Emotional influences on cognitive control: a missing factor in research on 'bilingual advantages'?



Anna K. Niedbala

Kellogg College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2018

Acknowledgements

I want to thank from the bottom of my heart the people who made it possible for me to complete this work through their unwavering professional and moral support. To list you all would take many many pages, so I will extend here instead a broad thank you to everyone I had the pleasure to work with in the Departments of Education and Experimental Psychology and to all my fellow Kellogg College Cornflakes and OUAC teammates.

I'd especially like to thank my research participants for volunteering their time in the name of science. Thank you to Draško Kašćelan who arranged participants and a testing space for me in Cambridge, and to Maren Stöber who did the same at LSE. Thank you to the students and staff at the schools who agreed to participate for going our of your way to help. And an especially emphatic thank you to Kedi Simpson for helping me out at the last minute when other plans fell through!

This research would not have been possible without the knowledge and guidance of my supervisors. Thank you to Victoria Murphy and Gaia Scerif for helping me become an independent researcher.

Thank you to the wonderful people who proofread this thesis and gave me invaluable feedback: Mae Zantout, Emma Abotsi, Annina Hessel, Michelle Van Tieghem, Martin Strohmeier, and Anette Niedbala.

Thanks to the lovely staff at Kellogg College and the Department of Education for your support, especially Kyra Smith, Shabaz Hussein, and Sarah O'Brien.

And to all the people who worked and trained and drank pints and debated life's big questions with me along the way: My coaches and teammates, especially Lesley, Kevin, Emma, Dan, Christina, Jenny, Heather, Maria, Sam, Andrew, Rowan, and Hugo. My office-mates Emma, Henriette, Annina, and Jaimie (thank you for the study breaks, snacks, advice, and humour!). All my family and friends: I'm blessed that there are far too many of you to name. Still I'd like to especially thank Matt, Ricky, Rachel, Claire, Ian, Gareth, Dan, Richard, Bruno, Flo, Tammo, Julia, Max, Sarah, Simon, Agne, Stephen, Claudia, Divya, Lavinia, Michelle, Sophie, Kat, Sam, Hannes, and of course Mom and Dad.

And most importantly, thank you to Martin, for being there through everything and for making it possible to even complete this DPhil. I could not have done any of it without you.

Funding

This research was made possible by a scholarship and bursaries from Kellogg College. Funding for participant incentives was also provided by the Department of Experimental Psychology.

Abstract

Over half of the world's population is multilingual with globalisation and migration ensuring a steady growth. This fact has inspired several flourishing research fields within the areas of emotion, cognition and linguistics. Yet, existing research on cognitive consequences of bilingualism has not taken into account the effect of emotion on this relationship. This dissertation addresses this gap. Based on the hypothesis that bilingualism leads to adaptations in the mechanisms underlying both emotion and cognitive control, we propose a framework that integrates theoretical and empirical perspectives from emotion, cognition, and linguistics research. To validate our approach, we devise three novel studies in order to test cognition–emotion interactions across different groups of monolinguals and bilinguals.

In Study 1, 24 English monolingual and 24 German–English bilingual adults completed a flanker-go/no-go (FNG) task. The participants viewed sets of three faces side-by-side in the centre of a computer screen and were instructed to press a button if the gender of the central neutral target face corresponded to the 'go' gender, which occurred in 66% of trials. The task-irrelevant 'flanker' faces expressed either positive, negative, or neutral emotion and were furthermore either gender-congruent or -incongruent to the target. In Study 2, 28 diverse bi- and multilingual adults completed the FNG task alongside 28 monolinguals. In addition, 29 monolinguals and 29 bilinguals completed a forced-choice flanker-only (FO) variant of the task. Study 3 repeated Study 2 with 10–11 year-old monolingual and English–French/German bilingual children. 36 bilinguals and 20 monolinguals completed the FNG task, while 35 bilinguals and 23 monolinguals completed the FO task.

Differences between the monolingual and bilingual groups were observed across all studies in the FNG task, but not in the FO task. Notably, the bilingual children also made significantly fewer errors in no-go trials across all conditions than their monolingual counterparts, providing novel evidence for a bilingual advantage in response inhibition. In Study 2, the multilinguals evidenced different response strategies than the other groups, suggesting that adaptations in cognitive control can vary based on the type of multilingual experience.

We interpret these findings based on an integrative theoretical framework of cognitive control, emotion, and bilingualism, and discuss their implications for research across all three fields.

Contents

List of Figures **xi**

List of Tables **xiii**

List of Abbreviations **xv**

1 Introduction **1**

1.1 Introduction 1

1.2 Contributions 3

1.3 Definitions and Scope 4

1.3.1 Bilingualism and Multilingualism 4

1.3.2 Cognitive Control 5

1.3.3 Emotion and Affect 5

2 Literature Review **7**

2.1 Introduction 7

2.2 Emotion–Language Interactions and Bilingualism 10

2.2.1 The Emotional Impact of Non-Native Languages 10

2.2.2 Effects of Emotion on Language Processing 11

2.2.3 Developmental Links Between Language and Emotion 14

2.2.4 Summary 17

2.3 Language and Cognitive Control 17

2.3.1 Cognitive Consequences of Bilingualism 17

2.3.2 Cognitive Control of Language 18

2.3.3 Types and Levels of 'Cognitive Control' 20

2.3.4 Summary 31

2.4 Cognition–Emotion Interactions 32

2.4.1 Models of Emotion and Cognitive Control 32

2.4.2 Cognition–Emotion Interactions in Development 40

2.4.3 Summary 43

2.5 Conclusion 44

3 Study 1 **47**
3.1 Background . . . 47
3.1.1 Cognitive Control of Emotion . . . 48
3.1.2 Experimental Measures of Emotion–Cognition Interactions . 50
3.2 The Present Study . . . 55
3.2.1 Research Questions . . . 55
3.2.2 Study Design . . . 56
3.2.3 Hypotheses . . . 58
3.3 Methodology . . . 60
3.3.1 Design . . . 60
3.3.2 Participants . . . 61
3.3.3 Materials . . . 62
3.3.4 Procedure . . . 63
3.3.5 Data Analysis Plan . . . 63
3.4 Results . . . 64
3.4.1 Pre-Processing and Stimulus Validation. . . . 64
3.4.2 Main Experiment . . . 66
3.5 Discussion . . . 71
3.6 Conclusion . . . 74

4 Study 2 **77**
4.1 Background . . . 77
4.1.1 Replicating Study 1 with Mixed Bilinguals . . . 78
4.1.2 Comparing the Flanker–Go/No-go and Flanker-only Tasks . 81
4.2 The Present Study . . . 84
4.2.1 Research Questions and Hypotheses . . . 85
4.3 Methods . . . 87
4.3.1 Experiment 1: Flanker–Go/no-go Task with Mixed Bilinguals 87
4.3.2 Experiment 2: Flanker-only task . . . 90
4.3.3 Data Analysis Plan . . . 91
4.4 Results . . . 92
4.4.1 Sample Demographics . . . 92
4.4.2 Experiment 1 . . . 92
4.4.3 Experiment 2 . . . 96
4.5 Discussion . . . 103
4.5.1 Experiment 1 Results . . . 103
4.5.2 Experiment 2 Results . . . 105
4.6 Conclusion . . . 108

5 Study 3 **113**
5.1 Background . . . 113
5.1.1 Maturation of Inhibition and Emotion Regulation Mechanisms 115
5.1.2 Bilingual Advantages in Childhood . . . 117
5.2 The Present Study . . . 118
5.2.1 Research Questions and Hypotheses . . . 118
5.3 Methods . . . 120
5.3.1 Experiment 1: Flanker–go/no-go task . . . 120
5.3.2 Experiment 2: Flanker-only task . . . 122
5.3.3 Data Analysis Plan . . . 122
5.4 Results . . . 123
5.4.1 Sample Demographics . . . 123
5.4.2 Experiment 1 . . . 123
5.4.3 Experiment 2 . . . 126
5.5 Discussion . . . 128
5.5.1 Experiment 1 Results . . . 128
5.5.2 Experiment 2 Results . . . 133
5.6 Conclusion . . . 135

6 General Discussion **137**
6.1 Summary and Comparison of Study Results . . . 138
6.2 Limitations . . . 141
6.3 Future Directions . . . 143
6.3.1 Bilingualism Research . . . 143
6.3.2 Cognitive Control Research . . . 144
6.3.3 Emotion and Emotion Regulation Research . . . 145
6.4 Concluding Remarks . . . 145

Appendices

A Data Analysis Details **149**
A.1 Model Selection and Hypothesis Testing . . . 149
A.1.1 Model Selection Process . . . 150
A.1.2 Data Distribution and Conversions . . . 153
A.1.3 Data Analysis Scripts . . . 153
A.1.4 Model Equations . . . 157
A.1.5 Planned Contrasts and *Post-Hoc* Tests . . . 163
A.2 Sample Size and Power Calculations . . . 167
A.2.1 *A-Priori* Sample Size Calculation . . . 167
A.2.2 *Post-hoc* Power Calculation . . . 168

A.3 Full Model Estimates 173
A.3.1 Study 1 173
A.3.2 Study 2 175
A.3.3 Study 3 178

B Supplementary Data Analysis **181**
B.1 Face Colour Contrast Analysis 181
B.1.1 Study 1 182
B.1.2 Study 2 182
B.1.3 Study 3 184
B.2 Additional Model Selection and Hypothesis Testing Scripts 185
B.2.1 Study 2 186
B.2.2 Study 3 198

References **205**

List of Figures

2.1 Executive Functions Model 22
2.2 Executive Functions Detail 23
2.3 Attention+EF Model 26
2.4 Cognitive Control Model 29
2.5 Cognitive Control Model with Emotion 31
2.6 Model of Cognitive and Emotion Control 39

3.1 Study 1 Hypotheses. 59
3.2 Study 1 Task. 60
3.3 Study 1 RTs. 66
3.4 Study 1 RT Flanker Effect. 68
3.5 Study 1 D-Prime Interactions. 69
3.6 Study 1 d' Contrasts. 70

4.1 Study 2 Task. 88
4.2 Study 2 Experiment 1 RTs. 94
4.3 Study 2 Experiment 1 RT Flanker Effect. 95
4.4 Study 2 Experiment 1 d'. 96
4.5 Study 2 Experiment 1 d' Contrasts. 97
4.6 Study 2 Experiment 2 RTs. 98
4.7 Study 2 Experiment 2 RT Flanker Effect. 99
4.8 Study 2 Experiment 2 Accuracy. 100
4.9 Study 2 Experiment 2 Accuracy Contrasts. 101
4.10 Study 2 Experiment 2 conflict adaptation effect. 102

5.1 Study 3 Task. 120
5.2 Study 3 Experiment 1 RTs. 125
5.3 Study 3 Experiment 1 RT Flanker Effect. 126
5.4 Study 3 Experiment 1 No-go Accuracy. 127
5.5 Study 3 Experiment 1 Go Accuracy. 128
5.6 Study 3 Experiment 1 False Alarm Rates. 130
5.7 Study 3 Experiment 1 Miss Rates. 131
5.8 Study 3 Experiment 2 RTs. 132

5.9 Study 3 Experiment 2 RT Flanker Effect. 133
5.10 Study 3 Experiment 2 RTs. 134
5.11 Study 3 Experiment 2 Accuracy. 135

A.1 Study 1 regression equation for selected model 158
A.2 Study 2.1 regression equation for selected LME model 158
A.3 Regression equation for d' accuracy models 159
A.4 Study 2.2 regression equation for selected RT model. 159
A.5 Study 2.2 regression equation for selected GLME model. 160
A.6 Study 3.1 regression equation for selected RT model. 160
A.7 Study 3.2 regression equation for selected RT model. 161
A.8 Study 3.1 regression equation for accuracy GLME model. 162
A.9 Study 3.2 regression equation for accuracy GLME model. 162
A.10 Study 1 *a priori* calculation. 167
A.11 Study 1 *post-hoc* calculation. 172

List of Tables

3.1 Study 1 Design. 61
3.2 Study 1 Mean Gender I.D. Accuracy. 65
3.3 Study 1 Mean Valence Ratings. 65
3.4 Study 1 Condition Means. 67
3.5 Study 1 RT ANOVA Table. 67
3.6 Study 1 d' ANOVA 70

4.1 S2 Participant Language Backgrounds 89
4.2 Study 2 Experiment 1 Means. 92
4.3 Study 2 Experiment 1 RT ANOVA. 93
4.4 Study 2 Experiment 1 d' ANOVA 95
4.5 Study 2 Experiment 2 Means. 96
4.6 Study 2 Experiment 2 ANOVA Table 98
4.7 Study 2 Experiment 2 Accuracy ANOVA Table 100

5.1 Study 3 Experiment 1 Descriptive Statistics. 124
5.2 Study 3 Experiment 1 RT ANOVA 124
5.3 Study 3 Experiment 1 Accuracy ANOVA 129
5.4 Study 3 Experiment 2 Descriptive Statistics. 129
5.5 Study 3 Experiment 2 RT ANOVA 130
5.6 Study 3 Experiment 2 Accuracy ANOVA 131

A.1 Study 1 fixed effects design matrix. 152
A.2 Study 1 fixed effects design matrix for planned contrasts 152
A.3 Study 1 fixed effects LRT. 157
A.4 Study 1 fixed effects design matrix with custom contrasts. 163
A.5 Study 1 predicted RTs. 168
A.6 Study 1 RT model estimates. 173
A.7 Study 1 $d\prime$ model estimates. 174
A.8 Study 2 Exp. 1 RT model estimates. 175
A.9 Study 2 Exp. 1 $d\prime$ model estimates. 175
A.10 Study 2 Exp. 2 RT model estimates. 176
A.11 Study 2 Exp. 2 accuracy model estimates. 176

A.12 Study 2 Exp. 2 conflict adaptation model 177
A.13 Study 3 Exp. 1 RT model estimates. 178
A.14 Study 3 Exp. 1 accuracy model estimates. 179
A.15 Study 3 Exp. 2 RT model estimates. 179
A.16 Study 3 Exp. 2 accuracy model estimates. 180

B.1 Type–III ANOVA table for the Study 1 RT model including a continuous covariate for face contrast. 183
B.2 Type–III ANOVA table for the Study 2.1 RT model including a continuous covariate for face contrast. 183
B.3 Type–III ANOVA table for the Study 2.2 RT model including a continuous covariate for face contrast. 184
B.4 Type–III ANOVA table for the Study 3.1 RT model including a continuous covariate for face contrast. 185
B.5 Type–III ANOVA table for the Study 3.2 RT model including a continuous covariate for face contrast. 185

List of Abbreviations

ACC . . . . . Anterior cingulate cortex.

ACH . . . . . Adaptive control hypothesis (Green & Abutalebi, 2013).

ANOVA . . . . Analysis of Variance.

AoA . . . . . Age of acquisition.

BOLD . . . . . Blood-oxygenation-level-dependent; a measure used in fMRI studies.

EC/EF . . . . Executive control / executive function.

EEG . . . . . Electroencephalogram.

ERP . . . . . Event-related potential. (N indicates a negative potential, P a positive potential, and the following numbers the elapsed time since the event, *e.g.* $N200$.)

fMRI . . . . . Functional magnetic resonance imaging.

FNG . . . . . Flanker–Go/No-go task.

FO . . . . . Flanker-only task.

(G)LME . . . (Generalised) linear mixed effects model.

IC . . . . . Inhibitory control.

L1, L2 . . . . . First/native language, second language.

PFC . . . . . Prefrontal cortex. (Lower-case prefixes indicate location, e.g. dorsolateral = dlPFC.)

RT . . . . . Reaction time.

WM . . . . . Working memory.

So, if you want to really hurt me, talk badly about my language. Ethnic identity is twin skin to linguistic identity—I am my language. Until I can take pride in my language, I cannot take pride in myself.

— Gloria Anzaldúa,
'How to tame a wild tongue', pp. 39.

1 Introduction

Contents

1.1 Introduction 1
1.2 Contributions 3
1.3 Definitions and Scope 4
1.3.1 Bilingualism and Multilingualism 4
1.3.2 Cognitive Control 5
1.3.3 Emotion and Affect 5

1.1 Introduction

A variation on a classic thought experiment in moral philosophy poses the following problem: Would you push a heavy man off a bridge, into the path of an oncoming train, in order to stop the train and save the lives of five people tied to the tracks? The answer, it turns out, depends on which language you speak: if it is not your mother tongue, you are far more likely to choose the 'utilitarian' approach, killing one to save five (Costa et al., 2014). Similarly, you are more likely to judge taboo situations as acceptable (Geipel et al., 2015), judge the goodness of situations based on outcomes rather than intentions (Geipel et al., 2016), and rely less on heuristic processing to make decisions (Keysar et al., 2012). In short, multilingualism has a demonstrable and compelling impact on the way we perceive and reason about the world.

This is a critical observation because over 50% of the world's population is multilingual. Globalisation and migration ensures that this number is steadily growing. Governments and institutions are increasingly confronted with questions relating to language pluralism: how important is foreign language education, and from what age should it be offered (Myles, 2017)? How can schools support children with different home languages (Granger, 2018; Poarch & Bialystok, 2017)? How should receiving societies integrate refugees (Beirens et al., 2007)?

In addressing these questions, the cognitive implications of multilingualism are an important consideration. Bilinguals have been shown to outperform monolinguals on various executive control tasks, a 'bilingual benefit' which is especially pronounced in children (Bialystok & Viswanathan, 2009). Individual differences in executive functions are also implicated in children's educational attainment (e.g. Bull & Scerif, 2001; Mulder et al., 2017) and social–emotional development (e.g. Duckworth & Seligman, 2005; Rueda et al., 2010) from an early age. This highlights a potential link between linguistic and emotional competence (Beck et al., 2012; Wolfe & Bell, 2007a).

