effects are found in both linguistic and non-linguistic tasks: following Meuter & Allport's finding of asymmetric switch costs in the linguistic domain, Kiesel et al. (2010) found that asymmetric switch costs also occur in familiar vs. unfamiliar tasks in non-linguistic settings, such as switching between a digit and a letter task This finding opens the possibility that linguistic tasks directly employ domain-general resources rather than domain-specific ones.

Furthermore, there is substantial evidence that executive control functions used in bilingual language processing and general executive control functions share a similar neural network: in a meta-analysis of neuroimaging studies of bilinguals, Abutalebi and Green (2007) argue that there is a network of cortical regions involved in the modulation of competition of L1 and L2 activation, including the prefrontal cortex, the inferior parietal cortex, and the anterior cingulate cortex, as well as different subcortical regions. Crucially, the inferior and lateral regions of the prefrontal cortex and the anterior cingulate regions are also involved in general executive control tasks (e.g., Braver; 2012). The prefrontal cortex is commonly associated with the employment of executive control, while the anterior cingulate cortex is associated with error monitoring.

Consistent with this notion, Deluca et al. (2019) found a range of neuroplastic changes in bilinguals' brains in a longitudinal study: in the space of three years, they identified significant increases in the volume of grey matter within the left cerebellum, white matter diffusivity in the frontal cortex, and reshaping of the bilateral hippocampus, amygdala and the left caudate. They also found that length of immersion and age of L2 acquisition are significant predictors of volumetric change in the cerebellum.

Taking premise 3 a step further, another approach considers the correlations of measures of intra-individual variation (IAV) in both linguistic control and executive control as indeces of shared systems (e.g., Pfenninger & Kliesch, 2023). IAV concerns behavioural differences within individual bilinguals across contexts and/or at different moments of development, as well as nonlinear developmental changes (Hickmann et al., 2018). Specifically, Unsworth (2015) claims that consistency of attention control is an important cognitive trait related to a number of cognitive abilities, such as L2 learning. Given that both applied linguists and cognitive psychologists (for instance, Mella et al., 2016; Rabbitt et al., 2001) depict IAV as a consistent trait among individuals, we anticipate a correlation between measures of linguistic control IAV and cognitive control IAV. IAV can be classified according to the time frame and tasks under consideration (Fagot et al., 2018).

This categorization includes:

- inconsistency, which pertains to fluctuations in performance within tasks (i.e., IAV across trials within cognitive and L2 tasks within a single assessment);
- dispersion, which relates to the influence of dondition and demands (i.e., cognitive and L2 IAV across various conditions within a single session); and
- variability, which signifies developmental changes (i.e., IAV within tasks observed in individual time-serial data from repeated observations).

4.3. The Current Study

While there are various supporting arguments for the notion of a bilingual advantage, the mixed evidence remains a challenge for generalized conclusions that the bilingual experience enhances structures beyond the linguistic domain. We propose that including more fine-grained detail of the bilingual experience may be able to shed some light on why some studies have found a bilingualism effect while others did not. Specifically, this study focusses on the impact of L2 immersion duration on the interplay between language control and executive control.

The present study investigates to what extent performance on tasks involving executive function may be ascribed to an increased duration of L2 immersion – i.e., an improvement of executive control performance as a result of training effects accumulated as part of the bilingual experience. We explore overall response times, switch/congruency costs, and mix costs on tasks relating to domain-specific language control and domain-general executive control in Chinese-English bilinguals as part of a longitudinal study spanning six months. At the beginning, participants had just started their university studies in the UK. The second and third sessions capture the development after three and six months respectively. We use linear-mixed effects models to account for further individual variation and aimed to recruit a homogenous group of bilingual participants to further limit the extent to which alternative effects could be introduced.

To examine the effects of bilingual experience on a domain-specific language switch mechanism, on one hand, and a domain-general switch mechanism on the other, we compare performance on a linguistic task vs. cognitive control tasks. The premise is that if linguistic and non-linguistic components share a switching mechanism; bilinguals, who utilize this system more frequently and thus train these mental faculties, should readily benefit from this training in domain-general tasks as well. Taking it one step further, our study also considers IAV in the different task-domains to see whether those who vary in one task also vary in the other. Here, the idea is that if the same or similar mechanisms are employed, then it ought to be likely we see some correspondence in IAV patterns, too.

Our methodology involves a dual-language picture naming task, a Simon Task (Simon, 1969), and a Flanker Task (Eriksen & Eriksen, 1974), including single-task and mixed task blocks. In the dual-language picture naming task, participants are instructed to name pictures in each language separately in single-tasks blocks, and to name the picture in one of the two languages in a pseudo-randomized manner in the mixed task block. In the general cognition-type single-task blocks, the directionality/location of the stimulus either matches the interfering information in one block (congruent) or contradicts (incongruent) it. In the mixed-task blocks, trials alternate between congruent and incongruent conditions.

4.3.1. Measures of language control and executive control

In this way, it is possible to gather different measures of language switch and executive control efficiency:

(Overall) response time (RT) reflects the overall average latency needed to process trials in various conditions. It is the main dependent variable and – in interaction effects with trial type, condition, and congruency – reflects average switch and mix costs, as well as inference scores, respectively.

Switch costs reflect the relative difficulty with which new, corresponding mental task sets are adopted in a mixed-task block. They are measured as the difference between similar (non-switch) and different (switch) trails. In our study, we investigate switch costs in the dual-language picture naming task to evaluate language switching ability.

Mix costs are a measure of the ability to monitor conflict between tasks and keep two task sets partially activated (Segal et al., 2021). They are usually measured as the difference between non-switch trials in the mixed block vs. single trials in a single-task block. In our study, we investigate mix costs in all tasks to evaluate the participants' ability to keep multiple task sets partially activated at any one time.

Interference scores are a measure of inhibitory control which reveal the relative, added difficulty with which incongruent trials are completed as opposed to congruent trials. They are typically measured as the difference between congruent and incongruent trials. In our study, we investigate interference scores in the Flanker and the Simon tasks to evaluate the participants' ability to inhibit conflicting information.

Language Balance describes the relative ease of access to one language as opposed to another and aids in understanding how the bilinguals' two languages interact with each other. Following Birdsong (2016), we computed a "between-language subtractive differentials" measure for language balance, involving performance score subtraction between languages. We calculated is usually calculated by subtracting the average response time for one language from the average response time of the other. In our study, language balance serves as a proxy for L2 use relative to L1 use and ease of access.

Switch Cost Asymmetry is another measure for language balance and is calculated by subtracting the average switch costs from one language (in our study, L1), from the other (L2).

We have introduced the term **mix cost asymmetry** in our study to reflect the discrepancy between mix costs in the L1 and the L2. We have calculated this measure by subtracting the average mix costs in the L1 from the average mix cost in the L2.

4.3.2. Measures of IAV

For all measures of IAV, we employed Lowie and Verspoor's (2019) coefficient of variation (CV), which is calculated by dividing the standard deviation by the mean of the given sample.

Inconsistency - reflecting the variation within a particular task demand within a single assessment – was calculated using means and SDs of Chinese and English trials in single-language condition (PNT) and congruent trials in the single-congruency condition (Simon, Flanker) – by participant for each session. In so doing, we aim to capture inconsistency within the the least complex trials, i.e., those without any additional task demands (e.g., switching, mixing) to control for any effects thereof.

Variability – reflecting developmental changes – was calculated using **across-session** means and SDs of Chinese and English trials in single-language condition (PNT) and congruent trials in the single-congruency condition (Simon, Flanker) – by participant.

Dispersion - relating to the influence of condition and task demands – was calculated using means and SDs across different task demands (Chinese/English; single/mixed; switch/non-switch; congruent/incongruent) within a single assessment – by participant by session. In so doing, we aim to capture the overall variation across task demands within a task, rather than merely relying on mean scores (for a discussion of this, see Birdsong, 2023; Pfenninger & Kliesch, 2023).

4.3.3. Hypotheses

We address the following research questions and hypotheses:

RQ1: Is there a bilingual advantage in that bilinguals outperform monolinguals in executive control tasks?

We hypothesize that bilinguals will consistently outperform monolinguals. If the bilingual experience enhances functions of executive control, then bilinguals will outperform monolinguals in general-cognition executive control tasks.

RQ2: If so, at which stage of the bilingual experience does it set in?

We expect there will be a significant effect of immersion duration on interference scores and mix cost. If the bilingual experience enhances functions of executive control through increased immersion duration, then bilinguals will become increasingly faster and better at inhibiting interfering information across the three measurements.

RQ3: Are bilingual individuals, who are better at switching between languages, also better at domain-general executive control tasks?

We expect that there is a correlation between individual performance on general-cognition executive control tasks and individual performance on the picture naming task, which means:

a) Those participants who have smaller naming latencies in the picture naming task also have shorter reaction times in the general-cognition executive control tasks, or: dual-language picture naming task RT predicts executive control tasks RT.

b) If language switch competency does enhance performance on domain-general executive control tasks, then those with higher switch competency should also exhibit greater competency in domain-general executive control tasks, or: dual-language picture naming task switch cost predicts executive control tasks interference scores.

c) If language mixing competency does enhance performance on domain-general executive control task, then those with higher mix competency should also exhibit greater competency in domain-general executive control tasks, or: dual-language picture naming task mix cost predicts executive control tasks mix cost.

RQ4: Is there a correlation between IAV measures in linguistic control and executive control tasks?

We hypothesize that there is a correlation between individual measures of IAV on general-cognition executive control tasks and on the picture naming task, which means that those who vary in one aspect of IAV in a task also vary in this aspect in another task. Specifically, if there is a connection between language control and executive control, then participants who (a) vary more within a single task at one time (i.e., those who are less consistent), those who (b) vary more within the same task across sessions (i.e., those with higher variability) and those who (c) vary more across different task demands at one time (i.e., those with higher dispersion), do so across linguistic and domain-general tasks.

4.4. Methods

4.4.1. Participants

A total of 30 Chinese-English bilingual adults participated in this study, ranging from 19 to 32 years of age (mean = 23.5 years). All Chinese-English bilinguals reported speaking some form of Chinese (26 Mandarin, 3 Cantonese, 1 Taiwanese) as their first language. Twenty participants identified as female, and ten as male. Most had only just arrived in the United Kingdom in the month prior to the first testing session to pursue university degrees. Six individuals took part in pre-sessional English courses, and sixteen stated that they were enrolled in an English language class at the point of first testing. The control group comprised 20 functionally monolingual English adults between the ages of 19 and 53 years (mean = 26.6 years) from the United Kingdom, of which thirteen identified as female and seven identified as male. All control participants had English as their first language. Eleven stated they had had some formal language education at school that they had not made use of since.

4.4.2 Study Design and Procedure

All participants completed a battery of different tasks, including a Flanker Task, a Simon Task, a dual-language picture-naming task and a language background questionnaire. Each of the tasks comprised four distinct components, as illustrated in Figure 16. Prior to engaging in the tasks, participants were exposed to eight practice trials designed to familiarize them with the task's requirements. The subsequent two components of each task were presented in a single-task format. In these single-task conditions, the executive control task featured a sequence of 20 congruent trials followed by 40 incongruent trials. In the language control task, participants completed 50 trials in Chinese followed by 50 trials in English. The concluding component of each task entailed a "mixed condition" block, wherein trials from the preceding blocks were incorporated in a semi-alternating manner. The order in which the three blocks was presented was fixed throughout the study. Doing so, we aimed to eliminate effects varying block orders may have had (see 2.2.4. Design of timed experiments).

The tasks were undertaken by participants across three distinct time points. The initial session occurred at the onset of the academic year in October 2016, followed by the second session spanning late January to early February 2017, and culminating with the third and concluding session in April 2017. Bilingual participants engaged in the complete set of tasks during all three sessions. Meanwhile, monolingual controls participated in the picture naming task during the first and third sessions. The participants completed the Flanker Task first, the Simon Task second, and the dual-language PNT last.

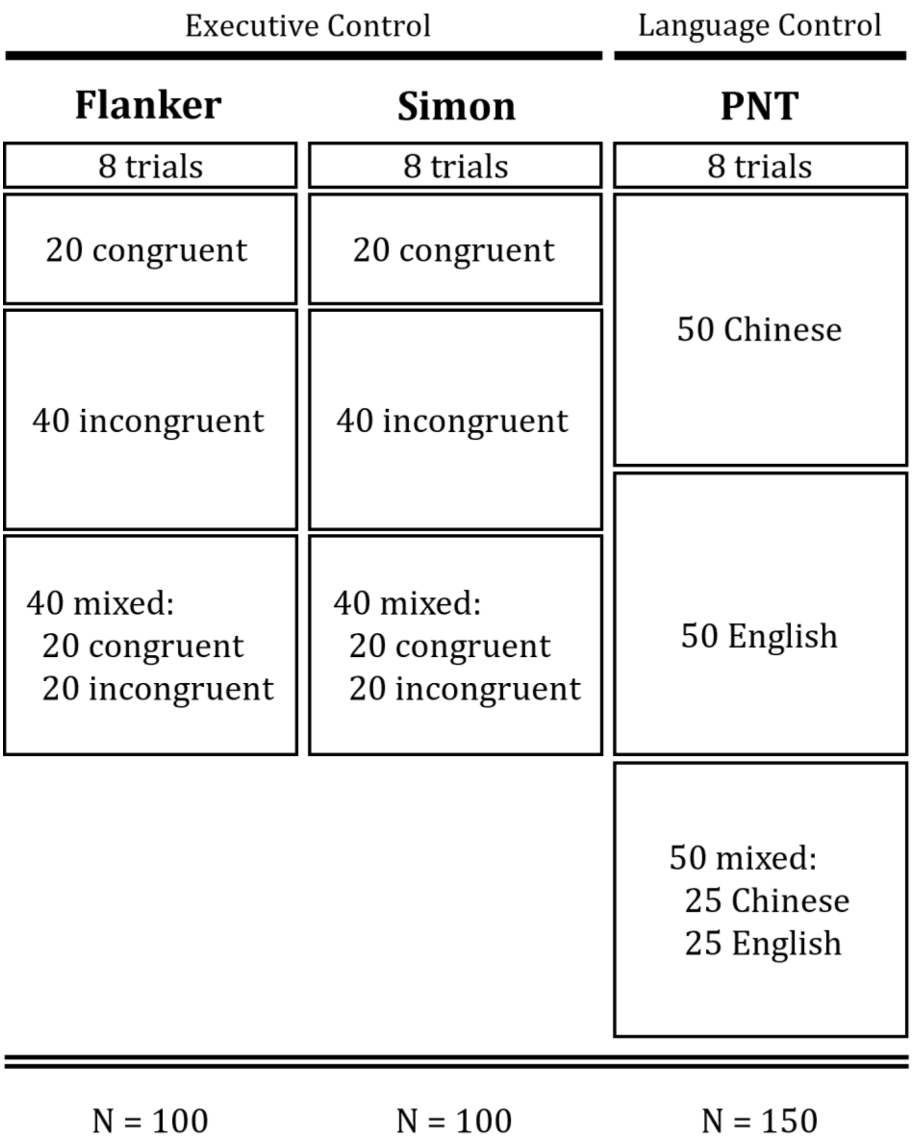


Figure 16: Study Design.

N refers to the number of experimental trials.

4.4.3. Experimental Tasks

The Simon Task and the Flanker Task have different conceptualizations within the Dimensional Overlap Model (Kornblum et al., 1990): for the Flanker Task, the Dimensional Overlap Model posits that it is an example of a task involving a Stimulus-Stimulus conflict. In such tasks, there is a conflict between items of the same stimulus (i.e., arrows) which originate and are resolved at the perceptual level. In contrast, for the Simon Task, the model posits that the task involves a stimulus-response conflict. In such instances, the conflict arises due to the stimulus dimension (here: stimulus location) being irrelevant to the response rule and the response dimension (Kornblum, 1994; Paap, 2019; Blumenfeld & Marian, 2014). Using these two different kinds of tasks allows us to better understand which kind of conflict monitoring mechanisms are employed in linguistic control.

4.4.3.1. Flanker Task

In the Flanker Task (Eriksen and Eriksen, 1974), participants respond to the direction of an arrow at the centre of the screen, which is flanked by two non-target arrows to either side. The direction of the non-target items either corresponds to the direction of the target item (congruent Flankers) or all show the opposite direction (incongruent Flankers). In this study, participants are asked to indicate the direction of the target stimulus by pressing the "3" key on the keyboard for left and the "9" key for right directionality.

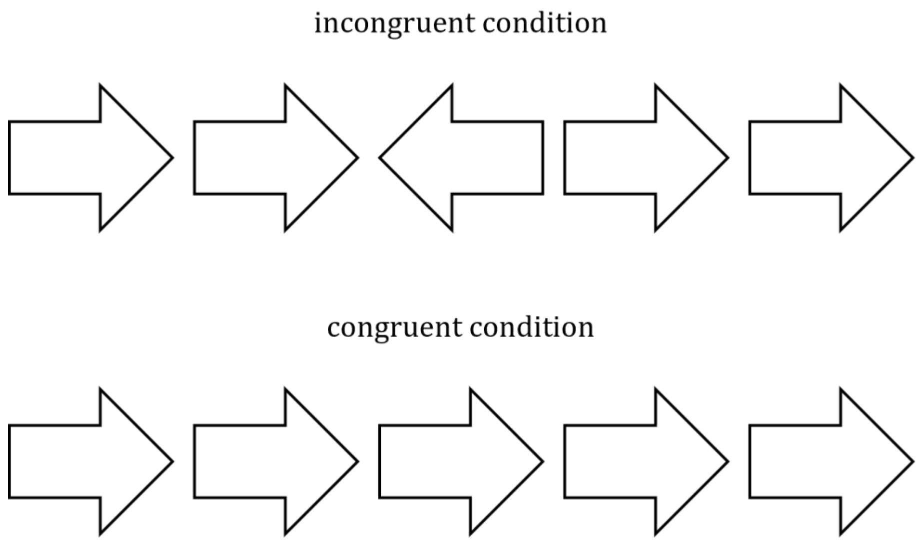


Figure 17: Incongruent and congruent conditions in the Flanker Task

4.4.3.2. Simon Task

The Simon Task (Simon and Rudell, 1967) involves suppressing spatial information in favour of colour information. Participants respond to the colour of the stimulus they see, which appears on either the left or the right side of the screen and are asked to press the left number “3” key whenever they see a red square and the right number “9” key when they see a green square, regardless of the side the stimulus appears on. Trials are "congruent" where the colour stimulus appears on the side of the screen that corresponds with the relative location of the response key (e.g., the red dot appears on the left side of the screen).

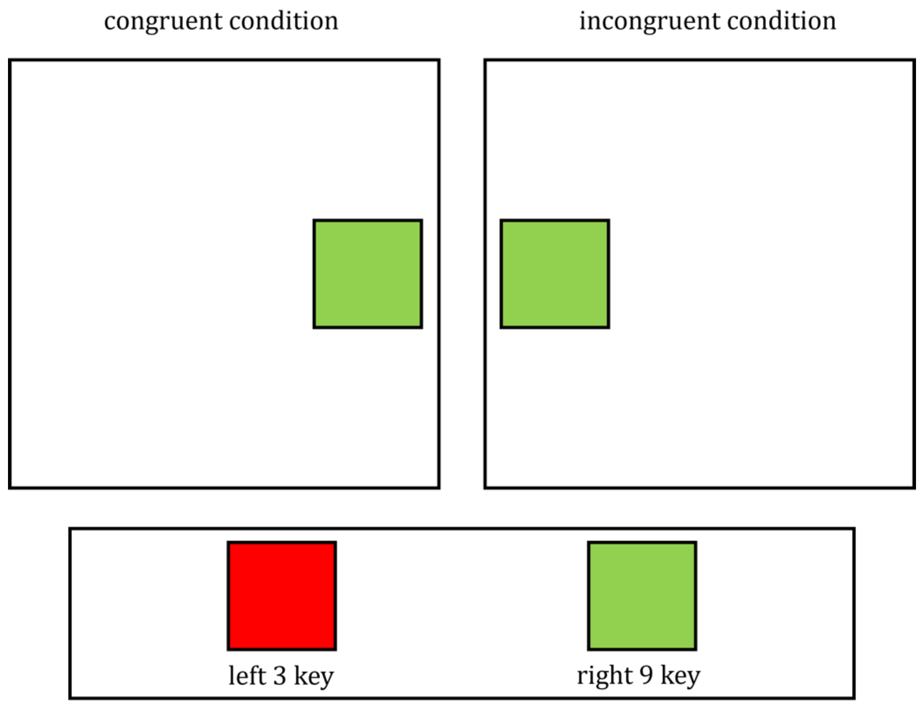


Figure 18: Congruent and incongruent conditions in the Simon Task

4.4.3.3. Dual-Language Picture Naming Task

In the dual-language picture naming task, bilingual participants engaged in naming items depicted in pictures as quickly as possible, using one of two languages. Oral responses were recorded as sound files. A coloured frame around each picture indicated whether the item should be named in Chinese (red) or English (blue). The experiment encompassed four consecutive blocks: an initial training block with 8 trials, a Chinese-target block comprising 50 trials, an English-target block also consisting of 50 trials, and a mixed block presenting both Chinese and English targets in semi-random order across 50 trials. This semi-random alternation ensured that approximately half of the trials were preceded by a trial in the other language ("switch"), while the other half followed a trial in the same language ("stay"). For monolingual participants, no such block distinction was applicable, as they exclusively underwent testing in English without any visual cues regarding response language. Each trial initiated with a 100ms acoustic signal, with the visual stimulus presented 500ms after the sound cue onset. Participants concluded each self-paced trial by pressing the space bar.

The task employed a randomized selection of pictures throughout the experimental blocks, drawn from a total pool of 520 images. This image collection was derived from Bates et al.'s (2003) study that examined timed picture naming across seven languages, providing insights into cross-linguistic variations in naming latencies and word frequencies.

4.4.3.4. Language Background Questionnaire

All participants completed a sociolinguistic background questionnaire as part of the first session. Questions included date of birth, gender, and knowledge of any second language. For the bilingual group, questions also included the date of arrival in the United Kingdom, order of acquisition of languages, any language test results (IELTS, TOEFL, university-internal language examinations), and whether they had taken part in any pre-sessional English classes (and if so, how many hours per week, and for how long), attendance of language classes (and if so, how many hours per week, for how long, and what kind of exercises are involved).

4.4.4. Data Analysis

4.4.4.1. Dual-Language Picture Naming Task

Using the voice recordings collected during the experiment, we manually measured reaction times and labelled responses in Praat (Boersma & Weenik, 2018). The coding system for response language and item accuracy was as follows: "ENG" represented English responses, and "CH" Chinese responses. Language accuracy was coded as "1" when the response language matched the target language and "0" when it did not. Item accuracy was coded as "1" when the item response matched the target response, and as "0" when it did not. Trials were classified as "non-switch" if the previous trial aimed to elicit a response in the same language as the current trial, and "switch" if the previous trial's target language differed from the current one.

Response times were computed from the commencement of the acoustic signal to the onset of the participant's reaction. To account for the temporal gap between signal initiation and visual stimulus presentation, 500ms were subtracted from the recorded time. Reaction times surpassing the mean plus two standard deviations (2770ms) were levelled, data from the initial trial within each block, along with trials featuring reaction times above 4000ms and below 250ms, and trials with inaccurate responses were excluded.

A comparatively large number of trials were excluded from analysis: 24.1% of all bilingual trials and 10.4% of all monolingual trials (Appendix 2.A: Data Exclusion Rates). Comparable studies only excluded around 5% of their data (e.g., Bonfieni, 2019).

Following Declerck and Philipp (2015), we calculated switch cost for bilinguals for each session and language by subtracting the average response times of the switch trials of the mixed block from the average reaction time of the non-stay trials. We calculated bilinguals' mix costs for each session and language by subtracting the average response times in the non-switch trials in the mixed-task condition and trials in the single-task condition of the corresponding language.

In accordance with Birdsong (2016), we computed a language balance metric using the method of "between-language subtractive differentials," which involves the subtraction of performance scores between the two languages. This means that for every participant and each session, we subtracted their average RT for the Chinese single-language block from their average RT for the English single-language block. Switch Cost Asymmetry is calculated by subtracting the average switch costs from the L1 from the L2. We have calculated mix cost asymmetry by subtracting the average mix costs in the L1 from the average mix cost in the L2. We calculated language balance, switch- and mix cost asymmetry for all participants and for each session individually.

4.4.4.2. Flanker and Simon Tasks

We removed the first trial of each block, missing values and responses with latencies beyond two standard deviations of the remaining data set (663ms - 4.13% for Simon, 5% for Flanker). Due to high accuracy in either task (> 95%), we only analysed response latencies.

4.4.4.3. Statistical Analysis

We run a series of linear mixed effects models to test our hypotheses using Jamovi and R (The Jamovi Project, 2022; Gallucci, 2019; R Core Team, 2021). For each analysis, we construct an optimal model by adding random and fixed factors on a step-by-step basis, retaining factors only when the more complex model is significantly superior to the reduced model. We used the Akaike Information Criterion (AIC) as an estimator of prediction error and conduct an ANOVA test for the models' goodness-of-fit relative to one another. Fixed and random factors were adopted to specifically test the predictions we have set out.

We split both the Flanker and the Simon datasets in two ways: the first set included both the bilingual and monolingual data and allows for comparisons between the two groups. Further, by creating a variable that captures both session and group, “GroupSession”, we could easily compare the bilingual performance across the three sessions with the monolingual controls’ outcomes. We compared the bilinguals’ overall RT, interference scores and mix cost to the monolinguals’. The resulting model descriptions are as follows:

Model 1: Flanker Monolinguals vs. Bilinguals. Code: RT ~ GroupSession*Condition + GroupSession*congruency + Condition*congruency + (1|participant), data = all flanker data.

Model 3: Simon Monolinguals vs. Bilinguals. Code: RT ~ GroupSession* condition* congruency + (1|participant), data = all Simon data.

For the second analysis, we excluded monolingual data and focussed on bilingual data, which allowed for clean testing of the effect Session, and additional predictor variables which were sourced from the outcomes of the picture naming task. We tested the overall impact on RT, the effect on interference scores (in an interaction with congruency), and effect on mix cost (in an interaction with condition) of the factors for per participant outcome variables for

Chinese single-language RT,	Chinese switch cost,	Chinese mix cost,
English single-language RT,	English switch cost,	English mix cost,
Language Balance.	Switch cost Asymmetry.	Mix cost asymmetry.

For the Flanker Task, language balance, Chinese switch cost and English switch cost significantly improved model fit, and for the Simon Task, language balance, “SwitchCostAsymmetry” and “MixCostAsymmetry” (in an interaction effect with Condition only) improved model fit. The resulting model descriptions are as follows:

Model 2: Flanker Bilinguals Language Control and Executive Control. Code: RT ~ Session * Condition + Condition * Congruency + LanguageBalance + ChineseSwitchCost + EnglishSwitchCost + (1|participant), data = all bilingual flanker data.

Model 4: Simon Bilinguals Language Control and Executive Control. Code: RT ~ Session + LanguageBalance + Condition * Congruency +Condition * MC-Asymmetry + Congruency * SC-Asymmetry + (1|participant), data = all bilingual Simon data.

4.5. Results

First, we will provide a brief overview of the results of the main findings of the dual-language picture naming task. Second, for both the Flanker and the Simon Task, we first provide results of the comparison between monolinguals and bilinguals in terms of overall RT, interference scores and mix cost. In a second step, we present results of the investigation of bilingual data with the addition of predictor variables sourced from outcomes of the picture naming task.

4.5.1. Dual-Language Picture Naming Task

The analysis of the picture-naming task revealed that the participants' L1 was still overwhelmingly active and easy to access in early stages of bilingual development, while L2 retrieval was associated with longer naming latencies ($p<0.001$). Our results also showed that the bilingual's L1 retrieval did not elicits significantly larger naming latencies than monolingual L1 retrieval ($p=0.69$). Figure 19 shows the bilinguals' performance of response times with respect to the monolinguals' overall performance and across sessions.

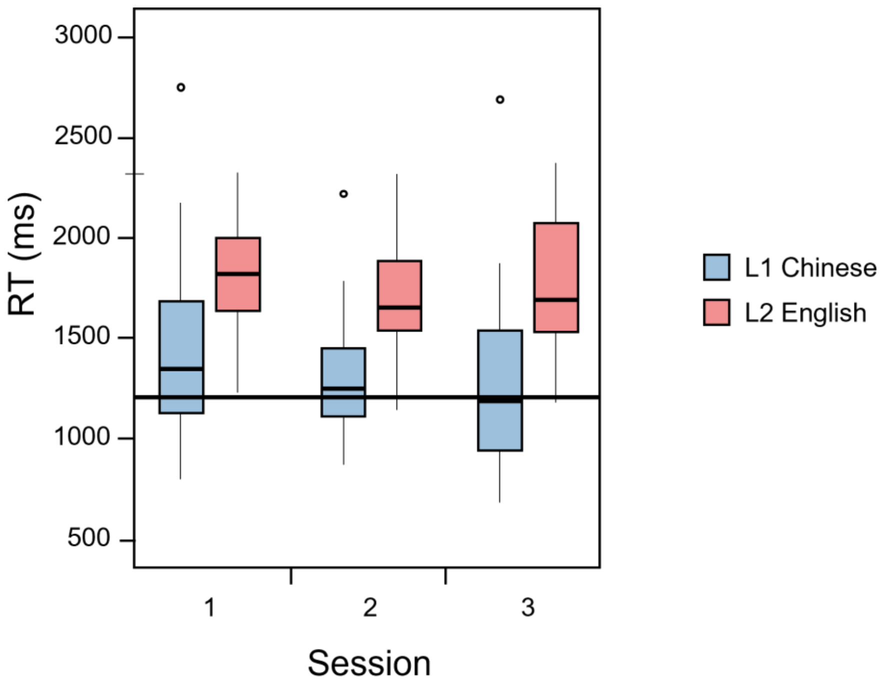


Figure 19: Average RT in the single-language PNT blocks by language and Session. Whiskers reflect extreme values that are no outliers (within 1.5x IQR). The line at RT=1186ms represents the monolingual response time average.

Figure 20 shows the magnitude of the bilinguals' switch costs into Chinese and English.

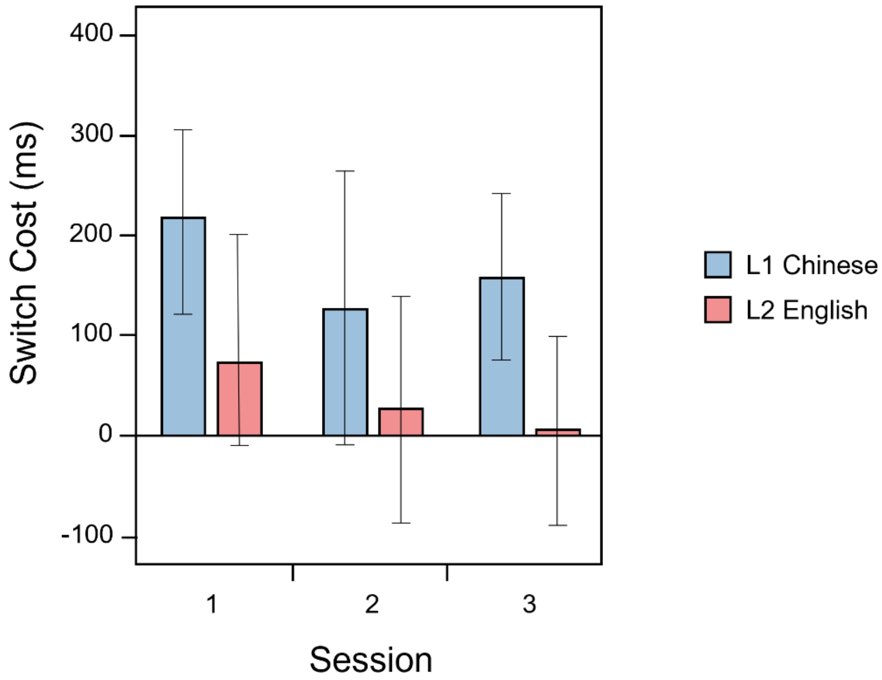


Figure 20: Switch cost into Chinese and English by Session.
Error bars reflect 95% Confidence Intervals.

The analysis of switch cost (the difference in RT between non-switch and switch trials in the mixed-task task type) revealed a significant main effect of trial type ($p<0.001$), echoing the finding that switch trials were slower than non-switch trials ($\beta=100.4$, $SE=17.2$, $p<0.001$), as well as a main effect of language ($p<0.001$), as Chinese items were named faster than English items ($\beta=-129.7$, $SE=37.2$, $p<0.001$), and a significant main effect of frequency ($p<0.001$), which indicates that more frequent words were named faster.

We have found that the mixed-language condition facilitated naming more than the single-language condition. Figure 21 shows the magnitude of the bilinguals' mix costs into Chinese and English.

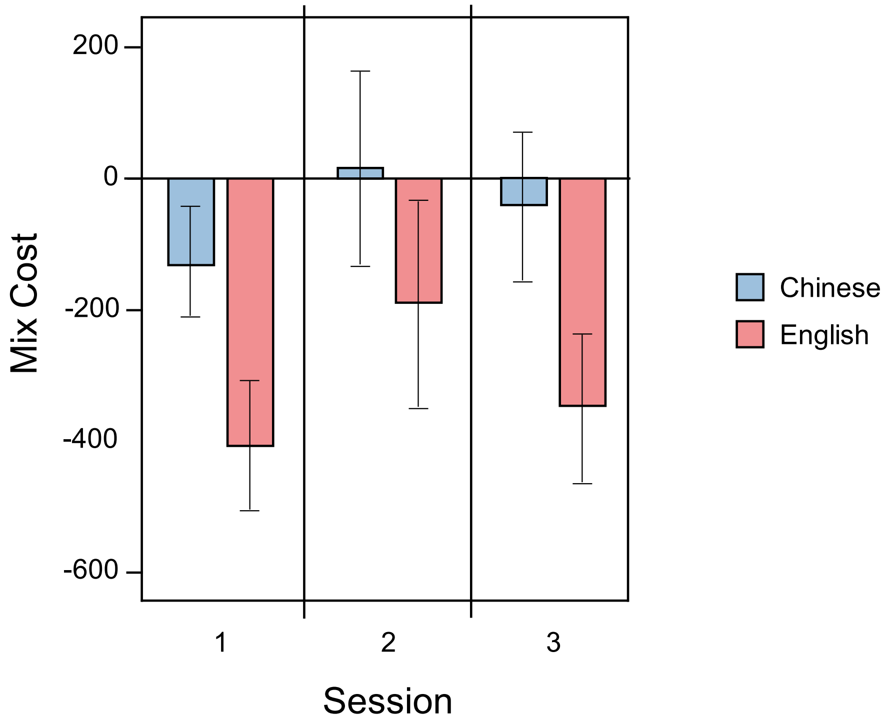


Figure 21: Mix cost of Chinese and English by Session.

Error bars reflect 95% Confidence Intervals.

We have found asymmetries between L1 and L2 in both switch cost and mix cost, and that these symmetries decline with increasing language balance. For details on the results, we refer the interested reader to 3.5. Results.

4.5.2. Flanker Task

4.5.2.1. Bilingual Development in Comparison with Monolingual Control

The analysis of all Flanker data revealed a significant main effect of GroupSession ($p< 0.001$), showing that in terms of overall RT, bilinguals significantly outperformed monolinguals in Session 3 (β=-30.1, SE=12.5, p=0.02), but not Session 1 or Session 2 (p=0.059 and p=0.054, respectively). A post-hoc Bonferroni analysis of GroupSession reveals that overall, bilingual RT does significantly differ between sessions 1 and 2 and session 3 ($p<0.001$ and p=0.001, respectively), but not between sessions 1 and 2.

```
Code: RT ~ GroupSession*Condition + GroupSession*congruency + Condition*congruency +
(1|participant), data = all flanker data.
```

Table 16: Flanker performance bilinguals vs. monolinguals by session

Fixed effects	Estimate	Std. error	P-Value
(Intercept)	394.93	6.35	<0.001
GroupSession BIL-1	-24.03	12.44	0.06
GroupSession BIL-2	-24.62	12.44	0.05
GroupSession BIL-3	-30.12	12.45	0.05
Congruency incongruent	62.14	1.03	<0.001
Condition single	-10.60	1.09	<0.001
GroupSession BIL-1: Congruency incongruent	-7.87	3.21	0.01
GroupSession BIL-2: Congruency incongruent	-12.51	3.22	<0.001
GroupSession BIL-3: Congruency incongruent	-11.19	3.25	<0.001
GroupSession BIL-1: Condition single	-3.71	3.23	0.3
GroupSession BIL-2: Condition single	-6.43	3.24	0.05
GroupSession BIL-3: Condition single	-11.06	3.28	<0.001
Congruency incongruent: Condition single	12.28	2.15	<0.001

There was also a significant main effect of Congruency, revealing that incongruent items elicited longer RTs than congruent items (β=62.1, SE=1.1, $p<0.001$) and a significant main effect of Condition ($p<0.001$), showing that single-condition trials were completed faster than mixed-condition trials (β=-10.6, SE=1.1, $p<0.001$). The addition of an interaction effect of Condition and Congruency significantly improved model fit ($p<0.001$), showing that the facilitation yielded through the single- condition is less for incongruent trials than for congruent trials.

The addition of a three-way interaction between GroupSession, Condition and Congruency did not significantly contribute to improved model fit ($p=0.6$), meaning that this effect is independent of GroupSession.

The model also shows a significant main effect of congruence ($p<0.001$), reflecting that incongruent trials were completed more slowly than congruent trials ($\beta=-48.7$, $SE=3.06$, $p<0.001$), and trial type ($p<0.001$), echoing the finding that switch trials elicited longer reaction times than non-switch trials ($\beta=55.8$, $SE=3.06$, $p<0.001$) and a main effect of Condition ($p<0.001$), showing that single-condition trials were completed slower than mixed-task trials ($\beta=44.3$, $SE=3.06$, $p<0.001$).

We found a reliable interaction between GroupSession and Congruency ($p<0.001$), which reveals that interference scores (RT in incongruent trials versus in congruent trials) significantly differ between functionally monolingual speakers and bilinguals across all three sessions (see Figure 22 and Figure 23). This shows that bilinguals are significantly better at handling interference than functionally monolingual speakers.

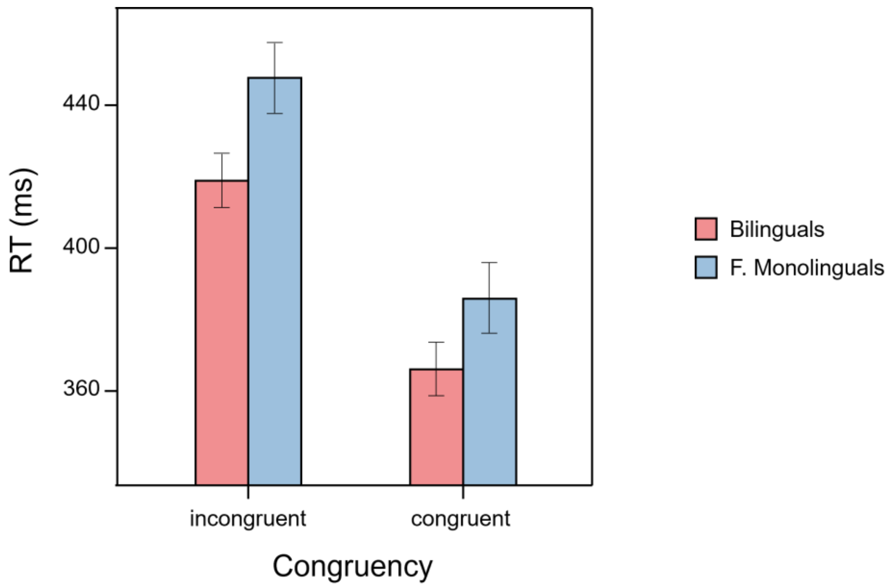


Figure 22: Flanker average RT by congruency by group.
Error bars reflect 95% Confidence Intervals.

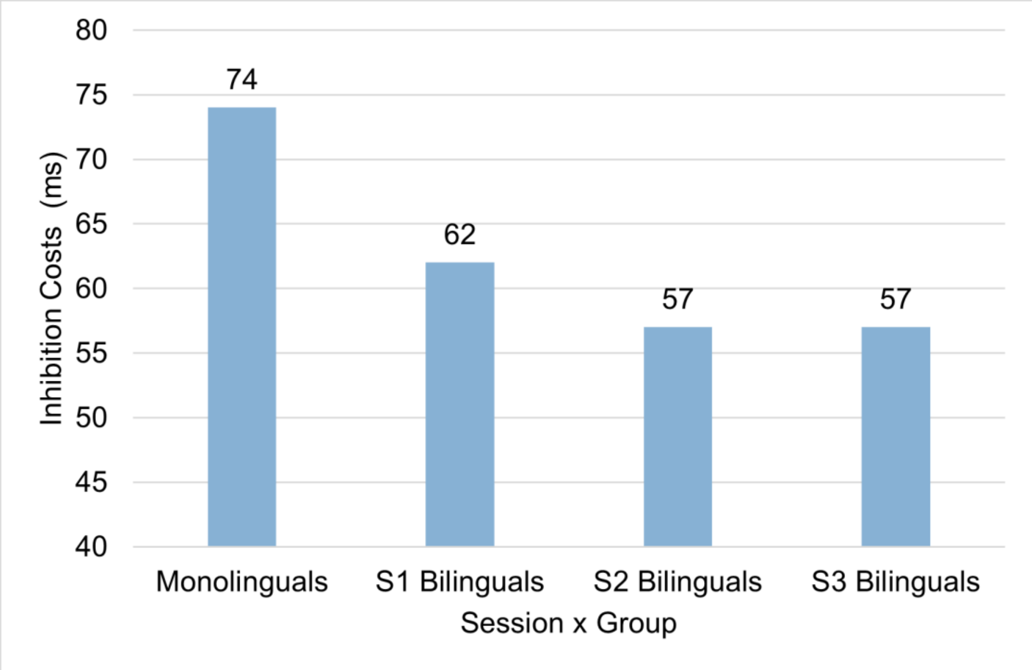


Figure 23: Flanker inhibition costs by Group and Session

Within bilinguals, interference scores did not significantly change across the sessions: in the model presented in Table 17, which only includes bilingual data, the interaction effect of Session and Congruency failed to reach significance. As such, the improved interference score does not improve with immersion duration.

Similarly, we have also found a significant interaction between GroupSession and Condition, which shows that the final two of the bilinguals' three sessions elicited significantly larger mix costs than the monolinguals' control session. Again, we tested the interaction of Session and Condition using model presented in Table 17, which only includes bilingual data. This time, there was a reliable interaction between Session and Condition ($p<0.001$), which shows that mix cost in session 3 was significantly larger than in session 1 (β =-9.7, SE=2.9; $p<0.001$).

Figure 24 illustrates how mix cost increases in bilinguals between Session 1 and Session 2 due to a greater facilitation effect of the single condition.

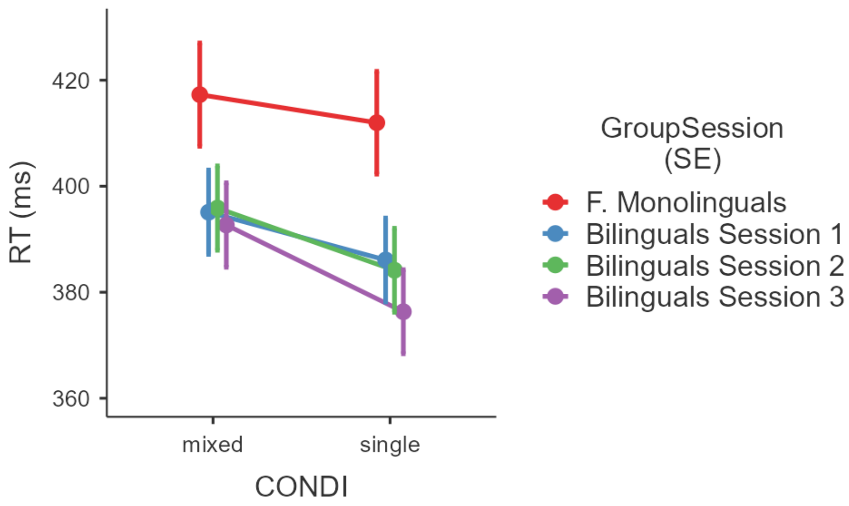


Figure 24: Flanker mix costs for monolinguals and bilinguals by session. Error-bars reflect standard errors.

4.5.2.2. Language Control Outcomes as a Predictor of Flanker Performance

The analysis of bilingual Flanker data with the addition of outcome data from the dual-language PNT yielded a significant effect of language dominance, showing that overall RT decreases with increasing L2 balance, and significant main effects of both languages' switch costs, showing that overall RT increases with increasing switch cost in both Chinese and English.

```
Code: RT ~ Session*Condition + Condition*Congruency + LanguageBalance +
ChineseSwitchCost+ EnglishSwitchCost + (1|participant),
data = all bilingual flanker data.
```

Table 17: Language control outcomes as a predictor of flanker performance

Fixed effects	Estimate	Std. error	P-Value
(Intercept)	387.83	8.24	<0.001
Condition single	-12.32	1.26	<0.001
Congruency incongruent	60.06	1.25	<0.001
Session 2	-2.79	1.68	0.097
Session 3	-3.42	1.60	0.05
Language balance	-0.02	0.01	0.002
Switch Cost into English	0.01	0.00	<0.001
Switch Cost in Chinese	0.01	0.00	0.001
Condition single: incongruent	11.28	2.50	<0.001
Condition single: Session 2	-4.36	3.06	0.155
Condition single: Session 3	-9.69	2.89	<0.001

We did not find any significant interaction effects of LanguageBalance, ChineseSwitchCost, EnglishSwitchCost or any other outcome variables on neither Congruency nor Condition, indicating that the selected language control outcomes do not predict interference scores, or mix cost.

4.5.3. Simon Task

4.5.3.1. Bilingual Development in Comparison with Monolingual Control

The analysis of all flanker data revealed a significant main effect of GroupSession ($p < 0.001$). While monolinguals did not outperform bilinguals in terms of overall RT in any of the three sessions, a post-hoc Bonferroni analysis revealed that there were significant differences between the bilingual sessions, as bilinguals became progressively faster over time.

```
Code: RT ~ GroupSession*Condition*congruency + (1|participant),
data = all Simon data.
```

Table 18: Simon performance bilinguals vs. monolinguals by session

	Fixed effects	Estimate	Std. error	P-Value
	(Intercept)	384.41	7.68	<0.001
	GroupSession BIL-1	5.28	15.09	0.728
	GroupSession BIL-2	-3.58	15.11	0.814
	GroupSession BIL-3	-14.52	15.10	0.341
	Condition single	-45.56	1.62	<0.001
	Congruency incongruent	21.09	1.62	<0.001
Mix cost	GroupSession BIL-1: Condition single	-10.90	4.73	0.05
	GroupSession BIL-2: Condition single	-11.75	4.84	0.05
	GroupSession BIL-3: Condition single	-14.75	4.81	0.01
Cong cost	GroupSession BIL-1: congruency incongruent	-6.46	4.73	0.172
	GroupSession BIL-2: congruency incongruent	-6.61	4.84	0.172
	GroupSession BIL-3: congruency incongruent	0.70	4.80	0.884
Cong*mix	Condition single: congruency incongruent	16.88	3.24	<0.001
	GroupSession BIL-1: Condition single: congruency incongruent	19.51	9.45	0.05
	GroupSession BIL-2: Condition single: congruency incongruent	15.47	9.67	0.110
	GroupSession BIL-3: Condition single: congruency incongruent	34.64	6.61	<0.001

We have also found a significant main effect of Condition ($p<0.001$), which showed that the single condition elicited faster reaction times than the mixed condition (β =-45.6, SE=1.6, $p=0.001$) and a significant main effect of Congruency ($p<0.001$), as incongruent trials elicited larger reaction times than congruent trials (β =21.1, SE=1.6, $p=0.001$).

As is the case in the Flanker Task, the significant interaction effect between Congruency and Condition shows that the facilitation yielded through the single condition is less for incongruent trials than for congruent trials (β =16.9, SE=3.2, p=0.001).

We have also found a marginally significant interaction effect of GroupSession and Condition (p=0.05), showing that bilinguals had larger mix costs than functional monolinguals across all three sessions (Figure 25). Within bilinguals, however, Condition did not significantly vary with Session, suggesting that bilingual mix cost did not significantly change throughout the testing period.

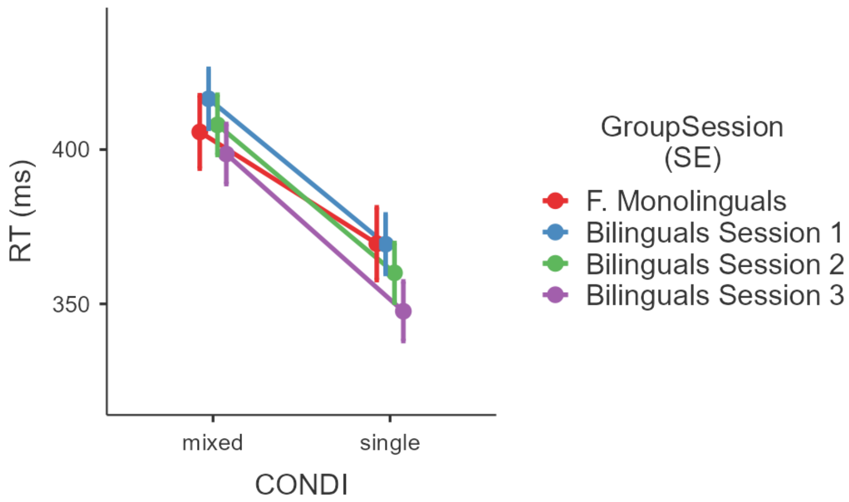


Figure 25: Simon mix cost for monolinguals and bilinguals per session. Error bars reflect Standard Errors.

The interaction effect of GroupSession and Congruency failed to reach significance but was retained in the model as a basis for the significant three-way-interaction between GroupSession, Condition and Congruency (p=0.004), which reveals that bilinguals have an additional facilitation effect of Congruency and single-condition in the first and third testing sessions, when compared to monolinguals (incongruent trials are significantly slower: β =19.5, SE=9.5, p=0.04 in Session 1, β =34.6, SE=9.61, p=0.001 in Session 3). The addition of the interaction effect of Session, Condition and Congruency failed to significantly improve the model in Table 19, where only bilingual data is entered, indicating that this additional facilitation effect does not significantly expand over time.

4.5.3.2 Language Control Outcomes as a Predictor of Simon Performance

The analysis of bilingual Simon data with the addition of outcome data from the dual-language PNT returned a marginally significant effect of Language Balance, showing that overall RT decreases with increasing L2 use. We also found a significant main effect for Switch Cost Asymmetry, which reveals that overall RT decreases as switches into Chinese become more costly as opposed to switches into English (Figure 25).

Code: RT ~ Session+ LanguageBalance + Condition*Congruency+Condition*MC-Asymmetry+Congruency*SC-Asymmetry + (1|participant),
data = all bilingual Simon data

Table 19: Language control outcomes as a predictor of Simon performance

Fixed effects	Estimate	Std. error	P-Value
(Intercept)	383.21	10.64	<0.001
Session 2	-10.89	2.47	<0.001
Session 3	-18.29	2.15	<0.001
Condition single	-50.01	1.87	<0.001
Congruency incongruent	21.09	1.88	<0.001
Switch cost asymmetry	-0.02	0.01	<0.001
Mix cost asymmetry	-1.58e-4	0.01	0.979
Language balance	0.02	0.01	0.05
Condition single: congruency incongruent	25.65	3.75	<0.001
Condition single: mix cost asymmetry	-0.02	0.01	0.01
Congruency incongruent: switch cost asymmetry	0.03	0.01	<0.001

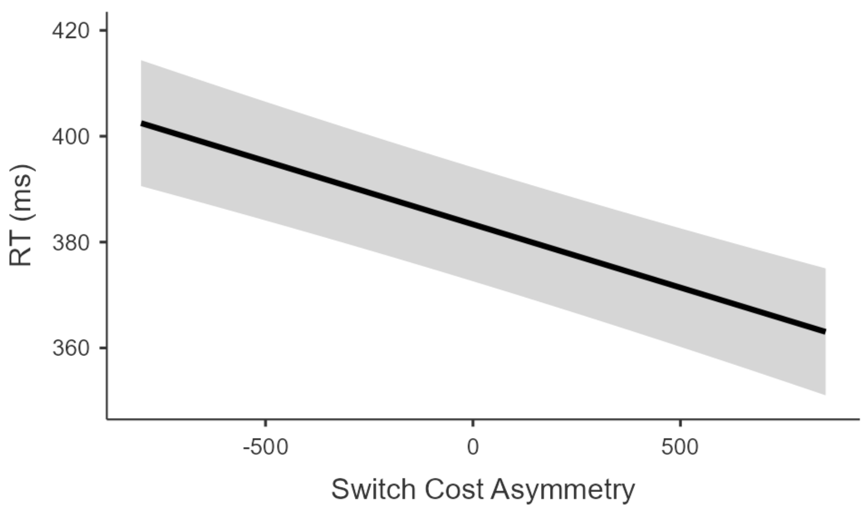


Figure 26: Linguistic switch cost asymmetry predicts Simon RT. Shaded area reflects Standard Errors.

Further, there was a significant interaction effect of Switch Cost Asymmetry on Congruency, as those participants with a larger, positive switch cost asymmetry (accessing English is easier as opposed to accessing Chinese) also have larger interference scores, although the overall reaction time was lower.

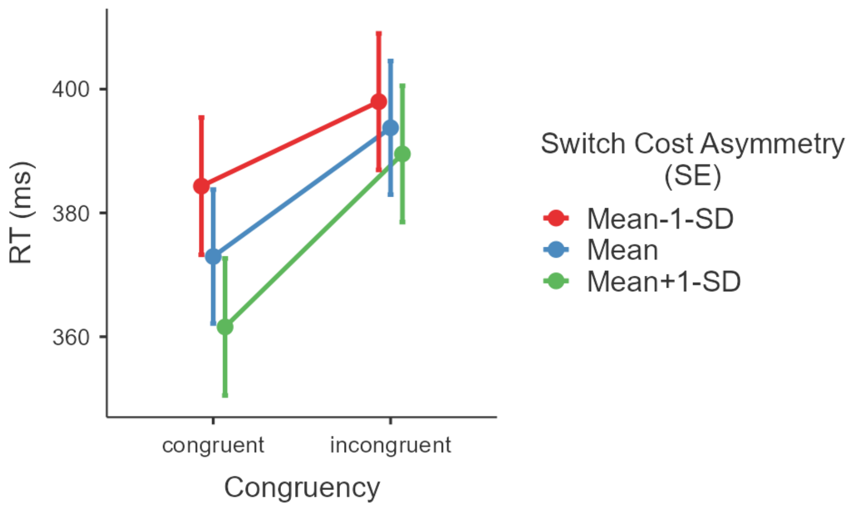


Figure 27: Simon interference scores by PNT switch cost asymmetry. Error bars reflect Standard Errors.

Mix Cost Asymmetry did not significantly improve model fit when entered as a fixed main effect but was retained to support the interpretation of the significant interaction effect of Mix Cost Asymmetry on Condition. Like the interaction effect of Switch Cost Asymmetry and Congruency, the latter shows that those participants with larger, positive mix cost asymmetry in the dual-language PNT (mixing in English is easier than in Chinese) also have a larger mix cost in the Simon Task.

4.5.3. Intra-Individual Variation

The correlational Spearman analyses of inconsistency, variability, and dispersion CV scores across the linguistic and cognitive tasks revealed that variability scores from the English PNT and the Flanker Task are significantly correlated ($\rho=-0.377$, $p=0.048$). This suggests that the less variable an individual was in the Flankers Task across the three sessions, the more variable they were in the English PNT across the three sessions. Other correlations almost reached statistical significance: Chinese and English inconsistency scores correlated almost significantly ($\rho=0.227$, $p=0.053$) and so did Flanker and Simon dispersion scores ($\rho=0.205$, $p=0.067$). It may be the case that a larger N would have led to more significant findings.

		CH-PNT	**EN-PNT**	**Simon**
Inconsistency within task within session	**EN-PNT**	$\rho=0.227$ $p=0.053$	-	-
	Simon	$\rho=0.108$ $p=0.372$	$\rho=-0.137$ $p=0.254$	-
	Flanker	$\rho=0.015$ $p=0.9$	$\rho=-0.016$ $p=0.896$	$\rho=0.139$ $p=0.215$
Variability within task across sessions	**EN-PNT**	$\rho=0.299$ $p=0.130$	-	-
	Simon	$\rho=-0.06$ $p=0.722$	$\rho=-0.205$ $p=0.305$	-
	Flanker	$\rho=-0.035$ $p=0.861$	**$\rho=-0.377$*** **$p=0.048$**	$\rho=-0.185$ $p=0.356$
		PNT		**Simon**
Dispersion across tasks within session	**Simon**	$\rho=-0.08$ $p=0.502$		-
	Flanker	$\rho=0.150$ $p=0.203$		$\rho=0.205$ $p=0.067$

4.6. Discussion

The aim of the present study was to longitudinally investigate the effect of linguistic control and executive control in Chinese international students pursuing a degree in the UK. We employed dual-language PNT, as well as Flanker and Simon Tasks at three different time points during their first year of stay in the United Kingdom.

We predicted that bilinguals would consistently outperform the baseline measure provided by the functional monolinguals in terms of i) overall RT, ii) interference scores and iii) mix cost. We also predicted that bilinguals would improve their performance with immersion duration (session). Evidence for a bilingual advantage in executive control tasks is mixed. In the Flanker Task, bilinguals significantly outperform the monolingual baseline measure in terms of overall RT in their third session. This suggests that bilinguals developed faster response times, but this could be both due to their L2 immersion experience and their repeated exposure to the task.

Bilinguals were also significantly better at handling interference as opposed to the monolingual measure, as evidenced by consistently smaller interference scores. We have, however, found significantly larger mix costs for bilinguals in sessions 2 and 3 as opposed to the functional monolinguals, which can be interpreted in different ways. First, this could mean that bilinguals are comparatively worse at using two mental task sets concurrently. However, considering the increased mix cost is a result of increased facilitation of the single-condition (as opposed to worse performance on the mixed-task condition), the increased mix cost more likely reflects that the bilinguals have reached a ceiling effect. The finding that mix cost increases significantly between the bilinguals' session 1 and session 3 supports this idea.

In the Simon Task, bilinguals did not outperform the measure of performance for the functionally monolingual speakers in terms of overall RT or interference score in any of the three sessions. Bilinguals mix costs were higher than those of the functional monolinguals. As with the mix cost in the Flanker Task, we propose that higher mix costs are in fact a reflection of better superior performance in the task overall, as it shows that bilinguals reach faster reaction times. Bilinguals also became faster at completing trials in the Simon Task, as overall RT significantly decreased for bilinguals between sessions 1 and 3. But interference scores and mix cost remained stable throughout.

The analysis of executive control tasks using outcomes from the dual-language PNT as predictors revealed that more language control outcomes affected performance of the Simon Task, as opposed to the Flanker Task. Switch cost asymmetry and L2 use had a significant effect on overall Simon RT, as the latter decreased with smaller switch cost asymmetry and higher L2 use. Further, those participants with a smaller PNT mix cost asymmetry also had a smaller Simon mix cost, and those participants with a smaller PNT switch cost asymmetry, again, had smaller Simon interference scores. As such, the language control outcomes correlate greatly with Simon outcomes. For the Flanker Task, L2 use, English switch cost and Chinese switch cost performance had a significant effect. Those with lower switch costs (in both Chinese and English) and higher L2 use are performing better in terms of overall RT on the Flanker Task. However, no PNT outcome had a significant impact on Flanker interference scores, or Flanker mix cost.

Correlational analyses on IAV measures showed a significant, negative correlation of variability scores from the English PNT and the Flanker Task significantly, suggesting that those who, across the three sessions, varied less in the Flankers Task, varied more in the English PNT. Some correlations – such as IAV correlations between Language Control measures and Executive Control measures – only approached significant levels. While it is a valuable approach to run correlational analysis on IAV measures, larger N – either through an increase in participants or testing sessions – is likely needed to obtain meaningful results (e.g., Pfenninger and Kliesch; 2023).

To avoid redundancy, an overview of the limitations of this analysis as well as of the study as a whole are presented in 6.3. Limitations. Specifically, the appropriateness of employing a functionally monolingual group as comparative measure as well as practise effects within the individual experimental sessions and between experimental sessions run are reflected upon in this subchapter.

4.7. Conclusion

The aim of the present study was to longitudinally investigate the interconnectedness between language control and executive control in the first year of L2 immersion in Chinese migrants in the UK. We have found that bilinguals did outperform monolinguals on several measures of executive control. However, immersion duration did not significantly contribute. Instead, language balance had a significant effect on several measures on language control and executive control.

5. Effects of Language Use on Language Control and Executive Control

A Longitudinal Study Within the First Six Months of L2 Immersion

Abstract

Bilinguals have varying demands when switching between their languages: while some use their languages in relatively more independent contexts (e.g., at work vs at home), others switch between languages many times during the day. The present study investigates how dual language use affects language control and executive control. Data on language use were obtained through an Activity Log Questionnaire. Experimental data were acquired by means of a dual-language picture naming task, a Simon task, and a Flanker task. Thirty Chinese-English bilinguals completed the battery of tests during three sessions over six months following their start of tertiary education in England. Results suggest that those who habitually code-switch have advantages in both language and executive control efficiency. Additionally, those with increased use of L2 in social contexts and academic contexts also performed better on several measures. Overall, our results support the notion that the type of language use affects language control and executive control capabilities and as such supports the Adaptive Control Hypothesis (Green & Abutalebi, 2013).

5.1. Introduction

There has been a growing interest in bilinguals' ability to switch between their languages with accuracy and ease. Previous research has demonstrated that bilinguals' languages are in a constant state of co-activation and that to access a target language selectively, bilinguals must actively inhibit the selection of the non-target language (e.g., Green, 1998; Poulisse & Bongaerts, 1994; de Bot & Schreuder, 1993; Kroll & Sunderman, 2003; Kroll et al., 2006). The process through which this competition is managed is referred to as both language control and linguistic control (Abutalebi & Green, 2007, 2008; Green and Abutalebi, 2013).

There is an ongoing debate over the extent to which language control translates to domain-general executive control. Executive control can be defined as a set of higher-order cognitive processes that are responsible for regulating and controlling behaviour. These processes involve the inhibition of irrelevant information to enable focus on the target information, such as suppressing irrelevant information, shifting between mental tasks and rules, and updating and maintaining representations stored in working memory (Miyake et al., 2000; Friedman & Miyake, 2004; Friedman et al., 2006).

Proponents of the bilingual advantage argue that bilinguals must exert control on their language production facilities, which gradually train the language selection mechanism – enabled by the brain's ability of neuroplastic transformation – and that this training transfers to non-linguistic functions of selective inhibition (Bialystok, 2009). Supporting the link between linguistic and non-linguistic control are studies such as Gollan et al. (2014), who found that individuals who encountered difficulty in language switching also found it challenging to transition between naming numbers and performing basic arithmetic operations.

Other studies have, however, pointed out that research on this topic is inconclusive. In a comprehensive meta-analysis of synthesizing data from 152 studies involving 891 effect sizes on adults, including unpublished data and various study-related factors, Lehtonen et al. (2018) initially find a slight advantage for bilinguals in inhibition, shifting, and working memory (effect size: $g = +0.06$ [0.00, +0.13]), which, however, disappeared after correcting for publication bias ($g = -0.07$ [−0.17, +0.04]). Similarly, in a meta-analysis of 46 studies on the matter, Noort et al. (2019) concluded that more than half the studies had positive results indicating that bilingualism enhances cognitive control tasks, while 28% reported conflicting outcomes and 17% exhibited opposite findings. Opponents of the bilingual advantage debate instead suggest that any advantages are likely due to external influences, such as individual differences, including pre-existing neurocognitive abilities (Lai et al., 2017), age (Massa et al., 2020; Bialystok et al., 2004), language exposure (Bonfieni et al., 2019) and proficiency (Abutalebi et al., 2013), cultural diversity (Samuel et al., 2018), or everyday interests – such as meditation (Teper and Inzlicht, 2013) or video games (Strobach et al., 2012), or linguistic factors – such as language similarity (Kirk et al., 2022) – rather than any specific effect of speaking two languages.