Beyond the classroom, a beneficial effect of bilingualism on executive control ability may have positive implications for society at large. Better executive control is associated with reductions in automatic racial stereotyping (Payne, 2005) and positively correlated with health, financial resources, and pro-social behaviour throughout the lifespan (Moffitt et al., 2011). In older adults, bilingualism is considered a potential protective factor against age-related cognitive decline (Guzmán-Veléz & Tranel, 2015).

In short, a detailed understanding of the relationship between multilingualism and cognition has potentially far-reaching implications for education and social policy. The rationale for such a relationship is well founded: as the pioneering researcher in this area writes,

> ...the bilingual mind must adapt to [a] mental configuration that includes jointly activated languages for communication to proceed. If the [neural] network recruited to manage attention to the two languages is part of a nonverbal attention or selection system, then the configuration is in place for bilingualism to impact the nature or quality of nonverbal cognitive processing and presumably the brain regions on which it is based (Bialystok, 2017, pp. 235).

In other words, based upon Bialystok's hypothesis that cognitive changes related to bilingualism are a form of "experience-dependent neuroplasticity" (2017, pp. 235), sustained experiences in various language contexts lead to adaptations in the brain regions required for those linguistic interactions.

However, while a growing body of research into this "nature and quality of nonverbal cognitive processing" in multilinguals and "the brain regions on which it is based" (Bialystok, 2017, pp. 235) exists, it has largely focused on individual cognitive control functions in isolation. This is not ecologically valid with regards to the contexts in which adaptations are posited to occur, as these involve social and emotional inputs alongside linguistic ones.

Emotion is an integrated and perhaps inextricable part of cognitive processing (Inzlicht et al., 2015; Pessoa, 2008, 2009). It is thus very likely that the compelling language effects on decision making discussed above can be further explained by cognitive control. Nevertheless, little research to date has directly addressed the question of how learning and using multiple languages impacts these cognitive–emotional processes. The aim of this dissertation is to synthesise approaches from linguistics, cognitive psychology, and emotion research in order to shed light on this increasingly relevant question.

1.2 Contributions

A large and controversial body of research links bilingualism to advantages in cognitive tasks ranging from inhibitory control to abstract planning. A recent review of this 'bilingual advantages' research described the current body of evidence as "hazy", marked by methodological inconsistencies (García-Pentón et al., 2016b). However, multilingual and multicultural participants also provide unique insights into the relationships between languages, social contexts, and cognition (Kroll & McClain, 2013). This is evidenced by the emergence of another increasingly popular research area focusing on the role of foreign languages in moral reasoning, decision making, and the perception of emotionality (Dewaele, 2008; Hayakawa et al., 2016; Pavlenko, 2008). Meanwhile, links between emotion and cognition have been well-established at both the neural and behavioural level, with many researchers arguing that the two cannot be characterised as separate processes at all (c.f. Pessoa, 2013).

However, the basic cross-disciplinary work needed to close the gap between these three areas—to integrate bilingualism, cognition, and emotion research—is virtually non-existent. The fields currently form three metaphorical circles of a Venn diagram in which the central triangle, the area where all three overlap, remains unfilled. At worst, the validity of the extant research is jeopardised by uncontrolled confounds arising from this gap. At best, insights from each research area can help bring clarity to unanswered questions in the other two fields.

This dissertation synthesises theoretical and empirical work to begin filling the gap at the centre of this disciplinary Venn diagram. It lays out a framework to represent relationships between emotional, cognitive, and linguistic mechanisms based on foundational theories in each field. Three empirical studies built on this framework are then presented which begin to investigate their interactions. Although grounded in cognitive theories, this research also provides the necessary first step to inform our understanding of multilingual learners, an increasing proportion of students across educational settings.

Chapter 2 reviews the relevant literature to bring the three fields together and introduces the unified framework. It identifies empirical gaps and proposes several avenues for cross-disciplinary research. Chapter 3 reports the results of a study addressing one of these gaps: the contribution of bilingualism to adults' ability to regulate task-irrelevant emotional information. Chapters 4 and 5 report follow-up studies with different populations and an additional experimental paradigm. Chapter 4 describes two experiments comparing monolinguals to a diverse sample of bilingual adults on two measures of executive control with emotional distractors. Chapter 5 reports on the same tasks run with monolingual and English–German/French bilingual schoolchildren aged 10–11 years. Chapter 6 discusses the overall findings of the three studies and their implications, and concludes the dissertation.

1.3 Definitions and Scope

1.3.1 Bilingualism and Multilingualism

As has been argued extensively, no two multilinguals are the same (Kroll & Chiarello, 2016), and thus no single term can characterise all types of language users involved

in this research. However, for the sake of consistency, this dissertation uses the term 'bilingualism' to denote the field of study. 'Bilingual' describes anyone who speaks two languages, whether from birth (*simultaneous bilingual*) or learned later (*sequential bilingual*), regardless of language dominance. First language (L1) denotes the first-acquired language or the self-identified mother tongue, while the second language (L2) refers to any later-learned, less dominant, or less-preferred language. 'Multilingual' is used as a collective term for anyone who speaks two or more languages in any combination. All of these factors influence the questions addressed in this dissertation, and so, while it seeks to draw conclusions about the impact of multilingualism on emotion and cognitive control, these are necessarily limited to the populations studied.

1.3.2 Cognitive Control

The term 'cognitive control' refers in a general sense to complex mental processing and problem solving in response to changing demands (Pessoa, 2008) and the ability to control awareness (Posner & Rothbart, 1998). It deliberately does not specify a particular theoretical framework such as *executive functions* (Miyake et al., 2000) or *executive attention* (Posner & Petersen, 1990) as there is some disagreement about which aspect(s) of cognitive control bilingualism influences (further discussion in Section 2.3). Rather, this literature review presents a model of cognitive control composed of inter-related processes from these frameworks.

1.3.3 Emotion and Affect

The term *emotion* refers to subjective experiences within the two-dimensional space of valence and arousal, where valence is defined as the negative or positive value of the experience and arousal as the intensity of the experienced emotion and associated physiological activation (Larsen & Diener, 1992). These experiences are sometimes labelled with the so-called 'universal emotions' of happiness, anger, fear, sadness, disgust, and surprise. Facial expressions of these six emotions are recognised cross-culturally, and there appear to be equivalent labels for them in many different languages (Ekman, 1992). *Affect* refers more generally to psychological states or properties, such as moods, emotions, and short-lived reactions, that are evaluative in nature, either positive or negative.

Es ist von jeher eine der wichtigsten Aufgaben der Kunst gewesen, eine Nachfrage zu erzeugen, für deren volle Befriedigung die Stunde noch nicht gekommen ist.

One of the Arts' most important tasks has always been to create a need, for whose fulfilment the hour has not yet come.

— Walter Benjamin,
'Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit', pp. 500.

2 Literature Review

Contents

2.1 Introduction 7
2.2 Emotion–Language Interactions and Bilingualism . . . 10
2.2.1 The Emotional Impact of Non-Native Languages 10
2.2.2 Effects of Emotion on Language Processing 11
2.2.3 Developmental Links Between Language and Emotion . 14
2.2.4 Summary 17
2.3 Language and Cognitive Control 17
2.3.1 Cognitive Consequences of Bilingualism 17
2.3.2 Cognitive Control of Language 18
2.3.3 Types and Levels of 'Cognitive Control' 20
2.3.4 Summary 31
2.4 Cognition–Emotion Interactions 32
2.4.1 Models of Emotion and Cognitive Control 32
2.4.2 Cognition–Emotion Interactions in Development 40
2.4.3 Summary 43
2.5 Conclusion 44

2.1 Introduction

Speaking multiple languages has been posited to improve 'cognitive control'—the skills that allow humans to respond to situations, plan, and solve complex problems. Experience-dependent neuroplasticity in the brain regions involved in language control is believed to cause adaptations that make general control skills more

efficient (Bialystok, 2017; Green & Abutalebi, 2013). While a large volume of research supports this hypothesis (see Bialystok, 2017, for a recent review), the findings have been heavily criticised (see e.g. Paap & Greenberg, 2013). Detractors argue that lack of clarity in the constructs being measured, and a resulting 'task impurity', lead to findings of 'bilingual advantages' that are impossible to generalise or replicate. (R. M. Klein, 2016; Paap et al., 2015; Paap et al., 2017; Paap & Sawi, 2014). In response to these criticisms, researchers have called for a stronger emphasis on the cognitive demands of day-to-day multilingual interactions, and how these may lead to measurable neural adaptations in non-linguistic cognition (Bialystok, 2017; Kroll & Bialystok, 2013; Kroll & Chiarello, 2016).

The multilingual experiences thought to drive this 'bilingual adaptation' process are not affectively neutral. In fact, detecting social cues and their emotional implications is an integral part of the leading theory of bi-/multilingual language control (Green & Abutalebi, 2013). Research in Applied Linguistics has shown that using a non-native language can drastically influence both perception of and response to emotional information. When bilinguals were asked to give subjective 'emotionality' ratings, judge fictional situations, or make moral decisions, outcome differences were consistently reported depending on whether a native or non-native language was used (Costa et al., 2014; Hayakawa et al., 2016; Keysar et al., 2012). However, Psycholinguistics studies indicate that processing of emotional words in L2 is comparable to L1 under some task conditions, suggesting that the cognitive demands of a task determine how influential emotional content is on these outcomes. A large body of research in Cognitive and Affective Neuroscience supports this theory: emotional and cognitive processes are closely linked both neurally and behaviourally (Pessoa, 2013). Some researchers go so far as to propose that cognitive control is driven by underlying emotional processes and the two are thus inseparable (Inzlicht et al., 2015). A question not addressed to date is whether day-to-day multilingual experiences and the social/emotional demands that come with them lead to adaptations in these well-established links between cognitive control and emotion.

Experience-dependent co-development of cognitive control, emotion, and language is nothing new theoretically: Developmental models of L1 acquisition, for example, posit that children acquire language through effortful interaction with the

world around them. They perceive, interpret, and integrate the linguistic, social, and emotional information presented to them into their own intentional states and then communicate these states (Bloom & Beckwith, 1989; Bloom et al., 2001). Thus, language, cognition, and emotion covary and develop together throughout childhood.

Nevertheless, outside of developmental science, research to date has either treated the above processes largely independently or focused on the relationship of one to another while leaving the third as a potential confound. Research on the interactions between emotion and cognition, for example, rarely reports the language background of participants, despite the fact that bilinguals comprise the global majority, and despite findings that bilingualism impacts cognitive control. Likewise, studies that investigate the impact of language on emotion processing rarely take cognitive control demands into account, even though these are strongly implicated in second language use.

This dissertation hypothesises that learning and using more than one language is associated with the same complex demands as L1 acquisition on both cognitive control and emotion. The putative increases in control efficiency conferred by bilingual experiences should thus extend to emotion processing as well. Investigating this question will contribute to a more nuanced and ecologically valid understanding of the contentious links between multilingualism and cognitive control. Without considering emotion, our understanding of bilingualism's influences on cognition is incomplete.

In order to address this gap, language, emotion, and cognitive control must be situated within one coherent framework. This chapter reviews significant research at the intersections of these three areas. Section 2.2 discusses the relationship between language and emotion from an applied linguistic (Section 2.2.1), psycholinguistic (Section 2.2.2), and developmental (Section 2.2.3) perspective. Section 2.3 reviews research on language and cognitive control, focusing on extant findings of bilingual advantages (Section 2.3.1), the cognitive control demands underpinning bilingual language use (Section 2.3.2), and the various theoretical frameworks of cognitive control (Section 2.3.3). Section 2.4 summarises the main theories and findings of cognition–emotion interactions as they apply to the experimental paradigms used in bilingual advantages research, with a particular focus on the role of emotional valence (Section 2.4.1).

2.2 Emotion–Language Interactions and Bilingualism

Literature investigating emotion–language interactions falls, broadly, into three categories: explorations of the emotional impact of L2 vs. L1; effects of emotional mood/meaning/content on language processing (both L1 and L2); and developmental links between linguistic, emotional, and social outcomes. This section reviews studies in each area and proposes a preliminary explanation for the role of cognitive control (which remains largely un-addressed) in each.

2.2.1 The Emotional Impact of Non-Native Languages

It is a well-established phenomenon that non-native languages feel subjectively less emotional. For example, according to multilinguals' self-reports, the perceived emotional force of curse words and taboo words is strongest in L1 and decreases in each additional language (Dewaele, 2004). Similarly, 'I love you' was reported to have the most emotional meaning in L1, according to questionnaire participants (Dewaele, 2008).

These subjective differences influence the way speakers interact with each other and communicate emotional content. Psychiatric patients may show reduced symptoms when speaking their second language compared to their first (see Paradis, 2008, for a comprehensive review of second language effects in psychiatric disorders). Furthermore, language switching has proved effective in helping bilingual patients recount traumatic memories (De Zulueta, 1984; Oquendo, 1996). Using a foreign language reduces framing effects and loss aversion in decision making (Keysar et al., 2012). It also increases focus on the outcome of a situation over the intention when judging the moral goodness of a situation (Geipel et al., 2016). Perhaps most compellingly, as described in the introduction, both English and Spanish participants were far more likely to choose the 'utilitarian' approach in the 'trolley problem' to save five people by pushing another person in front of a train if the problem was posed in their L2. Those using their L1 were less likely to want to intervene and push the person because the action was presumed to cause emotional distress (Costa et al., 2014; Hayakawa et al., 2016).

Finally, Hsu et al. (2015) provide neuroimaging evidence for emotionality differences. Their participants read emotional literature passages and rated their emotion content whilst undergoing fMRI. The emotional neural responses were more varied in L1 conditions, and positive emotions showed more activation and were rated as happier in the L1. There were no language differences in fearful or neutral passages.

Together, these findings are consistent with theories of grounded cognition, which have become a popular way of explaining language representation (Barsalou, 2010). Grounded cognition, also known as embodiment, suggests that items in memory, including words, are not independent 'symbols' that need to be retrieved, but are formed by a network of neural activation based on the modalities in which the items were encountered (Barsalou, 2008). Based on the above theory, L2 may be considered 'disembodied' in some circumstances, depending on proficiency, experience, context of use, etc., because it lacks the contextual experiences necessary to form sensory representations of words. Thus, emotion may not impact L2 processing in the same way it does L1 (Pavlenko, 2009).

However, the above literature investigates emotion–language interactions through participants' reactions to passages of text, which is a relatively complex level of processing. Psycholinguistic studies of language processing paint a more complicated picture of these interactions. Though not inconsistent with the grounded account, they suggest that emotionality differences cannot be simply explained by differences in language representation. Rather, they involve complex interactions between emotion and language control.

2.2.2 Effects of Emotion on Language Processing

Language processing is susceptible to affective influences in multiple ways. For example, words are recalled better during a surprise recall task following both deep and shallow processing tasks if they are emotion terms or taboo words (Caldwell-Harris & Ayçiçegi-Dinn, 2009). Adults differentiate English words from non-words in a lexical decision task more quickly if they have positive or negative valence than if they are affectively neutral (e.g. 'death' or 'sunshine' vs. 'shoe'; Kousta et al., 2009). These effects are usually attributed to the different ways emotional

and neutral words are represented in memory (e.g. Pavlenko, 2008), which supports the embodiment account.

However, emotion also impacts the processing of neutral language. For example, Verhees et al. (2015) and Chwilla et al. (2011) demonstrated that negative or positive mood (induced by film clips) influenced the way Dutch adults perceived sentences with verb-agreement violations or low close-probability nouns, as measured by differences in event-related potentials (ERPs) in the $P600$ and $N400$ time-windows, respectively.[1] Adults in a negative mood had smaller $P600$ effects (reflecting a reduced response to errors) following agreement errors and smaller $N400$ effects, reflecting lower semantic retrieval effort, in response to low-probability words compared with those in a positive mood, who showed the reverse effect. Verhees et al. (2015) also reported mood-related differences in earlier ERP time-windows, reflecting cognitive control effects. A larger $P1$ positivity was observed in response to sentences with errors in the positive mood condition but not in the negative mood condition. A trend toward larger $N1$ negativity in sentences with errors was observed when the task was difficult (judging sentence grammaticality) but not when it was easier (judging the text font).[2] This suggests that mood may influence early-stage sentence processing and attentional adaptation to different task demands, interacting with language control processes. Similarly, Gray (2004) found that in an n-back task where he manipulated both verbal demands (letters as stimuli) and spatial demands (stimulus location on screen), inducing a negative mood in participants enhanced verbal short term memory, while inducing a positive mood enhanced visual short term memory.

Studies conducted in participants' L2 further support the notion that emotion interacts with language in a more complicated way than simply through differences in mental representation. The word recall effects mentioned above (Caldwell-Harris & Ayçiçegi-Dinn, 2009), in which emotional and taboo words were more frequently remembered in a surprise recall task, were consistent across L1 and L2 in bilingual speakers. In addition, Opitz and Degner (2012) found comparable ERP effects in

[1] The $P600$ is a positive-going wave which begins ~600 ms post-stimulus-onset and generally has a larger amplitude when processing grammatical errors compared to correct sentences. The $N400$ (negative-going, ~400 ms) is associated with lexical processing and thought to reflect semantic retrieval.