Methodological differences between studies have also been pointed out to weigh heavily on the likelihood of achieving positive results (Baum & Titone, 2014; Noort et al., 2019; Giovannoli et al., 2020). Noort et al. (2019), for instance, highlight variation in the selection of bilinguals, the use of non-standardised tasks and a shortage of longitudinal studies. Ultimately, it is generally agreed that more research is needed to disambiguate the effect of the bilingual experience on domain-general executive control facilities.

One factor of particular interest is the effect of language use on language control and executive control: the needs of bilinguals concerning the transition between their languages differ; for instance, some bilinguals may only use one language in specific settings (e.g., at work compared to home), while others routinely switch within the same setting. It has been proposed that the advantage bilinguals have in linguistic and executive control could be linked to how often they use both languages during conversations in everyday life and in which situational contexts. Those who frequently switch between languages - and in particular, those who do so in more formal/work contexts, such as professional interpreters – may be more practised with monitoring processes. Therefore, they may display better linguistic control and executive control performance than those bilinguals who mix very little, who have distinct sociolinguistic environments for each language, and who use one or both languages in more informal contexts (Costa et al., 2009; Green, 2011; Green & Abutalebi, 2013; Green & Li, 2014).

While some studies have found support for this notion (e.g., Ooi et al., 2018; Han et al., 2022; Hartanto & Yang; 2020; Lai & O'Brien, 2020), others have not (e.g., Kałamała et al., 2020; de Bruin, et al., 2015). A potential cause for the discrepancies between findings could be external factors which modulate the effect that language use has on language control (and, by extension, may have on executive control), such as age, age of acquisition (AoA), L2 exposure, L2 proficiency and length of residence (LoR). Soveri et al. (2011), for example, found a link between AoA and the ability to inhibit a prepotent response: individuals who learned their second language at an earlier age had less of a Simon effect. Bilinguals with lower AoA and with equal fluency in both languages also displayed significantly lower mixing costs.

The present study aims to detangle the effects of bilingualism and external influences by investigating how dual language use affects language switch efficiency and executive control functions. We measure several outcomes relating to language -, and executive control in a cohort of Chinese-English adult bilinguals as part of a dual-language picture naming task, a Flanker and a Simon task, and responses to an Activity Log across three sessions. We measure overall reaction time (RT), switch cost, mixed cost, and interference scores, as well as hours spent interacting in Chinese and English, and language-switching habits to test whether - and if so, which - aspects of language use can predict any outcomes. Participants are in their first year of university in the United Kingdom. The initial session occurred shortly following the participant's arrival in the UK, followed by an evaluation after three months and, finally, six months after the initial session. In so doing, we aim to capture the effect of language use while controlling stringently for other measures such as age, age of acquisition, exposure, proficiency and LoR.

5.2. Literature Review

Language use is thought to significantly affect language switch ability in bilinguals and, by extension through the purported link between language control and executive control, forms part of the debate over the existence of a bilingual advantage in cognitive control. This section reviews selected literature on language use and its role in bilinguals' linguistic control and executive control performance outcomes and considers externalities which affect the relationships between these three aspects.

Language use describes how speakers employ language in everyday life and is strongly influenced by the contexts in which it is used, as well as by the cultural and social norms of the speaker. For instance, formal job interviews require a different register to informal social events. Similarly, scientists may use jargon when talking with colleagues over work but would ideally not do so when communicating their research to the public. In bilinguals, such different contexts may impact how they employ their two languages. Often bilinguals speak a different language at work with their colleagues than at home with their family and friends. However, mixing and switching between languages is not uncommon in families and communities where bilingualism involving the same language pair is the norm (e.g., Bosma & Blom, 2019).

It has been suggested that bilinguals who employ both languages in daily activities may be more adept at linguistic control than those who do not regularly mix their two languages and have more distinct linguistic environments in which they employ either language (Costa et al., 2009; Green, 2011; Green & Abutalebi, 2013; Green & Li, 2014). The idea is that constant mixing keeps both languages active in mind, thus equalising the likelihood of selection and, in this way, necessitating higher demands on real-time control and monitoring mechanisms as opposed to (predominantly) single-language use. Further, the more frequently a bilingual is confronted with situations that entail language mixing or switching, the more training the involved facilities receive, and the better they will perform in future situations, leading to measurable improvements to their linguistic control (Bialystok, 2009).

Green (2011) points out that while inhibition and selection must occur in bilingual language production, this selection must not necessarily take place at the same locus (also: Kroll et al., 2006; Abutalebi & Green, 2007). Instead, he argues that there may be two loci: one site, which manages selection of a language when languages are being mixed, and one which manages selection when the languages are being kept separate. He suggests that when languages are mixed, both languages remain activated and the selection of the target item happens only later in the processing timeline, i.e., on a lexical level. Hence, he proposes that in mixing contexts, the majority of the inhibition is exercised at this later, lexical stage. In contrast, when languages are used in distinct contexts, he proposes that the selection occurs at the place of an "overarching language schema" (see also: Meuter & Allport, 1999; Costa & Santesteban, 2004; Campbell, 2005), and that in this case, most of the inhibition processes happen at this level. He bases this notion on observations made in neurological studies, which have shown that language-switching tasks are associated with certain frontal and subcortical regions (Abutalebi et al., 2008; 2012; Luk et al., 2012).

De Groot and Christoffels (2006) coined the terms "local" inhibition to refer to the processes, or locus involved in late inhibition (competition on a lexical level) and "global" inhibition to refer to the processes, or locus, associated with early inhibition (competition on language-schema level). Figure 28 illustrates the differences between global and local selection, with thick lines representing connections that are co-activated and underlined items indicating non-target competitors.

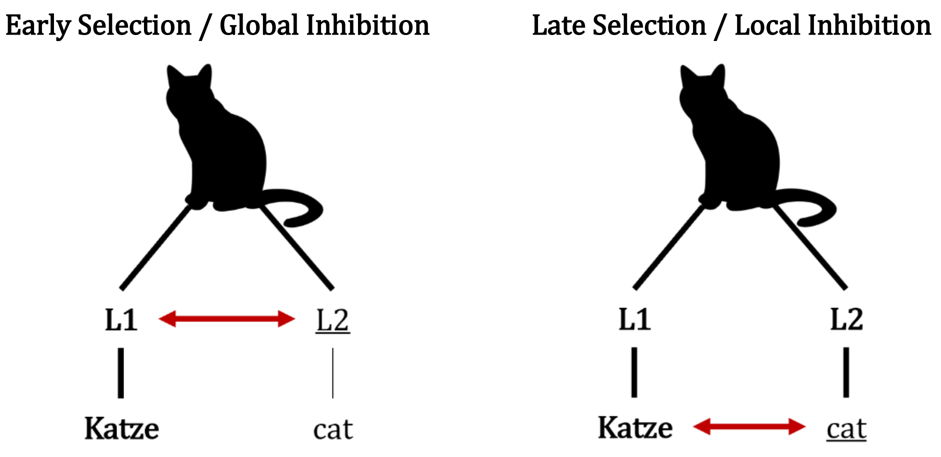


Figure 28: Global vs. local inhibition in bilingual lexical retrieval in production

Extending this idea, the Adaptive Control Hypothesis (ACH; Green & Abutalebi, 2013) and the Cognitive Process Model (CPM; Green & Li, 2014) consider the possibility that different contexts impose demands on different parts of the control mechanism. They distinguish between three interactional contexts of bilingual language use: in a single-language context, one language is used exclusively in one environment and the other language is used exclusively in another distinct environment (e.g., English at work, Spanish at home). In contrast to this, dual-language contexts involve both languages being used by speakers who switch between them during conversations, but without switching within an utterance. Finally, dense code-switching contexts involve frequent language switches within the same utterance, with morphosyntactic adaptation also often occuring as seen through examples such as French-Alsatian "*choisieren*" ('to choose'), or Tagalog-English "*ipagdadrive*" ('I will even drive') (Green & Abutalebi, 2013).

The hypotheses argue that the degree of language control and cognitive control will vary across these interactional contexts. While the ACH proposes that that inhibitory control and cognitive flexibility can adaptively change over time to meet recurring demands, the CPM proposes two different modes for the varied language usage: competitive mode and cooperative mode. In competitive mode, bilinguals use their languages separately and must make a conscious effort to suppress one language over the other so that it is in line with the target language. However, when bilinguals are engaged in intense language-switching, both languages are produced interchangeably within utterances, which means they can be used cooperatively – in the cooperative mode. While the competitive mode is associated with increased cognitive demand, monitoring, and inhibition, in cooperative mode, there is less need for constant monitoring or control of which language is being spoken.

Several studies provide evidence in support of both the ACH and the CPM. In a large sample size study (N=175) involving nine executive control tasks and encompassing multiple sessions, Hartanto and Yang (2020) found that those with an extended background in dual-language contexts needed less effort to switch between tasks than their counterparts who primarily spoke on a single-language basis. Furthermore, those who were regularly exposed to dense language-switching contexts were significantly better in goal maintenance and inhibitory control.

Ooi et al. (2018) investigated four different groups – monolingual speakers, non-switching late bilinguals, and non-switching early bilinguals from Edinburgh; as well as switching early bilinguals in Singapore – and outcomes of their executive control performance on two attentional control tasks. In so doing, they were able to discern effects of dual-language use and dense code-switching contexts, while being able to control for effects of early and late bilingualism. They applied two different tasks: the visual Attention Network Task (ANT), which combines elements of a cued reaction time task and a flanker task, and the auditory Elevator Task (a Test of Everyday Attention (TEA)), to tap into both visual and auditory processing, respectively. They found a correlation between interactional context and performance on conflict resolution: Singaporeans outperformed other groups in the ANT, while late bilinguals from Edinburgh had better outcomes in a TEA subtest. They conclude that interactional contexts affect attentional control differently.

Han et al. (2022) investigated the impact of language switching on cognitive switching and inhibition in bilinguals. Data was collected from 31 Mandarin–English bilingual adults by way of administering the Language and Social Background Questionnaire (Anderson et al, 2018) and the Bilingual Switching Questionnaire (Rodriguez-Fornells et al., 2012). Participants also completed verbal and nonverbal switching tasks – namely a verbal fluency task, a dual-language picture naming task and colour-shape switching task – and a Go/No-go test, which evaluated their domain-general inhibitory control. Their results showed that those who frequently engaged in language switching also performed better when it came to switching between Chinese and English and in nonverbal task switching. Moreover, participants with regular dense language-switching experience scored higher on the Go/No go test.

However, evidence from other studies have failed to provide support for the ACH. For instance, De Bruin et al. (2015) compared active and inactive older adult bilinguals with monolinguals on executive control, considering various factors such as socio-economic status, education, IQ, gender, and age. The Simon task revealed no difference between the groups in overall reaction times or the Simon effect. Furthermore, while raw switching costs varied between active (but not inactive) bilinguals and monolinguals in the task-switching paradigm study; no differences were seen between the groups on overall RTs or proportional switching/mixing costs.

Similarly, Kałamała et al. (2020) ran a large-scale correlational study which included 195 participants. They administered four executive control tasks (an antisaccade task, a Stroop task, a Go/No go task and a stop-signal task), measured language-switch frequency and intensity by means of a questionnaire, and controlled for a broad range of individual difference measures including general intelligence, musical performance habits, gaming habits, and L2 proficiency. Furthermore, they use both the latent variable approach – which identifies hidden patterns and relationships in data – and Bayesian estimation to understand whether language-switching was able to predict any of the executive control outcome measures. Their results do not support the Adaptive Control Hypothesis' prediction, as there was no relation between the intensity of dual-language context experience and response inhibition in bilinguals.

Overall, there is a wealth of research providing support for both the ACH and the CPM. However, other researchers such as de Bruin et al. (2015), and Kałamała et al. (2020) have not been able to support the predictions made by either ACH or CPM. As such, the evidence is as of now inconclusive. Contrasting the methodologies of the studies supporting the hypothesis as opposed to the studies contradicting it, it would seem that it is the stringency with which externalities are controlled for that restricts the likelihood of arriving at a supporting argument; with studies which include more stringent control measures arriving at a positive result with a decreased likelihood.

One such control measure considered by several studies (e.g., Kałamała et al., 2020) is language dominance. Language dominance is thought to modulate the effect of linguistic control on either language, as different dominance constellations may exert different demands on the control mechanism. Language dominance is strongly associated with each language's relative use and ease of access: the more a language is used relative to another, the more "proceduralised" the use will become, the easier it will be to access this language, the more of a "competitor" it may pose relative to the other language.

In bilinguals with low L2 use, i.e., bilinguals, who have just begun to immerse themselves in an L2-dominant environment, the L1 is still the default communicative system in the brain and has great ease of access, indexed by monolingual-like retrieval times. As the L2 builds up a presence in the bilingual's mind and use is more frequent, inhibitory forces come into effect to manage the nascent competition between the two language systems (Green, 1998). At this stage, it is often shown that the resulting competition is unequal, as the default L1 is disproportionately suppressed to allow activation of the weak lexical links of the L2 (e.g., Meuter & Allport, 1999).

As the bilingual becomes increasingly proficient in their L2, the two language's average overall RTs balance out, and the asymmetry of inhibition costs across the two language systems decreases (e.g., Meuter & Allport, 1999). In highly proficient L2 speakers, who have spent a prolonged period living in the L2 speaking environment – predominantly using their L2 –a pattern of "reverse dominance" may show (Birdsong, 2018). In reverse dominance, the L2 replaces the L1 as the default communicative system, and bilinguals may encounter difficulties accessing their L1 with the ease and speed they were accustomed to – in a process known as "language attrition". The decline in lexical retrieval ability is thought to be the first noticeable and the most immediate effect of language attrition (Schmid & Jarvis, 2014; Schmid & Köpke, 2009).

Neuroimaging research has demonstrated a difference between highly and less proficient bilinguals: those who are balanced in their bilingualism use the same brain regions for lexical access tasks in both languages, whereas unbalanced bilinguals require additional activity from frontal areas linked to general cognitive control (Abutalebi, 2008; Abutalebi & Green, 2007).

One method to retrieve a measure for relative language dominance is the calculation of “between-language subtractive differentials” (Birdsong, 2016). In this method, the difference between a bilingual’s performance (for instance, as measured in RT as part of a picture naming task) in a linguistic task in one language as opposed to their performance in the same linguistic task in their other language. The closer the discrepancy is to zero, the more balanced the bilinguals' languages are said to be (Birdsong, 2018).

Previously introduced as a factor affecting language dominance, L2 usage captures the qualitative and quantitative input speakers receive. Bonfieni et al. (2019) explore the effects of language proficiency and daily exposure on bilinguals' ability to switch between languages. The study included 83 participants aged 18-40, with an average age of 26.3, who were either Italian-English bilinguals living in Scotland or Italian-Sardinian bilinguals living in Sardinia. The researchers used a cued dual language picture naming task to measure naming latencies, switch costs and mix costs. The results showed that switching between two languages largely depends on L2 proficiency and daily L2 exposure, whereas daily L2 exposure was also found to impact language mixing.

AoA plays an extensive role in second language learning (Birdsong, 1999) and is strongly related to language dominance (Birdsong, 2014). AoA also impacts the bilinguals’ mental architecture, i.e., the cortical thickness of inferior frontal gyri (Klein, Mok, Chen, & Watkins, 2014) – indicating a larger number of synaptic connections (faster processing) in speakers with a lower AoA; cortical activation in relation to lexical access (Perani et al., 2003) – indicating a stronger activity; and language lateralization (Hull & Vaid, 2007) – indicating a more efficient allocation of cognitive resources. Soveri et al. (2011) found that age, along with age of L2 acquisition and everyday dual-language use all had an impact on various measures of executive control: bilinguals with a lower AoA displayed smaller Simon effects. In addition, those with a lower AoA and a balanced language use were also found to have significantly smaller task switching costs. Costa et al. (2006) examined the language switching task of highly proficient early Spanish-Basque bilinguals and late Spanish-English bilinguals and discovered no difference between switch costs in both groups, indicating a lack of effect of AoA. However, the latter group was composed of students enrolled in an interpreter school which may have led to an unequal comparison.

In conclusion, this literature review has presented the factor of language use in the sense of interactional contexts within the debate on bilingual cognitive advantage. It has shown examples of studies that find evidence for varying language use impacting language control and/or executive control and presented several external factors that modulate the effect of language use on the cognitive facilities at issue.

For this reason, we propose an increased focus on external factors whose effects on the way language use interacts with linguistic and executive control are not fully understood. This paper focusses on the relative usage frequency of either language and the situational contexts in which it is used as modulating factors of language control on the one hand, and executive control on the other. Prior to further motivating the current study, however, the remainder of this literature review explores which additional aspects may influence the effect that language use may have on language control and executive control, i.e., the constraints on their relationships.

Taken together, the varying support for the effect of interactional contexts could be due to the heterogeneity of the bilingual population, which is difficult to control for. To avoid the potential pitfalls of combining different types of bilinguals and assuming homogeneity, this study focuses on a singular type of bilinguals and considers their individual language exposure and language usage patterns in detail. In this way, we aim to accurately represent data without introducing noise or increasing the risk for Type II error.

5.3. The current study

This study examines the influence of language use on performance in language control and executive control tasks in Chinese-English bilingual students at the University of Essex (UK) cross-longitudinally across three sessions in a period of six months. The group are relatively homogenous in that at the time of first testing, most had only just arrived in the UK from China and in this way followed similar L2 learning trajectories and are exposed to the similar L2 environment. In this way, we aim to control for effects of LoR, language history, L2 exposure, and AoA.

Our methodology involves a dual-language picture naming task, Flanker and Simon tasks, and an activity log. The activity log was specifically designed to reflect typical activities of a university student. We analysed responses on the activity log by means of a principal component analysis and used computed components to calculate factor loadings for each participant for each session through a regression.

In the dual-language picture naming task, participants name pictures in either Chinese or English. In the Flanker and Simon tasks, the orientation and, respectively, the locus of stimuli either corresponded or conflicted with interferential information across congruent and incongruent conditions. Chinese and English trials, as well as congruent and incongruent trials, are presented first as part of single-language/congruency type blocks and thereafter, within a mixed-language/congruency type block, wherein the trial types are presented in a pseudo-alternating manner.

We measure overall response times, interference scores and switch and mix costs.

(Overall) response time (RT) reflects the overall average latency needed to process trials in various conditions. It is the main dependent variable and – in interaction effects with trial type, condition, and congruency – reflects average switch and mix costs, as well as inference scores, respectively.

Interference scores reflect the ability to suppress interfering information. In our study, we determined these scores through the discrepancy of congruent and incongruent trials across flanker and Simon tasks and use the outcome as a proxy to determine a participant's ability to attend to target information selectively.

Switch costs are representative of the relative difficulty with which new mental task sets are adopted. In our study, we measure switch costs by calculating the discrepancy between non-switch and switch trials in the mixed-language condition and use this outcome as a proxy to evaluate a bilingual's language-switching ability.

Mix costs are a metric to assess the capacity for concurrent task set activation (Segal et al., 2021). Typically, mix costs are calculated by subtracting the average RT of non-switch trials in single-task blocks from non-switch trials in mixed-task blocks. In our study, we calculate mix costs for both the dual-language picture naming task and the Flanker and Simon tasks.

Language Balance describes the relative ease of access to one language as opposed to another and aids in understanding how the bilinguals' two languages interact with each other. It is usually calculated by subtracting the average response time for one language from the average response time of the other. In our study, language balance serves as a proxy for L2 proficiency. In short:

Table 20: Summary of language control and executive control measures

	Language Control	Executive Control
Overall RT	X	X
Interference Scores		X
Switch Costs	X	
Mix Costs	X	X
Language Balance	(as control measure)	

We address the following research question and hypotheses:

How does language use affect linguistic control and executive control?

Hypotheses:

I. Those participants who regularly code-switch will perform better in switch trials and in mixed-language trials as opposed to those participants who do not.

Regular code-switchers are more practised in switching, which means they have developed more efficient mechanisms of dealing with conflicting linguistic representation. Thus, they will thus face less difficulty when switching and mixing languages.

II. Indices of language use which predict a change in performance in the linguistic control measures also do so for executive control measures.

If linguistic control utilizes the same cognitive resources as domain-general executive control, then an improvement in language control – effectuated through a language use that involves a significant degree of language switching – ought to also improve participant's domain-general executive control measures.

5.4. Method

5.4.1. Apparatus

All tasks were conducted in person in the Reaction Time lab of the Department of Language and Linguistics at the University of Essex. The Dual-Language Picture Naming Task was presented to the participants on a Windows computer using E-Prime with a Serial Response Box and voice trigger.

5.4.2. Participants

A total of 30 Chinese-English bilingual adults participated in this study, ranging from 19 to 32 years of age (mean = 23.5 years). All Chinese-English bilinguals reported speaking some form of Chinese (Cantonese, Mandarin, Taiwanese) as their first language. Twenty participants identified as female, and ten as male. Most had only just arrived in the United Kingdom in the month prior to the first testing session to pursue university degrees. Six individuals took part in pre-sessional English courses, and sixteen stated that they were enrolled in an English language class at the point of first testing.

5.4.3. Procedure

All participants completed three reaction time experiments – a picture naming task, a Simon task and a Flanker's task, a C-test, an activity log, and a questionnaire on language background, respectively.

Participants completed these tasks during three temporally distinct sessions. The first session took place at the start of the academic year in October 2016, the second one in late January and early February 2017, and the third and final one took place in April 2017. Bilinguals performed the picture naming task, the Simon and Flanker's tasks, as well as the activity log at all three sessions, and the C-test during sessions one and three. Controls completed the Simon and Flanker's tasks and the C-test only during the first session but took part in the picture naming task in sessions one and three. The language background questionnaire was administered to both groups during at the time of first testing.

All three timed tasks had four parts (Figure 29). All tasks were preceded by eight mock trials to accustom the participant to the task demands. The second two parts of each task were presented as single-task conditions. In the single-task conditions, a block of 20 congruent trials was followed by a block of 40 incongruent trials in the executive control task. In the language control task, 50 Chinese trials were followed by 50 English trials. The final part of each task was a "mixed-condition" block, where the previous blocks' trials feature in a semi-alternating manner. The order in which the three blocks was presented was fixed throughout the study. Doing so, we aimed to eliminate effects varying block orders may have had (see 2.2.4. Design of timed experiments).

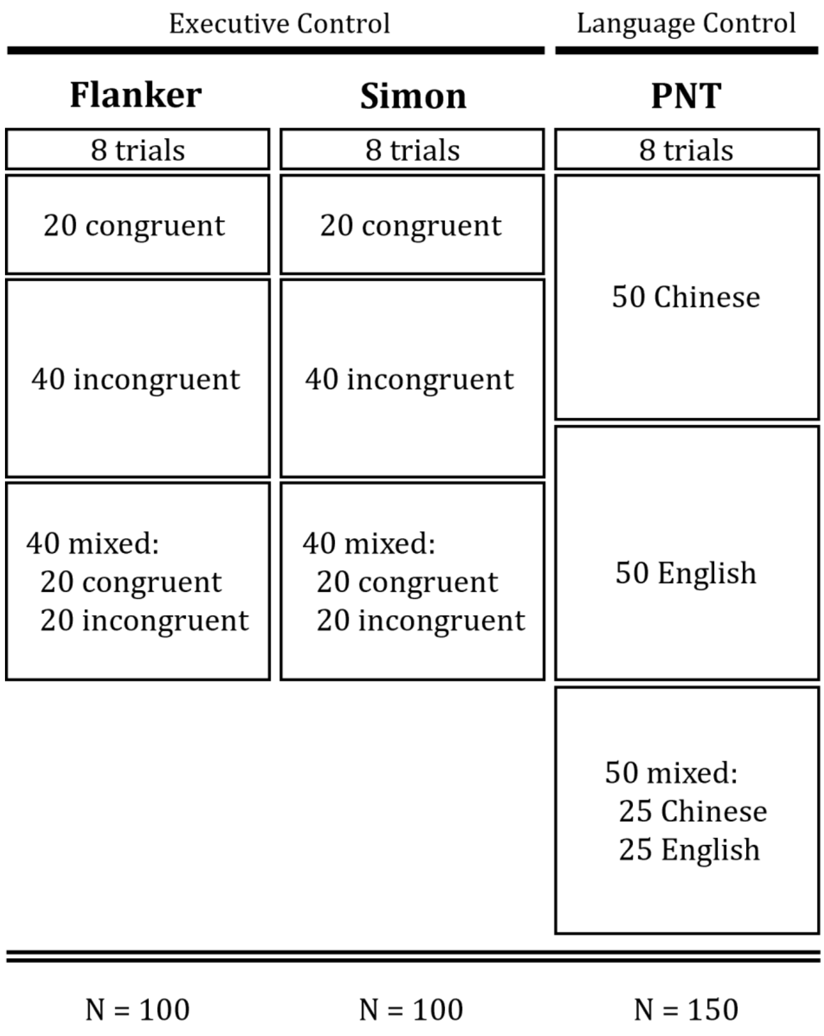


Figure 29: Study Design. N refers to the number of experimental trials.

Table 21 summarises the variables manipulated in the experiments, as well as their marked and unmarked categories and the measurements they enable.

Table 21: Summary of Variables in Experimental Tasks

Factor	Marked (subtrahend)	Unmarked (minuend)	Measurement (difference)
Congruency	incongruent	congruent	interference scores
Task Type	mixed congruency	single congruency	mix cost
Trial Type	switch	non-switch	switch cost
Language	L2 (English)	L1 (Chinese)	language balance

5.4.3.1. Dual-Language Picture Naming Task

In the dual-language picture naming task, bilingual participants were shown a picture and asked to name the item in one of two languages as fast as they could. Responses were given orally and recorded as a sound file. A coloured frame around the picture indicated whether the item should be named in Chinese (red) or English (blue). The experiment was split into four consecutive blocks: a training block consisting of 8 trials (4 with a blue frame to cue an English response and 4 with a red frame to cue a Chinese one), a Chinese-target block of 50 trials, an English-target block of 50 trials, and a mixed block with both Chinese and English targets alternating semi-randomly across 50 trails. The semi-random alternation is such that there are 25 trials of either language in an order whereby about half the trials were preceded by a trial of the other language ("switch"), and the other half was preceded by a trial of the same language ("stay"). For monolinguals, there was no such distinction between blocks, as they were only tested in English, and no visual cue to response language was included. Each trial begins with an acoustic signal of 100ms, with the onset of the presentation of the visual stimulus occurring 500ms after the onset of the sound cue. The self-paced trial ends with the participant pressing the space bar.

The pictures for the task were employed randomly across the experimental blocks and were selected from a battery composed of a total of 512 pictures. The battery is based on Bates et al.'s (2003) comparison of timed picture naming in seven languages, which provides information about differences in naming latencies and word frequencies cross-linguistically.

5.4.3.2. Flanker Task

In the Flanker task, participants respond to the direction of an arrow at the centre of the screen, which is flanked by two non-target arrows to either side. The direction of the non-target items either corresponds to the direction of the target item (congruent Flankers) or all show the opposite direction (incongruent Flankers). In this study, participants are asked to indicate the direction of the target stimulus by pressing the "3" key on the keyboard for left and the "9" key for right directionality.

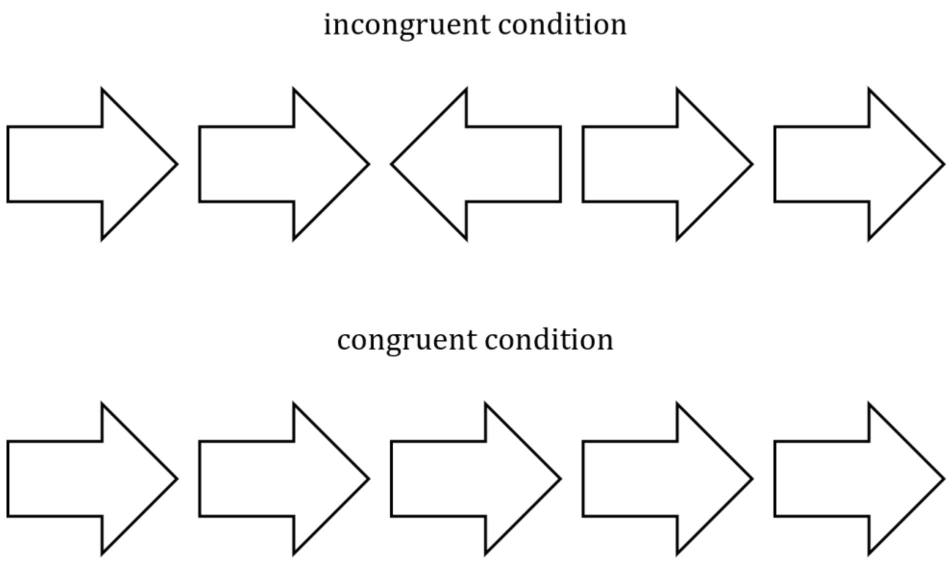


Figure 30: Incongruent and congruent conditions in the Flanker task

5.4.3.3. Simon Task

The Simon task (Simon & Rudell, 1967) involves suppressing spatial information in favour of colour information. Participants respond to the colour of the stimulus they see, which appears on either the left or the right side of the screen and are asked to press the left arrow key whenever they see a green dot and the right arrow key when they see a red dot, regardless of the side the stimulus appears on. Trials are "congruent" where the colour stimulus appears on the side of the screen that corresponds with the relative location of the response key (e.g., the red dot appears on the right side of the screen).

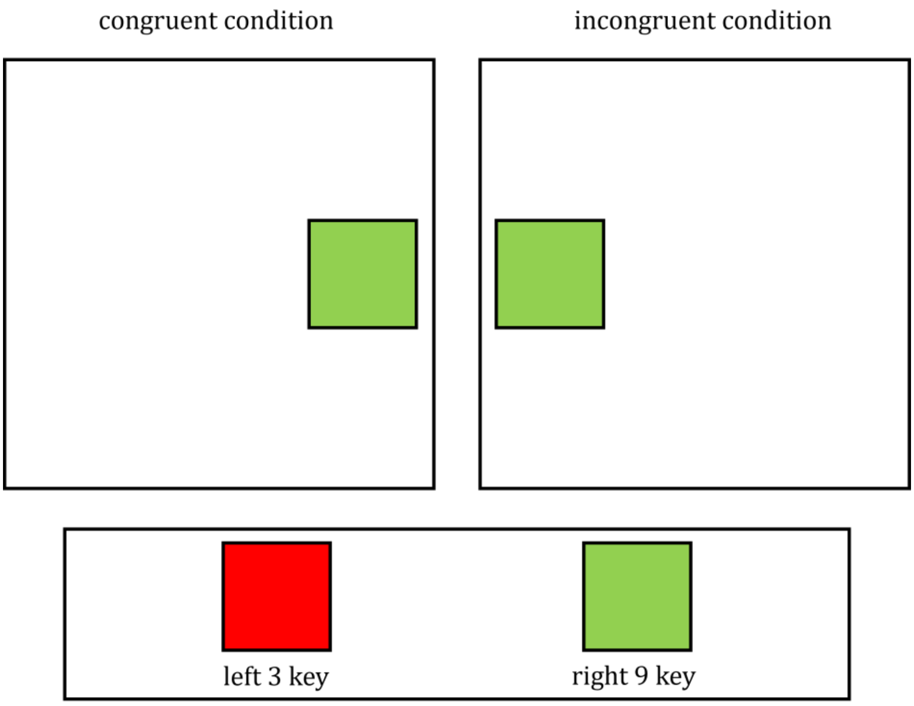


Figure 31: Congruent and incongruent conditions in the Simon task

For both the Flanker and the Simon task, we measure the time between stimulus onset and participant response and the accuracy of the response. The Simon Task and the Flanker Task have different conceptualizations within the Dimensional Overlap Model (Kornblum et al., 1990): for the Flanker Task, the Dimensional Overlap Model posits that it is an example of a task involving a Stimulus-Stimulus conflict. In such tasks, there is a conflict between items of the same stimulus (i.e., arrows) which originate and are resolved at the perceptual level. In contrast, for the Simon Task, the model posits that the task involves a stimulus-response conflict. In such instances, the conflict arises due to the stimulus dimension (here: stimulus location) being irrelevant to the response rule and the response dimension (Kornblum, 1994; Paap, 2019; Blumenfeld & Marian, 2014).

Using these two different kinds of tasks allows us to better understand which kind of conflict monitoring mechanisms are employed in linguistic control.

5.4.3.4. Activity Log Questionnaire

All bilingual participants completed a paper questionnaire asking about their everyday language use at each of the three sessions to reflect any changes in language use behaviour throughout the duration of the study. Questions include how many hours they spend pursuing different activities (e.g., class attendance, watching films, spending time with friends), and what percentage of this time is spent interacting in English as opposed to Chinese. The questionnaire also captures Likert-scale data about code-switching, e.g., whether the participant uses Chinese words or sentences when writing e-mails in English. Further, it investigates the proportion of English social contacts as opposed to Chinese ones, and whether the participant has returned to China between test sessions. The full catalogue of questions asked as part of the activity log questionnaire can be found in the Appendix 5: Full catalogue of questions contained in the Activity Log.

5.4.3.5. Language Background Questionnaire

All participants completed a sociolinguistic background questionnaire as part of the first session. Questions included date of birth, gender, and knowledge of any second language, the date of arrival in the United Kingdom, order of acquisition of languages, any language test results (IELTS, TOEFL, university-internal language examinations), and whether they had taken part in any pre-sessional English classes (and if so, how many hours per week, and for how long), attendance of language classes (and if so, how many hours per week, for how long, and what kind of exercises are involved).

5.4.4. Data Analysis

5.4.4.1. Dual-Language Picture Naming Task

Using the voice recordings collected during the experiment, we manually measured reaction times and labelled responses in Praat (Boersma & Weenik, 2018). Response language was coded "ENG" for English responses, and "CH" for Chinese responses. Item responses were coded "0" for any missed trials, i.e., where no response was attempted by the participant, "1" when the response exactly matched the target item, "2" when the response was deemed an appropriate synonym for the target (e.g., "kitten" for "cat"), "3" for responses which were semantically further removed from the target, for instance, distantly related responses (e.g. "dog" for "cat") or hypernyms (e.g., "animal" for "cat") and "4" for any responses that did not resemble the target in any way. Language accuracy was coded as "1" where the response language fit the target language, and "0" where it did not. Item accuracy was coded "1" where the item response was previously coded as either "1" or "2", and else "0". Trials were coded as "non-switch" when the previous trial aimed to elicit a response in the same language as the current trial, and "switch" where the previous trial target language differed from the current one.

Reaction times were measured from the onset of the acoustic signal to the onset of the participant response. 500ms were subtracted from the measured time to reflect the difference in time between signal onset and visual stimulus onset. We levelled reaction times higher than the sum of overall experiment's mean plus two standard deviations (2770ms) and removed data from the first trial in each block as well as trials with reaction times below 250ms, to prevent unusually long and short trials skewing the distribution of the averages, respectively. We also excluded inaccurate and missed trials (i.e., those where language or item accuracy was coded "0"), as well as trials with RTs beyond 4000ms (prior to levelling) for the analysis.

A comparatively large number of trials were excluded from analysis: 24.1% of all bilingual trials and 10.4% of all monolingual trials (Appendix 2.A: Data Exclusion Rates). Comparable studies only excluded around 5% of their data (e.g., Bonfieni, 2019).

5.4.4.2. Activity Log Questionnaire

Activities missing a majority of responses were discarded from further analysis (e.g., "attending a religious service/praying", "working"; see appendix 1). Where no response was given, we replaced missing values with the groups' session averages to be able to retain the participant's data while keeping the component uninfluenced. Absolute hours spent per Activity (or fractions hereof) were calculated for both languages. The final file contains absolute hours spent interacting in Chinese and English, as well as the variables on codeswitching, the proportion of contacts speaking different languages and whether the person has been back to China recently.

We conducted a principal component analysis to reduce the number of dimensions using SPSS. We started with an *oblimin* rotation to test for intercorrelation between factors. However, none of the factors were highly correlated with one another. As such, we proceeded with a *varimax* rotation, which forces components to be entirely uncorrelated, to facilitate interpretation. We chose to limit the count of principal components to seven, as they accounted for 57% of total variance. A full description of the rotated principal component matrix including correlations with individual activities is provided in Appendix 6: Rotated Component Matrix of the Principal Component Analysis. The principal components can be summarised as follows:

Component 1: Dual-language free time activities (C1 – 2L leisure). Activities include reading books or newspapers, watching TV and films, and interacting with friends in either language.

Component 2: Dual-language exam preparation (C2 – 2L academic). Activities include preparing for exams in either language, and working on assignments in either language.

Component 3: Dual-language code-switching experience (C3 – 2L switching). This component captures any code-switching habits.

Component 4: L1-focus academic supporting activities (C4 – L1 academic). Activities include out-of-class contact with teacher in Chinese and completing extra assignments in Chinese. Improving English vocabulary also loads highly on this component.

Component 5: L2-focus on social contacts (C5 – L2 social). Participants who score highly on this factor have fewer Chinese contacts, and instead more international contacts, and

English contacts. English class preparation and Chinese assignment writing also load highly on this component.

Component 6: L1-focus on social contacts (C6 – L1 social). Participants who score highly on this factor spend a lot of time interacting with Chinese contacts and less so with English contacts. Practising sports in English and improving English pronunciation also load highly on this component.

Component 7: L2-focus on core academic activities (C7 – L2 academic). Activities include working on assignments, attending classes and interacting with colleagues in English.

Based on these components, we created seven variables and computed factor loadings for each participant by session using a regression. In numbers, the factor loadings by participant can be summarised as follows:

Table 22: Summary of participants' factor loadings

	C1	C2	C3	C4	C5	C6	C7
Observations (N)	79	79 (78)	79	79 (78)	79	79	79
Median	-0.26	-0.25 (-0.26)	0.027	-0.10 (-0.15)	-0.13	-0.054	-0.24
Minimum	-1.23	-1.17	-1.89	-1.41	-2.02	-3.50	-1.79
Maximum	5.65	5.68 (2.50)	3.09	7.44 (1.22)	4.34	4.62	4.05

C2 – 2L academic and C4 – L1 academic each had one observation which lay several standard deviations removed from the remaining sample in a way which caused confounding results. Hence the decision was taken to discard these data points from the data set. The revised descriptive analyses for Components 2 and 4 are provided within Table 22 in brackets. Appendix 3 provides an overview of the development of the components by session.

The data was analysed and contextualised with the remaining data set on a session-by-session basis. This means no data on language behaviour from a previous session was used to analyse patterns in the linguistic or cognitive tasks of a consecutive session.

5.4.4.3. Language Background Questionnaire / IELTS

Any IELTS scores from the first-session language background questionnaire were extended to sessions 2 and 3. We used the group average (6.8) for those participants missing IELTS results as a proxy to replace missing values as we expected the participants' score to be largely similar and their inclusion allowed us to control for any differences between participants.

5.4.4.4. Statistical Analysis

We ran a series of linear mixed effects models to test our hypotheses using Jamovi 2.3 and R 4.1 (The Jamovi Project, 2022; Gallucci, 2019; R Core Team, 2021). For each analysis, we constructed an optimal model by starting with a full model of random coefficients, fixed factors, and interaction effects and removing individual effects on a step-by-step basis, retaining factors only when the more complex model is significantly superior to the reduced model. We used the Akaike Information Criterion (AIC) as an estimator of prediction error and conducted an ANOVA test for the models' goodness-of-fit relative to one another.

Fixed main factors entered in each analysis included task type (single vs. mixed), session (1, 2, 3), Activity Components 1 through 7, and IELTS scores and participant ID as random factor. In the analysis of the dual-language picture naming task, we also included trial-type (switch vs. non-switch), target language as fixed factor and item ID as random factor. In the analyses of Simon and flanker tasks, we included congruency (congruent vs. incongruent) as fixed factor. As our group was overwhelmingly homogenous in terms of age and gender, we refrained from entering these aspects into our analyses. To investigate effects on either mix cost, switch cost, interference scores or on a specific language, we added interaction effects of different main effects on one another.

5.5. Results

For each of the three tasks, we will first provide a brief overview of the main findings and as a second step present the results of the investigation of the activity log components, IELTs and C-test scores on the tasks' main outcomes (overall RT, switch cost, interference scores, and mix cost). For details on the PNT, Simon and Flanker results refer to 4.5 Results.

5.5.1. Picture Naming Task

The analysis of the dual-language picture naming task revealed that trials in the single language conditions elicited significantly higher RTs (β=196.4, SE=27.8, $p<0.001$). This contradicts our expectations, as we had expected the mixed condition to be more difficult to process. Refer to 4.6. Discussion for a discussion on these results. Trials which followed a switch - within the mixed language task type –elicited significantly larger RTs (β=91.7, SE=18.5, $p<0.001$). In line with our expectations, switches incurred an additional processing cost. English trials were significantly slower than Chinese trials (β=275.0, SE=27.5, $p<0.001$), suggesting that overall, the group preferred Chinese naming. Switch costs and mix costs were significantly lower in English as opposed to Chinese (both <0.001), evidencing an asymmetry for both switch and mix costs, and thereby suggesting that the bilinguals are relatively unbalanced and in the earlier stages of L2 acquisition. Finally, trials were significantly faster in session 2 as compared to session 1 (β=-67.3, SE=15.6, $p<0.001$), however, there was no significant difference of overall RT between sessions 3 and 1 (β=-19.8, SE=27.5, $p=0.47$). This unexpected result may be an indication of nonlinearity in second language development (e.g., Pfenninger & Kliesch, 2023).

Figure 32 shows overall RTs for non-switch and switch trials in the mixed-language task type and all trials in the single-language task type by language:

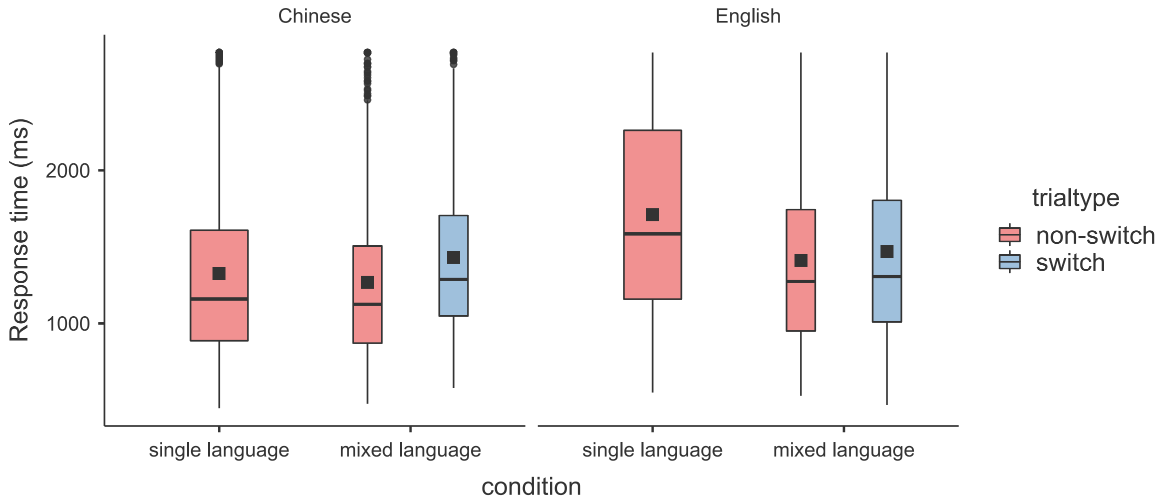


Figure 32: PNT response times overview.

Whiskers reflect extreme values that are no outliers (within 1.5x IQR).
Black squares represent means.

The addition of activity components as covariates revealed that six out of the seven activity components influenced either overall RT, switch cost or mix cost.

Formula: (lmer) RT ~ 1 + C1 – 2L leisure + C2 – 2L academic + C3 – 2L switching + C5 – L2 social + C6 – L1 social +C7 – L2 academic + Task Type + IELTS + language + session + Trial Type + Task Type:C1 – 2L leisure + Task Type:C3 – 2L switching + Trial Type:C3 – 2L switching + language: Task Type + language: Trial Type + language: C3 – 2L switching + language: C2 – 2L academic + C5 – L2 social:language+ (1|item)+(1|Participant)

Table 23: Model for activity components' effects on dual-language PNT

Overview	Fixed effects	Estimate	St. Error	P-Value
Intercept	(Intercept)	1496.4	52.0	< .001
Main effects	C1 – 2L leisure	-24.9	12.1	0.05
	C2 – 2L academic	-44.7	13.0	< .001
	C3 – 2L switching	-35.5	15.5	0.05
	C5 – L2 social	-26.8	10.0	0.01
	C6 – L1 social	-43.8	10.2	< .001
	C7 – L2 academic	-36.3	8.2	< .001
	Task Type single	196.4	27.8	< .001
	Trial Type switch	91.7	18.5	< .001
	IELTS	-245.1	100.3	0.05
	Language: English	275	27.5	< .001
	Session 2	-67.3	15.6	< .001
	Session 3	-19.8	27.5	0.47
Task type effects (mix costs)	C1 – 2L leisure * Task type single	23.1	11.0	0.05
	C3 – 2L switching * Task type single	37.3	14.7	0.05
Trial type effects (switch costs)	C3 – 2L switching * Trial Type switch	36.7	18.0	0.05
Language effects	C3 – 2L switching * language English	53.4	11.2	< .001
	C2 – 2L academic * language English	98.1	14.5	< .001
	C5 – L2 social * language English	-89.5	11.7	< .001
	Language English * Task Type single	266.5	55.6	< .001
	Language English * Trial Type switch	-151.5	36.8	< .001

Participants loading highly on any component – other than C4 – L1 academic, which failed to improve this model (and all following models) – had significantly lower overall RTs (Table 23). Notably, participants who load highly on C2 – 2L academic and C4 – L1 academic saw the highest reduction in RT, while participants who load highly on C1 – 2L leisure had comparatively smaller reductions. Figure 33 visualizes the individual components' effect sizes and standard errors.

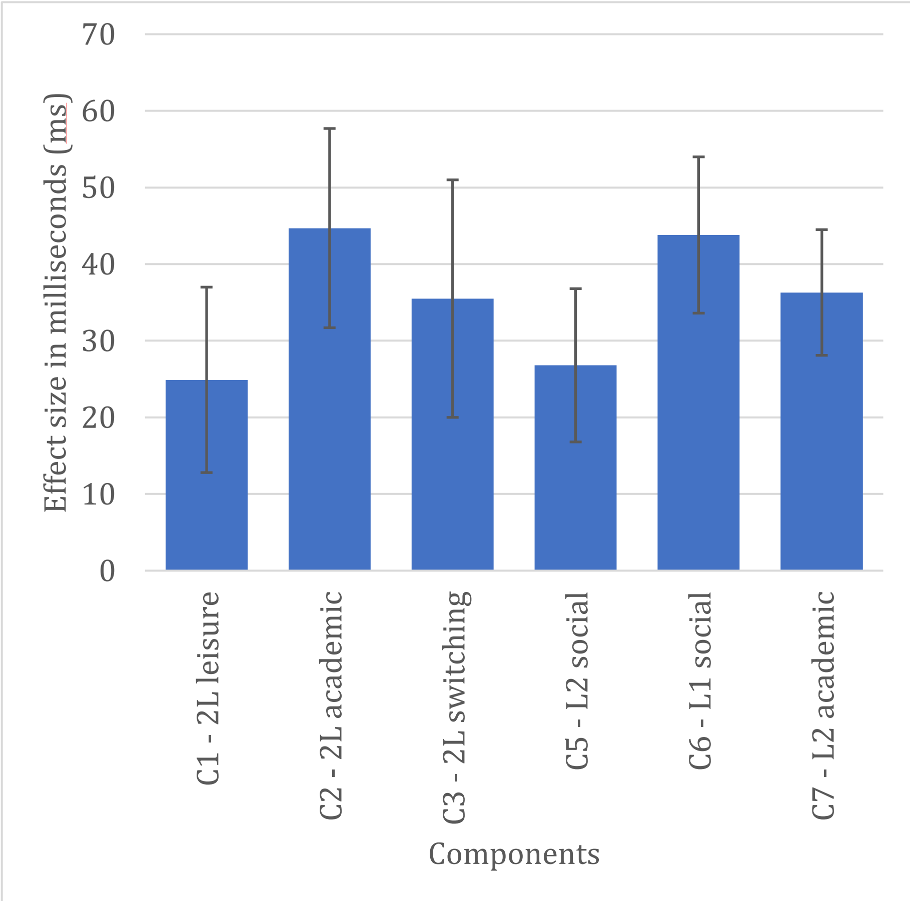


Figure 33: PNT components' effect sizes and standard errors on response time. Error bars reflect 95% Confidence Intervals.

Participants with higher IELTS scores had lower overall RTs. However, IELTS scores did not impact neither mix cost nor switch cost, as both interaction effects were non-significant. Further, the effect of IELTS was independent of language, again as there was no significant interaction effect between IELTS scores and language.

There were significant interaction effects of condition and Activity C1 – 2L leisure ($p<0.05$) as well as of task type and Activity C3 – 2L switching ($p<0.05$), which revealed that those who loaded highly on the components had a stronger facilitation effect in the mixed condition as opposed to the single condition, leading to higher mix costs (Figure 34 and Figure 35).

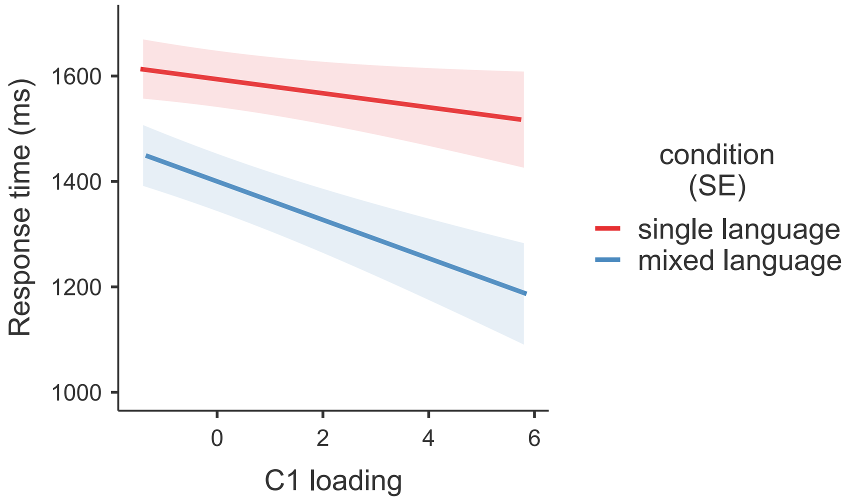


Figure 34: The interaction effect of C1 loading and condition in the PNT. Shaded areas reflect 95% confidence intervals.

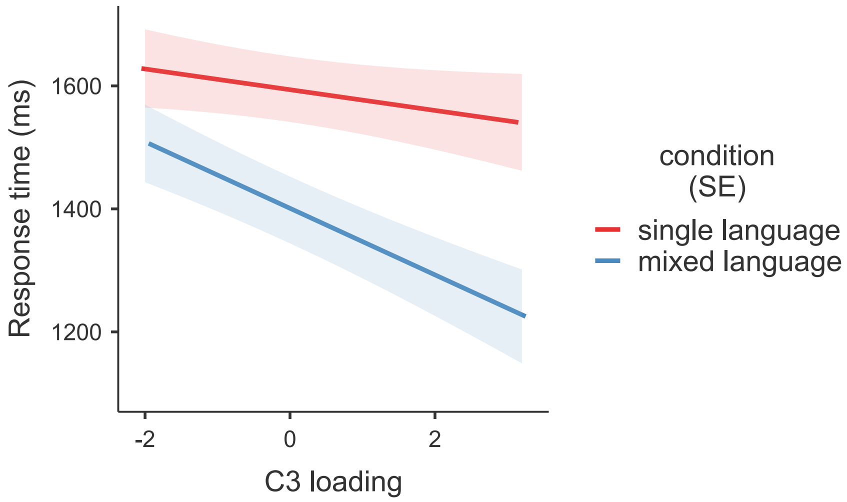


Figure 35: The interaction effect of C3 loading and condition in the PNT. Shaded areas reflect 95% confidence intervals.

A post-hoc simple effects analysis revealed that these higher mix costs are a result of the RT in the mixed condition – i.e., the condition that is "at conflict" in the sense that it is the condition we have expected to change – decreasing as component loads increase. In our discussion, we will refer to this kind of effect as "conflict facilitation".

Similarly, a significant interaction of Activity C3 – 2L switching (dual-language code-switching habits) and trial type reveals that those participants who regularly engage in code-switching have larger switch costs. However, a post-hoc simple effects analysis revealed that the increased switch cost is a result of disproportionally decreasing RTs in the – non-conflicting – stay trial type as opposed to the switch trial type, as is illustrated in Figure 36. We shall henceforth refer to effects like these as "non-conflict facilitation" to distinguish between conflict facilitation effects like those seen above in the analysis of mix costs.

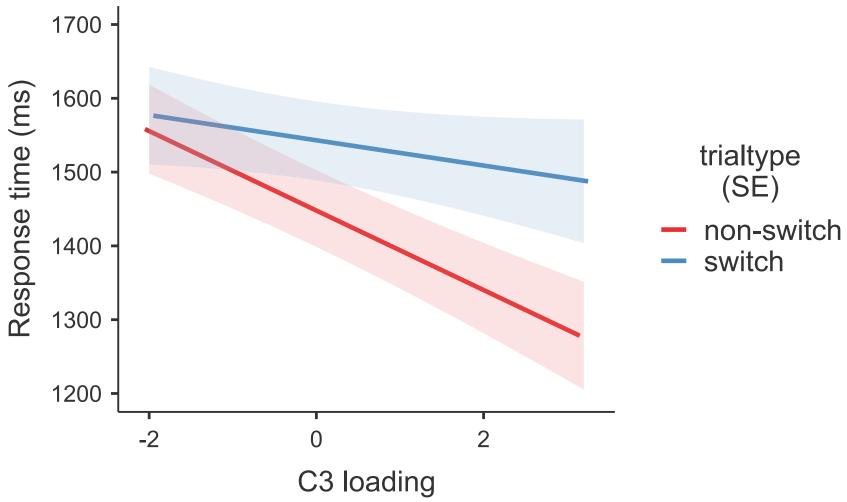


Figure 36: The interaction effect of C3 loading and trial type in the PNT.
Shaded areas reflect 95% confidence intervals.

The significant interaction of C3 – 2L switching and Language shows that participants who score high on the component respond faster in Chinese trials than in English trials ($< .001$), as illustrated in Figure 37.

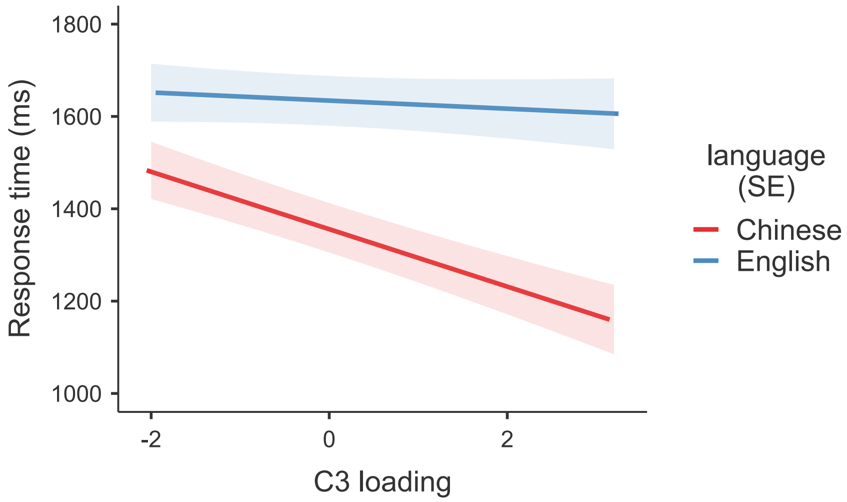


Figure 37: The interaction effect of C3 loading and language in the PNT.
Shaded areas reflect 95% confidence intervals.

Similarly, those who scored highly on activity C2 – 2L academic also showed a greater cost when responding in English than in Chinese. This significant interaction shows that Chinese responses become significantly faster with an increasing score on Activity C2 – 2L academic while English responses remain largely unchanged (Figure 38).

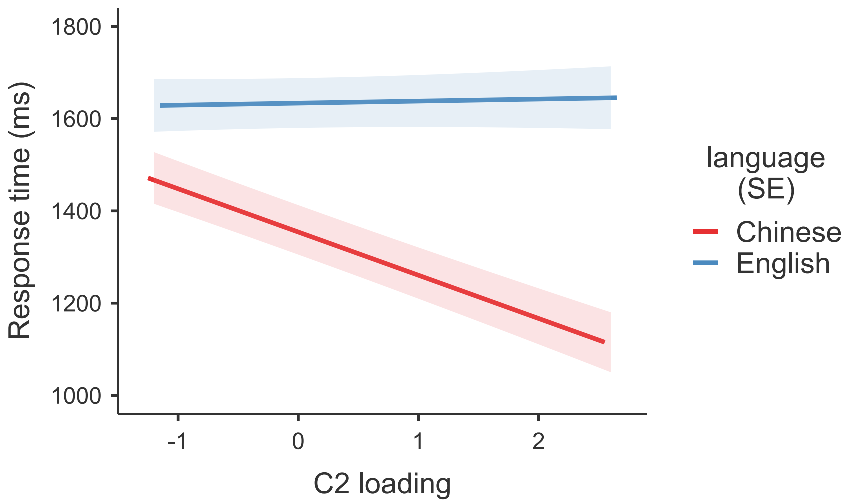


Figure 38: The interaction effect of C2 loading and language in the PNT.
Shaded areas reflect 95% confidence intervals.

Conversely, a significant interaction effect between Activity C5 – L2 social (focus on L2 contacts) and language reveals that those who load highly on the component had significantly lower RTs in English (β =-71.5, SE=11.9, $p<0.001$), while their RT in Chinese largely unchanged (β =17.9, SE=11.3, $p<0.112$; Figure 39).

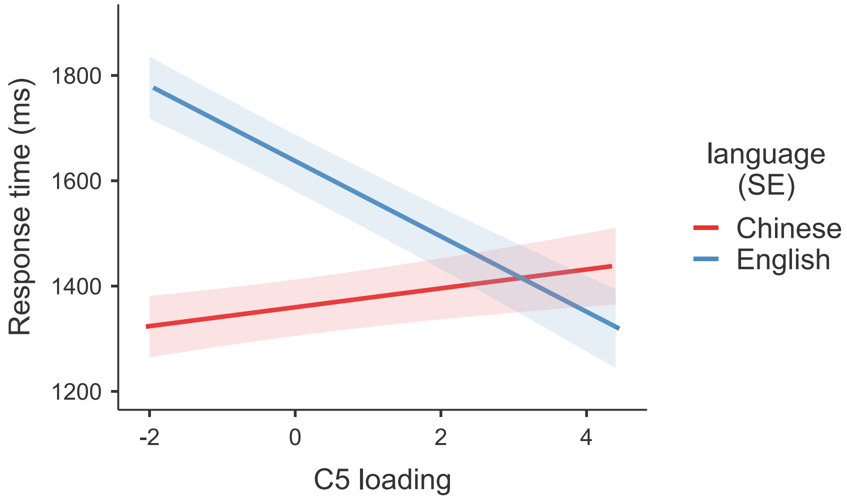


Figure 39: The interaction effect of C5 loading and language in the PNT. Shaded areas reflect 95% confidence intervals.

5.5.2. Simon Task

The analysis of the Simon task revealed that, as expected, trials in the – conflicting – mixed -congruency condition elicited significantly higher RTs than those in the single-congruency condition ($\beta=49.2$, $SE=2.0$, $p<0.001$), congruent trials elicited significantly lower RTs than incongruent trials ($\beta=22.4$, $SE=2.0$, $p<0.001$); and that overall RT, but unexpectedly not interference scores or mix cost, became significantly lower with each session ($p<0.001$). For a discussion of these results, refer to 4.6. Discussion. Figure 40 shows overall RTs for congruent and incongruent trials in the mixed-congruency task type and single-congruency task type.

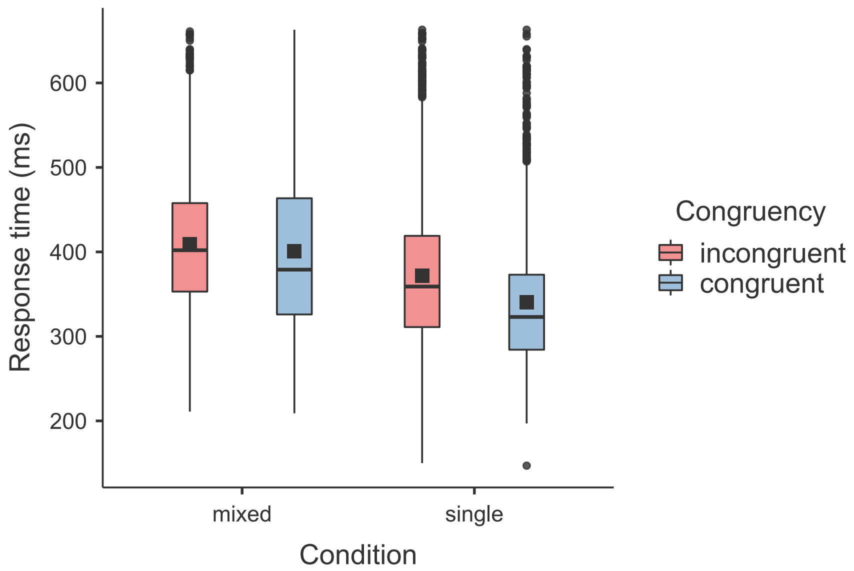


Figure 40: Simon Task reaction times overview.

Whiskers reflect extreme values that are no outliers (within 1.5x IQR). Black squares represent means.

Adding activity components as covariates revealed four activity components affected Simon Task outcomes.

Formula: (lmer) RT ~ 1 + C3 – 2L switching + C5 – L2 social + C7 – L2 academic + CONDI + Session + Congruency + TaskType:C5 – L2 social + Congruency:C7 – L2 academic + TaskType:Congruency + (1 | Part)

Table 24: Model for activity components' effects on Simon Task

Overview	Fixed effects	Estimate	St. Error	P-Value
Intercept	(Intercept)	384.7	11.2	< .001
Main effects	Congruency	22.4	2.0	< .001
	Task Type	49.2	2.0	< .001
	Session1	-6.6	2.6	0.05
	Session2	-19.1	2.8	< .001
	C3 – 2L switching	-5.4	2.4	0.05
	C5 – L2 social	-4.0	1.4	0.01
	C7 – L2 academic	-2.2	1.3	0.1
Task type effects (mix costs)	Task Type ✻ C5 – L2 social	-4.7	1.9	0.05
Congruency effects (interference scores)	Congruent1 ✻ C7 – L2 academic	7.7	1.8	< .001
	Congruency ✻ Task Type	-25.3	4.0	< .001

Participants who scored high on C3 – 2L switching and those who scored highly on C5 – L2 social performed significantly faster on overall RT (β=-5.4, SE=2.4, p<0.023 and β=-4.0, SE=1.4, p<0.005, respectively). C7 – L2 academic failed to improve model fit significantly, but its main effect was retained in the model to facilitate the interpretation of interaction effects involving it.

Participants who scored high on C5 – L2 social (focus on L2 contacts) benefitted from significantly reduced RT in the mixed-congruency Task Type (β=-6.33, SE=1.85, $p<0.001$), leading to decreased mix costs (Figure 41). RT in the single-congruency Task Type did not change significantly (β=-1.63, SE=1.60, $p<0.308$).