[2] $P1$ (~125-175 ms) is thought to index perceptual processing and $N1$ (~175-225 ms) attentional processing (Verhees et al., 2015).

both L1 and L2 for pleasant/unpleasant compared to neutral words during word reading. An early posterior negativity (EPN) of similar amplitude occurred after the onset of emotional words in both languages, but its latency was longer in L2. Ponari et al. (2015) were able to show that emotional words were identified more quickly in L2 lexical decision tasks, mirroring findings in L1 (Kousta et al., 2009). This happened regardless of the participants' L1, L2 age of acquisition (AoA), or context and frequency of L2 use. In a lexical decision task including primes, however, Altarriba and Canary (2004) found differences between monolingual and bilingual participants. Prime words were displayed for 200 ms before being replaced by a high-arousal target word. High- and moderate-arousal primes reduced lexical decision latencies for the targets relative to emotionally-unrelated primes. However, this 'affective priming' effect was reduced in Spanish-English bilinguals compared to English monolinguals. The authors argue that, as the groups performed comparably with neutral primes, this result reflects group differences in processing of the high-arousal words in particular (Altarriba & Canary, 2004). However, an alternative explanation is that the higher-arousal primes did not interact with attentional processing in the same way for bilinguals and monolinguals. In other words, the primes may have been processed differently at the level of selective attention, rather than appraised and processed differently at the cognitive level (see Section 2.4).

These studies illustrate that the nature and difficulty of a task appear to play a larger role in any resulting differences between monolinguals and bilinguals than do L2 proficiency and AoA. Hayakawa et al. (2016) suggest that in addition to emotional 'disembodiment', the increased cognitive effort required for L2 processing may interfere with the emotional impact of text. In other words, cognitive control resources are limited, and both language processing and emotion perception draw on these resources. No work to the authors' knowledge has investigated links between the difficulty of language processing and the perception of its emotionality. So far, only studies investigating linguistic and emotional development in childhood shed further light on this relationship.

2.2.3 Developmental Links Between Language and Emotion

Bloom's *Intentionality Model* proposes that a child acquires language through three processes: First, interaction with people and objects; second, the desire to communicate by expressing mental states and intentions to these people/about these objects; and third, the effort required to integrate communication into these interactions (Bloom et al., 2001). This process is fundamentally tied to a child's emotional, social, and environmental interaction. Bloom and Beckwith (1989) conducted a longitudinal study of children's affect expressions, which they defined as changes in vocalisations, facial expression, or body posture, and how these related to word learning. Twelve infants, equated on language achievement, were compared during lab observations upon reaching two stages of vocabulary development: 'First Words' at a mean age thirteen months, and 'Vocabulary Spurt' at a mean age nineteen months. Bloom and Beckwith (1989) found that at First Words, the infants' affect expression decreased before word articulation, and was most intense immediately following these word utterances, but that the two did not overlap. There was more temporal convergence of affect expression and word articulation in the Vocabulary Spurt stage, but the words the children produced with accompanying affect in this stage were usually positive in valence, low in intensity, high-frequency, and among the earliest learned. The authors propose that the infants' increased affective expression immediately before and after word production may indicate that word learning is facilitated by the child's motivation to express contents of her awareness which give rise to emotions. Bloom et al. (2001) expanded on these findings, continuing to observe the same sample of children throughout their first two years of life to establish connections between their play behaviour, language development, and affect expression. Their cumulative findings suggest that the three processes are closely inter-related and share cognitive resources.

Bloom's model may help explain other research that links emotional and linguistic competence in older children. Beck et al. (2012), for example, gave 7-9 year-old German children a battery of tests for emotional (EC) and linguistic competence (LC). EC tests included assessments of the children's expressive knowledge of emotion vocabulary words, facial emotion recognition, identifying basic emotions in fictional scenarios, and awareness of mixed emotions in these same scenarios. LC

tests involved evaluating children's elicited narratives for structure and evaluative devices, measures of literacy, verbal fluency, and receptive vocabulary. Correlation analyses showed that most of these individual factors were highly correlated with each other, but still distinct in the individual constructs measured. Confirmatory factor analysis suggested that the EC and LC abilities were related on a higher-order level. Of the nine factors, receptive vocabulary, literacy, emotion knowledge, and awareness of mixed emotions had the strongest correlations with, and heaviest loadings on, the common 'language and emotional competence' factor. The authors suggest two possible interpretations of these findings: One is that lexical-semantic concept development facilitates emotional concept development (c.f. Lindquist et al., 2015). The other is that both linguistic and emotional competences rely on a common conceptualisation mechanism.

The latter interpretation supports the Bloom et al. (2001) model, and is backed up by recent evidence that multilingual exposure improves socio-cognitive outcomes. For example, infants' perspective taking, a skill closely associated with word learning according to socio-pragmatic accounts such as Bloom's (Gollek & Doherty, 2016), was improved by even limited foreign language exposure (Liberman et al., 2017). Bilingual preschoolers were better at identifying a speaker's emotions in natural speech than their monolingual peers (Yow & Markman, 2011).

Merely linking both linguistic and emotional concept development to a common overlying conceptualisation mechanism does not, by itself, explain how the two interact. It does, however, help explain related research findings linking children's language skills to emotion regulation and social competence measures. For example, Conboy et al. (2015) demonstrate that joint attention, which they measured by recording 42 month-old English monolingual infants' gaze shifts between a social partner and a reference object, predicted the infants' ability to detect a Spanish phoneme contrast after only 5 hours of social interaction with a Spanish tutor, as indexed by ERP responsiveness. Similarly, English–Mandarin bilingual and English–Mandarin–L3 trilingual 4–5 year-old children outperformed English monolinguals in a task measuring their ability to detect referential cues, such as eye gaze and pointing, during novel word-to-object mapping (Yow et al., 2017). In a group of Mandarin–English bilingual preschoolers, social competence measures assessed by

teacher questionnaires were positively associated with skills in both languages and with performance on emotion regulation tasks (Ren et al., 2016).

Adding cognitive control measures as an additional explanatory factor for emotional development, Wolfe and Bell (2007b) investigated the relationship between children's cognitive skills, language ability, and temperament. The latter refers to a child's innate emotional reactivity and their self-regulation of this reactivity (c.f. Rothbart, 1989). Children were tested once at 8 months and once at 4.5 years of age. Their cognitive performance was measured by tests of working memory and inhibitory control (WMIC), and language ability was assessed by the Peabody Picture Vocabulary Test. Their parents completed an Infant/Child Behavioural Questionnaire. At 8 months, infants (N=50) completed a WMIC task based on the A-not-B task (Diamond & Doar, 1989): they were trained to predict the location of a toy hidden by the researcher by looking to the correct side (right or left), and were rewarded by verbal feedback on their guess. According to the authors, the task required infants to update their memory of the toy's location through several displacements and to inhibit looking towards a previously rewarded location. The WMIC tasks administered at 4.5 years old (N=25 after attrition) were a day-night Stroop-like task in which the children were instructed to say "day" when shown a dark card with a moon on it and "night" when shown a light card with a sun on it, as well as a 'yes-no' task where they were instructed to say "yes" when the experimenter shook her head and "no" when the experimenter nodded. Wolfe and Bell (2007b) found that temperament in infancy, specifically the Soothability dimension, predicted both WMIC and language performance at 4.5 years old more than infant WMIC task performance did. This relationship was mediated by childhood approach/anticipation behaviours, such as the ability to sit still when excited.

Findings with older children and adolescents also demonstrate that the temperamental trait of Effortful Control (EC) correlates with executive attention (Rueda et al., 2005), as well as with literacy and with verbal measures of intelligence (Checa & Rueda, 2011). Furthermore, all of the above factors contributed to peer-reported likeability and sociability ratings in 12-year-old students (Checa et al., 2008). In other words, an emotionally stable temperament, executive control ability, linguistic performance, and social integration are all positively correlated

throughout childhood. It is therefore reasonable to assume that they remain linked in adulthood, and that experience-induced adaptations in one dimension also have an impact on the others.

2.2.4 Summary

Together, these findings point to cognitive control of language, attention, and behaviour as a decisive factor mediating the relationship between language and emotion. The following sections therefore investigate the links between these skills and cognitive control in more depth.

2.3 Language and Cognitive Control

2.3.1 Cognitive Consequences of Bilingualism

Significant research has focused on the cognitive consequences of bilingualism, particularly the work of Bialystok and colleagues (2009; 2017). There is mounting evidence that bilinguals across the lifespan perform better on certain (generally lab-based) cognitive control tasks, as discussed below. Among older adults, bilinguals present fewer symptoms of cognitive decline and dementia relative to age-matched monolinguals, suggesting that bilingualism contributes to cognitive reserve (Bialystok et al., 2012).

These findings suggest either that attaining proficiency in multiple languages is associated with better cognitive control skills, or that the cognitive demands of bilingualism also improve domain-general cognitive control skills. Although the first possibility cannot be ruled out, developmental findings render this unlikely. For example, Bialystok and Barac (2012) tested emerging bilingual children in French and Hebrew immersion programmes. They found a dissociation between *metalinguistic awareness*, the ability to reason about language, and executive control task performance. Higher proficiency in French/Hebrew was associated with the former, while length of time in the immersion programme was associated with the latter.

Some studies have also reported bilingual *dis*advantages, especially in verbal fluency (e.g. Sandoval et al., 2010) and free recall (e.g. Fernandes & Moscovitch,

2002). These disadvantages are often attributed to differences in language proficiency, especially due to bilingual children's smaller relative vocabularies and bilingual adults' slower lexical retrieval, and in many cases they disappear when controlling for these factors (see Bialystok, 2009). For example, performance in a memory and proactive interference task, which required participants to recall lists of words from a particular category while ignoring interference from previously encountered categories, was shown to be related to vocabulary size regardless of language background. Bilinguals had fewer intrusions from previous categories when vocabulary was controlled (Bialystok & Feng, 2009). Bialystok and Feng (2009) argue that both child and adult bilinguals in their study compensated for smaller vocabularies with better executive control.

Apart from language proficiency, often subtle changes in task design, such as the proportion of conflicting to facilitatory trial types (Costa et al., 2009), influence whether bilinguals perform differently from monolinguals. Some argue that an abundance of tasks which ostensibly measure the same construct, but whose outcomes do not necessarily correlate with each other, make it hard to draw definitive conclusions from studies on bilingual advantages (Paap & Greenberg, 2013). Others, however, argue that this view ignores the nuances associated with bilingual experience by attempting to reduce its influence to a specific construct (Kroll & Bialystok, 2013). A multifaceted approach, which does not focus on a unitary 'bilingual advantage' but investigates its potential influence across a wide range of cognitive tasks, can help address this criticism. This begins with understanding the executive demands of bilingual language control and isolating the individual processes involved.

2.3.2 Cognitive Control of Language

According to the Adaptive Control Hypothesis (ACH; Green & Abutalebi, 2013), there are seven different control processes involved in bilingual language use. Depending on the language context, which can be single-language, dual-language, or 'dense', i.e. mixed language, the bilingual must:

1. Maintain communicative goals, such as remaining in the appropriate language (*goal maintenance*).

2. Monitor for potential conflicts with these goals, such as lexical competitors or environment noise (*conflict monitoring*).
3. Detect salient cues that signal the need to change behaviour, such as the arrival of a new interlocutor who speaks a different language (*salient cue detection*).
4. Inhibit a current behaviour, e.g. speaking in language A (*behavioural/response inhibition*).
5. Disengage from the current task schema (*task switching*).
6. Engage a new task schema (*task switching*).
7. Plan multilingual utterances opportunistically to maximise communicative efficiency. (This applies specifically to intra-utterance code-switching, and is thus only relevant in specific populations of bilinguals from 'dense' language contexts where everyone speaks both languages.)

Green and Abutalebi (2013) as well as Bialystok (2017) argue that sustained participation in multiple language contexts leads to neural adaptations in the brain networks that underlie these eight control processes. Neural mechanisms of bilingual language control have been pinpointed by numerous fMRI studies. Abutalebi (2008) concluded from a meta-analysis of imaging studies of bilingual syntactic and lexico-semantic processing that, in addition to sharing the neural mechanisms involved in L1 processing, using an L2 recruits the same neural mechanisms as general cognitive control tasks. In particular, there is increased activation during L2 relative to L1 processing in three domain-general executive control areas: the left dorsolateral prefrontal cortex (dlPFC), which is involved in implementing control processes (MacDonald et al., 2000); the anterior cingulate cortex (ACC), which is involved in monitoring performance (Bush et al., 2000); and the basal ganglia, which are involved in action selection (Stocco et al., 2010). Similarly, Ye and Zhou (2009b) review evidence for shared neural mechanisms, comprising a distributed 'control network' of regions in the ventral and dorsal PFC, ACC, parietal cortex and basal ganglia, that is involved in tasks of language control, such as sentence comprehension and lexical selection, bilingual language control, and non-linguistic cognitive control.

Studies also largely support the hypothesis that the specific processes identified by the ACH are more efficient in bilinguals: Bilingual advantages have been reported

for (1) goal-maintenance (Colzato et al., 2008), (2) conflict monitoring (Costa et al., 2008), and (3) interference suppression (Mercier et al., 2014). To the authors' knowledge, behavioural evidence for a bilingual performance advantage in either (4) salience detection or (5) response inhibition is currently absent from the literature. However, fMRI evidence suggests that bilinguals use different neural mechanisms than monolinguals to complete a behavioural inhibition stop-signal task (Costumero et al., 2015). Specifically, they rely on a a 'salience detection' network consisting of the inferior frontal gyrus, insula, and dorsal ACC (Costumero et al., 2015). Research on task schemas (6) is generally limited to psycholinguistic studies (e.g. lexical retrieval; Dijkstra & van Heuven, 2002). Although task switching studies have also reported bilingual advantages (Garbin et al., 2010), these may be limited to updating stimulus-response mappings (Hernández et al., 2013).

However, there is still considerable disagreement over whether and how hypothesised 'bilingual adaptations' (Bialystok, 2017) are observable in non-linguistic contexts, or whether the evidence discussed above is due to confounding factors in task design or sampling (Paap & Sawi, 2014). Qualitative differences in brain activation networks do not necessarily correlate with quantitative differences in task performance. In order to address this controversy, more clarity is needed in the constructs underlying 'domain-general cognitive control' and their mechanisms.

2.3.3 Types and Levels of 'Cognitive Control'

Researchers generally agree that 'cognitive control' refers to higher-order thought processes required to solve problems, plan, select strategies, and direct attention. There are several different, though conceptually overlapping, frameworks which have informed research into these processes. This dissertation does not argue for one framework over another. However, given that bilingualism research has been subject to criticism for the variability in study findings caused by often subtle changes in task design, it is important to outline these frameworks and the relationships and differences between them.

2.3.3.1 Executive Functions

At the level of specific control processes, a model that has gained considerable popularity in the bilingualism literature is Miyake et al.'s (2000) tripartite framework. Miyake et al. (2000) compared data from numerous commonly-used executive control tasks hypothesised to measure three distinct but inter-related executive functions (EFs):

1. 'Updating' working memory (WM) contents (defined as explicitly manipulating WM rather than simple recall)

2. 'Inhibition' of prepotent responses

3. 'Shifting' between tasks or dimensions.

The contribution of each function to an overall model of executive control was assessed using confirmatory factor analysis. Results showed that a 'full three-factor model' in which the functions were moderately correlated but independent constructs provided the best fit to the data. Thus, updating, inhibition, and shifting can be considered three separable but interrelated skills contributing to the general construct of executive control, as visualised in Figure 2.1. As the authors themselves note, this is not an exhaustive account of executive control. Further studies are needed to address capabilities which are not accounted for by this framework, but are commonly ascribed to executive control, such as dual-task coordination and more 'complex' processes like planning, as well as to investigate whether more basic processes underlie each function (Miyake et al., 2000).

Of these three EFs, inhibition has received the most attention within the bilingualism literature due to its key role in theories of language control (Green, 1986; Green & Abutalebi, 2013). Both updating and shifting have received some attention as well (see for example Garbin et al., 2010; Hernández et al., 2012; Hernández et al., 2013; Morales et al., 2013). Most researchers hypothesise that in order to select words for production in the target language, the bilingual must inhibit competitors in the non-target language (e.g. Kroll et al., 2008). The asymmetrical latencies observed when bilinguals switch between a more dominant L1 and a less dominant L2 suggest that more effort is required to inhibit the former, providing one point of evidence in favour of inhibition (Meuter & Allport, 1999).

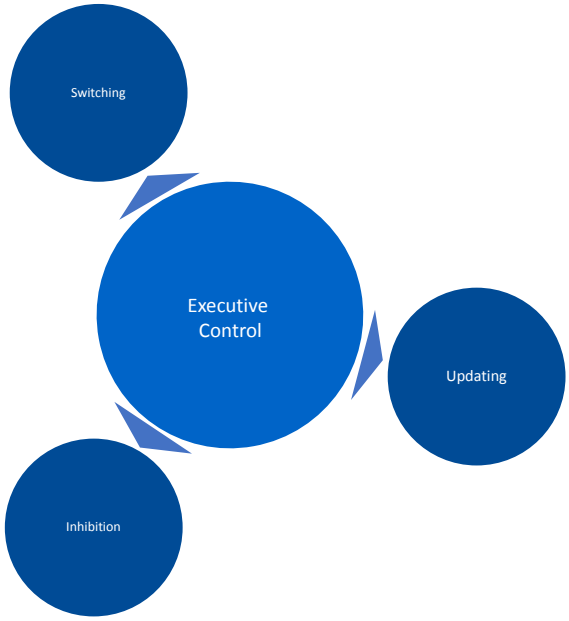

Figure 2.1: The three executive functions (EFs) that contribute to executive control according to Miyake et al.'s framework (2000).

However, the level at which this inhibition occurs is more contested. Some theories suggest that it occurs at the lexical level as interference suppression between lexical competitors, and some favour response inhibition at later, articulatory stages (see Spalek et al., 2013, for a summary).

Inhibition itself is also a complicated construct, with Nigg (2000) identifying no fewer than eight different types/systems of inhibition. He divides these into three classes, executive, motivational, and automatic, which differ in their task demands and neural correlates. The ACH suggests that for bilinguals, two executive inhibition subtypes are required during language interactions: 1) interference suppression, in order to control distractions from the environment, and 2) response inhibition, to switch between languages in response to environmental cues (Green & Abutalebi, 2013). The former, however, is closely linked to attention and conflict monitoring (see below). Furthermore, goal maintenance, cue detection, and switching involve aspects of what Nigg (2000) terms motivational inhibition: the "bottom-up or limbic→cortical interruption of ongoing behaviour or suppression of behavioural response due to fear or anxiety in the presence of immediate novel social situation or cues for punishment" (pp. 238).

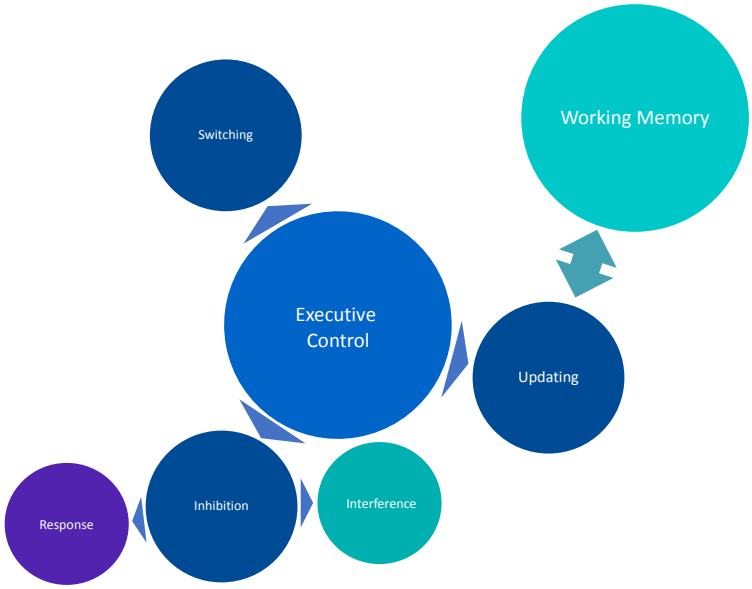

Figure 2.2: An extension of Figure 2.1 indicating how executive functions link to WM (teal bubble) and the sub-functions of inhibition relevant to bilingual language control.

Experimental evidence supports the notion that bilingualism improves interference suppression, but not response inhbition, in non-linguistic tasks. For example, Martin-Rhee and Bialystok (2008) compared bilingual and monolingual children's performance on the Simon task, designed to measure interference suppression, and two Stroop-like interference tasks designed to measure inhibition of a prepotent response. In the Simon task, participants must respond to the colour of a shape presented on screen with a left or right keypress. The shape can appear in the centre of the screen (neutral condition) or to the left or right. If the side of presentation is the same as the key corresponding to the shape's colour, the trial is congruent, while the side of presentation conflicts with the side of the keypress in incongruent trials. Participants must thus ignore the irrelevant dimension of stimulus location. RTs are reliably slower to incongruent trials and faster to congruent trials, an observation known as the Simon effect (Simon & Berbaum, 1990). To measure response inhibition, Martin-Rhee and Bialystok (2008) used a variant of the Stroop task in which children were asked to say 'day' in response to an image of the moon and 'night' to an image of the sun, requiring them to suppress the prepotent tendency to name the picture. Bilingual children were faster than their monolingual peers on the Simon task, but no differences were found on the day-night Stroop

task. To test the claim that this null effect was due to a bilingual disadvantage in picture naming due to slower lexical retrieval which masked a response inhibition advantage, Martin-Rhee and Bialystok (2008) conducted a second study in which they used two variations of the Simon task. In the 'univalent' version, an arrow was presented in the centre of the screen and children had to either indicate its direction (in congruent blocks) or the opposite direction (in incongruent blocks), as indicated by instructions before each block. This was intended to trigger conflict between the prepotent stimulus-response mapping of responding to the direction of the arrow and the task goal of answering correctly. In the 'bivalent' condition, the arrow appeared on the left or right of the screen, causing conflict between two stimulus dimensions, direction and location. Again the bilingual children were faster on the bivalent task, but no differences were found on the univalent task. The authors conclude that bilinguals are better at interference suppression, but have no advantage for response inhibition.

While this conclusion is supported by behavioural findings in adults, both fMRI studies (e.g. Costumero et al., 2015; Rodríguez-Pujadas et al., 2014) and ERP studies (e.g Fernandez et al., 2013; Moreno et al., 2014) have identified differences between monolinguals and bilinguals in the brain networks recruited for response inhibition tasks, specifically in the ACC and other networks related to language control. These findings can, however, also be interpreted in terms of conflict monitoring (Rodríguez-Pujadas et al., 2014).

The lack of consistency in findings from tasks all ostensibly meant to measure inhibition may be explained by protracted development of executive functions as well as a lack of clarity in the construct itself. Using a similar approach to Miyake et al. (2000), Brydges, Reid et al. (2012) modelled results from a battery of EF and intelligence tests with a sample of 7–9-year-old children. They found that rather than the three-factor model, a unitary EF factor provided the best fitting model in both age groups, which they interpret as evidence that EFs differentiate into diverse functions later in development as the PFC matures. However, in a recent update to their model which focused on cognitive underpinnings rather than on individual functions, Miyake and Friedman (2012) note that once the contribution of a 'common EF' construct was parcelled out, inhibition did not contribute any additional unique

variance in performance across a range of different EF tasks even in adults. In other words, inhibition was nearly 100 % correlated with the common EF construct.

In light of recent criticism of the lack of replicability in bilingual advantage studies (Paap, 2014), researchers have called for a move away from inhibition as a basis for bilingualism research (see Bialystok, 2017, for a thorough discussion of this issue). Bialystok (2017) suggests that the findings from inhibition-focused studies can also be interpreted in terms of selective attention, and points to the attention literature as a preferable framework for further study on bilingual advantages.

2.3.3.2 Executive Attention

Within attention research, executive control has been considered an attentional process alongside alerting and orienting (see Figure 2.3; Posner & Petersen, 1990; Posner & Rothbart, 1998). It is assumed to control the processes of selection, conflict resolution, and error detection (Norman & Shallice, 1986). The 'central executive' control system of capacity-limited attention is also central to Baddeley's (2003) widely-accepted model of WM. The construct overlaps heavily with the executive functions described above: for example, the flanker, Stroop, and Simon tasks, often used to measure inhibition, are also considered experimental measures of executive attention. They have been shown to rely on a brain network involving the ACC and lateral frontal lobe regions (Posner & Rothbart, 1998). Figure 2.3 provides a visual summary of the relationship between attention and EF based on the way these constructs are conceptualised in the literature.

As discussed in Section 2.2.3, links between executive attention and language performance, particularly in relation to temperamental effortful control, are well-established in developmental research. There is some evidence for a relationship between bilingualism and attentional processes as well. First, exposure to multiple languages has been shown to change infants' gaze behaviour: 8-month-old bilingual infants showed a preference for the mouth region of faces, compared with monolingual infants who preferred the eye region (Ayneto & Sebastián-Gallés, 2017). A study compared infants' native and non-native phoneme discrimination, vocabulary size, and non-verbal executive control performance on an object retrieval task. Native contrast discrimination was positively correlated with the executive control task, while non-native contrast discrimination was negatively correlated with it (Conboy

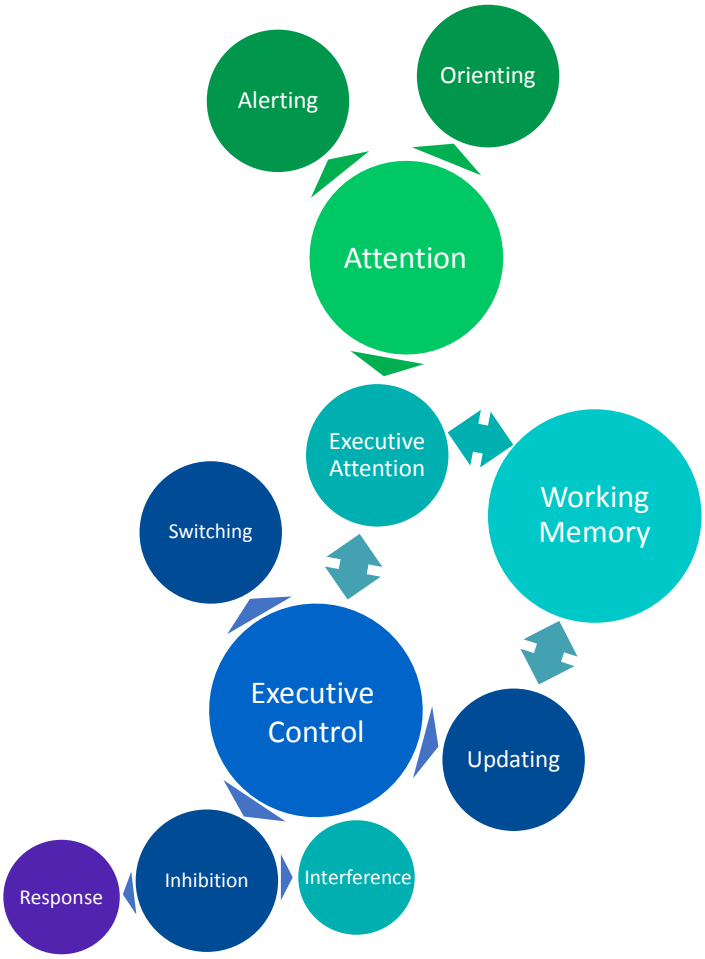

Figure 2.3: A model of the attention and executive systems underlying cognitive control and their overlap. Arrows and bubble colours denote the hypothesised relationships between the concepts: Bidirectional arrows indicate overlapping constructs, while triangles indicate sub-functions. Green denotes attention, blue denotes EF, and teal an overlap between the two.

et al., 2008). The authors suggest that these results reflect infants' ability to separate relevant from irrelevant information in the speech stream. Astheimer et al. (2014) conducted an ERP study with 5-year-old children during which attention probes were inserted during a listening task. The study found that early, attention-related ERP waves were more negative when the probes occurred 100 ms after word onset than in any other location, suggesting that the children allocated attention specifically to the onset of words. Furthermore, the degree of negativity in this measure predicted the children's performance on a complex grammaticality judgement task, but not on simpler linguistic or non-linguistic tasks, pointing to a possible relationship between attention and metalinguistic awareness in this population.

Recent studies have also associated attention and bilingualism in children. Sorge et al. (2017) showed that attention ability as assessed by parent and teacher questionnaires, as well as bilingualism, predicted better performance on three EF tasks in 8–11 year-old children: stop-signal, flanker, and frogs matrices (a test of spatial working memory based on the Corsi blocks test; Berch et al., 1998). Attention was the strongest predictor of performance on the stop-signal task, while bilingualism was the strongest predictor of performance on the flanker task.

Bialystok et al. (2017) also showed that bilingualism interacts with ADHD in young adults. Bilingual participants with ADHD performed worse than monolinguals with ADHD, monolingual-neurotypical, or bilingual-neurotypical young adults in a stop-signal task, suggesting that the effects of ADHD are more severe for bilinguals in this task. This supports Bialystok's (2017) argument that bilingual advantages may be due to attentional adaptations rather than other sub-functions of executive control.

Finally, as discussed in Section 2.2.3, executive attention measures correlate with temperamental traits related to emotion control, such as Effortful Control and Soothability, providing another potential link between bilingualism and emotion. Posner and Rothbart (1998) argue that this has developmental foundations: attentional orienting towards interesting stimuli reduces negative affect expression in infants, but this returns once the stimulus is removed (Harman et al., 1997), suggesting that it is held at bay during orienting but does not subside. Posner and Rothbart (1998) suggest that this may arise from caregivers' soothing practices, as it is common to calm a distressed infant by directing its attention towards a novel stimulus. As the ACC is a known locus for pain and distress perception, the authors suggest that these early behaviours may lead to the integration of executive attention and emotion in the ACC as the infant learns emotion regulation. This account is very similar to Bloom's (2001) intentionality model discussed in Section 2.2.3. In addition, a posited central role of the ACC in integrating executive attention and emotion regulation aligns with a third framework of cognitive control, based around detecting and resolving conflict.

2.3.3.3 Conflict Monitoring

Cognitive control tasks which require a resolution between competing stimuli or responses induce conflict: 'any disagreement or discrepancy between mental representations, response tendencies, or actual behaviour' (Inzlicht et al., 2015, pp. 127). Conflict monitoring theory (Botvinick et al., 2001) proposes that cognitive control processes are based on conflict detection and resolution. A 'conflict monitor', whose neural substrate is believed to be in the ACC, detects potential conflicts and relays them to the PFC, which engages the appropriate executive function(s) to resolve them (Botvinick et al., 2004). For example, in a flanker task, there is a conflict between the response tendency initiated by the target and that initiated by the flanker, and executive processes such as interference suppression and response selection are needed in order to respond correctly. Conflict can be conceptualised as a signal that engages attention, and executive functions are deployed to resolve it. This relationship is visualised in Figure 2.4.

Evidence of conflict monitoring processes comes from ERP analyses of the inhibition-related $N2$ and the error-related negativity (ERN), a negative wave peaking approximately 100 ms following incorrect responses on experimental tasks (Yeung et al., 2004). Using an ERP study to test the predictions of a model simulation based on conflict, Yeung et al. (2004) argue that both potentials reflect conflicts between possible responses: the $N2$ prior to the correct response, and the ERN following an error. The adjustment of control processes in response to conflict is supported by the conflict adaptation, or 'Gratton', effect (Gratton et al., 1992). In interference suppression tasks such as the flanker task, incongruence effects are attenuated by previous incongruent trials, suggesting that the level of control allocated to suppressing flanker interference is increased by the perceived conflict in the prior trial.

The involvement of the ACC in detecting conflicts mirrors the performance monitoring role ascribed to this brain region during language control (see Section 2.3.2 above). Indeed, the ACC is highlighted in bilingualism studies as a likely neural substrate for bilingual adaptations (Abutalebi et al., 2011). As posited in the ACH, bilinguals experience conflict between goals, environmental inputs, and mental representations during communication (Green & Abutalebi, 2013). This appears to translate to better performance on tasks which require resolution of

conflict between various types of stimuli. Experimental studies support the idea that bilinguals perform better on both linguistic (Moreno et al., 2010) and non-linguistic (Costa et al., 2008; Festman & Münte, 2012; Poulin-Dubois et al., 2011) conflict resolution tasks. Ye and Zhou (2009a) frame their discussion of bilingual advantages specifically in terms of conflict detection and resolution as a shared mechanism for language and cognitive control.

Figure 2.4 presents a full model of cognitive control as it is conceptualised in this dissertation, with the partially overlapping constructs of attention and EF presented in Figure 2.3, engaged by a conflict signal, and producing a behavioural response or changing the focus of attention.

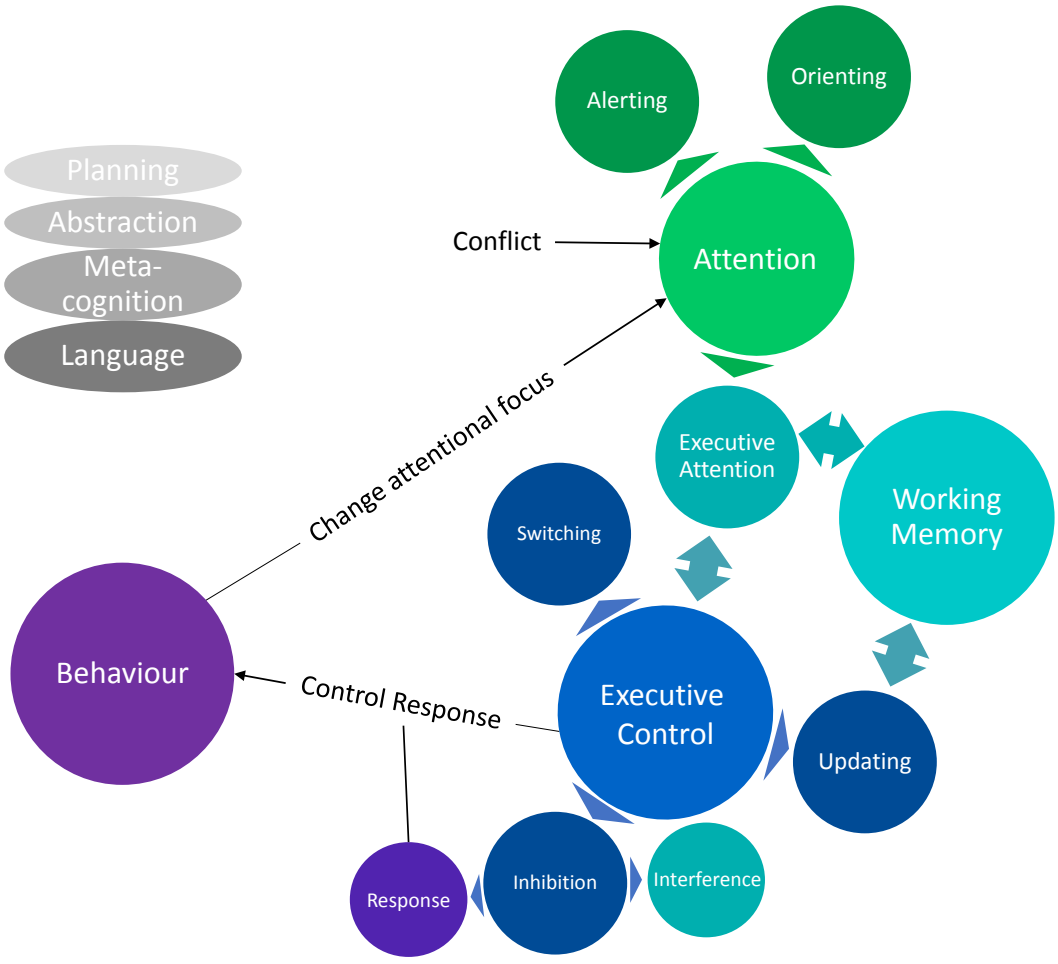

Figure 2.4: A model of the cognitive control system (attention, in green, and EF, in blue) triggered by conflict and resulting in a behavioural response (purple). Bidirectional arrows and teal or indigo colours denote overlapping constructs. The grey ovals denote higher-level functions ascribed to cognitive control but beyond the scope of this dissertation.