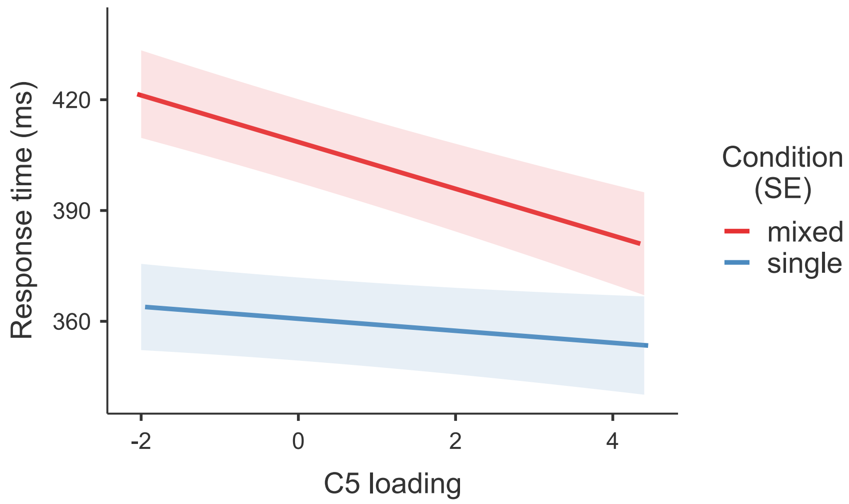


Figure 41: The interaction effect of C5 loading and condition in the Simon task. Shaded areas reflect 95% confidence intervals.

Participants who scored highly on C7 – L2 academic (L2-focus on core academic activities) had a significant benefit in reacting to congruent trials (β=-6.02, SE=1.71, p<0.001) but not incongruent trials (β=1.68, SE=1.52, p<0.270), leading to higher interference scores (Figure 42).

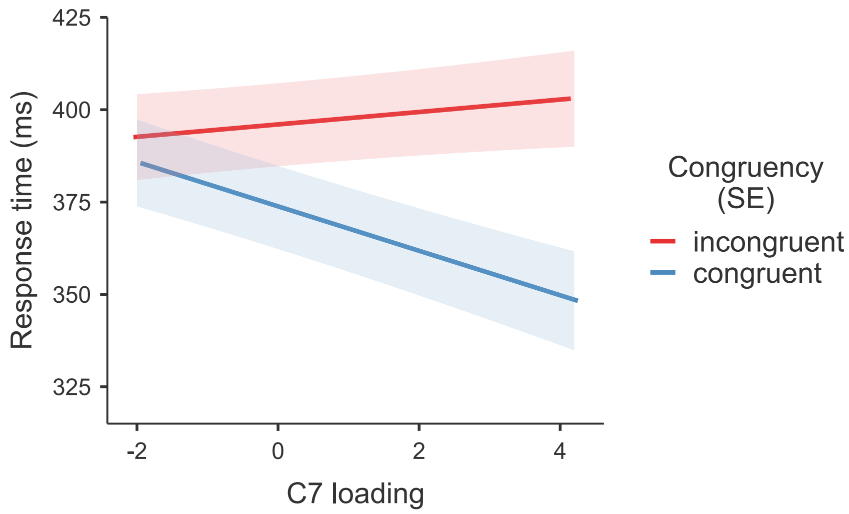


Figure 42: The interaction effect of C7 loading and congruency in the Simon task. Shaded areas reflect 95% confidence intervals.

5.5.3. Flanker Task

The analysis of the Flanker task revealed that, as expected, trials in the – conflicting – mixed-congruency condition elicited significantly higher RTs than those in the single-congruency condition (β =10.0, SE=1.3, p<0.001) and that incongruent trials elicited significantly lower RTs than congruent trials (β =53.6, SE=1.3, p<0.001). Finally. overall RT, but surprisingly not interference scores or mix cost, decreased with each session (p<0.001). For a discussion of these results, refer to 4.6. Discussion. Figure 43 shows overall RTs for congruent and incongruent trials in the mixed-congruency task type and single-congruency task type.

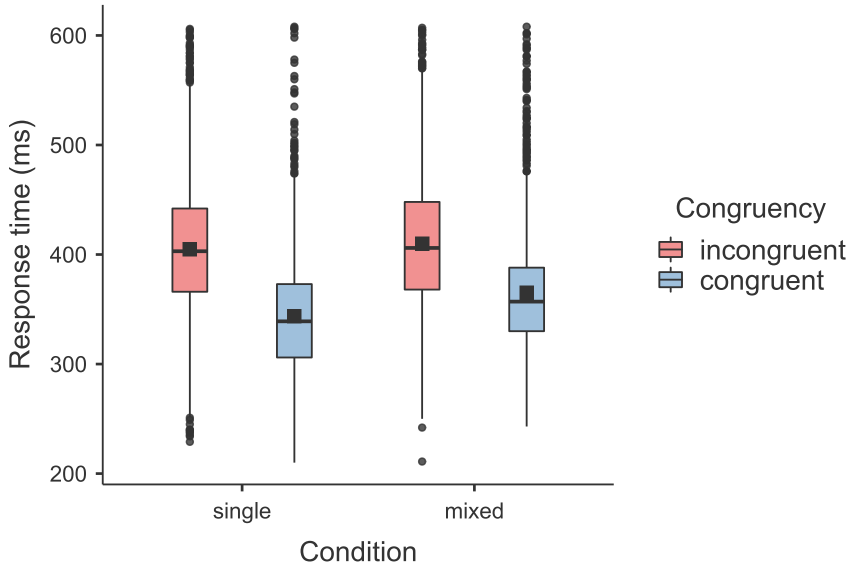


Figure 43: Flanker Task reaction times overview.

Whiskers reflect extreme values that are no outliers (within 1.5x IQR).
Black squares represent means.

The addition of activity components as covariates revealed that six out of the seven components influenced either overall RT, switch cost or mix cost.

Formula: (lmer) RT ~ 1 + C1 – 2L leisure + C2 – 2L academic + C3 – 2L switching + C5 – L2 social + C6 – L1 social + C7 – L2 academic + Session + Task Type + Congruency + Task Type: Congruency + C7 – L2 academic: Task Type + C5 – L2 social: Task Type + C2 – 2L academic: Congruency + Congruency:C3 – 2L switching + Congruency:C5 – L2 social + Congruency:C7 – L2 academic + (1|Participant)

Table 25: Model for activity components' effects on Flanker Task

Overview	Fixed effects	Estimate	St. Error	P-Value
Intercept	(Intercept)	384.1	8.3	< .001
Main effects	C1 – 2L leisure	6.8	1.7	< .001
	C2 – 2L academic	5.1	1.4	< .001
	C3 – 2L switching	-1.8	1.7	0.3
	C5 – L2 social	-0.8	1.2	0.5
	C6 – L1 social	7.0	1.2	< .001
	C7 – L2 academic	-2.4	1.0	0.01
	Session1	1.0	1.8	0.6
	Session2	-22.7	3.2	< .001
	Task Type Mixed	10.0	1.3	< .001
	Congruency Incongruent	53.6	1.3	< .001
Task type effects (mix costs)	Task Type * Congruency	-13.5	2.6	< .001
	C5 – L2 social * Task Type	3.0	1.5	0.05
	C7 – L2 academic * Task Type	4.5	1.3	< .001
Congruency effects (interference scores)	C2 – 2L academic * Congruency	3.8	1.6	0.05
	C3 – 2L switching * Congruency	-4.5	1.4	0.01
	C5 – L2 social * Congruency	-3.3	1.6	0.05
	C7 – L2 academic * Congruency	2.8	1.3	0.05

Participants who load highly on C1 – 2L leisure, C2 – 2L academic, C6 – L1 social and C7 – L2 academic had significantly higher overall RTs. C3 – 2L switching, and C5 – L2 social did not significantly affect RT, but were retained as main effects in the model to support the interpretation of any interaction effects they are involved in.

Participants who score high on C5 – L2 social have a larger facilitation effect for the unmarked single-congruency Task Type (β =-2.29, SE=1.36, p=0.09) as opposed to the mixed-congruency Task Type (β =0.75, SE=1.49, p=0.62), leading to increased mix costs (β =3.04, SE=1.54, p=0.049).

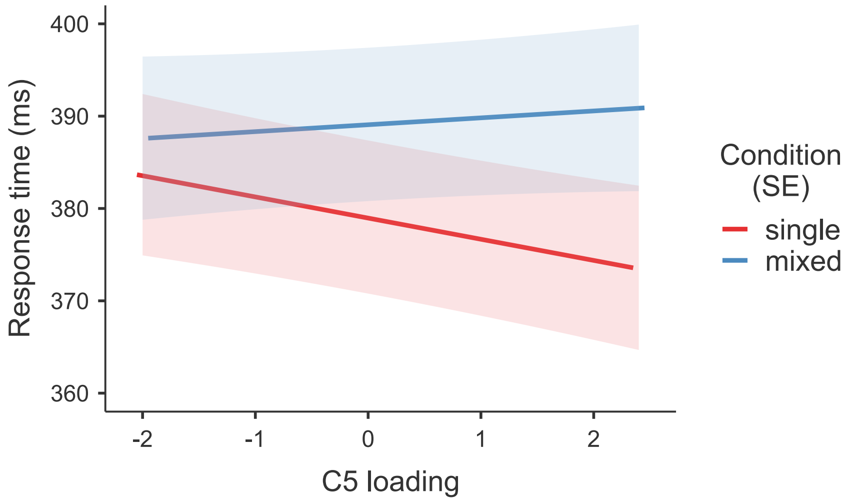


Figure 44: The interaction effect of C5 loading and condition in the Flanker task. Shaded areas reflect 95% confidence intervals.

Mix costs also increased with higher loadings on C7 – L2 academic (β =4.53; SE=1.25; p=0.001). RTs in the single-task Task Type dramatically decrease with increased loadings onto C7 – L2 academic (β =-4.7, SE=1.09; p=0.001), while RTs in the mixed-task Task Type remain largely unchanged (β =-0.17; SE=1.20; p=0.883).

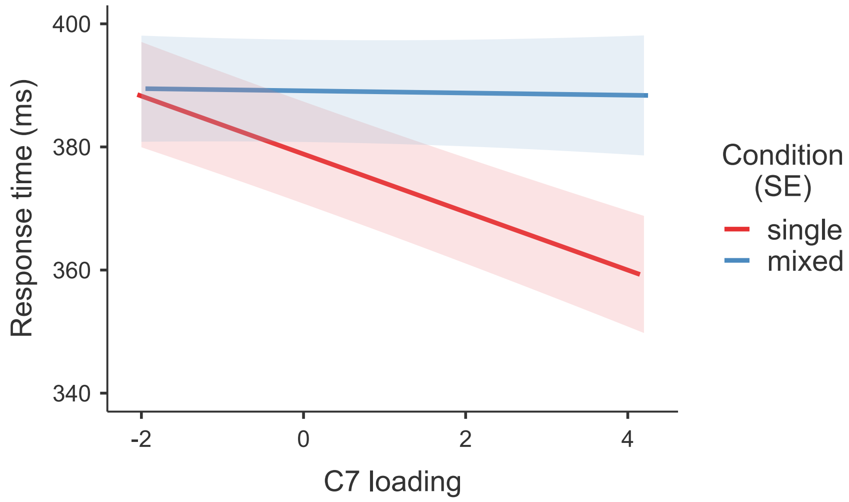


Figure 45: The interaction effect of C7 loading and condition in the Flanker task. Shaded areas reflect 95% confidence intervals.

Interestingly, the effect of C7 – L2 academic on Task Type is mirrored in the analysis of interference scores. Again, the difference between incongruent and congruent trials increases with increased loading onto C7 – L2 academic. This difference is due to a disproportionate facilitation effect of the non-conflicting congruent items (β=-3.85; SE=1.22; p=0.002) as opposed to incongruent items (β=-1.04, SE=1.09, p=0.338), and can as such be interpreted as an increased cost effect due to non-conflict facilitation.

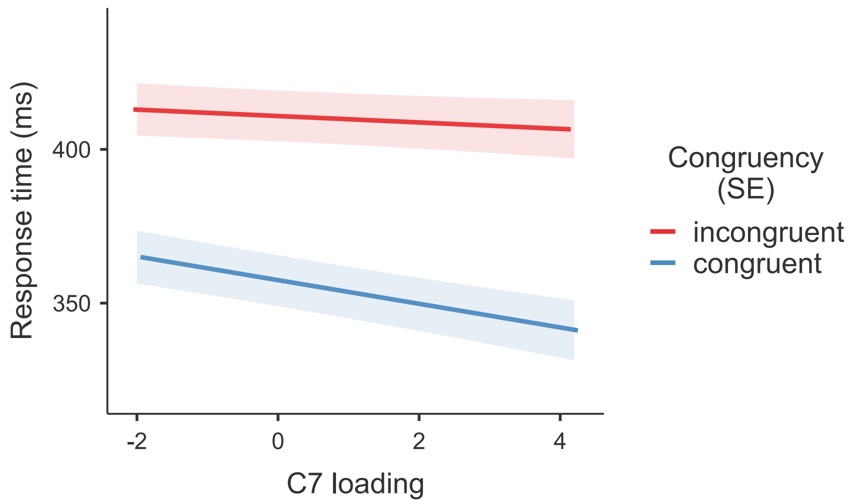


Figure 46: The interaction effect of C7 loading and congruency in the Flanker task. Shaded areas reflect 95% confidence intervals.

Participants with a high loading on C2 – 2L academic had slower reaction times on both congruent and incongruent trials and were disproportionately slower in the incongruent trails (β =7.0; SE=1.6; p=0.001) as opposed to the congruent trials (=3.2; SE=1.73; p=0.064). As such, the higher a participant loaded on C2 – 2L academic, the higher the inference scores (β =3.8; SE=1.22; p=0.02). Figure 47):

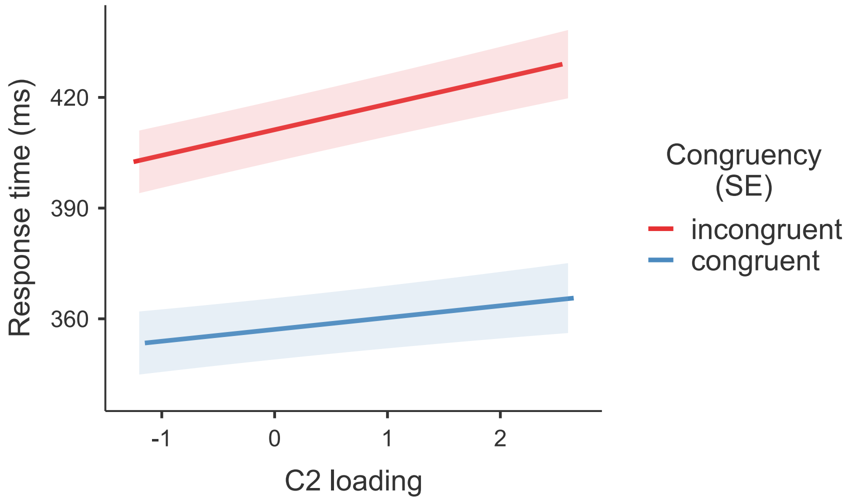


Figure 47: The interaction effect of C5 loading and congruency in the Flanker task. Shaded areas reflect 95% confidence intervals.

Participants who load highly on C3 – 2L switching and participants who load highly on C5 – L2 social benefitted from lower interference scores (β=-4.5, SE=1.4, p=0.002; β=-3.3, SE=1.6, p=0.038, respectively). In both cases the RT of the conflicting incongruent condition reduces (C3 – 2L switching: β=-4.5, SE=1.8, p=0.023; C5 – L2 social: β=-2.4, SE=1.4, p=0.072) while the RT in the unmarked, congruent condition remains statistically unchanged (C3 – 2L switching: β=0.5, SE=1.9, p=0.808; C5 – L2 social: β=0.9, SE=1.5, p=0.6).

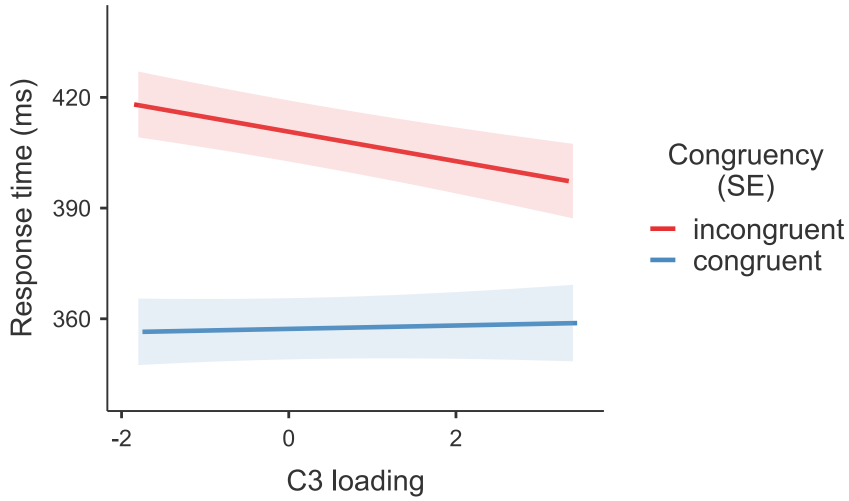


Figure 48: The interaction effect of C3 loading and congruency in the Flanker task. Shaded areas reflect 95% confidence intervals.

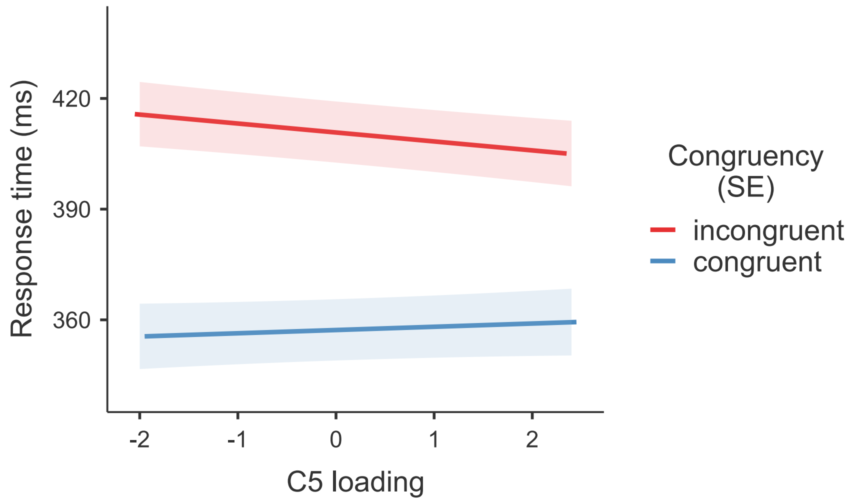


Figure 49: The interaction effect of C5 loading and congruency in the Flanker task. Shaded areas reflect 95% confidence intervals.

Table 26 summarises the findings of the results section. Results relating to LC have been greyed out while results relating to EC have been left while to aid orientation.

Table 26: Summary of Findings Across Tasks

Effect on	Task	1 (L1/L2)	2 (L1/2)	3 (L1/2)	4 (L1)	5 (L2)	6 (L1)	7 (L2)
Overall RT	PNT	↓	↓	↓	-	↓	↓	↓
	Simon	-	-	↓	-	↓	-	-
	Flanker	↑	↑	-	-	-	↑	↓
Interference scores	Simon	-	-	-	-	-	-	↑*
	Flanker	-	↑	↓	-	↓	-	↑*
Language English (Chinese)	PNT	- -	- (↓)	- (↓)	- -	↓ -	- -	- -
Mix Cost	PNT	↑*	-	↑*	-	-	-	-
	Simon	-	-	-	-	↓	-	-
	Flanker	-	-	-	-	-	-	↑*
Switch Cost	PNT	-	-	↑*	-	-	-	-

In Table 26, arrows with an asterisk (*) indicate instances in which the change in cost (difference) is due to the non-conflicting subtrahend (congruent trial average, stay trial average, single-task trial average) increasing or decreasing disproportionately to the conflicting minuend (incongruent trial average, switch trial average, mix-task trial average). As such, in instances with an upward-facing arrow "↑*", costs increase, but RT of the minuend decreases.

5.6. Discussion

In this study we considered the effect of language use on language control and executive control. We set out to test two hypotheses: first, whether participants, who regularly code-switch, will perform better in switch trials and in mixed-language trials as opposed to those participants who do not. Second, whether indices of language use which predict a change in performance in the linguistic control measures also do so for executive control measures. Taken together, most of the seven components of language use had significant impacts on the performance in this study's tasks.

Overall, we have found that those participants who regularly code-switch (i.e., participants who scored highly on dual-language use C1 – 2L leisure, C2 – 2L academic and C3 – 2L switching) had lower overall RTs, lower RTs in Chinese (C2 – 2L academic and C3 – 2L switching), but higher mix costs (C1 – 2L leisure and C3 – 2L switching) and switch costs (C3 – 2L switching). As for the higher mix costs, post-hoc simple effects analyses revealed that these resulted from disproportionate facilitation effects in the conflicting mix condition as opposed a deterioration of performance in the mix condition. For switch costs, another post-hoc simple effects analysis revealed that the increase in costs was due to a disproportionate decrease in RT in non-conflicting stay trials.

As such, we may conclude that our data provides some evidence that regular language switchers – those who loaded highly on C1 – 2L leisure and C3 – 2L switching – are better at mixing languages and that, interestingly, those who scored highly on C3 – 2L switching also have a significant facilitation in non-switch trials, which is disproportionately larger compared to the facilitation in switch trials.

We hypothesized these three components, which positively impacted LC in terms of overall RT and mix costs, would also do so for measures of EC performance. However, only C3 – 2L switching influenced the Simon task performance in that it reduced overall RT, and the Flanker task performance in that it reduced interference scores. These are effects, however, that are not echoed in the PNT results.

C4 – L1 academic reflects the time participants spend on out-of-class contact with teacher in Chinese and completing additional assignments in Chinese as well as improving English vocabulary and was found not to affect any outcome measures in any of the three tasks.

Similarly, those who often use their L2 in social contexts had a similar effect on executive control in that overall RT decreases for the Simon task only, and interference scores in the Flanker task only. However, also Simon mix costs decreased with increasing social L2 use. In contrast, social L2 use did not affect language control in any way.

C7 – L2 academic (using L2 in academic settings) had a large impact on performance in the flanker task, in that it affected RT, interference scores and mix cost: RT decreases overall, and disproportionately so in single-task conditions and congruent trials, resulting in higher mix costs and interference scores, respectively. While C7 – L2 academic has several effects on Flankers, it has comparatively little effects on the Simon task: only the effect on interference scores is replicable in the Simon results.

Interestingly, most components significantly reduced overall RT in the Picture Naming task, suggesting that most forms of language use are beneficial in reducing lexical access latencies.

As such, we conclude that our data does not provide evidence for transfer effects, i.e., habitual language switching effectuating changes to the language control domain transferring to the same changes in the executive control domain generally.

One interpretation of these results is that there are different sub-mechanisms driving the executions of different tasks: the inhibition of interfering language on the one hand, and the inhibition of flanking/visuo-spatial information on the other, (i) may at least partially managed by different sub-mechanisms and that (ii) individual language use impacts these sub-mechanisms differently. In Figure 50, we have illustrated different configurations the different sub-mechanisms may have in relation to one another.

If there were a complete overlap of the systems that support both the operations for the language control task and the executive control tasks (as in “full overlap”), then any type of improvement would likely result in an improvement of both. But in our data, this only holds to a limited extent.

Instead, what we see is that some measures of language use (e.g., use of L2 in social contexts, C5 – L2 social) affect measures of both systems. As such, a case can be made that there may be some sub-mechanisms that support both LC and EC mechanisms, and it is the intersection of this overlap that is susceptible to change as a result of specific language use. This finding is reflected in previous studies, such as Ooi (2018), who found that interactional contexts – as experienced by two different groups of bilinguals – impacted their performance in two tasks differently – as outlined in this study's literature review.

Overall, different impacts on sub-mechanisms could also account for varying evidence for the bilingual advantage in the past literature. Varying, non-standardized tasks across the literature may have tapped into the effects of bilingual language use – and bilingualism generally – on slightly different sub-mechanisms, which may overlap with other sub-mechanisms to smaller, or larger extents.

However, as a counterargument, one could also argue that it is not just this intersection, but rather the entire system (i.e., brain as whole) that receives an advantage as a result of different habits (e.g., increased L2 social contacts), which then engenders these observable changes in processing. In the same way, it could also be that an underlying characteristic of the individuals who load high on the component (for instance, social extroversion – in the case of high loadings on C5 – L2 social (L2 social contacts)) drives this change in the first place, as opposed to language use.

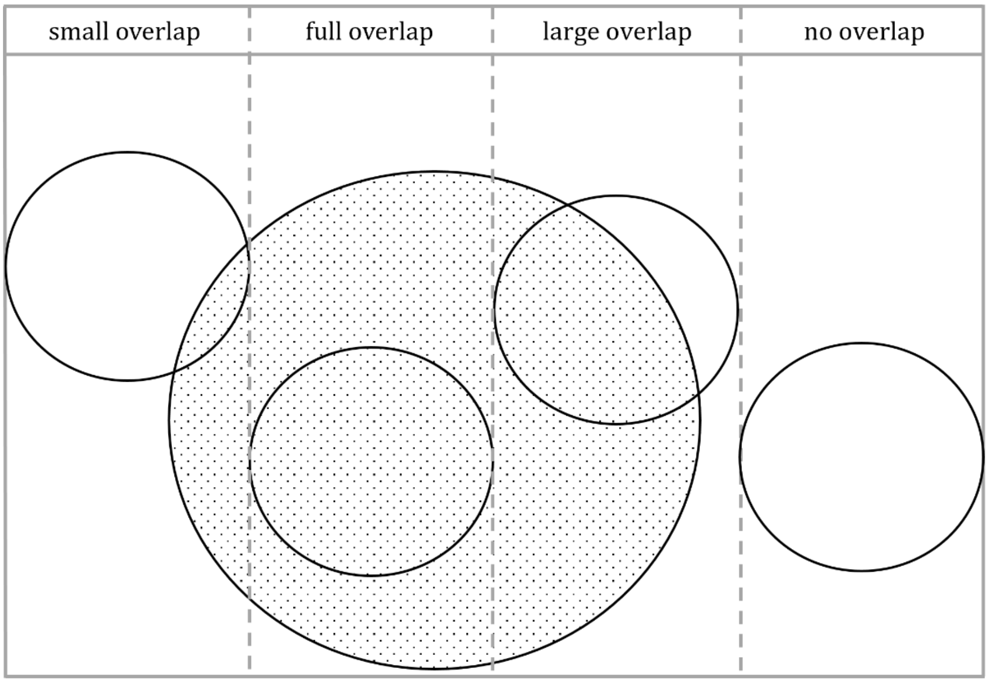


Figure 50: Potential configurations of LC and EC (sub)systems

Ultimately, this line of reasoning leads back to our initial observation that several factors seem to have an impact on the way on both language control and executive control and that it is difficult to tease apart different effects. In our introduction, we have mentioned gaming (Strobach et al., 2012) and meditation (Teper & Inzlicht, 2013). Further, regular aerobic exercise has been found to increase performance on several measures of executive control as well (Guiney & Machado, 2013; for a review). As a consequence of this evidence, this study cannot exclude an alternative explanation for the observed results.

Some of the studies cited as part of the literature review attempt to control for such externalities (e.g., Ooi et al. (2018) control for participants' video gaming habits). Overall, however, there are wealth of other externalities which seem to have an impact on the mind's functions, which are, as of now, still uncontrolled for. As such, future studies may consider even larger dataset including richer contexts, by expanding on the volume of questionnaires and/or the number of experimental tests and/or by including other observations (see following paragraph).

Further to such externalities, we acknowledge that our activity log and its subsequent analysis falls short of being able to successfully differentiate between the different types of bilinguals as described by the ACH. Specifically, in classifying participants into single-language context and dual-language context bilinguals as well as in quantifying the density of the code-switching behaviour – discussed below. First, by offering a multitude of different sub-contexts to respond to and asking participants to do so in hours rendered the aggregation of these sub-contexts to form larger context categories difficult. Here, it would have been more instructive to ask the participants to provide larger categories and have the participant respond directly to whether these tended to be single-language or dual-language contexts for them. Second, by running a PCA, the sub-context bins of different languages further became entangled within principal components, in a way that rendered it not possible to understand how much of either language was being used in different environments.

Another problem is in questionnaires on the nature of the bilinguals language use which painting an abstracted picture of the authentic reality of use. Respondent's subjectivity and generalizations both skew and reduce the full picture. While some studies attempt to control for the inter-reliability of self-reported responses (e.g., Bonfieni et al., 2019); these reports are still a secondary source of information. In this way, the remaining data on language use misses many nuances and is subjective. Hofweber et al. (2016) for instance, further describe nuances of interactional contexts. Within the umbrella term of "code-switching", they introduce the terms "alternation", "insertion L1->L2", and "insertion L2->L1" to the already discussed term "dense code-switching". An alternation is when the utterance has a clear switch point (e.g., Italian-German: "Ich kann nicht zur Arbeit || perché devo andare al dottore" (*I can't go to work, because I have to go to the doctors'*)), whereas insertions are introductions of single-word in an utterance of a predominantly different language (e.g., the insertion of German "Tierschutzverein" (animal shelter) into an English utterance: "If you want a pet, you should visit the Tierschutzverein first"), similar to word borrowing. In a majority of previous studies, it is not clear which exact switches bilinguals utilize most, and therefore the ecological validity of questionnaires is currently lacking. Ideally, therefore, future research on the effect of language use should combine insights from corpuses on real life interactions.

To avoid redundancy, an overview of the limitations of this analysis as well as of the study as a whole are presented in 6.3. Limitations. Specifically, the appropriateness of employing a functionally monolingual group as comparative measure as well as practise effects within the individual experimental sessions and between experimental sessions run are reflected upon in this subchapter.

5.7. Conclusion

In conclusion, this study reflects an advantage experienced by frequently code-switching bilinguals over those who use their languages in more separate contexts in a relatively homogenous group of late bilinguals in their first year of study in an L2 environment. Our results provide mixed support for both the Adaptive Control Hypothesis (Green & Abutalebi, 2013) and the Cognitive Process Model (Green & Li, 2014). Furthermore, an increase in L2 use in social and academic contexts was associated with improved performance across several measures of both linguistic and executive control, highlighting the importance of L2 exposure and quantity of use.

6. General Discussion

This chapter functions as a concluding element of the dissertation in which we revisit the purpose of the individual analyses within this dissertation, summarise the findings and contributions of the current study and evaluate their efficacy in light of our initial aims. Finally, we discuss limitations and offer ideas for directions for future research.

Our study set out to explore the interrelationships of bilingual lexical access, executive control and differing interactional contexts in a group of late adult bilinguals in their first year of study abroad in the UK using a dual-language PNT, Simon and Flanker tasks and Activity Logs administered thrice over the course of six months. To this effect, the starting point of this thesis was to consider the development of bilingual lexical access during the early stages of L2 immersion by analysing naming latencies in a dual-language picture naming task. The next step was to consider the ways in which the bilinguals' executive control functions – as assessed through the Simon Task and the Flanker Task – were impacted by changes in the participants' language control. Finally, we studied the participant's language use habits through the Activity Log Questionnaire to explore to what extent differences in language use habits had an impact on the interplay of language control and executive control during the period at issue.

These developments and their respective impacting factors were mainly investigated separately from each other in the article chapters provided in Chapters 3, 4 and 5. In this section, we will consider the broader picture of the relationship between language control and executive control based on the findings of the thesis. We will argue that changes in executive control functions are likely due to a complex interplay of numerous factors whose individual impacts are not sufficiently understood yet. Further, we propose that to capture this intricate complexity more comprehensively, forthcoming research could benefit from exploring methodologies involving advanced pattern recognition.

6.1. Summary of the findings

In Chapter 3, we investigated the development of bilingual naming latencies, switch costs and mix costs during the initial stages of L2 immersion to see if and how the lexical accessibility of either language changes. We have found that, contrary to predictions, the bilinguals' L1 RT significantly decreases over time while L2 RT does not. We have found that overall switch cost significantly decreased between sessions 1 and 3.[21] However, this decrease was independent of language: switch cost into English decreases between session 1 and consecutive sessions, but not significantly so, while switch cost into Chinese is lowest in session 2. In line with observations by Meuter and Allport (1999), we did, however, find evidence for asymmetric switch costs, as switches into the Chinese L1 were significantly more costly than into the English L2. We have found a significant interaction effect between trial type, language, and language balance, which suggests those bilinguals who are more proficient in the L2 have a smaller switch cost asymmetry. We conclude that L2 proficiency strongly impacts switch cost asymmetry and since L2 proficiency did not significantly increase during the testing period, switch cost asymmetry did not recede, either. Curiously, we found no evidence for mix cost. Instead, mixing languages was found to facilitate naming, as responses were 170ms faster on average, suggesting that with two language systems/lexica co-activate, neither is globally inhibited, which renders naming faster overall. We question whether global inhibition exerts a higher cognitive load than local inhibition. Further, it seems as though this phenomenon is specific to less proficient individuals, as increasing proficiency decreased the facilitation effect. We also find a mix "cost" (or rather: "facilitation") asymmetry, with mixing in the English L2 being significantly more facilitative than in the L1. Curiously, like the switch cost asymmetry, this facilitation strongly depends on L2 proficiency.

In Chapter 4, we explored the development of congruency costs and mixed costs in Simon and Flanker tasks during the initial stages of L2 immersion to understand if and how the performance in executive control tasks changes. We have found that the evidence for a bilingual advantage in executive control tasks is mixed. It was only in the third session of the Flanker task that bilinguals significantly outperformed monolinguals in terms of overall RT. We did find, however, that in the Flanker task, bilinguals had

[21] The possibility of this finding being a result of being a practise effect is discussed in 6.3. Limitations.

significantly smaller congruency costs, suggesting that bilinguals are better than monolinguals at inhibiting conflicting information. These findings failed to replicate in the Simon task, as bilinguals did not outperform monolinguals in terms of overall RT or congruency cost in any of the three sessions. Interestingly, bilinguals were found to have a significantly larger mix cost than monolinguals in both the Simon and the Flanker tasks. However, in both instances, the increase in mix cost was due to the bilinguals performing significantly faster in the single task condition as opposed to the mix task condition.

In a second step, we tested whether any outcomes of the dual-language picture naming task can predict performance on either of the executive control tasks to understand the relationship between language control outcomes and executive control performance. Several outcomes of the picture-naming task were able to predict performance in both the Simon and the Flanker tasks. For instance, those with a lower switch cost asymmetry and more symmetric access to either language in the dual-language picture naming task also had lower overall RTs in the Simon task. Further, those participants with a smaller PNT mix cost asymmetry also had a smaller Simon mix cost, and those participants with a smaller PNT switch cost asymmetry, again, had smaller Simon congruency costs. For the Flanker task, L2 proficiency, English switch cost and Chinese switch cost performance had a significant effect. Those with lower switch costs (in both Chinese and English) and higher L2 proficiency are performing better in terms of overall RT on the Flanker task. However, no PNT outcome had a significant impact on Flanker congruency cost or Flanker mix cost. Finally, we investigated participant's intra-individual variation (IAV) on both the linguistic control and executive control tasks to understand whether there were any patterns between IAV in one domain and the other. Contrary to expectations, we found that, across sessions, participants who varied less in the Flankers task varied more in English naming latencies.

In Chapter 5, we investigate the role of language use on language control and executive control, respectively, to understand how the amount of exposure to certain activities in the two languages shapes the performance on either task. Our results offer partial corroboration for both the Adaptive Control Hypothesis (Green & Abutalebi, 2013) and the Cognitive Process Model (Green & Li, 2014). Additionally, frequent use of a second language in social and academic settings was associated with better performance across various measures of both linguistic and executive control, highlighting the importance of L2 exposure and quantity of use.

6.2. Implications

The findings of the current study have several implications regarding the research of early development of bilingual lexical access, the bilingual advantage debate, as well as methodological designs. In the following, we consider the following:

(i) how different stages in the development of language control may impact investigations surrounding executive control,

(ii) whether our findings can be classed as a bilingual advantage in terms of quantity,

(iii) whether the comparison with monolinguals – here as in other studies – is justified, and

(iv) whether our experimental group was sufficiently homogenous to allow for the exclusion of confounds while questioning whether homogeneity is constructive in establishing ecological validity.

Finally, we discuss the impact of categorical thinking in this thesis as in the bilingual advantage debate as a whole and explore to what extent different approaches – namely, chaos-theoretical constructs and computational models – can provide a different point of view on the debate.

Our studies have implications for the timeframes at which we may expect differences to the language control system to occur, which may then – according to Bialystok (2009) – present an adaption challenge to the cognitive skills, eventually resulting in changes to executive functioning: our results have shown that language control mechanisms are present as early as the first month of L2 immersion, as evidenced by high switch costs into the L1 from the first session onward. This suggest that the increased processing requirements faced by bilinguals have set in at least as early as the onset of L2 immersion.

We have shown that switch and mix cost asymmetries are closely linked to language balance: as the L2 gains in relative use, the competition between the two language systems seems to equalize. This finding could mean that the costs associated with switching or mixing decrease as bilinguals become increasingly efficient at dual-language management – a linear development. At the same time, it could also mean that as the two language systems become more equal in associative strength, the strength with which the (dominant) non-target language needs to be inhibited decreases and thus decreasing the need for active inhibition mechanisms – a non-linear development. There is likely a combination of these developments involved whose impacts are difficult to entangle – particularly as they are likely further entangled in other aspects such as switching habits. As an implication for further debate on the topic of the bilingual advantage, however, this finding highlights the usefulness in the inclusion of language balance, switch and mix costs as well as it could point toward different developmental stages exerting different demands on the system, which may follow a non-linear trend.

This study has implications on the term “advantage” in bilingual executive control. Overall, our studies’ results are most fittingly reflected in the title of Costa et al.’s (2009) study “On the bilingual advantage: now you see it, now you don’t”. While we did find some instances, in which bilinguals arguably “performed better” as a group than their monolingual counterparts, this “advantage” depended a lot on which task we were looking at (Simon, Flanker), and which metric (overall RT, mix cost, inference score) was at issue. In several places, this ambiguity as to what is classed as success or advantage led to mixed interpretations. For instance, while Chapter 4 bilinguals present significantly better inference scores (than monolinguals) in the Flanker Task, none such benefit was found in the Simon Task. Further, bilinguals did not outperform monolinguals in any aspect in the Simon Task.

Overall, this lack of clarity under what circumstances we can class something as “advantage” leads to the observation that the likelihood of a positive findings is rather subjective. Paap et al. (2015) support this notion, arguing that bilingual advantages either do not exist, or are limited to specific – and perhaps even random – circumstances. Adding to this subjectivity is De Bruin et. al.’s (2015) finding that between the years of 1999 and 2012 there has been a publication bias in favour of studies publishing a positive finding on the bilingual advantage. Overall literature might be skewed towards favouring the presence of the bilingual advantage and some researchers may have thus been motivated to interpret an objectively mixed result as a positive result instead as a consequence of confirmation bias.

Even for results where a significant improvement is given - such as Flanker inference scores – the question remains at which threshold one can consider this difference in reaction times an “advantage” in its own right. In our results we showed that when bilinguals “outperformed” monolinguals in inference scores, then this was usually only by a few milliseconds (8 – 13ms). Despite this statistical significance, this proposed advantage (1-3% change to the average overall RT) seems rather negligible to the pragmatic mind. To this end, we also question to what extent this quantity would present an advantage outside of psychological assessment. For example, where conflicting visuo-spatial information needs to be inhibited in real-life situations such as driving a car or reading.

The observation that the likelihood of positive findings is subjective in the sense that it seems to depend on **what** is being considered comes hand-in-hand with the question of **who** is being considered. There is an ongoing debate in bilingualism research debating the validity of comparing measures of psycholinguistics tasks between mono- and bilingual participants (e.g., Rothmann et al., 2023; Kroll et al., 2012; De Groot & Kroll, 2014). The basic notion is that mono- and bilinguals are too different to allow for a meaningful comparison, as bilingualism fundamentally shapes the brain in various ways – for instance, by introducing the requirement for inhibitory control to activate the target language selectively. Supporters of this notion argue that any comparisons result in less meaningful comparisons, as two distinctively different elements are being compared. Further, given the high prevalence of bilingualism in the world – at least over half the world's population (Grosjean, 2021) – supporters of this notion argue that bilingualism can be considered the "norm" or "evolutionary default" and that instead, monolingual speakers are the "exception".

This view somewhat turns the debate over the "bilingual advantage" on its head, as it assumes bilingualism as the norm and monolingualism as a more limited use of the naturally more potent language facilities. Supporters of this notion argue that comparisons among different groups of bilinguals are more helpful in understanding underlying processes and appreciating the complexities of the bilingual mind, for instance, by comparing groups of bilinguals of different ages, language pairs, or language balance. Others argue that a distinction between mono- and bilinguals is a valuable initial step to establish that there are any differences. In a second step, it is argued, the locus of the differences can then be examined in the context of bilingual language processing as opposed to language processing more generally.

Our current study understands itself as an exploratory study, which utilises more coarse distinctions between group: the (functionally) monolingual group of English speakers was included to provide a brief exploratory insight into the differences between them and the bilingual Chinese-English speakers. Resource limitations meant that it was not possible to recruit a better matching, and more homogeneously monolingual, participant group. We acknowledge that through the heterogeneity introduced by variables such as the previous L2 exposure reported by just over half of the English native speakers, we are unable to exclude potential confounds in our analysis. Moreover, there may be significant cultural discrepancies between the two groups, as bilinguals were raised in China while the functional monolinguals grew up in the UK. This discrepancy there is an additional layer of complexity affecting the comparability of the two groups (e.g., Samuel et al., 2018). Overall, these discrepancies between the two groups mean that the meaningfulness of any found association is inevitably reduced (Rothmann et al., 2023). Ideally, we would have liked to recruit a control group from China. This ideal group would differ from the experimental group only in L2 immersion duration. Age, gender, sociolinguistic background as well as other potential confounds (mentioned in later chapters) would ideally have been matched between the two groups. On top, we would have liked to measure performances of the within the same number of sessions to be able to account for any practise effects.

Concurrently, effects within the bilingual group should not be underestimated. The aim of the stringent recruitment policy was to create a sample which was as homogenous as possible to avoid the impact of potential confounds. The implementation of linear mixed models allowed for a comprehensive investigation of both unique and shared effects. However, in each of the studies – specifically Chapters 4 and 5 – we have acknowledged that there may be several other effects may have influenced participants' performance within the bilingual group. For instance, meditation (Teper & Inzlicht, 2013) and video games (Strobach et al., 2012), for which we do not take any measures in this study. Also, while "homogeneity" allows for the interpretation of effects between groups, the choice to focus on a small subset of the bilingual population restricts the ecological validity of our results. The question remains to what extent a slightly different subset of the bilingual population would have incurred different results.

To summarise, this discussion of implications has explored the notion of different development stages of language control and its implications for executive control, the underlying differences between mono- and bilingual language processing, questioned whether these differences can be considered an "advantage" in a more applied sense and finally, considered alternative explanations for the observed variations between participants. Overall, this discussion painted a convoluted picture on the question of the bilingual advantage. Having discussed the above points individually in their respective subchapters, this final subchapter – within the 6.2. Implications section – collates and further abstracts these points to form a basis for a proposition.

6.3. Limitations

This forthcoming chapter will cover the methodological limitations we have encountered in the process of conducting the current study. Specifically, we will explore the results and discuss the implications of the PNT Data Audit (Appendix 2: PNT Data Audit) and discuss the possibility of repetition effects – as opposed to effects relating specifically to the bilingual experience – in the executive control tasks.

The PNT data audit revealed that a comparatively large number of trials were excluded from analysis: 24.1% of all bilingual trials and 10.4% of all monolingual trials (Appendix 2.A: Data Exclusion Rates). Comparable studies only excluded around 5% of their data (e.g., Bonfieni, 2019). While reasons varied (e.g., RT > 4000ms, skipped trails), the overall impression this finding portrays is that participants found our presentation of the task more difficult or less engaging than the presentations provided in other studies. Potential reasons for this discrepancy in exclusion rates could be that that the PNT was presented to participants after the Simon Task and the Flanker Task, and that participants did not receive compensation for their participation. The finding that bilinguals tended to complete trials increasingly slowly within a session would support this interpretation (Appendix 2.B: Training Effects). It could have been the case that as the experiment went on, participants became increasingly disengaged with the task and consequently became slower in naming. However, this interpretation cannot be verified due to the lack of counterbalancing in the study's design.

Our results show lower RTs in the mixed condition as opposed to the single condition (Figure 14: Mix cost of Chinese and English by Session.) and a lower trial exclusion rate within the mixed block than in within single-language blocks (Appendix 2.A: Data Exclusion Rates). A possible explanation for this effect could be an unequal distribution of word frequencies (Appendix 2.C: Word Frequency Distributions): the analysis of word frequency distributions revealed that the frequencies of words employed in the different lists for the different blocks did not prove comparable, as distributions varied across blocks and sessions (Appendix 4: Lists Used in the PNT By Block And By Session).

Overall, mixed blocks featured significantly higher average word frequencies. Further, the order in which the blocks were presented was not counterbalanced with the mixed task always being presented last (see 2.2.4. Design of timed experiments). This could have given way to practise effects, whereby participants become better at completing a task with increased exposure to it. For this reason, we believe that there is a positive likelihood that the assessed mixed benefit may be an artefact of skewed experimental parameters.

Nonetheless, a small number of other studies have found notions of mix benefits under certain circumstances – e.g., De Bruin et al. (2018) in voluntary language switches. What is more, we found several links to mix effects and other effects such as language balance (3.5.4. Language Balance) and language use metrics (5.5.1. Picture Naming Task), suggesting that the results may be systemic after all. One possible reason why few other studies have found a mix benefit could be the time of observation. Our investigation was carried out at a very early stage of L2 immersion, while others consider populations with a much higher LOR – for instance, 3.7 years in Bonfieni et al. (2019)'s study. Supporting this notion is our finding that mix "cost" asymmetry reduces as bilinguals become more balanced: as overall English response times decrease, the effect of mixing no longer provides an increased facilitation effect. Perhaps a mix benefit is a characteristic of very early L2 immersion.

As we cannot tell the development of effects such as the above from the time span at the centre of the current investigation, this limited time frame also represents a limitation to this study. Another way in which the time frame of the study may have limited analysis is in the exploration of intra-individual variation (4.5.3. Intra-Individual Variation). Here, two correlational analysis – between inconsistency scores of EN-PNT and CH-PNT as well as dispersion scores between Flanker and Simon – have narrowly missed the 0.05 significance level. Considerations on the level of intra-individual variation typically require a large N – either achieved through the number of participants or the number of sessions – to allow for a meaningful analysis of patterns. It is possible that the above correlations would have reached statistical significance if the study had included further sessions – or participants.

The consideration of time and the timeframe of the study also raises questions about training elements. Our study assessed the same group of bilinguals three times over six months to understand the impact of L2 immersion on experimental outcomes. Although we re-administered the PNT with monolingual controls in the third session, we refrained from administering the Simon and Flanker tasks multiple times and instead used the monolinguals' performance on their first iteration as a baseline measure for comparisons with bilinguals. This absence of data for monolingual executive control tasks in the second and third sessions complicates the interpretation of effects of training through iterative assessment or L2 immersion. It would have been instructive to include monolinguals in the executive control assessments for sessions 2 and 3 as well, to facilitate the interpretation of the individual effects.

Further to the limitations outlined here, 5.6. Discussion provides further considerations on the limitations specific to Chapter 5. Specifically, the discussion explores the limitations of the use of self-reporting questionnaires and the difficulties in quantifying language switching habits.

6.4. Future Directions

We argue that a recurring "fallacy" in the debate on the bilingual advantage is the employment of categorical thinking. Categorisation provides practical advantages in conducting research as it allows the operationalisation of these categories as factors in research design analysis. However, breaking down complex systems and dynamics can lead to oversimplification, disregarding the nuanced interactions and intricacies present in real-world contexts. For instance, at the level of individual speakers, categorical thinking becomes evident in the classification of individuals as either "monolingual" or "bilingual". Above, we have explored the critical aspects of such categorisation, citing researchers who argue that both populations are so strikingly different that a direct comparison between the groups is inadequate.

The likes of Kroll et al. (2012) and De Groot and Kroll (2014) argue that, instead, a more nuanced approach ought to be taken, focussing on effects within the bilingual continuum. Different researchers have taken this approach and investigated, for instance, the effects of different L2 environments within a group of bilinguals (e.g., Ooi et al., 2018). This dissertation features time and language use distinctions. While these approaches are extremely useful in shedding light on the influence of one – or arguably, many – additional features, there will always be several dimensions left accounted for. For instance: the restfulness of last night's sleep (Dahl, 1996) to the proximity to the next assignment deadline (Krabbe et al., 2017), down to one's childhood experiences

(Friedman et al., 2009) and genetic makeup (Braver et al., 2010; Barnes et al., 2011; Reuter et al., 2007)[22]. In a very liberal allusion to "Heisenberg's Uncertainty Principle" (1927), it could be said that – using currently available methods – the focus on one aspect inevitably comes at a loss of another.

Moving from exploring the complexity between-speaker variation, the same sort of categorical thinking tends to be employed to refer to processes of the mind. For instance, in research that refers to **the** system that regulates language and **the** system that is responsible for executive control. Again, there are some researchers who do take a more nuanced approach in suggesting that these systems may have different aspects to them – for instance, in arguing that there are "local" and "global" inhibition systems in language control (de Groot & Christoffels, 2006). In this dissertation, we recognize that there are different aspects of executive control (e.g., inhibition, shifting) and that different cognitive skills are required in the three experiments we explored.

However, while arguably more nuanced, we argue that these labels still fall critically short of capturing the full complexity of the intricate processes occurring in the mind. Consequently, effects may become difficult to interpret. For instance, in our results to Chapter 5 we discovered that different variables of language use had different impacts on metrics relating to language control and executive control. Without an even more nuanced approach to which specific circumstances engender these effects, these results cannot be understood. We argue that the categorical thinking bins cognitive processes in such a way that do not allow for meaningful interpretation.

As a response to these findings and this problem, in our 5.6. Discussion, we introduced the idea that there may be several subsystems – each supporting different cognitive aspects on a small scale – and that certain groups of these subsystems may be differently co-involved in both language control and executive control. In this view, there is not "one" language control system, or executive control system that may or may not share certain overlaps. Instead, there is a multitude of small-scale operators which each support certain aspects of language control and executive control under certain conditions. Taking inspiration from biolinguistics and computational linguistics, this bottom-up view moves away from the more traditional top-down approach and tries to

[22] It is interesting to note at this point that those studies, which manage to control for several factors do not find evidence for a link between language switching habits and executive control (e.g., Kałamała et al., 2020; de Bruin, et al., 2015).

link the nuanced results we see in this dissertation – but also in the research on this debate generally (e.g., Paap et al. (2015) – increasingly to the underlying architecture and mechanism of the brain.

The brain is estimated to possess around 85 billion neurons (Azevedo et al., 2009). These groups of neurons fire if they reach a certain chemically induced action potential (Bean, 2007). While it is not yet fully understood how exactly these neurons induce a certain behaviour, some research suggests that neurons fire at different frequencies depending on the specific task at hand. For instance, Georgopoulos et al. (1982) shows how a single motor neuron shows different activation patterns if a lever is pulled in one direction as opposed to another. Crucially, action potentials are incredibly fast and fire several hundreds of times during a second (idem). In this sense, it is a continuous action as opposed to a discrete on/off process. For this reason, there may be several different patterns of activation and as such, cognitive processes are perhaps more characteristic of a chaotic system – containing patterns that never repeat.

While it is not currently common or possible to directly link linguistic processes to these individual operators[23], we argue that appreciating this underlying abundance of small systems assists in understanding human cognition. For instance, in the understanding that there is an intricate interplay of factors, which may present nonlinear and even unpredictable patterns. This shift from rigid cause-and-effect explanations to holistic understanding highlights the significance of small changes leading to major outcomes. We also acknowledge that the currently available and employed methodologies are not yet sufficient to capture the full complexity of this vast diversity of internal factors – as well as the multitude of external influences. However, it may be worth to consider employing methods that more closely resemble this underlying processing system, such as neural networks.

Neural networks can uncover insights that linear models might miss due to their non-linear nature and complex architecture. While linear models assume linear relationships between input and output variables, neural networks consist of interconnected layers of

[23] A difficult theoretical issue in linking linguistic processes to the cognitive architecture is that linguistics and neuroscience do not use a common measure to describe cognitive and linguistic processes (Poeppel, 2012; 2016). While fundamental concepts of representation in linguistics include the terms language, lexica and lexical items; in neuroscience, the fundamental elements are neurons and synapses. The neurological pendants to the linguistics units are – as of now – still poorly understood (Boeckx, 2013).

nodes ("neurons") that process data through non-linear activation functions. This enables them to learn complex relationships and hierarchies within the data, allowing for the recognition of intricate patterns, subtle nuances, and interactions that linear models may overlook. Neural networks can automatically extract relevant features from raw data and do not rely on manual feature engineering. Recently, computational psycholinguists successfully reaped the advantages of neural networks to detect ADHD (Deng et al., 2022) and dyslexia in children (Haller et al., 2022), and to infer native and non-native reading comprehension (Reich et al., 2022). Another interesting approach in this field is the generation of synthetic data, which forego the problem of data scarcity (e.g., Prasse et al., 2023). We suggest that in the future similar research methodologies could be implemented to understand the processing differences – and any potential processing advantages – between bilinguals and monolinguals while accounting for individual differences.

6.5. Concluding remarks

To conclude, this study explored the relationship of linguistic control and executive control in a cross-longitudinal study of bilingual development in early stages of an L2 immersion context. We have explored the development of bilingual lexical access in the first months of L2 immersion and investigated the relationship between language control and executive control – as well as the impact of secondary factors on this relationship – during this period. Notwithstanding its limitations outlined above, this study aims to have contributed to understanding the interplay of language control and executive control in an early L2-immersion context.

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of neurolinguistics*, *20*(3), 242-275.

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and cognitive processes*, *23*(4), 557-582.

Abutalebi, J., Annoni, J. M., Zimine, I., Pegna, A. J., Seghier, M. L., Lee-Jahnke, H., ... & Khateb, A. (2008). Language control and lexical competition in bilinguals: an event-related fMRI study. *Cerebral cortex*, *18*(7), 1496-1505.

Abutalebi, J., Della Rosa, P. A., Ding, G., Weekes, B., Costa, A., & Green, D. W. (2013). Language proficiency modulates the engagement of cognitive control areas in multilinguals. *Cortex*, *49*(3), 905-911.

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), 2076-2086.

American Psychological Association. (n.d.). Lexical Access. In *APA dictionary of psychology*. Retrieved January 10, 2024, from https://dictionary.apa.org/lexical-accessAnderson, J. A., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior research methods*, *50*(1), 250-263.

Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., ... & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, *513*(5), 532-541.

Bak, T. H., Long, M. R., Vega-Mendoza, M., & Sorace, A. (2016). Novelty, challenge, and practice: The impact of intensive language learning on attentional functions. PloS one, 11(4), e0153485.

Bamidis, P. D., Vivas, A. B., Styliadis, C., Frantzidis, C., Klados, M., Schlee, W., ... & Papageorgiou, S. G. (2014). A review of physical and cognitive interventions in aging. *Neuroscience & Biobehavioral Reviews*, *44*, 206-220.

Barnes, J. J., Dean, A. J., Nandam, L. S., O'Connell, R. G., & Bellgrove, M. A. (2011). The molecular genetics of executive function: role of monoamine system genes. *Biological psychiatry*, *69*(12), e127-e143.

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

Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics*, 35(5), 857-894.

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

Bean, B. P. (2007). The action potential in mammalian central neurons. *Nature Reviews Neuroscience*, *8*(6), 451-465.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development*, 81(6), 1641-1660.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and cognition, 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., Craik, F. I., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological science in the public interest*, *10*(3), 89-129.

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

Birdsong, D. (2014). Dominance and age in bilingualism. *Applied Linguistics*, *35*(4), 374-392.

Birdsong, D. (2016). Dominance in bilingualism: Foundations of measurement, with insights from the study of handedness. In C. Silva-Corvalán & J. Treffers Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 85–105). Cambridge University Press.

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in psychology*, 9, 81.

Birdsong, D. (2023) Whither Bilinguals, Natives, and Variability? A Commentary on "The Neurocognitive Underpinnings of Second Language Processing: Knowledge Gains From the Past and Future Outlook". *Language Learning*.

Birdsong, D. (Ed.). (1999). *Second language acquisition and the critical period hypothesis*. Routledge.

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Bilingualism: Language and Cognition*, 17(3), 610-629.

Boeckx, C. (2009). *Language in cognition: Uncovering mental structures and the rules behind them*. John Wiley & Sons.

Boeckx, C. (2013). Biolinguistics: Facts, fiction, and forecast. *Biolinguistics*, *7*, 316-328.

Boersma, P. & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 10th June 2019 from http://www.praat.org/.

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta psychologica*, *193*, 160-170.

Bosma, E., & Blom, E. (2019). A code-switching asymmetry in bilingual children: Code-switching from Dutch to Frisian requires more cognitive control than code-switching from Frisian to Dutch. *International Journal of Bilingualism*, *23*(6), 1431-1447.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. Trends in cognitive sciences, 16(2), 106-113.

Braver, T. S., Cole, M. W., & Yarkoni, T. (2010). Vive les differences! Individual variation in neural mechanisms of executive control. *Current opinion in neurobiology*, *20*(2), 242-250.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, 1(1)

Campbell, J. I. (2005). Asymmetrical language switching costs in Chinese–English bilinguals' number naming and simple arithmetic. Bilingualism: Language and Cognition, 8(1), 85-91.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. Studies in second language acquisition, 26(2), 227-248.

Costa, A., & Santesteban, M. (2004). Bilingual word perception and production: two sides of the same coin?. Trends in cognitive sciences, 8(6), 253.

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

Dahl, R. E. (1996, March). The impact of inadequate sleep on children's daytime cognitive function. In *Seminars in pediatric neurology* (Vol. 3, No. 1, pp. 44-50). WB Saunders.

Davidson, D. E. (2010). Study abroad: When, how long, and with what results? New data from the Russian front. *Foreign Language Annals*, 43(1), 6-26.

De Bot, K., & Schreuder, R. (1993). Word production and the bilingual lexicon. The bilingual lexicon, 191, 214.

De Bruijn, E. R., Dijkstra, T., Chwilla, D. J., & Schriefers, H. J. (2001). Language context effects on interlingual homograph recognition: evidence from event-related potentials and response times in semantic priming. Bilingualism: Language and Cognition, 4(2), 155-168.

De Bruin, A., Bak, T. H., & Della Sala, S. (2015). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *Journal of Memory and Language*, *85*, 15-26.

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

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

De Groot, A. M., & Christoffels, I. K. (2006). Language control in bilinguals: Monolingual tasks and simultaneous interpreting. Bilingualism: Language and Cognition, 9(2), 189-201.

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

DeKeyser, R. M. (2007). Study abroad as foreign language practice. Practice in a second language: *Perspectives from applied linguistics and cognitive psychology*, 208-226.

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

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. Proceedings of the National Academy of Sciences, 116(15), 7565-7574.

Deng, S., Prasse, P., Reich, D. R., Dziemian, S., Stegenwallner-Schütz, M., Krakowczyk, D., ... & Jäger, L. A. (2022, September). Detection of ADHD based on eye movements during natural viewing. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 403-418). Cham: Springer Nature Switzerland.

Dewey, D. P. (2008). Japanese vocabulary acquisition by learners in three contexts. Frontiers: *The Interdisciplinary Journal of Study Abroad*, 15, 127-148.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in second language acquisition, 26*(2), 249-273.

DuFon, M. A., & Churchill, E. (Eds.). (2006). *Language learners in study abroad contexts* (Vol. 15). Multilingual Matters.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. Science, 270(5234), 305-307.

Eriksen, B. A.; Eriksen, C. W. (1974). Effects of noise letters upon identification of a target letter in a non- search task. Perception and Psychophysics. 16: 143–149.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Félix-Brasdefer, J. C. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community. *Language learning*, 54(4), 587-653.

Foster, P. (2009). Lexical diversity and native-like selection: The bonus of studying abroad. In *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 91-106). London: Palgrave Macmillan UK.

Freed, B. F. (1995). Second language acquisition in a study abroad context. *Second Language Acquisition in a Study Abroad Context*, 1-361.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. Journal of experimental psychology: General, 133(1), 101.

Friedman, N. P., Corley, R. P., Hewitt, J. K., & Wright Jr, K. P. (2009). Individual differences in childhood sleep problems predict later cognitive executive control. *Sleep*, *32*(3), 323-333.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. Psychological science, 17(2), 172-179.

Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module]. Retrieved from https://gamlj.github.io/.

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., ... & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. Frontiers in psychology, 5, 11.

References

Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., & Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, *2*(11), 1527-1537.

Giovannoli, J., Martella, D., Federico, F., Pirchio, S., & Casagrande, M. (2020). The Impact of Bilingualism on Executive Functions in Children and Adolescents: A Systematic Review Based on the PRISMA Method. *Frontiers in psychology*, 11, 574789.

Gollan, T. H., Kleinman, D., & Wierenga, C. E. (2014). What's easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. Journal of Experimental Psychology: General, 143(6), 2167.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. Journal of memory and language, 58(3), 787-814.

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

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

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

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

Green, D. W. (2003). The neural basis of the lexicon and the grammar in L2 acquisition. The interface between syntax and the lexicon in second language acquisition, 1, 197-208.

Green, D. W. (2011). Language control in different contexts: the behavioral ecology of bilingual speakers. Frontiers in psychology, 2, 103.

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

Guiney, H., & Machado, L. (2013). Benefits of regular aerobic exercise for executive functioning in healthy populations. *Psychonomic bulletin & review*, *20*(1), 73-86.

Haller, P., Säuberli, A., Kiener, S. E., Pan, J., Yan, M., & Jäger, L. (2022). Eye-tracking based classification of Mandarin Chinese readers with and without dyslexia using neural sequence models. *arXiv preprint arXiv:2210.09819*.

Han, X., Li, W., & Filippi, R. (2022). The effects of habitual code-switching in bilingual language production on cognitive control. *Bilingualism: Language and Cognition*, 25(5), 869-889. doi:10.1017/S1366728922000244

Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. Cognition, 150, 10-19.

Hartanto, A., & Yang, H. (2020). The role of bilingual interactional contexts in predicting interindividual variability in executive functions: A latent variable analysis. *Journal of Experimental Psychology: General*, *149*(4), 609.

Heikoop, K. W., Declerck, M., Los, S. A., & Koch, I. (2016). Dissociating language-switch costs from cue-switch costs in bilingual language switching. Bilingualism: Language and Cognition, *19*(5), 921-927.)

Heisenberg, W. (1927-03-01). "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik". *Zeitschrift für Physik*. 43 (3): 172–198.

Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch?. Current Directions in Psychological Science, 10(5), 164-168.

Hickmann, M., Veneziano, E., & Jisa, H. (Eds.). (2018). Sources of variation in first language acquisition: Languages, contexts and learners. John Benjamins.

Hofweber, J., Marinis, T., & Treffers-Daller, J. (2016). Effects of dense code-switching on executive control. *Linguistic approaches to bilingualism*, *6*(5), 648-668.

Hull, R., & Vaid, J. (2007). Bilingual language lateralization: A meta-analytic tale of two hemispheres. *Neuropsychologia*, *45*(9), 1987-2008.

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

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, 21(1), 60-99.

Kałamała, P., Szewczyk, J., Chuderski, A., Senderecka, M., & Wodniecka, Z. (2020). Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis. *Cognition*, *204*, 104373.

Keijzer, M. 2007. *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada*. PhD dissertation, Vrije Universiteit Amsterdam.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. Psychological bulletin, 136(5), 849.

Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. Language and Speech, 32(1), 1-23.