Beyond functioning as a trigger for cognitive control, conflict monitoring has

been widely linked to emotion processing and regulation. Emotional stimuli receive prioritised attention because they are motivationally significant (Weinberg & Hajcak, 2010), and may thus engage executive functions in the same way as conflict signals. The two may be even more closely linked, however. Inzlicht et al. (2015) argue that any form of conflict involves so-called 'integral affect', which is affect arising from the nature of the task itself, even if the stimuli are emotionally neutral. Negative or aversive integral affect generated by conflict may drive the mobilisation of executive control resources to quickly resolve it (Inzlicht et al., 2015). In support of this theory, differences in emotional response to conflict have been shown to predict differences in cognitive control measures. For example, Hobson et al. (2014) found that ERNs were reduced and error rates increased during a go/no-go task[3] when participants were explicitly instructed to reappraise their emotional states between blocks.

Inzlicht et al. (2015) hypothesise that cognitive control is not initiated by conflict itself, but by the integral affect generated by this conflict. In other words, *cognitive control is driven by emotion*. If this can be substantiated, then no model of cognitive control is complete without factoring in emotional influences, and neither is our understanding of the impact of bilingualism.

However, demonstrating that cognitive control depends on emotion is not the aim of this dissertation. Rather, the studies reported here aim to measure the impact that emotional stimuli have on commonly utilised measures of executive control, and whether this impact differs between monolingals and bilinguals. This would provide a foundation for future work to examine this relationship in more detail. Emotion, even conceptualised as a separate process, still demonstrably influences executive functions, as discussed in Section 2.4 below. Navigating daily life requires coordination between innate processing/response tendencies for emotions, such as attention capture and fight-or-flight responses, and other demands on the cognitive control system, such as attending to task-relevant stimuli or responding in a way that meets situational goals. Figure 2.5 incorporates emotion and its influences on the cognitive control system into the model as presented in Figure 2.4 above. Research supporting this relationship is reviewed in the next section.

[3]The go/no-go task is a response inhibition task where participants are asked to respond to one stimulus/stimulus type whilst withholding responses to all others: see Section 3.1.2.2 for more detail.

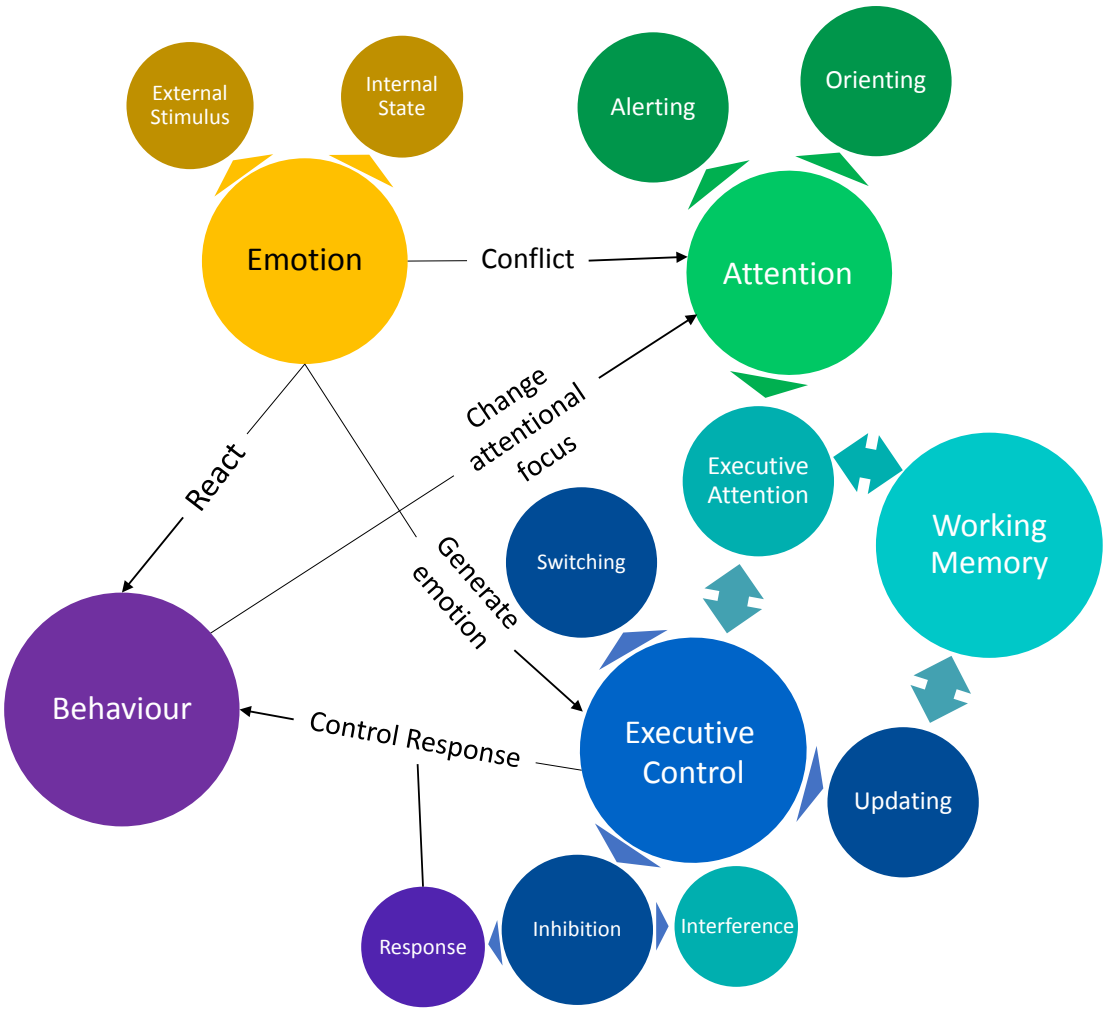

Figure 2.5: An extension of Figure 2.4 that includes the influence of emotion (both from external stimuli and internal states, in yellow) on cognitive control.

2.3.4 Summary

The studies reviewed in this section indicate that the theoretical assumptions behind cognitive control, and by extension the effects of bilingualism on cognitive control, remain unclear. A cautious approach is necessary in order to bring clarity to often disparate findings. This approach should be informed by an interdisciplinary theoretical framework: 1) Insights into (bilingual) language control should be drawn from psycholinguistic studies. 2) Evidence from neuroimaging should be used to identify neural substrates where language-related adaptations could occur and to inform our understanding of the functional relationship between emotion and cognitive processing. 3) Selection of experimental tasks should be grounded in models of executive control. An essential consideration when designing studies are

the processes by which bilingual adaptations are posited to occur. They depend on individual factors such as language context, ability, and frequency of use. The ACH places bilingual language use in a context of social interaction. This means that interpersonal social and emotional factors are involved and also subject to adaptation. Tasks selected to measure bilingual adaptations must consider all of these factors, including the role of emotion.

2.4 Cognition–Emotion Interactions

While considerable attention has been devoted to the subjective differences in emotionality between bilinguals' languages and the acquisition processes underlying them (as discussed in Section 2.2), there have been few studies devoted to the cognitive control mechanisms involved in processing emotional language. This section examines the relationship between emotion and cognitive control, emphasising control functions that are also relevant to bilingualism research.

2.4.1 Models of Emotion and Cognitive Control

It is commonly accepted that emotional stimuli draw and hold attention more effectively than neutral stimuli. This is considered an adaptive response to biologically relevant information that aids the organism's survival (Brosch et al., 2010). A number of attention tasks such as the attentional blink (e.g. Lim et al., 2008), visual search (see Frischen et al., 2008), and flanker tasks (e.g. Grose-Fifer et al., 2013) all show clear influences of emotional states or stimuli on behavioural outcomes. However, the effect of emotion on cognitive performance varies significantly by task, load, and individual differences. A number of models and theories conceptualise this relationship. The *dual competition* model (Pessoa, 2009), for example, describes how and why emotion and cognition interact during processing (see Section 2.4.1.1 below). The *Model of the Cognitive Control of Emotion* (Ochsner et al., 2012, MCCE,) summarises cognitive mechanisms for regulating emotional responses (Section 2.4.1.2). The appetitive–aversive view of emotions (Section 2.4.1.3) treats them as motivators for approach or avoidance behaviour (Fredrickson, 2004; Lang et al., 1997; Schwarz & Clore, 1996). Factoring these formulations into an integrated

framework of cognitive control and emotion provides a critical first step for studying their relationship to bilingualism.

2.4.1.1 The Dual Competition Model

According to Pessoa's (2009) *dual competition* model, cognitive control is a capacity-limited resource. The affective significance of stimuli or emotional/motivational states creates competition for this resource. At the perceptual level, it enhances sensory representations (c.f. Vuilleumier et al., 2001). At the executive control level, it claims attentional priority (c.f. Egner & Hirsch, 2005). Pessoa (2010) proposes that visual perception of an emotional stimulus produces 'multiple waves' of activation from the visual cortex to the amygdala, ACC, insula, and parts of the PFC, allowing for quick processing of the stimulus. Reciprocal connections also allow for the top-down regulation of this process. Pessoa (2010), for example, describes a circuit including the pulvinar, regions of the parietal cortex (associated with attentional processes), PFC, ACC, and insula that may play a role in determining the biological and behavioural relevance of the stimulus and modulating its processing stream accordingly.

The dual competition model predicts that emotional stimuli enhance cognitive control performance when they are relevant to the task because they increase the amount of processing resources allocated to solving the task. However, task-irrelevant stimuli, and all fear-related stimuli, are predicted to disrupt performance because they consume resources which would otherwise be available for the task. Similarly, motivational states enhance behaviour that maximises reward, at the cost of other demanding behaviours (Pessoa, 2009). These predictions are largely upheld in studies of executive control–emotion interactions.

For example, Kanske and Kotz (2010b, 2011b) incorporated emotional stimuli into a flanker task to investigate their impact on executive attention. Adult native German speakers were instructed to indicate the red or green ink colour of a centrally-presented target word. The same word was displayed above and below the target in either congruently in the same colour or incongruently in the opposite colour. The words were either neutral or negative in valence. Negative words, such as 'terror', reduced response latencies on incongruent trials relative to neutral words such as 'table', and the negative trials were accompanied by a greater $N200$ negativity.

The modulation of this early potential by word valence suggests a fast integration of emotion and conflict processing (Kanske & Kotz, 2010b). These findings were also replicated with positive words (Kanske et al., 2011). Furthermore, negative incongruent trials were also accompanied by greater activation in the ventral ACC (Kanske & Kotz, 2011b) and a greater ERP conflict negativity (Kanske et al., 2011), which were further enhanced in participants high in temperamental effortful control and low in subclinical depression and anxiety scores, as measured by questionnaires (Kanske & Kotz, 2012). Emotional words, regardless of valence, recruited additional control resources that facilitated conflict processing. Together, these findings suggest the ventral ACC is a potential locus for the integration of conflict resolution and emotion processing, particularly for negative emotions. Effortful control traits seem to enhance this integration, while anxious and depressive states moderate it.

The effects of emotional stimuli are reduced when the task demands increase and there is more competition for control resources. For example, Fox et al. (2012) displayed task-irrelevant fear-conditioned angry faces above or below the central stimulus in a letter-version of the flanker task. An increase in reaction times to fear-conditioned stimuli when the flanker letters were easier to distinguish from the targets compared to the perceptually more difficult trials suggested that attention was effectively directed away from the distracting faces in the high-load condition (Fox et al., 2012). Fox et al. (2012) also found that in the low-load condition, the fear-conditioned distractor had the opposite effect in participants high in trait anxiety, leading to faster reaction times. This again supports Pessoa's (2009) hypothesis that affective states enhance goal-directed performance. In this case, heightened fear-sensitivity due to trait anxiety enhanced attentional avoidance of the fear-conditioned stimuli.

These findings raise questions about the level of automaticity involved in perception of emotional (visual) stimuli. Pessoa and colleagues also demonstrated that brain regions involved in processing emotional faces were only active during an experimental task if attentional demands were sufficiently low, suggesting that attention can gate the processing of emotional stimuli very early in perception (Pessoa et al., 2002). In other words, emotion processing appears to be subject to top-down regulation from the earliest stages. The MCCE (Ochsner et al., 2012) further elaborates on such top-down emotion control mechanisms.

2.4.1.2 The MCCE

In their *Model of the Cognitive Control of Emotion* (MCCE), Ochsner et al. (2012) outline the levels of processing at which cognitive control can be deployed to control the impact of emotional stimuli or states. They differentiate between attention (awareness of a stimulus), appraisal (valuation of the stimulus as appetitive or aversive), and responding. Perception of a stimulus occurs first and recruits the posterior and dorsolateral PFC to direct attention to salient stimuli. Appraisal occurs in the 'emotion generation' system of the amygdala, ventral striatum, and insula, which evaluate arousal/threat, reward value, and stimulus-related bodily states, respectively. Appraisal leads to an emotional reaction or response. Thus, cognitive control of emotion can be deployed at three levels:

1. At the *attentional* level, where selective attention controls which stimuli are allowed to enter awareness (and/or signal potential conflict).

2. At the *cognitive* level, where the stimulus' meaning is 'appraised' (i.e. its conflict with behaviour/survival goals is evaluated), and the initial appraisals can be consciously changed (reappraisal; i.e. determining that the stimulus is not relevant to the current task/behaviour, despite generating some form of conflict).

3. At the *behavioural response* level, where emotional reactions and expressions are regulated (e.g. resisting a flight response; maintaining a neutral facial expression).

D. Mitchell et al. (2007) provide evidence for top-down modulation of emotion processing by attention and cognitive control networks dependent on task-relevance. Using a variant of the affective Stroop task, they superimposed words onto images of neutral or fearful faces. They then used fMRI to compare participants' blood-oxygenation-level-dependent (BOLD) responses in emotion and cognitive control networks under three task conditions. In the 'task-relevant' condition, participants identified the gender of the face. In the 'low-load' condition, they named the case of the text (upper/lower-case), while in the 'high-load' condition, they counted the number of syllables in the word. Amygdala and medial PFC reactivity to fearful faces was much higher in the task-relevant gender-identification condition

than in either of the load conditions, and further decreased relative to load level. Conversely, activity in the dlPFC[4] and other pre-frontal regions linked to attentional control increased with processing load. The authors suggest that the neural circuits observed respond to the task-relevance of stimuli and adjust their perceived salience according to task demands, in this case down-regulating emotional reactivity to favour the more difficult processing conditions.

Interestingly, D. Mitchell et al. (2007) also reported a condition × valence interaction in the ACC, which was more active in trials with fearful faces, in both the gender-identification and the syllable-counting condition but not in the low-load upper/lower-case condition. The authors suggest that this may be related to response conflict, but note that it is difficult to explain the absence of ACC activation in the low-load case. One potential explanation lies in Pessoa's (2009) argument that fear-related stimuli command additional control resources and are thus more disruptive to cognitive tasks. The low-load condition may not have been attentionally demanding enough for the fearful faces to interfere, while the more demanding syllable-count condition may have required down-regulation of the fearful faces by the ACC. In the gender condition, the fearful faces were task-relevant, potentially leading to enhanced representation in the ACC. Processing of the facial expression may also have interfered with gender identification.

Behavioural evidence suggests that the neural mechanisms that coordinate emotional and non-emotional cognitive processes are not functionally distinct from the domain-general control mechanisms implicated in bilingualism research. In a comprehensive review and meta-analysis, Schmeichel and Tang (2014) demonstrate that better performance on (neutral) inhibitory control and working memory tasks is correlated with emotion regulation functions identified by the MCCE. Their review focused specifically on three processes: 1) *Reappraisal*, or re-assessing an initial emotional response. 2) *Subjective Responding*, or reacting to recalled emotional memories. (Functions 1 and 2 operate at the 'emotion generation' stage within Ochsner's (2012) 'cognitive' level, see item 2 above.) 3) *Expressive Suppression*, which refers to an individual's ability to monitor inappropriate emotional expressions (this occurs at Ochsner's 'response' level). Across a number of reviewed studies,

[4]The dlPFC is frequently linked to executive functions, particularly conflict processing (Botvinick et al., 2004), as well as to emotional reappraisal (Ochsner & Gross, 2005) and EF/emotion integration (Albert et al., 2010).

participants' performances on experimental inhibitory control measures, such as the stop-signal task, and tests of working memory monitoring and updating, such as the operation-span task, were positively associated with these three skills (Schmeichel & Tang, 2014).

2.4.1.3 Appetitive–Aversive Response Tendencies

An influential theory proposed by Lang et al. (1997) suggests that positive or negative valence is appraised as a reward or danger signal, and thus associated with behavioural approach ('appetitive') and withdrawal ('aversive') systems, respectively. Arousal indexes the degree of activation of these two systems. In a similar vein, Schwarz and Clore (1996) suggest that valence is treated as information about a task, and that negative affect triggers more detailed, bottom-up, survival-oriented processing, while positive affect can broaden attention and causes more global, heuristic processing. Negative stimuli may signal response conflict, and negative mood the need for more analytical processing, thus disrupting or enhancing control performance (R. L. Mitchell & Phillips, 2007). Positive affect-related broadening of attention, while not an immediate survival-oriented adaptive response, encourages activities such as creativity, exploration, and play, which may lead to improved cognitive control both evolutionarily and over the lifespan, as proposed by the *broaden-and-build* theory (Fredrickson, 2004). However, the effects of emotions on task performance are eliminated if participants' attention is drawn to the cause of the perceived emotion (Schwarz & Clore, 1996), perhaps reflecting reappraisal processes.

Go/no-go tasks are often used to measure the influence of positive vs. negative stimuli on approach (go) and withdrawal (no-go) behaviour. Goldstein et al. (2007) performed an fMRI study using a linguistic go/no-go task with positive, negative, or neutral words. The participants were instructed to respond to words in normal font with a button press, but to withhold responses to words printed in italic. The right dlPFC as well as the dorsal ACC showed the highest degree of activation in negative no-go trials measured by BOLD signals, which the authors argue indicates a functional convergence related to the integration of response inhibition and negative stimulus processing in this region. Albert et al. (2010) performed a similar go/no-go task using EEG. They superimposed letters on a positive, negative, or neutral background image and instructed participants to respond to only one of the letters.

They observed a larger P3 potential, which has been associated with response inhibition in numerous imaging studies (c.f. Baumeister et al., 2014), in positive no-go trials relative to negative or neutral trials. While this appears to contradict Goldstein et al. (2007), the differences in task demands provide an explanation. In the Goldstein et al. (2007) study, the experiment was split into *go* blocks (100 % go trials) and *no-go* blocks (33.7 % no-go trials). The no-go blocks require closer monitoring than the go blocks to override the prepotent go response, and thus demand more attentional resources. The addition of emotional stimuli places further strain on attention, and as reported by Kanske and Kotz (2011b), the ACC appears to be particularly active in negative conflict contexts. The difference in locus—ventral vs. dorsal—may be due to the difference between interference suppression in the flanker task and response inhibition in the go/no-go task. As discussed by Posner and Rothbart (1998), developmental factors may influence the way these processes interact with emotion in the ACC. In the Albert et al. (2010) study, the background image varied between blocks, but remained consistent throughout the block, creating a longer-lasting emotional context. According to Lang, Bradley, and Cuthbert's 1997 theory, the positive block would bias approach (go) behaviour, and the negative block withdrawal (no-go) behaviour, facilitating positive go reactions over all other conditions, as reflected in the behavioural effects.

Hare et al. (2005) found similar results when emotional facial expressions constituted a response dimension. They conducted an emotional go/no-go task with fearful, happy, and neutral faces. Here, reaction times were slower when the targets were fearful faces, suggesting that the fearful targets created an aversive state. Negative go trials were accompanied by greater amygdala activation, and the degree of activation correlated with response time, suggesting that stronger appraisal responses increased withdrawal behaviour. Participants also committed more false alarms when withholding responses to positive faces, suggesting that positive faces increased approach behaviour. These trials were associated with greater activity in the caudate nucleus, which is associated with impulse control, and activity in this region was negatively correlated with false alarm rate (Hare et al., 2005), suggesting that it improves response inhibition in the presence of positive stimuli. This is broadly reconcilable with Pessoa's predictions about the influence

of motivational states, which may influence executive functions to maximise reward or minimise punishment (Pessoa, 2009).

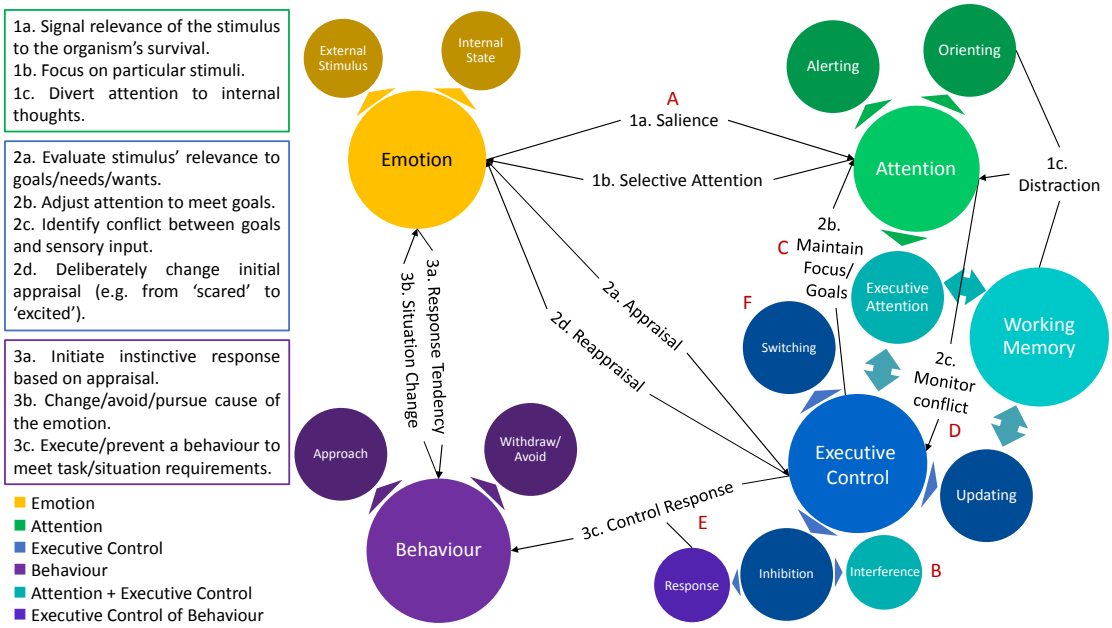

Figure 2.6: A model combining cognitive and emotional control functions into one processing stream. Boxes and arrows represent processes within the MCCE, and their colours correspond to the levels of processing denoted by the associated nodes. Bidirectional arrows and mixed colours (teal, indigo) denote overlapping constructs, while single arrows denote sub-functions. Letters denote processes which are hypothetically subject to bilingual adaptation according to the ACH. A = Salient Cue Detection, B = Interference Control, C = Goal Maintenance, D = Conflict Monitoring, E = Behavioural Inhibition, F = (Dis-)Engaging Task Schemas.

In summary, emotion influences the processing stream from the earliest perceptual levels, and this influence can be modulated through attentional gating, executive functions, and/or behaviour control. Evidence suggests that individual differences in general cognitive control functions are associated with better outcomes in emotion regulation tasks. As discussed in Section 2.3.2, the ACH (Green & Abutalebi, 2013) outlines eight cognitive control processes that are required to navigate a bi-/multilingual environment. These processes also map onto domain-general functions. While possible bilingual adaptations to these functions have been experimentally tested, the extent to which such experience-dependent adaptations influence emotion processing is unknown. To provide a basis for future research in this area, Figure 2.6 situates the ACH's functions within a combined model

of cognitive and emotional control.[5]

There is also evidence for specific mechanisms that coordinate cognitive and emotional processing, such as circuits that determine the behavioural relevance of a stimulus and enhance/inhibit its representation accordingly. These are more difficult to reconcile with studies of bilingual adaptations to executive functions. However, as within language development research (Section 2.2.3), studies of cognition–emotion interactions in childhood and adolescence can help disentangle the various mechanisms involved.

2.4.2 Cognition–Emotion Interactions in Development

Converging evidence suggests that in addition to interactions between emotion and executive control, there are specific mechanisms that regulate these interactions. This interpretation is supported by developmental studies which indicate that children are more affected by emotional stimuli in cognitive tasks compared to adults. According to the developmental mismatch hypothesis, sub-cortical brain structures that drive emotions and impulsive behaviour mature before the pre-frontal cortex that houses control processes needed to regulate them (see e.g. Casey et al., 2008; Somerville et al., 2010). The differences between children and adults in emotional compared to neutral cognitive control tasks can thus provide crucial insights into the mechanisms of emotion regulation.

Grose-Fifer et al. (2013), for example, compared adults (age 23-35) and adolescents (11-17) using pictures of happy, fearful and neutral faces. The groups completed two flanker tasks: in the non-emotional task they discriminated a central letter (e.g. HHSHH), while in the emotional task, they identified the facial expression of a centrally displayed face flanked by images of the same person with either the same or a different expression. In the non-emotional condition, younger adolescents were less accurate and slower in incongruent conditions than both older adolescents and adults, indicating that their conflict resolution ability had not fully matured. In the faces task, both younger and older adolescents were less accurate than adults and showed larger interference effects, although the main effect of congruence for both groups was significant. This indicated that the control mechanisms responsible

[5]The ACH's eighth control function, opportunistic planning, has been omitted as it is beyond the scope of this dissertation.

for relevance detection and adjusting stimulus salience discussed above develop later than conflict resolution. The flanker interference effect was significant for happy targets flanked by negative faces, but not vice-versa, suggesting that negative faces are more distracting than positive and/or produce higher conflict. These results mirror findings by Fenske and Eastwood (2003), who showed that flanker congruency effects are smaller when target faces have negative expressions relative to both neutral and positive faces, and also smaller for neutral than for positive faces. This corresponds with the theories discussed above which suggest that negative affect narrows the scope of attention.

Tottenham et al. (2011) compared children (mean age=9 years), adolescents (mean age=15) and adults (mean age=24) in the same emotional go/no-go task used by Hare et al. (2005, discussed in Section 2.4.1.3 above). They found that false alarm rates were significantly higher in general, and especially for both positive and negative emotional no-go faces, compared to neutral no-go faces. The false alarm rates decreased with age: adults committed fewer errors than adolescents, who committed fewer errors than children. False alarms to emotional no-go stimuli showed a steeper decline with age relative to overall false alarms. Adolescents had the fastest reaction times to emotional go faces, though adolescent males showed a speed-accuracy trade-off that led to slower RTs when emotional faces were the no-go condition. Like the studies by Goldstein et al. (2007) and Albert et al. (2010) discussed above, these findings suggest that emotion, regardless of valence, interferes with response inhibition. They also provide further evidence that the mechanism underlying the integration of emotion and response inhibition has a protracted development and that better performance by adults cannot simply be explained by better response inhibition ability. It is interesting to note that only positive faces increased false alarm rate in adults completing the same task (Hare et al., 2005, see above), while the fearful faces also produced more false alarms in this study. This indicates that the emotional stimuli interfered more with response inhibition overall for the children, perhaps reflecting a larger strain on attentional resources. However, positive targets with negative flankers produced more interference than vice-versa for similar age groups in the flanker task discussed above (Grose-Fifer et al., 2013). Differences between positive and negative stimuli appear to be task-dependent in children and adolescents, which suggests that it is

not merely behaviour modification mechanisms that mature late. It may be that the Tottenham et al. (2011) findings were due to two separate processes: in the positive condition, the children were not able to override the dominant approach behaviour due to less mature impulse control mechanisms in the caudate nucleus, but in the negative condition, the fearful faces interfered with children's' ability to detect the task-relevance, or *salience*, of the emotion as a no-go cue.

Hwang et al. (2014) provide neuroimaging evidence for the maturational delay of attention networks involved in boosting the salience of task-relevant stimuli and down-regulating emotional reactivity. They compared children (10-14 years old), adolescents (14-18), and adults (18+) performing an affective Stroop task similar to the task used by D. Mitchell et al. (2007) discussed above. In 'task trials', participants saw a positive, negative, or neutral distractor picture, followed by a grid of numerical digits. They were required to state how many digits were shown, regardless of the numerical value. Congruent trials matched in the number and value of the digits shown (e.g. four 4s), while incongruent trials mismatched (e.g. five 3s). In baseline 'view trials', participants saw a blank screen. There were no behavioural differences between the groups. However, children showed a greater activation of the dorsal ACC and middle and superior temporal gyrus in task trials compared with adolescents and adults, especially in incongruent trials. The authors suggest that the children required greater dorsal ACC activation to achieve comparable behavioural results to the older groups due to decreased functional integrity in this region. The children also had increased activation of the left amygdala in task trials with negative distractor images relative to neutral ones, indicating that they were less effective at priming task-relevant information and thus less able to suppress negative emotional distractors (Hwang et al., 2014). The left amygdala was also more active in both positive and negative emotional trials across all trial conditions for the children, suggesting generally greater emotional responsiveness. Finally, greater functional connectivity between the left amygdala and a network of frontal and temporal structures in negative task trials was evident in the children, and to a lesser degree in adolescents, relative to the adults. The authors suggest that this reflects the younger participants' attempts to compensate for increased amygdala responses with additional attentional control in these conditions.

Together, these findings point to a particular maturational delay in relevance-detection and selective attention mechanisms.

2.4.3 Summary

To summarise, cognitive control–emotion interactions appear to occur on three levels:

1. At the level of selective attention, emotional stimuli receive prioritised perceptual and attentional processing, but this is conditional on available attentional resources (and modulated by emotion-related states such as trait-anxiety).

2. One or more relevance-detection mechanisms appear to adjust the perceived salience of stimulus representations based on task demands, and these mechanisms appear to rely on connectivity between the amygdala, ACC, and task-dependent regions of the PFC.

3. At the level of executive functions, positive or negative stimuli (or emotional states) generate corresponding emotions that may interfere (if task-irrelevant) with other cognitive control processes. These emotions can also lead to approach or withdrawal behaviours that must also be over-ridden when they conflict with task demands.

As noted in Section 2.3.3.3, some researchers have recently argued that cognitive control itself can be seen as an emotional process, as specific features of a task may give rise to 'integral affect' (Inzlicht et al., 2015). For example, conflict between stimuli, action tendencies, or the present state and a desired future state gives rise to a transient aversive emotional reaction. Inzlicht et al. (2015) argue that this experience of conflict recruits control in the same way that emotional stimuli receive enhanced processing, and that this is why negative stimuli facilitate conflict resolution. Posner and Rothbart (1998) point out that the ACC is fundamentally involved in pain processing, and that early childhood soothing behaviours aimed at distracting the child from unpleasant states may be at the root of the integration of emotion and attention. Pessoa (2013) argues that beyond appraisal, the amygdala is fundamentally a relevance detector, and arguably most human behaviour is driven in some way by relevance, whether with regards to survival, goals, or current tasks.

2.5 Conclusion

Bilinguals appear to process emotional language differently in their L2, but only in subjective reports and more difficult language tasks. Emotion and language interact in children's language development and are closely linked to cognitive and social/emotional outcomes. Emotion and cognitive control interact at the level of attentional resources and cognitive changes, and the neural mechanisms that determine this resource allocation appear to mature late in adolescence. Speaking a second language also impacts cognitive control, and recent evidence suggests that this effect is also at the level of attentional control. Where is the link?

Bloom's (2001) Intentionality Model proposes that children acquire language through a complex process of interacting with the world around them, driven by the desire to communicate intentions and mental states, and require cognitive effort to integrate interaction and communication. In other words, emotional states both cause the child to initiate a verbalisation and make it more difficult to complete this verbalisation successfully. This process may help explain the intricate links between cognitive, emotional, social, and verbal outcomes in young children (Checa et al., 2008). More specifically, this may explain why linguistic, emotional, and cognitive processes all seem to include an overlapping neural basis in the ACC and dlPFC.

As outlined by Green and Abutalebi (2013), the process of learning a second language, whether simultaneously to a first language or as an adult, and using it effectively to communicate, places an additional demand on these networks. However, neither the process of language learning nor the context of L2 communication is affectively neutral. The desire to express mental states and interact with others is the same in any language, although different social and learning contexts may moderate this desire. The demands of using a second language, however, require additional cognitive effort, leading to effortful integration similar to that observed by Bloom in pre-verbal infants. It therefore stands to reason that if being bilingual leads to neural adaptations in the control processes needed to communicate, and if these control processes involve domain-general mechanisms, the adaptations would influence emotional processes as well.

A promising starting point for investigating this hypothesis is at the level of executive attention, which has been implicated across all three areas of research

reviewed here, particularly with regard to relevance detection. Two interesting findings that may begin to explain the compelling bilingual decision making effects described in the Introduction are the fact that emotion regulation circuits moderate loss aversion (Sokol-Hessner et al., 2012) as well as moral judgements (Decety et al., 2011). However, before such questions are answered, detailed information about the relationship between cognitive control, emotion processing, and bilingual experience is needed.

I don't want to be at the mercy of my emotions. I want to use them, to enjoy them, and to dominate them.

— Oscar Wilde,
'The Picture of Dorian Gray', pp. 111.

3 Study 1

Contents

3.1 Background 47
3.1.1 Cognitive Control of Emotion 48
3.1.2 Experimental Measures of Emotion–Cognition Interactions 50
3.2 The Present Study 55
3.2.1 Research Questions 55
3.2.2 Study Design 56
3.2.3 Hypotheses 58
3.3 Methodology 60
3.3.1 Design 60
3.3.2 Participants 61
3.3.3 Materials 62
3.3.4 Procedure 63
3.3.5 Data Analysis Plan 63
3.4 Results 64
3.4.1 Pre-Processing and Stimulus Validation. 64
3.4.2 Main Experiment 66
3.5 Discussion 71
3.6 Conclusion 74

3.1 Background

Empirical studies have consistently linked multilingualism to advantages in cognitive control, even in non-linguistic tasks (Bialystok, 2017). As part of their Adaptive Control Hypothesis (ACH), Green and Abutalebi (2013) suggest that

these advantages arise because the mental processes that set parameters for task completion must adapt in response to different language contexts. The demands of these contexts, Green and Abutalebi (2013) suggest, may lead to neural adaptations that make the required control processes more efficient overall.

Neither interactional contexts nor internal parameters for task completion are affectively neutral. However, while a growing body of research links bilingualism to differences in performance on measures of executive control ranging from conflict monitoring (Costa et al., 2008) to task switching (Hernández et al., 2013) to deployment of attention (Sorge et al., 2017), there is little research investigating its influence on control of emotions. Ecologically valid investigation of these posited cognitive advantages must consider their emotional implications as well. Indeed, recent research suggests that subjective emotional impact (Dewaele, 2008), moral judgements (Geipel et al., 2015), and decision biases (Keysar et al., 2012) are all influenced by the language used. The cognitive explanations for these findings are, however, not yet clear. In fact, no research to date has investigated the role of emotion in bilingual cognitive control. The aim of this study is, therefore, to determine whether bilinguals differ from monolinguals in their cognitive control performance in the presence of emotional distractors. The following sections explain how cognitive control is deployed to regulate emotion and how this may influence performance on executive tasks.