Kirk, N., Declerck, M., Kemp, R., & Kempe, V. (2022). Language control in regional dialect speakers – monolingual by name, bilingual by nature? *Bilingualism: Language and Cognition*, 25(3), 511-520. doi:10.1017/S1366728921000973

Klein, D., Mok, K., Chen, J. K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: a cortical thickness study of bilingual and monolingual individuals. *Brain and language*, *131*, 20-24.

Klimova, B., Valis, M., & Kuca, K. (2017). Cognitive decline in normal aging and its prevention: a review on non-pharmacological lifestyle strategies. *Clinical interventions in aging*, 903-910.

Köpke, B., & Keijzer, M. (2019). Introduction to psycholinguistic and neurolinguistic approaches to language attrition., in M. S. Schmid, and B. Köpke (eds), The Oxford Handbook of Language Attrition, Oxford Handbooks

Köpke, B., & Schmid, M. S. (2004). Language attrition: The next phase. *First language attrition: Interdisciplinary perspectives on methodological issues*, *28*(1).

Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop-and Simon-like stimuli. Psychological research, 56(3), 130-135.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility--a model and taxonomy. *Psychological review*, 97(2), 253.

Krabbe, D., Ellbin, S., Nilsson, M., Jonsdottir, I. H., & Samuelsson, H. (2017). Executive function and attention in patients with stress-related exhaustion: perceived fatigue and effect of distraction. *Stress*, *20*(4), 333-340.

Krakauer, J. W., & Shadmehr, R. (2006). Consolidation of motor memory. *Trends in neurosciences*, *29*(1), 58-64.

Kroll, J. F., & Ma, F. (2017). The bilingual lexicon. The handbook of psycholinguistics, 294-319.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilinguals memory representations. Journal of Memory and Language, 33, 149–174.

Kroll, J. F., & Sunderman, G. (2003). Cognitive processes in second language learners and bilinguals: The development of lexical and conceptual representations. The handbook of second language acquisition, 104129.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. Bilingualism: Language and cognition, 9(2), 119-135.

Lai, C.L., Lau, Z., Lui, S.S., Lok, E.Y., Tam, V., Chan, Q., Cheng, K., Lam, S.M., & Cheung, E.F. (2017). Meta-analysis of neuropsychological measures of executive functioning in children and adolescents with high-functioning autism spectrum disorder. *Autism Research, 10*.

Lai, G., & O'Brien, B. A. (2020). Examining language switching and cognitive control through the adaptive control hypothesis. *Frontiers in Psychology*, *11*, 1171.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., De Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. Psychological Bulletin, 144(4), 394.

Lenneberg, E. H. (1967). The biological foundations of language. Hospital Practice, 2(12), 59-67.

Levelt, P. (1989). Language production.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. Psychological Science, 20(12), 1507-1515.

Lövdén, M., Ghisletta, P., & Lindenberger, U. (2005). Social participation attenuates decline in perceptual speed in old and very old age. *Psychology and aging*, *20*(3), 423.

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2012). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and cognitive processes*, *27*(10), 1479-1488.

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.

Mack, M. A. (1983). Psycholinguistic Consequences of Early Bilingualism: A Comparative Study of the Performance of English Monolinguals and French-English Bilinguals in Phonetics, Syntax and Semantics Experments (Doctoral dissertation, Brown University).

MacWhinney, B., & Kroll, J. F. (2005). A unified model of language acquisition. Handbook of bilingualism: Psycholinguistic approaches, 4967, 50-70.

Mägiste, E. (1979). The competing language systems of the multilingual: A developmental study of decoding and encoding processes. Journal of Verbal Learning and Verbal Behavior, 18(1), 79-89.

Mägiste, E. (1985). Development of intra-and interlingual interference in bilinguals. Journal of Psycholinguistic Research, 14(2), 137-154.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. Proceedings of the National Academy of Sciences, 97(8), 4398-4403.

Mansens, D., Deeg, D. J. H., & Comijs, H. C. (2018). The association between singing and/or playing a musical instrument and cognitive functions in older adults. *Aging & Mental Health*, *22*(8), 970-977.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. Bilingualism: language and cognition, 11(1), 81-93.

Massa, E., Köpke, B., & Yagoubi, R.E. (2020). Age-related effect on language control and executive control in bilingual and monolingual speakers: Behavioral and electrophysiological evidence. *Neuropsychologia, 138*.

Mateos-Aparicio, P., & Rodríguez-Moreno, A. (2019). The impact of studying brain plasticity. *Frontiers in cellular neuroscience*, *13*, 66.

Meara, P. (2004). Modelling vocabulary loss. Applied linguistics, 25(2), 137-155.

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Structural plasticity in the bilingual brain. Nature, 431(7010), 757-757.

Mella, N., Fagot, D., & De Ribaupierre, A. (2016). Dispersion in cognitive functioning: Age differences over the lifespan. Journal of Clinical and Experimental Neuropsychology, 38, 111–126.

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

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

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cognitive psychology, 41(1), 49-100.

Mora, J. C. (2008). Learning context effects on the acquisition of a second language phonology. *A portrait of the young in the new multilingual Spain*, 241-263.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. Psychological science, 22(11), 1425-1433.

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. Developmental science, 10(6), 719-726.

Murman, D. L. (2015, August). The impact of age on cognition. In *Seminars in hearing* (Vol. 36, No. 03, pp. 111-121). Thieme Medical Publishers.

Ooi, S. H., Goh, W. D., Sorace, A., & Bak, T. H. (2018). From Bilingualism to Bilingualisms: Bilingual experience in Edinburgh and Singapore affects attentional control differently. *Bilingualism: Language and Cognition*, *21*(4), 867-879.

Paap, K. (2019). The bilingual advantage debate: Quantity and quality of the evidence. *The handbook of the neuroscience of multilingualism*, 701-735.

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.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265-278.

Paradis, M. (2004). A neurolinguistic theory of bilingualism (Vol. 18). John Benjamins Publishing.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. Language attrition: Theoretical perspectives, 33, 121-133.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. The bilingual mental lexicon: Interdisciplinary approaches, 70.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., & Fazio, F. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human brain mapping*, *19*(3), 170-182.

Pfenninger, S., & Kliesch, M. (2023). Variability as a functional marker of second language development in older adult learners. *Studies in Second Language Acquisition*, 1-27

Pierce, L. J., Klein, D., Chen, J. K., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. Proceedings of the National Academy of Sciences, 111(48), 17314-17319.

Poarch, G. J., & Van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. Journal of experimental child psychology, 113(4), 535-551.

Poeppel, D. (2016). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Understanding Cognitive Development*, 34-55.

Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience*, 32(41), 14125-14131.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied linguistics*, 15(1), 36-57.

Prasse, P., Reich, D. R., Makowski, S., Ahn, S., Scheffer, T., & Jäger, L. A. (2023, May). SP-EyeGAN: Generating Synthetic Eye Movement Data with Generative Adversarial Networks. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications* (pp. 1-9).

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.

R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2022-01-01).

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

Ramos, S., García, Y. F., Antón, E., Casaponsa, A., & Duñabeitia, J. A. (2017). Does learning a language in the elderly enhance switching ability?. Journal of Neurolinguistics, 43, 39-48.

Reich, D. R., Prasse, P., Tschirner, C., Haller, P., Goldhammer, F., & Jäger, L. A. (2022, June). Inferring native and non-native human reading comprehension and subjective text difficulty from scanpaths in reading. In *2022 Symposium on Eye Tracking Research and Applications* (pp. 1-8).

Reuter, M., Ott, U., Vaitl, D., & Hennig, J. (2007). Impaired executive control is associated with a variation in the promoter region of the tryptophan hydroxylase 2 gene. *Journal of Cognitive Neuroscience*, *19*(3), 401-408.

Ringbom, H., & Jarvis, S. (2009). The importance of cross-linguistic similarity in foreign language learning. The handbook of language teaching, 106-118.

Rodriguez-Fornells, A., Krämer, U. M., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology*, *2*, 388.

Rohwedder, S., & Willis, R. J. (2010). Mental retirement. *Journal of Economic Perspectives*, *24*(1), 119-138.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition.

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological psychology*, *54*(1-3), 35-54.

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International neuropsychological Society*, *16*(5), 754-760.

Samuel, S., Roehr-Brackin, K., Pak, H., & Kim, H. (2018). Cultural effects rather than a bilingual advantage in cognition: A review and an empirical study. *Cognitive Science*, *42*(7), 2313-2341.

Schaie, K. W. (1994). The course of adult intellectual development. *American psychologist*, *49*(4), 304.

Schmid, H. J. (2017). A framework for understanding linguistic entrenchment and its psychological foundations.

Schmid, M. S. (2011). Language attrition. Cambridge University Press.

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, 17(4), 729-748.

Schmid, M. S., & Köpke, B. (2009). L1 attrition and the mental lexicon. The bilingual mental lexicon: Interdisciplinary approaches, 209-238.

Schmid, M. S., & Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. Linguistic Approaches to Bilingualism, 7(6), 637-667.

Schmid, M. S., & Yilmaz, G. (2021). Lexical Access in L1 Attrition—Competition versus Frequency: A Comparison of Turkish and Moroccan Attriters in the Netherlands. Applied Linguistics, 42(5), 878-904.

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

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

Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. Oxford university press.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. Linguistic approaches to bilingualism, 1(1), 1-33.

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.

Statistical Consulting Group, UCLA. (n.d.). Introduction to SAS. UCLA: Statistical Consulting Group. Retrieved January 24th, 2024 from https://stats.oarc.ucla.edu/sas/modules/introduction-to-the-features-of-sas/

Steinhauer, K., & Kasparian, K. (2020). Brain Plasticity in Adulthood—ERP Evidence for L1-attrition in Lexicon and Morphosyntax After Predominant L2 Use. Language Learning, 70(S2), 171-193.

Steinkrauss, R., & Schmid, M. S. (2017). Entrenchment and language attrition.

Stine-Morrow, E. A., Payne, B. R., Roberts, B. W., Kramer, A. F., Morrow, D. G., Payne, L., ... & Parisi, J. M. (2014). Training versus engagement as paths to cognitive enrichment with aging. *Psychology and Aging*, *29*(4), 891.

Strobach, T., Frensch, P.A., & Schubert, T. (2012). Video game practice optimises executive control skills in dual-task and task switching situations. *Acta psychologica, 140 1*, 13-24 .

Teper, R., & Inzlicht, M. (2013). Meditation, mindfulness and executive control: the importance of emotional acceptance and brain-based performance monitoring. *Social cognitive and affective neuroscience, 8 1*, 85-92.

The jamovi project (2022). *jamovi*. (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org.

Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. Intelligence, 49, 110–128.

Van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., ... & Lim, S. (2019). Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? A systematic review. Behavioral Sciences, 9(3), 27.

Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of meta-analyses. Neuroscience & Biobehavioral Reviews, 26(7), 849-857.

Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition*, 14(3), 412-422.

Yilmaz, G., & Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. The Mental Lexicon, 7(3), 249-274.

Zuelsdorff, M. L., Koscik, R. L., Okonkwo, O. C., Peppard, P. E., Hermann, B. P., Sager, M. A., ... & Engelman, C. D. (2019). Social support and verbal interaction are differentially associated with cognitive function in midlife and older age. *Aging, Neuropsychology, and Cognition*, *26*(2), 144-160.

Appendices

Appendix 1: Model Summaries

Appendix 1.A: The Development of Bilingual Lexical Access

The models of this study explore different aspects of the dual-language PNT. The first model uses only single-language blocks and excludes Session 2 to allow for a clean comparison between bi- and monolinguals. The second and third models focus on switch- and mix costs, respectively, and use specific subsets to facilitate the interpretation of effects on the costs by avoiding nested contrasts. Each analysis employed *Participant|Session* as a random factor, which allows for the participants' performance to vary over session, and *Item|Language*, which lets the random effect of the item vary across languages. All random factors significantly contributed to the model fit in each of the analyses. Every factor was coded as simple contrast and the coefficients *frequency* and *percentage of excluded trials* were scaled as centred.

Model for bilingual and monolingual single-task naming latencies
Code: (lm, REML) RT ~ 1+ GroupLanguage*Session + frequency + percentage of excluded trials + (1+Session|participant) + (1+GroupLanguage|item),
data = subset of non-switch trials, excluding Session 2

Model for bilinguals' switch costs in Chinese and English
Code: (lm, REML) RT ~ trial_type*language*language_balance + trial_type*Session frequency + (1+language|item) + (1+Session|participant),
data = subset of bilinguals' mixed-task trials

Model for bilinguals' mix costs in Chinese and English
Code: (lm, REML) RT ~ task_type*language*language_balance + task_type*Session + task_type*frequency + task_type*percentage of excluded trials + (1+language|item) + (1+Session|participant), data = subset of bilingual non-switch trials

To illustrate the interplay of switch cost asymmetry, mix cost asymmetry and language balance, we included an analysis which takes language balance as dependent variable and session, switch- and mix cost asymmetry as predictors, and participant ID as a random factor. The model is as follows:

Appendix 1: Model Summaries

Model for bilinguals' language balance

Code: (lm, REML) LanguageBalance ~ Session + SC-ASYM + MC-ASYM,
data = summary of PNT outcomes per participant by Session

Appendix 1.B: Language Control Measures Predict Executive Control Performance

The models of this study explore different aspects of the two tasks assessing executive control, the Flanker Task and the Simon Task. We split both the Flanker and the Simon datasets in two ways: the initial subset encompassed both bilingual and monolingual data, enabling comparisons between these groups. We introduced the variable "GroupSession" to combine session and group information. This facilitated comparisons of bilingual performance across three sessions with the outcomes of monolingual controls. The models we build on these subsets were as follows:

Model for Monolinguals vs. Bilinguals: Flanker

Code: RT ~ GroupSession * Condition + GroupSession * Congruency + Condition * Congruency + (1|participant), data = all Flanker data.

Model for Monolinguals vs. Bilinguals: Simon

Code: RT ~ GroupSession* condition* congruency + (1|participant),
data = all Simon data.

The second subset excluded monolingual data to focus on the bilinguals. This approach allowed for a clean examination of the impact of the "Session" factor and additional predictor variables, derived from the results of the picture naming task. The models we build on these subsets were as follows:

Language Control and Executive Control: Flanker

Code: RT ~ Session * Condition + Condition * Congruency + LanguageBalance + ChineseSwitchCost + EnglishSwitchCost + (1|participant),
data = all bilingual flanker data.

Language Control and Executive Control: Simon

Code: RT ~ Session + LanguageBalance + Condition * Congruency +Condition * MC-Asymmetry + Congruency * SC-Asymmetry + (1|participant),
data = all bilingual Simon data.

All factors were coded as simple contrasts. For each analysis, we employed *1|Participant* as a random factor, which allows for the participants' performance to vary.

Appendix 1: Model Summaries

Appendix 1.C: Effects of Language Use

For the three models in Chapter 5, the different activity components (C1-C7) were entered into the models as coefficients, which were scaled as "centred". All factors were coded as simple contrasts unless otherwise stated below the individual models.

Model for dual-language PNT

Code: (lmer) RT ~ 1 + C1 – 2L leisure + C2 – 2L academic + C3 – 2L switching + C5 – L2 social + C6 – L1 social +C7 – L2 academic + Task Type + IELTS + language + session + Trial Type + Task Type:C1 – 2L leisure + Task Type:C3 – 2L switching + Trial Type:C3 – 2L switching + language: Task Type + language: Trial Type + language: C3 – 2L switching + language: C2 – 2L academic + C5 – L2 social: language+ (1|item)+(1|Participant)

The factor "Condition" was coded as Helmert contrast to facilitate the interpretation of the effect and interaction effects.

Model for Simon Task

Code: (lmer) RT ~ 1 + C3 – 2L switching + C5 – L2 social + C7 – L2 academic + Condition + Session + Congruency + TaskType:C5 – L2 social + Congruency:C7 – L2 academic + TaskType: Congruency + (1 | Participant)

The factor "Congruency" was coded as Helmert contrast to facilitate the interpretation of the effect and interaction effects.

Model for Flanker Task

Code: (lmer) RT ~ 1 + C1 – 2L leisure + C2 – 2L academic + C3 – 2L switching + C5 – L2 social + C6 – L1 social + C7 – L2 academic + Session + Task Type + Congruency + Task Type: Congruency + C7 – L2 academic: Task Type + C5 – L2 social: Task Type + C2 – 2L academic: Congruency + Congruency:C3 – 2L switching + Congruency:C5 – L2 social + Congruency:C7 – L2 academic + (1|Participant)

The factors "Congruency" and "Condition" were coded as Helmert contrasts to facilitate the interpretation of the effects and interaction effects.

Appendix 2: PNT Data Audit

To better understand whether, and if so, to what extent, the data obtained through the dual language PNT is skewed or contains anomalies that may impact the data's accuracy or reliability, we conducted a data audit. To this end, we present analyses on data exclusion rates, training effects and word frequency distributions.

Appendix 2.A: Data Exclusion Rates

We excluded a relatively large percentage of trials from our final dataset. Overall, 24.1% of all bilingual trials, and 10.4% of all monolingual trials were excluded. The reasons for exclusion include inaccurate responses (9.4% across the entire dataset), "skipped" trials, where no response was given by the participant (10.1%), and reaction times falling beyond the minimum or maximum outlier cut-off points (7.7%). Table 27 shows a model for the validity of trials.

Formula: (glm) Validity ~ Language*Tasktype + Language*Tasktype + Session * Trialtype + frequency, data = summary of PNT outcomes per participant by Session

Table 27: Model for validity of trials.

Fixed effects	Estimate[25]	Std. error	P-Value
(Intercept)	1.8	0.2	<0.001
Language Chinese	1.5	0.1	<0.001
Trial type switch	-0.2	0.1	0.03
Task type mixed	0.9	0.1	<0.001
Session 2	-0.9	0.2	<0.001
Session 3	0.1	0.2	0.6
Frequency	0.3	0.1	<0.001
Trial type switch: Language Chinese	-0.7	0.2	0.002
Session 2: trial type switch	-0.5	0.2	0.008
Session 3: trial type switch	-0.4	0.2	0.06
Task type mixed: Language Chinese	-1.1	0.3	<0.001

An analysis of data validity revealed a main effect of language (<0.001), with Chinese trials less often excluded than English trials (β=1.5, SE=0.1, p<0.001), Session (<0.001), as a significantly larger number of trials were excluded in Session 2 (β=-0.9, SE=0.2, p<0.001) with respect to Session 1, but not in Session 3. Significantly more

[25] Valid responses were coded "1" and invalid responses "0". With a binary outcome variable, the choice analysis is a generalised linear mixed model (GLM). GLM estimates show log odds as opposed to the percentage change of 100%.

trials were excluded in the switch trial type as opposed to the non-switch trial type (β=-0.2, SE=0.1, $p<0.03$), but trials in the mixed-language task were less often excluded than those within single-language tasks (β=0.9, SE=0.1, $p<0.03$).

Significant interaction effects between trial type and language and task type and language reveal two things: first, that switching into Chinese caused more invalid trials than if the trials had been non-switch, while validity in English remained largely undeterred ($p_{bonferroni} = 1$). Second, that English validity benefitted more from the mixed-language task type as opposed to Chinese ($p_{bonferroni} = 0.5$). A significant interaction effect between Session and trial type further amplifies difficulties of language switching in the second Session.

Language balance also had a small but significant effect on validity but was removed from the final model to allow its convergence.

Appendix 2.B: Training Effects

To understand whether the mixed-language facilitation effect is due to a confounding effect of order within the experimental blocks, we built a linear mixed-effects model including information on which order a specific trial had within its block. To this effect, we reintroduced the first trial of each block and created a variable called "Order-in-Block" which assigns the ordinal value of each trial per block.

Formula: (lm, REML) RT ~ Order-in-Block*Block + Frequency*Block + Session + Language, data = subset of bilingual data including first trials

Table 28: Model to investigate training effects.

Fixed effects	Estimate	Std. error	P-Value
(Intercept)	1475.96	54.078	<.001
Language: Chinese	-126.12	40.139	0.002
Session 2	-79.56	26.868	0.008
Session 3	-104.33	29.016	0.001
Block: English single	237.48	56.084	<.001
Block: CH-EN mixed	-4.17	35.504	0.907
Frequency	-46.9	10.267	<.001
Order-in-Block	1.49	0.381	<.001
Block: English single*frequency	-36.38	22.576	0.107
Block: CH-EN mixed*frequency	-53.05	19.971	0.008
Block: English single*Order-in-Block	3.19	0.973	0.001
Block: CH-EN mixed*Order-in-Block	1.67	0.839	0.046

The model revealed that Order-in-Block had a highly significant effect on RT. Other than a training effect, however, succeeding trials turned out to become significantly slower overall ($\beta=1.5$, $SE=0.4$, $p<0.01$). Further, the significant interaction effect of Block and Order-in-Block revealed that the effect of Order-in-Block is different across blocks. Figure 51 visualises these differences. In the Chinese single block, RT stays constant throughout. In contrast, English trials slow down fastest ($\beta=3.2$, $SE=1$, $p<0.01$) and succeeding trials in the mixed block less so ($\beta=1.7$ $SE=0.8$, $p<0.05$).

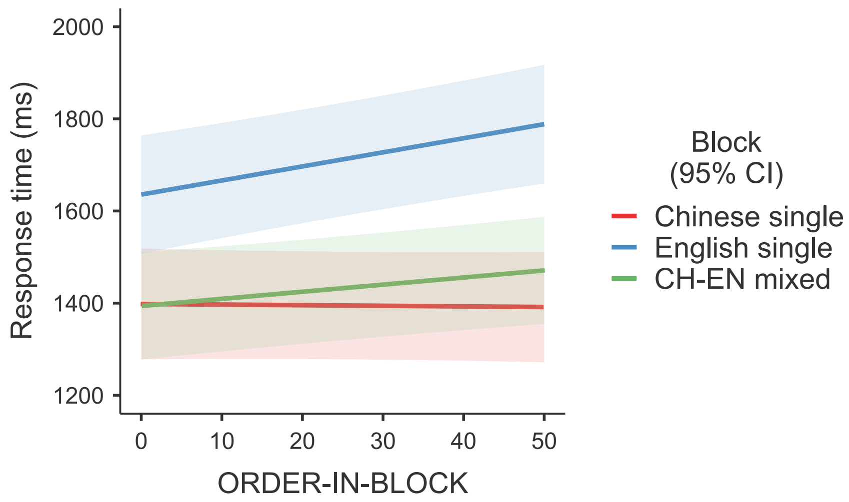


Figure 51: Order-in-block effects on response time by block.
Shaded areas reflect 95% confidence intervals.

An interpretation of these results could be that there is no training effect and instead participants tire in the English and mixed block as trials progress. Crucially, the mixed block has significantly lower RTs even though it follows the English block in which trials generally slowed down over time.

Appendix 2.C: Word Frequency Distributions

As previously mentioned in 2.2.1. (Dual-Language) Picture Naming Task, while the order of individual words on lists was randomized across experiments, the same lists were employed for the individual blocks for each of the three sessions. Put differently: there was a dedicated list for each of the three blocks in each of the three sessions. For this reason, it was decided to investigate whether these lists have comparable word frequency distributions. Figure 52 shows the word frequency distributions by block and by session.

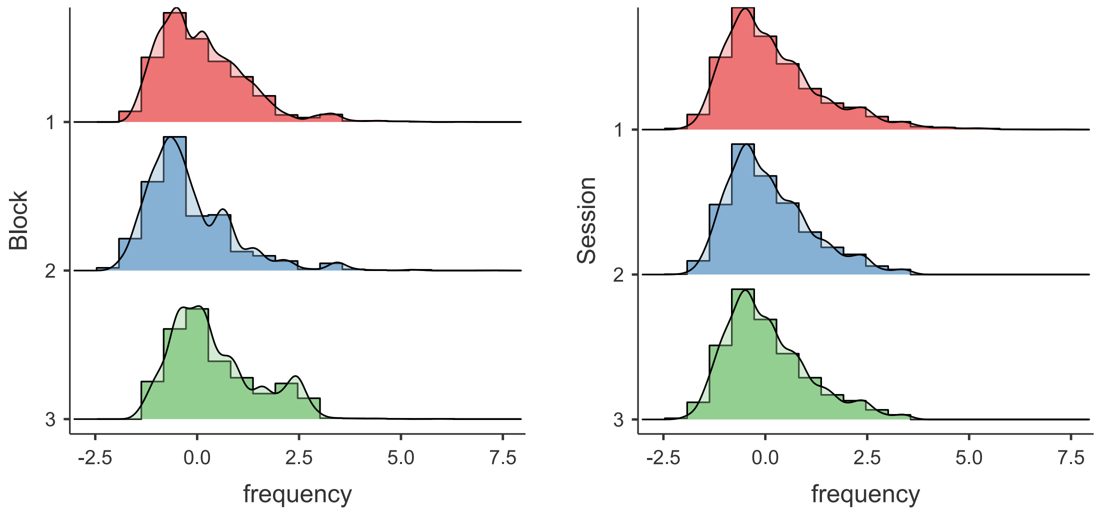


Figure 52: Relative word frequency distributions across blocks and sessions

An investigation of relative word frequency of stimuli by block revealed that there are significant discrepancies in the stimuli's word frequency distributions. Table 29 shows the results of a linear mixed model analysis using word frequency as dependent variable. The linear mixed model for frequency as dependent variable reveals that both session, block, and the interaction effect of block*session are highly significant. This suggests that word frequency of the stimuli was not uniformly distributed across the data set and instead varied greatly both between blocks and sessions. The mixed block had – on average – the highest word frequencies, while the English block (the second) had the lowest. Figure 53 visualises these findings. Crucially, this finding may provide some insight as to why our study – unlike most others in the field – has found a mix benefit as opposed to a mix cost.

Formula: (lm, REML) Frequency ~ 1 + Block*Session + (1|OBJID) data = subset of all bilingual trials including first trials

Table 29: Model for word frequency distributions of stimuli.

Fixed effects	Estimate	Std. error	P-Value
(Intercept)	0.0874	0.04732	0.065
English Block	-0.3301	0.11621	0.005
Mixed Block	0.3028	0.11532	0.009
Session 2	-0.1364	0.01124	<.001
Session 3	-0.1277	0.00990	<.001
English Block*Session 2	0.0459	0.02875	0.111
Mixed Block*Session 2	0.1198	0.02474	<.001
English Block*Session 3	0.0415	0.02505	0.098
Mixed Block*Session 3	0.1134	0.02230	<.001

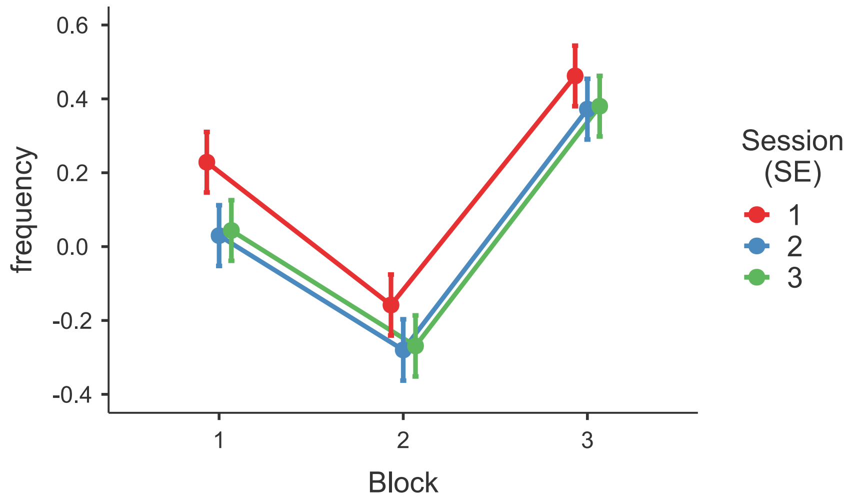
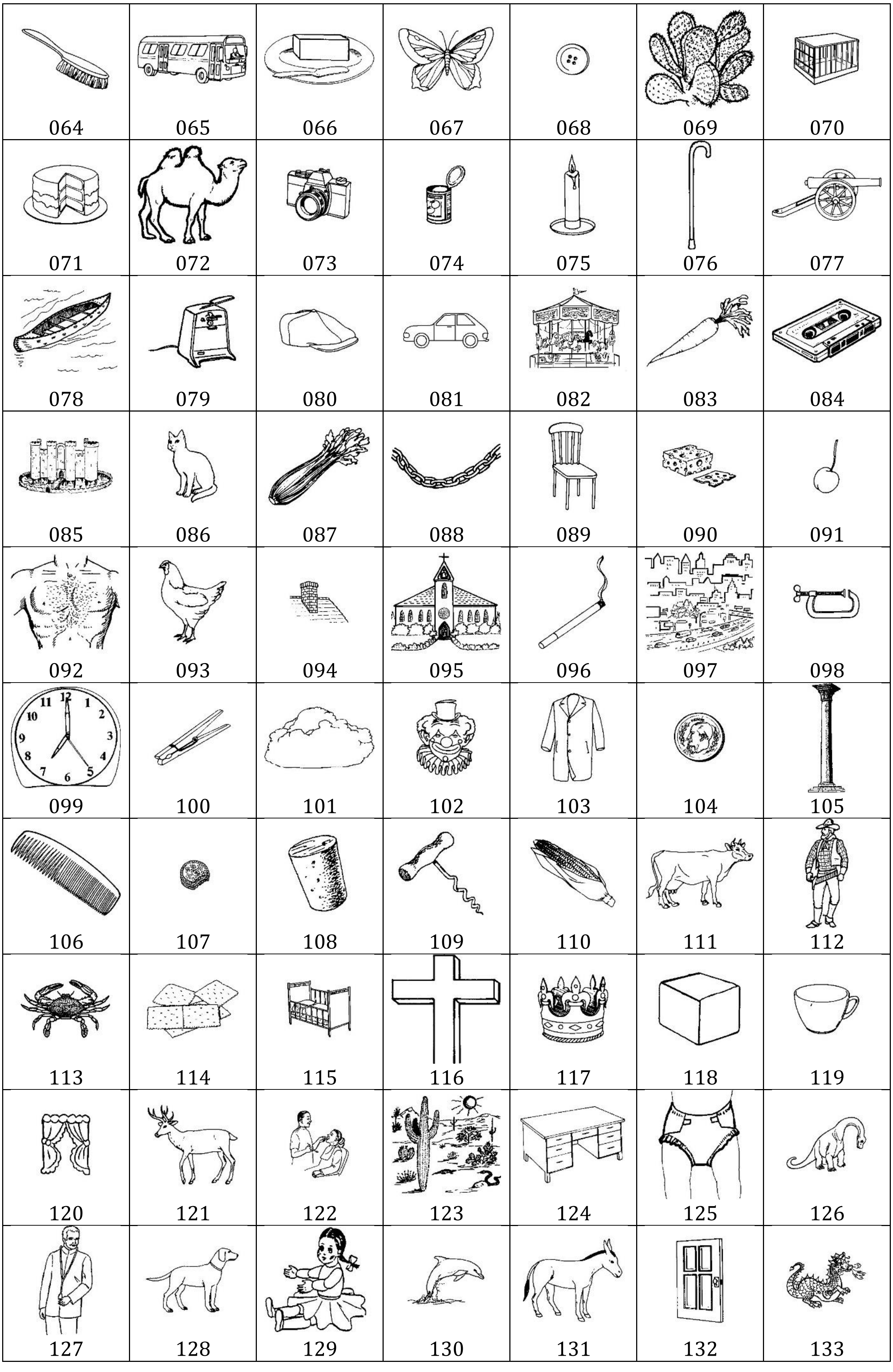


Figure 53: Relative word freuquencies by block by session.