3.1.1 Cognitive Control of Emotion

According to Ochsner et al.'s (2012) Model of the Cognitive Control of Emotion (MCCE), emotion control can be broken down into five strategies or processes:

1. Situation selection: voluntarily avoiding situations that may give rise to unwanted emotions.

2. Situation modification: voluntarily leaving emotional situations.

3. Attention deployment: controlling the focus of attention toward goal-relevant/away from emotion-inducing input, *e.g., selective attention.*

4. Cognitive change: modifying the initial emotional response ('emotion generation') that accompanies the *appraisal* of a stimulus or state (i.e., the determination of its valence and arousal) to a more favourable one; *e.g., reappraisal.*

5. Response modulation: inhibiting an emotional response when it gives rise to inappropriate action tendencies; *e.g., expressive suppression* (controlling emotional facial expressions).

The model's first two strategies (items 1–2 above) involve avoiding the cause of an emotion, and thus precede the cognitive control stages. Attentional selection, cognitive (re-)appraisal, and response selection, then, make up the three stages at which cognitive control can be deployed to regulate emotion. These emotion regulation skills are highly correlated with non-affective executive function measures: reappraisal and expressive suppression, for example, can be predicted by inhibition and working memory updating (Schmeichel & Tang, 2014). In other words, neural mechanisms for emotion control overlap with those for domain-general cognitive control (Pessoa, 2008).

The way emotions interact with general control functions varies depending on valence, arousal, and task demands, however. Emotional stimuli tend to bias processing in their favour because emotion is tied to biological salience (i.e., relevance to the organism's survival), and thus captures enhanced attention (Pessoa, 2009). Thus, task-relevant emotional stimuli improve task performance, while task-irrelevant stimuli disrupt it. However, when task processing demands are high, task-irrelevant emotional stimuli have a reduced effect on performance relative to less demanding conditions (Fox et al., 2012; D. Mitchell et al., 2007). Research on the impact of emotional stimuli on cognitive control tasks suggests that there is a unique mechanism for relevance detection which controls attention and enhances the representation of stimuli or features relevant to the task (Egner & Hirsch, 2005; D. Mitchell et al., 2007). Assuming emotion and executive control share the same capacity-limited processing resources (the 'dual competition model', Pessoa, 2009), this relevance-detection mechanism involves both emotion regulation (selective attention, cognitive reappraisal of emotional distractors as irrelevant) and executive control (updating working memory representations of relevant stimuli, controlling interference from irrelevant ones). In addition, positive stimuli tend to induce approach (reward-seeking) states, while negative stimuli induce withdrawal (avoidance) states (Lang et al., 1997). Therefore, in tasks where responses are made, response modulation/inhibition is needed to override mismatching valence-dependent action tendencies.

In other words, cognitive control of emotion at the different levels of Ochsner's (2012) model (i.e. attention deployment, cognitive change, and/or response selection) interacts with different levels of domain-general cognitive control (i.e. attentional selection, interference suppression, and response selection), depending on the attentional and executive demands of a task, the relevance of emotional dimensions, and the valence of the emotion involved.

Bilingualism has been demonstrated as a predictor of enhanced ability to coordinate several control functions (Costa et al., 2009; Costa et al., 2008) and maintain task goals in the presence of interference (Colzato et al., 2008). These skills may translate to a bilingual advantage in regulating emotional distractors during non-linguistic cognitive control tasks. A task that explicitly manipulates attentional and executive demands, as well as the task-relevance of emotional information, is necessary to test this hypothesis.

3.1.2 Experimental Measures of Emotion–Cognition Interactions

3.1.2.1 The Flanker Task

The Eriksen flanker task (Eriksen & Eriksen, 1974) and its variants are commonly used to measure participants' ability to focus on relevant stimuli in the presence of distractions. In the traditional version of the task, an arrow is presented in the centre of a computer screen, flanked to the left and right by additional arrows. These 'flankers' point either in the same direction as the central cue (e.g. ←←←←←, the 'congruent' condition) or in the opposite direction (←←→←←, 'incongruent'). Participants must indicate the direction of the central arrow. Reaction time (RT) measures consistently show faster responses to congruent and slower responses to incongruent trials (the 'flanker effect'). This suggests that flanker features are integrated into the task when they are helpful and require inhibition when unhelpful, leading to a behavioural cost (Eriksen & Eriksen, 1974).

Bilingual participants have demonstrated advantages over monolingual controls in flanker task studies, although results are inconsistent: while some report a reduced flanker effect, some report faster overall RTs in congruent as well as incongruent conditions (see Costa et al., 2009). As Costa et al. (2009) point out, this may be due to differences in experimental design, such as the frequency of

congruent/incongruent trials and their order (blocked vs. random), which change the executive demands of the task. Thus, a reduction in the flanker effect may reflect bilinguals' enhanced interference suppression, while faster overall RTs instead reflect more efficient selective attention to relevant stimulus dimensions.

The flanker effect has also shown susceptibility to emotional interference/facilitation. In a verbal stimulus variant of the task, adult participants were asked to name the colour of a target word. The same word was displayed directly above and below it, in the same (congruent) or a different colour (incongruent). Words with both positive and negative emotional valence *reduced* response latencies to incongruent trials in comparison to affectively neutral words, suggesting that emotional meaning facilitated interference suppression of the colour dimension (Kanske & Kotz, 2010b, 2011b, 2012; Kanske et al., 2011). The authors interpret this effect as a prioritisation of conflict resolution in emotional situations, driven by a portion of the ventral ACC only active in incongruent emotional trials (Kanske & Kotz, 2010a, 2011b) and reflected in the amplification of the N200, the earliest conflict-related ERP, by emotional valence (Kanske & Kotz, 2010b, 2011a). However, arousal varied between the positive/negative and neutral stimuli in these studies; it is thus not clear whether the effects are due to valence or arousal. Furthermore, conflict occurred only in the colour dimension, so colour-incongruent trials still matched on the emotion dimension, potentially contributing to the facilitation of conflict resolution in emotional relative to neutral colour-incongruent trials. Finally, tasks involving verbal stimuli introduce language-processing confounds when testing bilingual participants, the aim of the current study.

In another adaptation of the flanker task, Grose-Fifer et al. (2013) used pictures of faces with happy, fearful and neutral facial expressions. Adults (23–35) and adolescents (11–17) completed two flanker tasks: a non-emotional task (discriminating a central letter, e.g. HHSHH) and an emotional task (detecting emotional expressions on the central face, e.g. ☺☹☺). The older adolescents (>15) performed comparably to adults in terms of RT and accuracy in the letter task, indicating that non-emotional interference suppression matures by this age (Rueda et al., 2004, a commonly accepted conclusion, see e.g.). In the faces task, all adolescents were less accurate and showed larger flanker interference effects in their RTs that adults (although the main effect of congruence was significant for both groups).

RTs were slower to happy target faces with fearful flankers, but not vice-versa, suggesting that negative faces are more distracting than positive. This is in line with Pessoa's (2009) conclusion that fear-related stimuli are more salient, and thus more disruptive to cognitive processes. The difference in performance between the neutral and emotional tasks for the adolescent group supports the theory that neural mechanisms for selective attention to emotional stimuli are later to mature than those involved in interference suppression in the neutral task. A limitation of the above task by Grose-Fifer et al. (2013) is that emotion (operationalised by facial expression) is the response dimension, and thus task-relevant. Previous studies have found that both task-relevance (D. Mitchell et al., 2007) and the attentional demands of the task (Pessoa et al., 2002) strongly influence the effects of emotional stimuli on both neural and behavioural responses. Zhou and Liu (2013) manipulated the response dimension in a flanker task so that participants responded to the colour (red/blue), gender, or emotion (happy/fearful) of faces which displayed two of the three dimensions (either colour and gender or gender and emotion). They found that congruency effects for each respective dimension were significant only when it was the relevant response dimension; in other words, attention was required to process emotional facial expressions. (Notably, the effect of task-relevance was strongest for colour, suggesting that this dimension was most strongly modulated by attention.) As a result, it is difficult to tell whether Grose-Fifer et al.'s (2013) findings are due to emotional interference with the selective attention or the interference suppression demands of the flanker task.

An additional limitation of the flanker task is that it does not provide an index of the effects of emotional stimuli on response selection. As noted above (Section 3.1.1), emotions are known to induce particular response tendencies—positive stimuli bias approach behaviour, while negative stimuli lead to avoidance behaviour (Lang et al., 1997). These tendencies may interact with the effects of positive and negative flankers and targets, leading to, e.g., slower RTs in negative and faster RTs in positive conditions. Findings from the go/no-go task can help test the effects of emotional stimuli on response selection.

3.1.2.2 The Go/No-Go Task

Like the flanker task, the go/no-go task has been used to measure the influence of emotional stimuli on relevance detection. However, rather than resolving interference between competing visual stimuli, the participant must override one dominant response tendency—making a response, such as pressing a button, in the presence of a frequent 'go' stimulus—when a less frequent 'no-go' stimulus is presented. The go/no-go task is thus considered a measure of behavioural response inhibition, which is highly correlated with the MCCE's construct of behaviour modification (Ochsner et al., 2012; Schmeichel & Tang, 2014).

To test the overlap of response inhibition and behaviour modification, Tottenham et al. (2011) compared children (mean age=9 years), adolescents (mean age=15) and adults (mean age=24) in a go/no-go task using happy, neutral, and fearful faces as response targets. The participants were asked to respond to one type of facial expression while withholding responses to the other two. False alarms (responses to 'no-go' facial expressions) were significantly more frequent when the no-go faces were happy or fearful compared to neutral. These false alarm rates decreased with age: children committed more errors than adolescents, who committed more errors than adults. False alarms on emotional no-go faces showed a steeper decline with age relative to overall false alarms. In addition, adolescents had the fastest RTs to emotional faces in go trials, and a speed-accuracy trade-off was observed in adolescent males whereby RTs were slower when the no-go condition was emotional. These findings suggest that emotion, regardless of valence, interferes with response inhibition. As in the case of interference suppression (see Section 3.1.2.1), the mechanism underlying the integration of emotional and behavioural response inhibition is late to mature, and better performance by adults cannot simply be explained by overall response inhibition ability. It is interesting to note that, in another sample of adults completing the same task described above (Tottenham et al., 2011), only happy faces increased false alarm rates (Hare et al., 2005). This is the result predicted by theories that positive emotion biases approach behaviour. The difference between the studies may indicate that emotional stimuli interfered more with children's response inhibition overall—perhaps reflecting a larger strain on attentional resources due to the coordination of two simultaneous demands, detecting the response dimension and appraising the salience of the

represented emotion. It may be that Tottenham et al.'s (2011) findings were due to two separate processes: negative valence interfered with childrens' ability to detect the task-relevance of the expression as a stop cue and withhold their response, while positive valence interfered with their ability to override the prepotent approach (go) behaviour due to less mature impulse control mechanisms. However, as emotion was again the task-relevant response dimension, it is difficult to separate these processes.

Unlike Zhou et al.'s (2013) findings with the flanker task, changing the response dimension of the face go/no-go task from emotional expression to gender still results in a significant effect of the (now irrelevant) emotion: Zhang and Lu (2012), for example, found shorter response latencies and higher accuracy on go trials when participants responded to the gender of both positive (happy) and negative (angry and fearful) faces compared to neutral faces. They found no differences between emotional and neutral faces in no-go trials. However, the proportion of go/no-go trials was equal, reducing the inhibition demands of the task. Forced-choice gender identification tasks (i.e. without a no-go condition) with emotional faces, meanwhile, show higher accuracy only for happy faces, but faster RTs only for fearful faces (D. Mitchell et al., 2007).

Bilingual advantages have not been reported on the go/no-go task, unlike the flanker task (see Bialystok, 2017, for a discussion of bilingual performance on both tasks). However, neural differences between monolinguals and bilinguals have been observed in brain networks related to language control in both fMRI studies (Fernandez et al., 2013) and ERP studies (Costumero et al., 2015) using the go/no-go task. Notably, Costumero et al. (2015) observed greater modulation of the so-called 'salience network' (comprising the dorsal anterior cingulate cortex, inferior frontal cortex, and frontal operculum) in bilinguals relative to monolinguals in response to no-go trials. This suggests that the bilinguals used a different strategy than the monolinguals to complete the same task, perhaps reflecting more efficient detection of the salient response dimension (colour of the stimuli) in this group. This has interesting implications for how bilinguals' performance may differ from monolinguals' on an emotional version of the go/no-go task where emotion is not the response dimension, and where the demands of coordinating response inhibition and interference from emotional stimuli may reveal group performance differences not observed in simpler versions of the task.

3.2 The Present Study

The above findings demonstrate that the emotion regulation mechanisms identified by Ochsner (2012) are active in laboratory measures of cognitive control: 1) At the perceptual level, selective attention gates the stimuli or stimulus attributes that enter awareness; 2) at the cognitive level, emotional stimuli receive prioritised processing, enhancing or interfering with performance depending on task demands; 3) at the behavioural level, emotional (particularly positive) stimuli interfere with response control.

The language control processes identified by Green and Abutalebi (2013) heavily overlap with these mechanisms: bilinguals must, for example, deploy selective attention to detect salient cues that signal the need to change languages and ignore interference from the environment, use executive control to switch between languages, maintain communicative goals, and plan utterances, and inhibit a prepotent behaviour (e.g. speaking the native language) in order to switch languages. Bilingual social interactions also require sophisticated emotion perception and regulation abilities alongside language control. This implies that any neural adaptations in domain-general control processes resulting from bilingual experience must involve emotional as well as cognitive control. A visualisation of how these processes overlap in attention, executive control, and behavioural control mechanisms is included in Figure 2.6 in Chapter 2 of this dissertation.

3.2.1 Research Questions

While the impact of bilingualism on executive control functions (and, to a lesser degree, on attention) has been extensively studied (see Bialystok, 2017, for a recent review), the role of emotion relative to these functions has not been investigated. This study addresses this gap with the following questions:

1. Are bilinguals better at regulating task-irrelevant affective information relative to monolinguals?
2. How do bilinguals differ from monolinguals in their processing of irrelevant affective information at the level of:
   (a) selective attention to task-relevant information?
   (b) interference suppression?

(c) behavioural response inhibition?

3.2.2 Study Design

To address the above questions, task designs, stimuli, and response dimensions from the experimental measures of cognition–emotion interactions discussed in Section 3.1.2 were modified to address the limitations identified. In particular, the experimental task was optimised to investigate whether—and if so, how—bilinguals differ from monolinguals at distinguishable levels of the processing hierarchy.

3.2.2.1 Task-relevance of emotional stimuli

Task-relevance is a determining factor in the disruption or facilitation of cognitive control processes by emotional stimuli (Pessoa, 2008, 2009). (See also Section 2.4.1 of the Literature Review.) To determine how bilinguals differ from monolinguals in their regulation of irrelevant emotional distractors, this study used a flanker task with neutral faces as response targets and neutral or emotional faces as flankers. Participants identified the gender of the target face. Thus, unlike the studies reviewed in Section 3.1.2.1, emotion was not a response-relevant dimension of the stimuli. This allowed a direct comparison between the effects of irrelevant emotional flankers in gender-congruent conditions, with no conflict in the task-relevant response dimension, and in gender-incongruent conditions, where the task-relevant feature of the flankers conflicted with that of the target.

3.2.2.2 Task complexity

While a flanker task provides a well-validated measure of interference suppression, indexed by the RT difference between congruent and incongruent trials, Pessoa (2008) suggests that more complex tasks may be necessary to avoid ceiling effects when measuring cognition–emotion interactions in adults. As noted in Sections 3.1.1 and 3.1.2.2, coordination of multiple task demands is also a common feature in tasks that reveal bilingual advantages. To increase the complexity of the flanker task in this study, a go/no-go component was added. Participants were thus required to not only suppress interference from flankers [a 'cognitive change' demand in Ochsner's (2012) taxonomy], but also to inhibit responses to one target type (a 'behaviour modification' demand). This also allowed the measurement of emotional

valence effects on response inhibition through false alarm rates (incorrect responses on no-go trials when no response is required).

A combined FNG (affectively neutral) paradigm has previously been used in both ERP (Brydges, Clunies-Ross et al., 2012) and simultaneous EEG–fMRI (Baumeister et al., 2014) studies, though only with monolingual participants. The studies revealed separate neural processes associated with attention [no-go N2 in conjunction with deactivation in the default mode network (DMN)—thought to underlie goal-oriented action and attention allocation], interference suppression (longer latencies and different scalp topographies in both N2 and P3 in incongruent compared to no-go conditions), and response inhibition (no-go P3 in conjunction with activation in the left anterior insula and inferior frontal gyrus—linked to response inhibition in numerous studies). This evidence for dissociable neural processes underlying attention, interference suppression, and response inhibition further validates the combined task choice for our proposed monolingual and bilingual samples. In addition, the latencies for the reported ERPs provide insight into the processing sequence: for example, the observed N2 negativities in the Brydges, Clunies-Ross et al. (2012) study peaked between 350–400 ms for incongruent–congruent conditions, while N2 negativity in no-go conditions peaked at 250–300 ms, suggesting that response inhibition occurs earlier than interference suppression[1]. While the stimuli used in the present study differ from those used by Brydges, Clunies-Ross et al. (2012) and Baumeister et al. (2014), their findings provide a basis for analysing the level (attention, interference suppression, or response inhibition) at which any observed differences between monolinguals and bilinguals might occur.

3.2.2.3 Emotional valence

Both negative and positive flankers were used in the present study, with arousal held constant and minimal. Including both positive and negative flankers made it possible to identify whether the task-irrelevant emotion dimension is evaluated (*appraised*) or ignored: as discussed in Section 3.1.1, positive and negative emotion can have opposite (or 'bipolar'; Lewis et al., 2006), effects on response tendencies

[1]The authors note that this finding is at odds with commonly accepted models of inhibition (i.e. Friedman & Miyake, 2004), but is feasible given the relative complexity of each process: while response inhibition involves stopping all responses, interference suppression requires response selection whilst also inhibiting distractors (Brydges, Clunies-Ross et al., 2012).