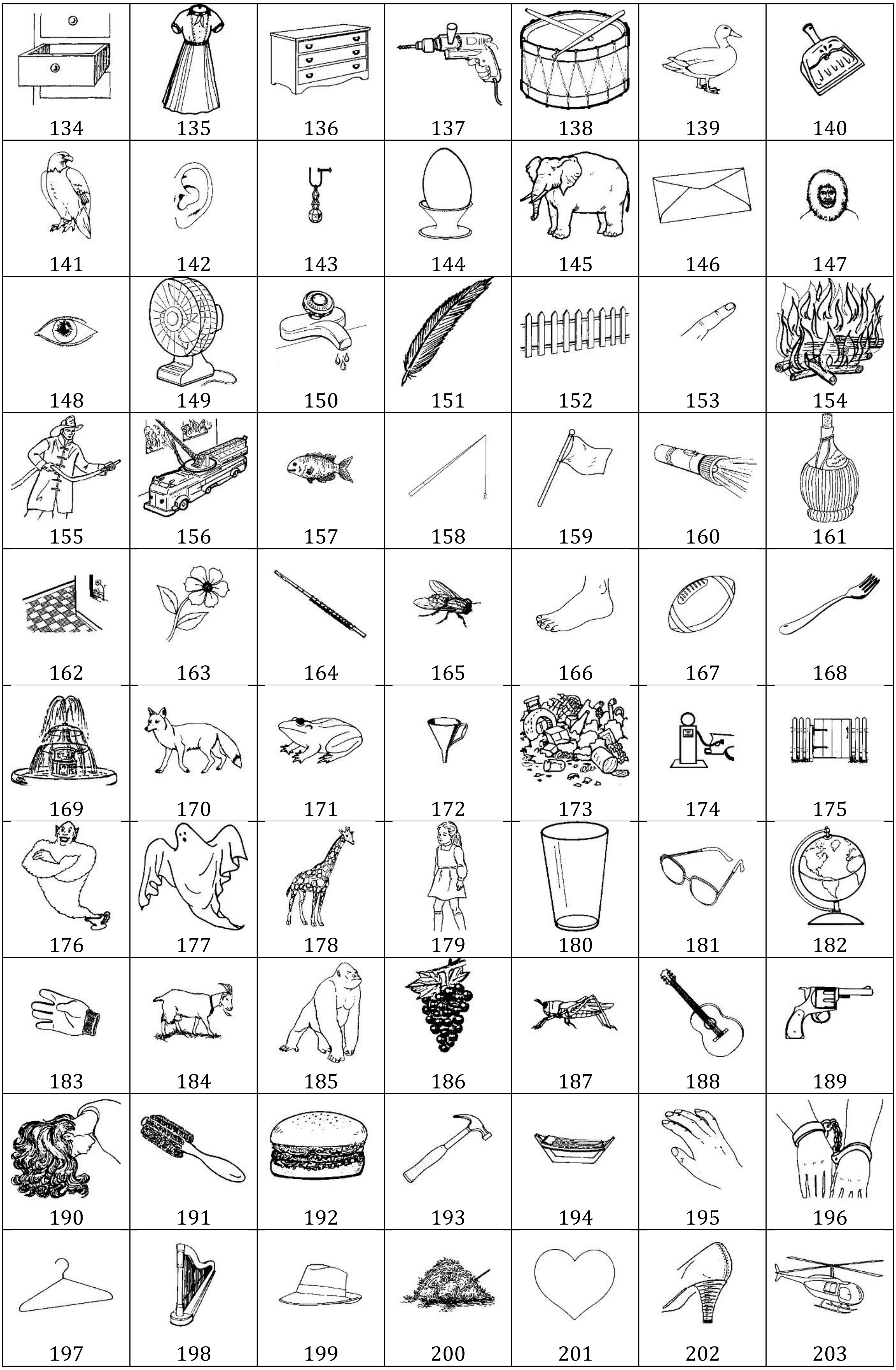
Error bars reflect Standard Errors.

Appendix 3: Pictures Used in the PNT

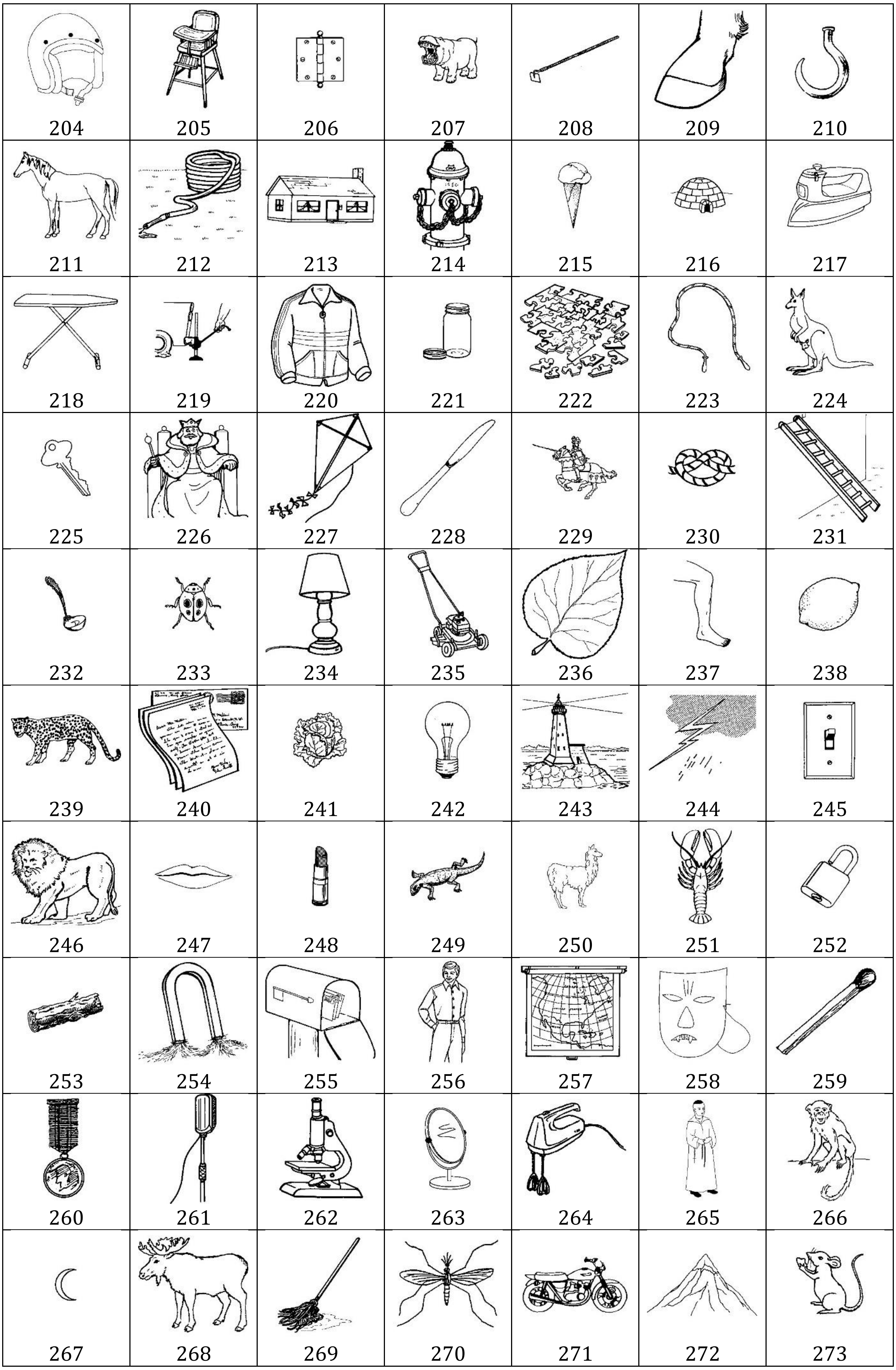
Appendix 3: Pictures Used in the PNT



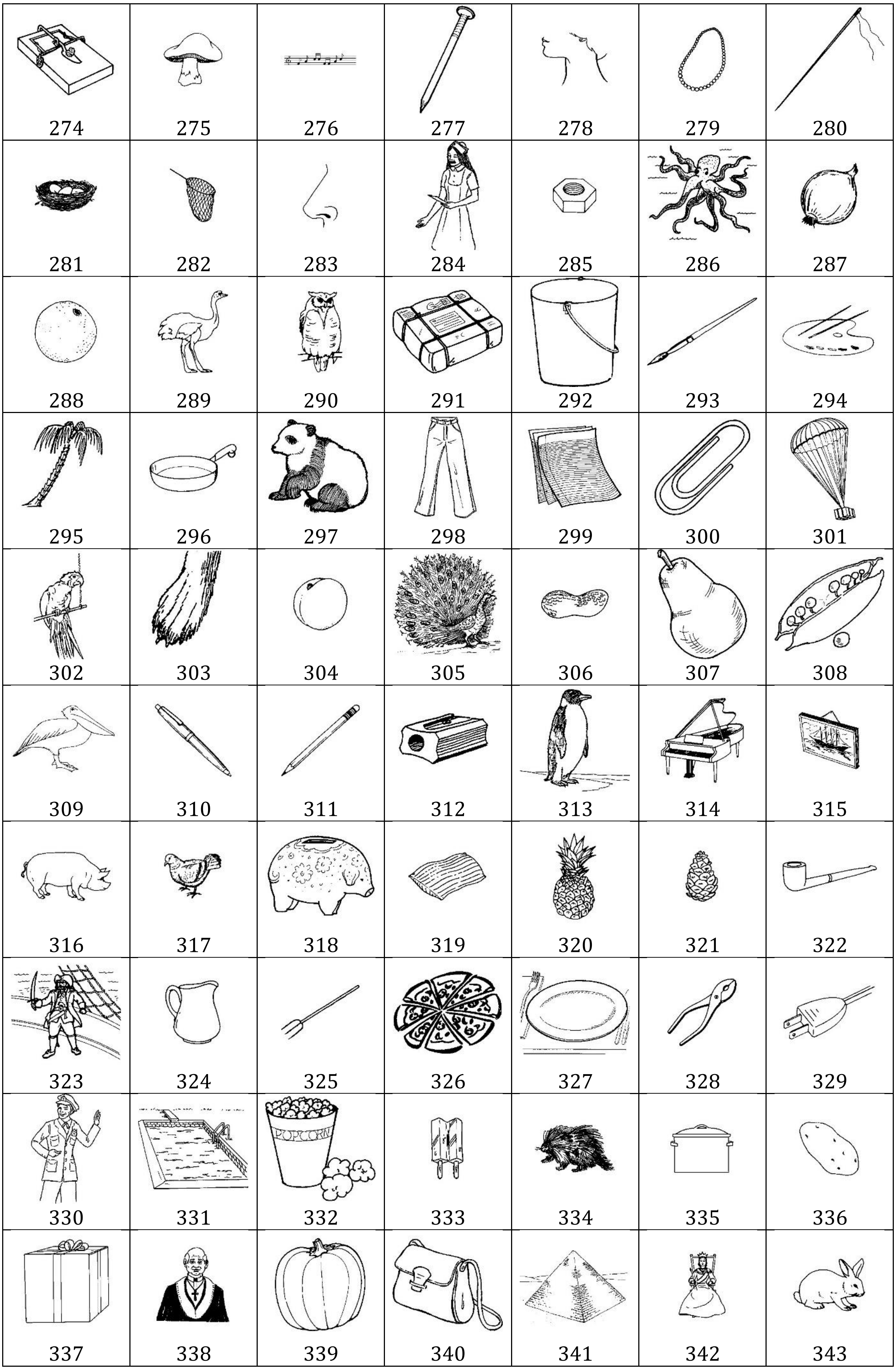
Appendix 3: Pictures Used in the PNT



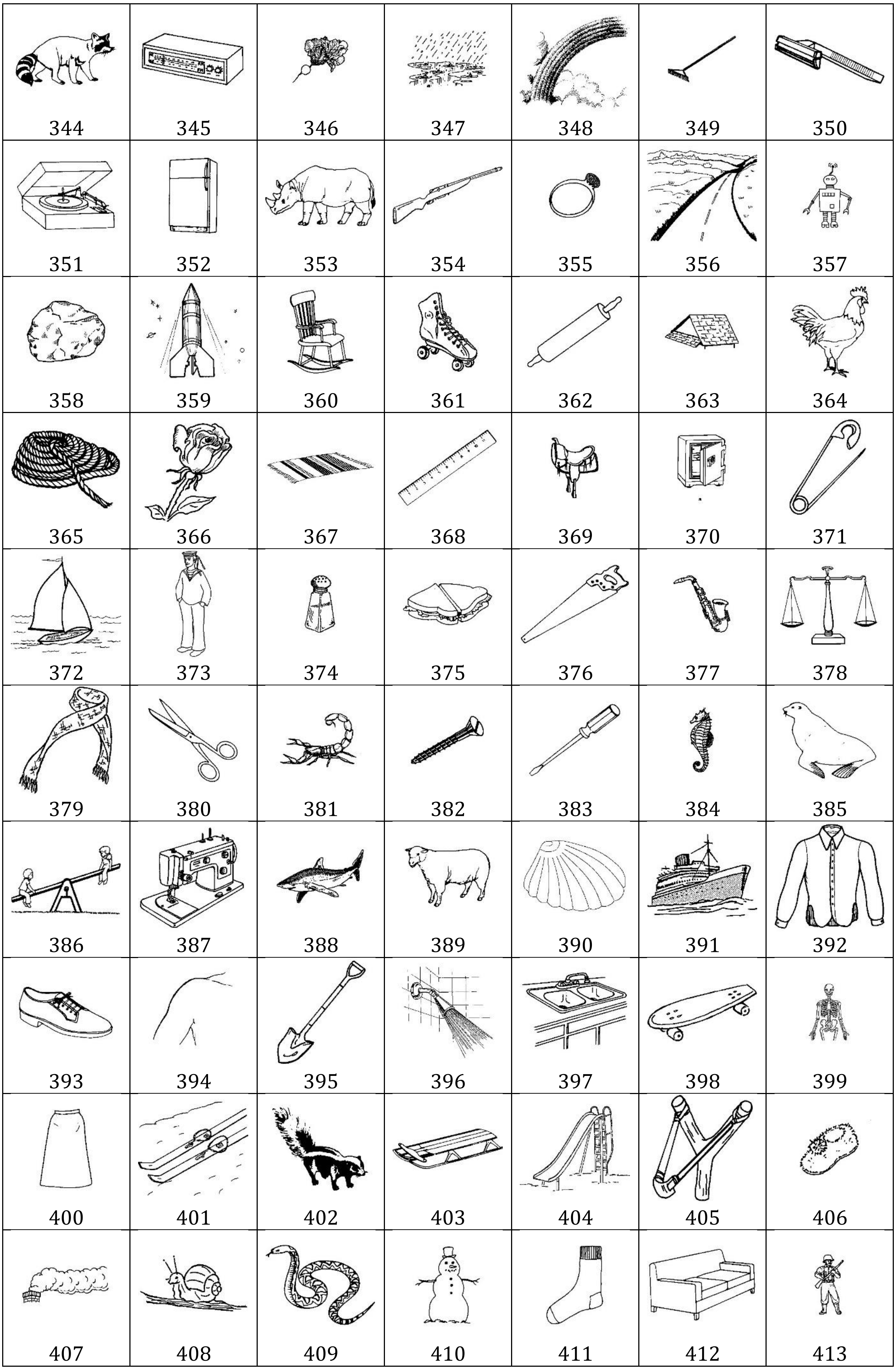
Appendix 3: Pictures Used in the PNT



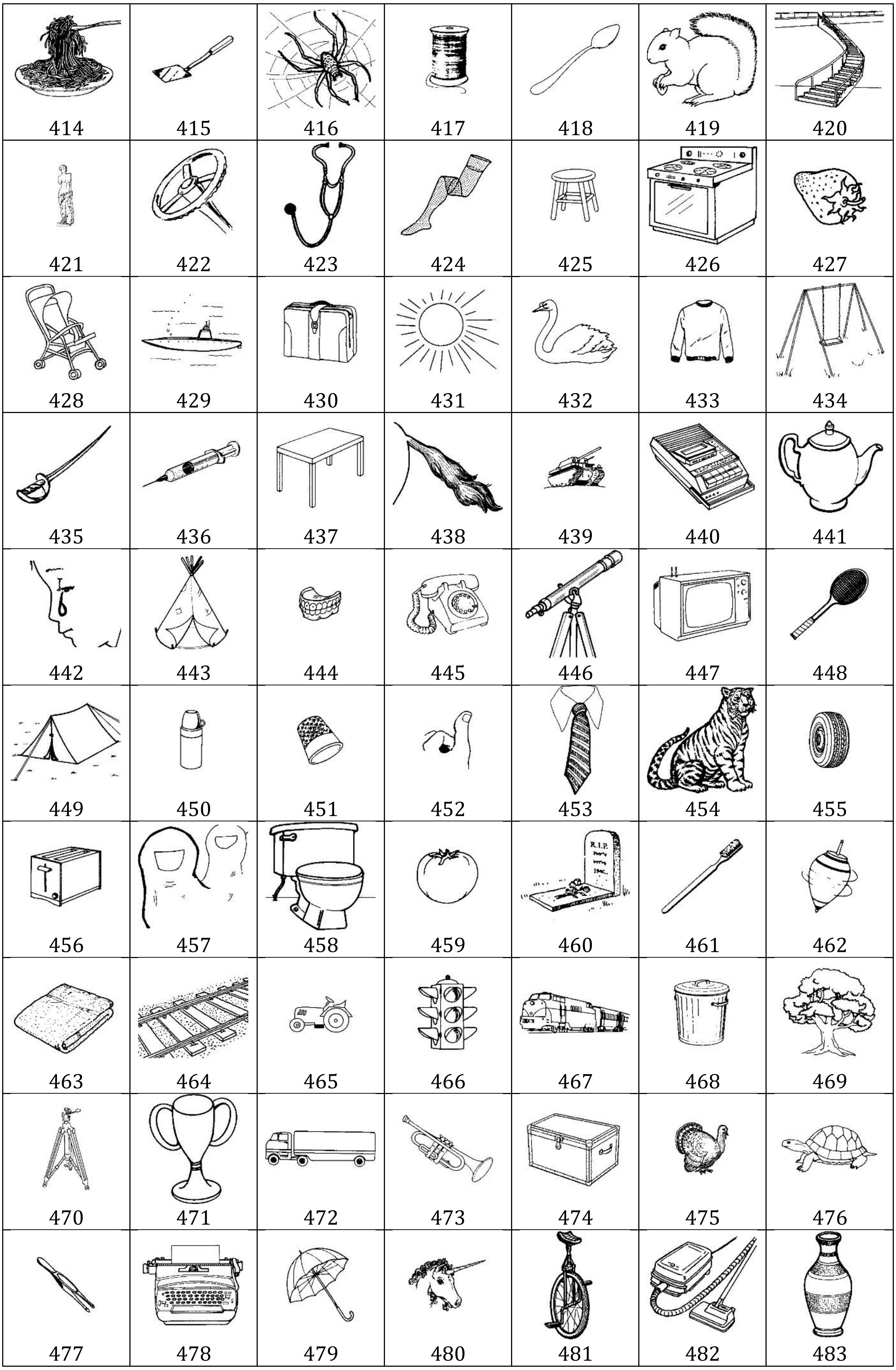
Appendix 3: Pictures Used in the PNT



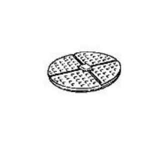
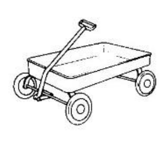
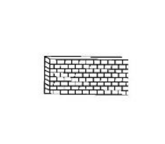
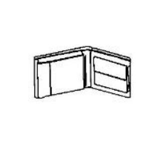
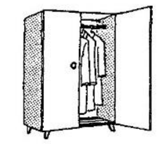
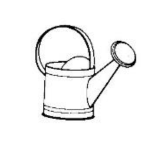
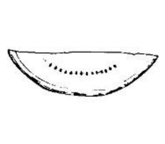
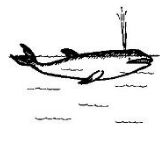
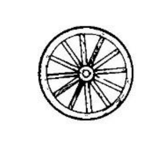
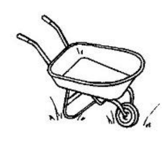
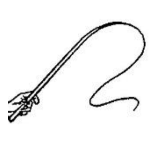
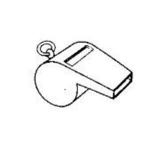
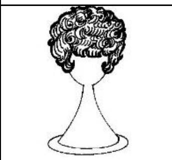
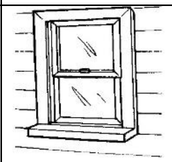
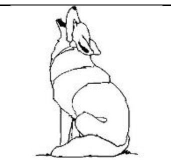
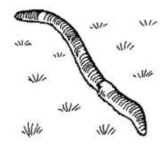
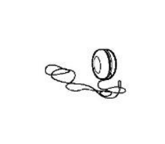
Appendix 3: Pictures Used in the PNT



Appendix 3: Pictures Used in the PNT



Appendix 3: Pictures Used in the PNT

 484	 485	 486	 487	 488	 489	 490
 491	 492	 493	 494	 495	 496	 497
 498	 499	 500	 501	 502	 503	 504
 505	 506	 507	 508	 509	 510	 511
 512	 513	 514	 515	 516	 517	 518
 519	 520					

Appendix 4: Lists Used in the PNT By Block And By Session

Appendix 4: Lists Used in the PNT By Block And By Session

Session 1		
Block 1: Chinese	Block 2: English	Block 3: Mixed
432 swan	055 bow	283 nose
322 pipe	480 unicorn	032 bat
237 leg	299 paper	153 finger
044 binoculars	269 mop	409 snake
337 present	508 wig	178 giraffe
503 wheel	185 gorilla	507 whistle
416 spider	307 pear	314 piano
341 pyramid	139 duck	083 carrot
217 iron	092 chest	054 bowl
004 alligator	459 tomato	467 train
286 octopus	367 rug	180 glass
340 purse	033 bathtub	159 flag
452 thumb	245 light switch	359 rocket
433 sweater	253 log	179 girl
462 top	124 desk	499 spiderweb
169 fountain	256 man	052 boot
347 rain	095 church	331 pool
306 peanut	231 ladder	050 bone
329 plug	190 hair	292 bucket
400 skirt	129 doll	133 dragon
439 tank	010 fish tank	072 camel
505 wheelchair	345 radio	130 dolphin
196 handcuffs	103 coat	313 penguin
300 paperclip	420 stairs	045 bird
069 cactus	327 plate	068 button
404 slide	240 letter	408 snail
117 crown	373 sailor	356 road
093 chicken	035 beard	244 lightning
302 parrot	066 butter	057 boy
342 queen	402 skunk	318 piggybank
094 chimney	490 bricks	170 fox
510 window	485 violin	120 curtains
078 canoe	091 cherry	353 rhinoceros
135 dress	435 sword	348 rainbow
272 mountain	441 teapot	126 dinosaur
315 picture	291 package	202 heel
220 jacket	218 ironing board	486 volcano
421 statue	379 scarf	387 sewing machine
085 castle	426 stove	276 music
249 lizard	039 bug	165 fly
096 cigarette	330 policeman	495 washing machine
423 stethoscope	422 steering wheel	390 shell
513 witch	394 shoulder	442 tear
388 shark	136 dresser	061 bride
193 hammer	007 antlers	413 soldier
381 scorpion	403 sled	006 ant
021 badge	504 wheelbarrow	364 rooster
506 whip	141 eagle	386 seesaw
493 walrus	460 grave	407 smoke
212 hose	055 bow	406 slipper

Appendix 4: Lists Used in the PNT By Block And By Session

Session 2					
Block 1: Chinese		Block 2: English		Block 3: Mixed	
476	turtle	195	hand	479	umbrella
425	stool	380	scissors	275	mushroom
089	chair	332	popcorn	056	box
411	sock	375	sandwich	163	flower
242	lightbulb	355	ring	181	glasses
157	fish	267	moon	138	drum
189	gun	009	apple	065	bus
444	teeth	228	knife	099	clock
473	trumpet	349	rake	060	bread
030	barrel	186	grapes	496	watch
031	basket	376	saw	116	cross
247	lips	296	pan	455	tire
236	leaf	396	shower	365	rope
447	tv	339	pumpkin	075	candle
449	tent	088	chain	020	backpack
211	horse	005	anchor	145	elephant
017	baby	216	igloo	352	refrigerator
498	watermelon	252	lock	257	map
446	telescope	374	salt	437	table
197	hanger	397	sink	119	cup
053	bottle	463	towel	316	pig
071	cake	500	well	224	kangaroo
258	mask	440	tape recorder	024	ball
111	cow	029	barbecue	210	hook
042	bench	369	saddle	484	vest
378	scale	241	lettuce	204	helmet
213	house	351	record player	271	motorcycle
239	leopard	278	neck	064	brush
059	bra	377	saxophone	448	tennis racket
434	swing	122	dentist	134	drawer
290	owl	127	doctor	282	net
427	strawberry	209	hoof	058	branch
509	windmill	288	orange	051	book
131	donkey	362	rolling pin	067	butterfly
454	tiger	385	seal	171	frog
284	nurse	082	carousel	445	telephone
070	cage	475	turkey	166	foot
034	bear	235	lawnmower	453	tie
259	match	382	screw	478	typewriter
087	celery	383	screwdriver	012	arrow
016	ax	254	magnet	357	robot
172	funnel	274	mousetrap	248	lipstick
366	rose	200	hay	026	banana
429	submarine	451	thimble	458	toilet
177	ghost	424	stocking	183	glove
281	nest	112	cowboy	222	puzzle
456	toaster	251	lobster	360	rocking chair
144	egg	023	balcony	182	globe
152	fence	325	pitchfork	113	crab
077	cannon	195	hand	419	squirrel

Appendix 4: Lists Used in the PNT By Block And By Session

Session 3					
Block 1: Chinese		Block 2: English		Block 3: Mixed	
395	shovel	305	peacock	142	ear
184	goat	401	skis	148	eye
114	crackers	377	saxophone	025	balloon
151	feather	344	raccoon	128	dog
384	seahorse	309	pelican	311	pencil
414	spaghetti	362	rolling pin	037	bed
358	rock	207	hippo	106	comb
277	nail	268	moose	132	door
040	bell	443	tepee	201	heart
501	whale	488	wagon	168	fork
483	vase	260	medal	167	football
363	roof	101	cloud	073	camera
062	bridge	205	highchair	043	bicycle
155	fireman	457	toe	393	shoe
319	pillow	262	microscope	343	rabbit
121	deer	176	genie	081	car
295	palm tree	336	potato	310	pen
398	skateboard	465	tractor	431	sun
515	woman	002	acorn	086	cat
471	trophy	370	safe	418	spoon
047	wood	304	peach	368	ruler
512	wing	324	pitcher	203	helicopter
022	bag	371	safety pin	266	monkey
430	suitcase	492	walnut	146	envelope
076	cane	334	porcupine	469	tree
187	grasshopper	285	nut	227	kite
123	desert	137	drill	102	clown
230	knot	229	knight	461	toothbrush
294	paint	161	wine	246	lion
350	razor	477	tweezers	041	belt
048	boat	517	wrench	399	skeleton
049	bomb	118	block	279	necklace
154	fire	303	paw	063	broom
232	ladle	208	hoe	234	lamp
097	city	206	hinge	110	corn
293	paintbrush	108	cork	361	roller skate
038	bee	264	mixer	519	zebra
198	harp	105	pillar	149	fan
115	crib	194	hammock	320	pineapple
438	tail	491	wallet	263	mirror
405	slingshot	250	llama	084	tape
338	priest	015	asparagus	226	king
308	peas	036	beaver	410	snowman
389	sheep	079	can opener	011	arm
323	pirate	270	mosquito	074	can
417	thread	301	parachute	273	mouse
289	ostrich	280	needle	221	jar
297	panda	415	spatula	516	worm
481	unicycle	140	dustpan	243	lighthouse
287	onion	305	peacock	392	shirt

Appendix 5: Full catalogue of questions contained in the Activity Log

Activitiy log

Instructions: When you fill in this questionnaire, please think of your use of English and Chinese ***over the past month***. Please indicate on average how many hours per week you have spent doing the following activities. Please give a percentage of how much of this was done in English, in Chinese or in another language.

		language of activity		
	hours per week (average)	% English	% Chinese	other (please specify)
Part A: Study				
class preparation				
class attendance				
out-of-class contact with teacher (e.g. office hours)				
preparing for exams				
working on assignments: background reading				
working on assignments: mathematical/statistical work				
working on assignments: writing				
working on assignments: other				
improving your English: vocabulary				
improving your English: grammar				
improving your English: pronunciation				
Work (please specify nature of job):				
time spent on job				
Part B: Social activities				
time spent in the company of family/friends/colleagues outside of study/work				
time spent interacting with family/friends/colleagues **who are in the UK** on social media or on the phone				
time spent interacting with family/friends/colleagues **who are in China** on social media or on the phone				
time spent interacting with family/friends/colleagues **who are in another country** on social media or on the phone				

Appendix 5: Full catalogue of questions contained in the Activity Log

Other activities				
	hours per week (average)	% English	% Chinese	other (please specify)
watching TV or films				
reading online				
reading books or newspapers				
emailing				
writing your diary				
other writing (e.g. poetry, blogs, essays other than for your study)				
attending a religious service, praying				
sports, working out (please specify: ____)				

Part C: Code-switching	Never	Rarely	Sometimes	Often
When you speak Chinese with friends or family, do you ever use English words or sentences?				
When you use Chinese on social media, do you ever use English words or sentences?				
When you write emails in Chinese, do you ever use English words or sentences?				
When you speak English with friends or family, do you ever use Chinese words or sentences?				
When you use English on social media, do you ever use Chinese words or sentences?				
When you write emails in English, do you ever use Chinese words or sentences?				

Part D: other information	
Have you been back to China since the last time you filled in this questionnaire?	yes/no
If yes, for how long?	weeks
What proportion of your friends in the UK are	
a) Chinese	
b) English	
c) other (please specify)	

Appendix 6: Rotated Component Matrix of the Principal Component Analysis

Appendix 6: Rotated Component Matrix of the Principal Component Analysis

	Component						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Chinese reading online	0.787						
Chinese watching TV or films	0.744					0.334	
English reading books or newspapers	0.743						
English watching TV or films	0.725						
English reading online	0.714			0.389			
Chinese time spent in the company of family/friends/colleagues outside of study/work	0.671					0.553	
English time spent interacting with family/friends/colleagues who are in the UK on social media or on the phone	0.610						
Chinese reading books or newspapers	0.600		0.447				
Chinese time spent interacting with family/friends/colleagues who are in China on social media or on the phone	0.581	0.301				0.322	
English emailing	0.432						
English preparing for exams		0.859					
Chinese preparing for exams		0.800					
Chinese working on assignments: background reading		0.659				0.363	
English class attendance		0.595					0.426
Chinese class attendance		0.546					
Chinese sports, working out (please specify: ___)		0.482		0.303		0.359	
English out-of-class contact with teacher (e.g. office hour consultation)		0.413		-0.378			
When you use English on social media, do you ever use Chinese words or sentences?			0.770				
When you write emails in Chinese, do you ever use English words or sentences?			0.737				
When you use Chinese on social media, do you ever use English words or sentences?			0.665		-0.367		
When you speak English with friends or family, do you ever use Chinese words or sentences?			0.541				
When you write emails in English, do you ever use Chinese words or sentences?			0.525				
When you speak Chinese with friends or family, do you ever use English words or sentences?			0.497				

Appendix 6: Rotated Component Matrix of the Principal Component Analysis

	Component						
	1	2	3	4	5	6	7
Chinese out-of-class contact with teacher (e.g. office hour consultation)				0.866			
Improving English Vocabulary				0.773			
Chinese working on assignments: other				0.746		0.356	
Proportion friends Chinese					-0.754		
Proportion friends other					0.676		
Chinese working on assignments: writing				0.311	0.581		
English class preparation		0.381			0.547		0.312
Have you been back to China since you came to the UK/since the last time you filled in this questionnaire?					-0.430		
Proportion friends English					0.344		
Chinese time spent interacting with family/friends/colleagues who are in the UK on social media or on the phone						0.647	
English sports, working out						0.627	
Improving English Pronunciation			0.301		0.367	0.507	
English time spent interacting with family/friends/colleagues who are in China on social media or on the phone	0.444					-0.455	0.311
Chinese emailing			0.328			0.380	
English working on assignments: writing							0.782
English working on assignments: background reading							0.640
Chinese class preparation		0.465					-0.479
English time spent in the company of family/friends/colleagues outside of study/work							0.441
English working on assignments: other		0.377		0.307		0.370	0.426