(approach/withdrawal). (See also Section 2.4.1.3 of the Literature Review.) If appraisal occurs, the prepotent tendency to approach positive stimuli and avoid negative ones should manifest in higher false alarm rates in positive and higher miss rates in negative conditions. If, however, the effect of both positive and negative flankers is the same, then it can be inferred that emotion regulation has taken place at an earlier processing stage.

3.2.3 Hypotheses

The effects of emotion at different levels of cognitive processing have, thus far, primarily been investigated in monolingual participants (or participants whose language background was not reported). However, reported bilingual advantages from affectively neutral studies allow some predictions for the present study. Bilinguals have shown an advantage in selective attention to task- or goal-relevant stimulus features and reactive inhibition of irrelevant information (Colzato et al., 2008). Selective attention to relevant input is also required to navigate bilingual environments (Green & Abutalebi, 2013). Furthermore, although behavioural differences between bilinguals and monolinguals are not typically found in the go/no-go task, Costumero et al.'s (2013) findings suggest that bilinguals process no-go stimuli differently from monolinguals at a neural level. The additional task complexity from flankers and emotional valence in our study may result in detectable behavioural differences.

> Hypothesis 1: A main effect of congruence (faster RTs and higher accuracy in gender-congruent compared to -incongruent flanker trials) was expected across all conditions for both groups.

As described in Section 3.1.2.1, both monolinguals and bilinguals consistently show a behavioural cost due to interference suppression demands when flankers conflict with targets on the response dimension, leading to a 'flanker effect' (slower RTs in incongruent trials).

In line with previous findings of a bilingual advantage in suppressing interference from task-irrelevant stimuli (Costa et al., 2009), we also expected group effects.

> Hypothesis 2: Bilinguals were predicted to have faster RTs and higher accuracy overall, as well as smaller flanker effect (faster RTs and higher accuracy in incongruent trials compared to monolinguals), manifesting in (a) a main effect of group and (b) a group × congruence interaction.

Both positive and negative flankers were predicted to interact with suppression of gender interference, resulting in valence effects on both RT and accuracy. Happy faces generally produce faster reaction times and higher accuracy, but also more false alarms, due to the prepotent tendency to approach positive stimuli (Hare et al., 2008).

> Hypothesis 3: We expected slower RTs in positive and negative compared to neutral emotional trials (a main effect of valence on RT). Accuracy in go trials, and false alarm rates in no-go trials, were both expected to be higher in positive compared to negative and neutral trials as well.

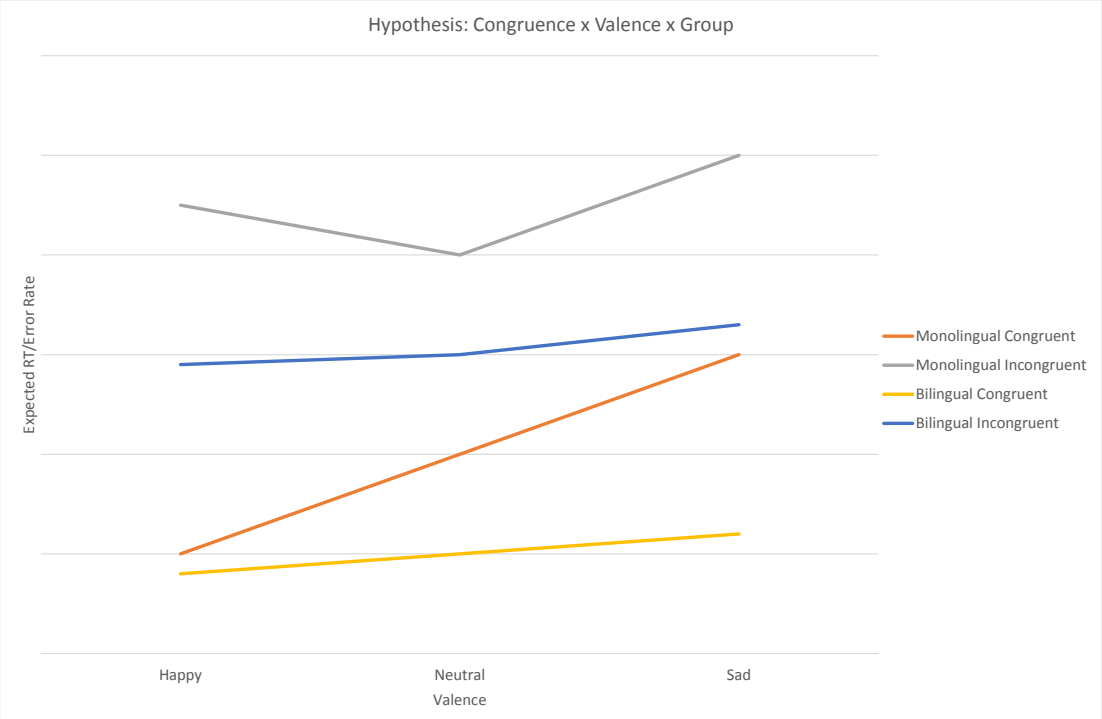

Figure 3.1: Hypotheses: The proposed direction of the Group × Congruence × Valence interaction effect on RTs and accuracy.

> Hypothesis 4: Crucially, bilinguals were expected to demonstrate more efficient selective attention to gender, resulting in faster RTs in congruent and slower RTs in incongruent conditions with no effect of valence. They were also expected to be more accurate and commit fewer false alarms in emotional trials. Monolinguals, in contrast, were expected show more interference from positive and negative gender-incongruent flankers, resulting in slower RTs and lower accuracy, particularly in incongruent negative trials. In congruent trials, monolinguals were expected to be fastest and most accurate with positive flankers.

These hypotheses are summarised by Figure 3.1.

3.3 Methodology

3.3.1 Design

Participants completed a combined FNG task. They viewed three faces side-by-side in the centre of a computer screen (see Figure 3.2) and were instructed to respond when the perceived sex (male/female) of the central face matched their assigned go condition. A $2 \times 2 \times 3$ design with language groups (monolingual/bilingual) as a between-subjects factor and flanker congruence (flankers of the same or opposite sex) $\times$ flanker valence (happy/sad/neutral expression) as within-subject factors was used. Two fully crossed stimuli lists (one with female faces as the go condition, one with males) were constructed. Each model occurred once in each condition as a target (6 times total, always neutral) and once in each condition as a flanker (6 times total, twice with each facial expression). The list was divided into 18 blocks of 12 trials (18 trials per condition), and no model occurred twice within the same block, to control repetition effects. This design is summarised in Table 3.1.

Pilot data showed that no-go accuracy was near ceiling when the proportion of

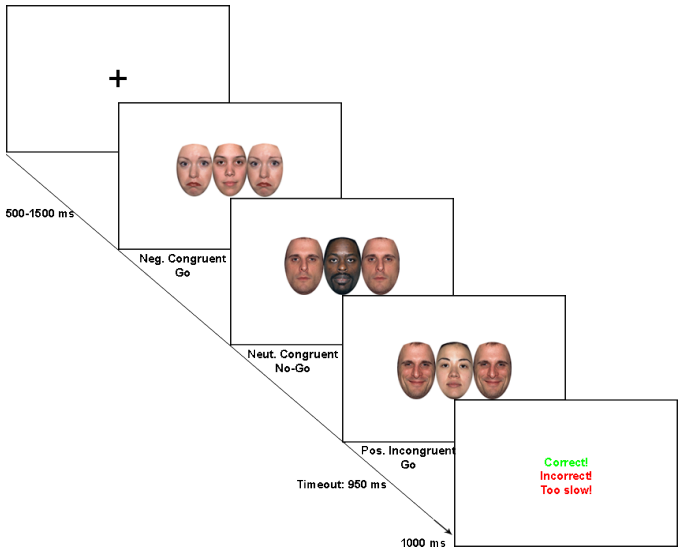

Figure 3.2: Experimental trial sequence. (Go condition = female).

go and no-go trials was equal (mean=0.97, pilot N=2). The % of no-go trials was adjusted to 33%, following the convention in the literature, by replacing two no-go targets in each block with go targets. No-go accuracy was reduced in subsequent pilots (mean=0.85, pilot N=2). The number of incongruent flanker trials was also reduced to 42% by replacing one incongruent flanker in each block with a congruent

Table 3.1: Experimental design. Go = Respond, No-Go = Withhold response, Target = central face, Flanker = peripheral faces, Flanker Expression = Facial expression of flankers.

Group	Target	Flanker	Flanker Expression
Monolingual	Go (66%)	Congruent (58%)	Calm
			Happy
			Sad
		Incongruent (42%)	Calm
			Happy
			Sad
	No-Go (33%)	Congruent (58%)	Calm
			Happy
			Sad
		Incongruent (42%)	Calm
			Happy
			Sad
Bilingual	Go (66%)	Congruent (58%)	Calm
			Happy
			Sad
		Incongruent (42%)	Calm
			Happy
			Sad
	No-Go (33%)	Congruent (58%)	Calm
			Happy
			Sad
		Incongruent (42%)	Calm
			Happy
			Sad

flanker—while a lower proportion of incongruent flankers has been shown to increase the flanker effect, a bilingual advantage has been observed only at a nearly balanced congruent/incongruent ratio (Costa et al., 2009). In the final lists, each model of the go sex was used 8 times as a target and 7 times as a flanker, while each model of the no-go sex occurred 4 times as a target and 5 times as a flanker.

3.3.2 Participants

The target populations for this study were English monolinguals and German-English bilinguals who learned English as a second language in school, all adults between 18 and 35 with no psychiatric history and no fluency to B1 level or above in a second/third language. A convenience sample was drawn from students and staff at the University of Oxford and the London School of Economics.

Participants for the study were recruited through posters in department and college common rooms, posts on social media, and personal contacts, as well as

through the Oxford Psychology Research (OPR) participant recruitment scheme. Inclusion criteria were listed in the study description, and participants were asked to only sign up if they met all of the criteria.

The final sample consisted of 24 German-English bilinguals and 24 English monolinguals between 18-35 years of age. The bilinguals were German native speakers (German or Austrian) who had learned English in school, reported a very high level of English fluency, and had studied or worked in an English-speaking country for at least six months. The participants were University students and to comply with admissions standards, they must pass an international English exam at a high standard (minimum of 7.0 for the IELTS, 100 for the TOEFL, or 185 for the CPE), ensuring a baseline level of proficiency at a minimum of high-intermediate ability. Language background data, including age of acquisition (AoA), relative frequency of use, and self-rated proficiency in speaking, listening, reading, and writing was collected using the Language History Questionnaire (Li et al., 2014). The Meara Yes/No vocabulary test (Meara & Miralpeix, 2015) and the German version of LexTALE, a similar vocabulary size test (Lemhöfer & Broersma, 2012), were used to establish a more objective assessment of baseline language proficiency. The Yes/No test has been shown to provide a reliable estimate of overall language proficiency and is quick and easy to administer (P. Meara, personal communication). It was thus the most appropriate proficiency measure within the time constraints of the study.

3.3.3 Materials

36 pictures of faces (18 female) were selected from the NimStim picture set (Tottenham et al., 2009). Images of their low-arousal (closed mouth) happy, sad, and calm facial expressions were selected (because of potential negativity bias in neutral faces, e.g. Somerville et al., 2004, calm facial expressions were used as neutral stimuli). Only happy/sad faces whose recognisability rating was >60% were used (this was impossible for the more ambiguous calm expressions). Piloting revealed that accuracy was at ceiling in the go/no-go task (mean=0.99 for both congruent and incongruent, pilot N=3). The images were therefore cropped so that just the faces were visible, which reduced accuracy slightly.

3.3.4 Procedure

At the start of each trial, a black fixation cross (+) appeared in the centre of the screen for 500–1500 milliseconds. This inter-stimulus interval varied randomly to prevent participants from falling into a response rhythm. Three faces were then presented close together in the centre of the screen, subtending approx. 4.8 deg visual angle (297×160 pixels, 99×160 each). The target face was always neutral. A second face was presented as a flanker on both sides of the target. The flanker could be either neutral (calm facial expression), sad, or happy and either congruent (same sex as the target) or incongruent. The small size was intended to slow disambiguation between target and flankers, following Alguacil et al. (2013). Piloting showed that performance was at ceiling when the faces were larger (large mean=0.99, pilot N=3 vs. small mean = 0.94, pilot N=4). Participants were instructed to press the Space key (marked with a blue sticker) if the centrally presented face appeared to be male/female (go condition alternated between participants).

Following the main experiment, participants provided valence ratings on a 7-point Likert scale (1=Very Negative, 7=Very Positive) for each face and completed a forced-choice speeded sex-identification task to ensure that subjective ratings and sex identifications matched the assumed values.

The experiment was administered using Psychophysics Toolbox extensions (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli, 1997) for MATLAB R2016b, running on a Dell Precision m4600 laptop with a 15.6-inch display.

3.3.5 Data Analysis Plan

(Generalised) linear mixed effects models [(G)LMEs] were used to make inferences about the collected data. Compared with the more widely used ANOVA, the LME is more effective for handling missing data (e.g. participant errors) and can account for the random variability due to repeated measures from both participants and items in one model (Baayen et al., 2008). Both the participants (convenience sample of students) and items (sample of faces) were subsamples intended to reflect a larger population and thus constitute random factors of variability. Since all participants responded to all items, these random factors are fully crossed.

A hypothesis-driven forward model-selection approach was used to find the most parsimonious random effects structure for the uncontrolled variability in the

data (Bates et al., 2015). Once the random effects were specified, the fit of a model containing the hypothesised fixed effect (a three-way interaction between the independent variables group, congruence, and valence) was tested using a Likelihood Ratio Test (LRT). Following the recommendations of Luke (2017), once the final model was selected, the significance of the independent variables was tested using a Satterthwaite approximation for degrees of freedom.

Full details of the data analysis process, including power calculations, model fitting and selection, and the respective scripts, can be found in Appendix A.

3.4 Results

3.4.1 Pre-Processing and Stimulus Validation.

A total sample of 52 participants (27 monolingual, 25 bilingual) took part in the experiment. Of these, three monolinguals were excluded from analysis due to previous fluency in other languages (two reported C1 proficiency in French, one childhood exposure to Gujarati). One bilingual's data was excluded due to programme malfunction, and one due to a subsequently diagnosed (and voluntarily disclosed) mental health condition.

The results of the sex identification task were analysed and faces with accuracy rates below 75% were removed from analysis to avoid the undue influence of poorly classified targets. 13 pilot participants completed gender identifications for each face in each emotional condition. This pilot task was reduced to neutral faces only in the main experiment to reduce the total experiment time. Results are shown in Table 3.2. One model (Model number 8 from the NimStim set, Tottenham et al., 2009) was classified with below 75% overall, resulting from $< 70\%$ accuracy in the sad condition and $< 75\%$ in the neutral condition, and so all trials which contained Model 8 as a target, neutral flanker, or negative flanker were removed. Model 12 was also classified with less than 70% accuracy in the sad condition and was therefore removed as a flanker. This resulted in a total of 570 trials being removed. Valence ratings were analysed to ensure the perceived flanker valence matched the expected valence. Mean ratings for each condition are shown in Table 3.3. Ratings generally matched the expected valence of the faces, although happy faces were rated as more different from neutral (3.5) than were sad faces. However, there

Table 3.2: Participants' mean accuracy on sex identification task.
* Removed from analysis.

Model	Mean Accuracy (SD)	Model	Mean Accuracy (SD)
1	0.986 (0.116)	23	0.961 (0.196)
2	0.986 (0.116)	24	1 (0)
3	0.919 (0.275)	25	0.96 (0.197)
5	0.806 (0.399)	26	0.987 (0.115)
6	0.987 (0.115)	27	1 (0)
7	0.945 (0.229)	28	0.986 (0.116)
8*	0.723 (0.451)	29	0.986 (0.116)
9	0.947 (0.225)	30	0.973 (0.163)
10	0.96 (0.197)	31	0.987 (0.115)
11	0.907 (0.293)	32	0.986 (0.116)
12	0.918 (0.277)	33	0.986 (0.116)
13	1 (0)	34	0.986 (0.116)
14	0.849 (0.36)	35	0.937 (0.246)
15	0.973 (0.162)	36	0.972 (0.165)
16	0.987 (0.115)	37	0.917 (0.278)
17	0.986 (0.117)	38	0.973 (0.163)
18	0.959 (0.199)	39	0.973 (0.163)
19	0.973 (0.162)	40	0.977 (0.15)
20	0.96 (0.197)	41	0.987 (0.115)
21	0.973 (0.163)	42	0.92 (0.273)
22	1 (0)	43	0.946 (0.228)

was considerable variability in participants' individual ratings. This may have been due to different interpretations of the task instructions (participants were instructed to rate "how positive or negative" they considered the face on a 7-point scale). An additional random effects term was added to the LME to account for this variability between different flankers.

Table 3.3: Participants' mean valence ratings.

Emotion	Mean Rating (SD)
Sad	2.53 (1.02)
Calm	3.84 (0.85)
Happy	5.43 (0.91)

3.4.2 Main Experiment

Table 3.4 and Figure 3.3 provide a summary of the mean RTs on hit (correct go) trials across conditions. There was a highly significant main effect of Congruence on RT, $F(1, 5800) = 12.66, p < 0.001$. The main effect of Gender was also significant, $F(1, 43.63) = 7.76, p < 0.01$. However, as can be seen in Figure 3.4, this is qualified by a significant three-way interaction between Group, Congruence, and Flanker Valence, $F(2, 5951) = 5.81, p = 0.003$.

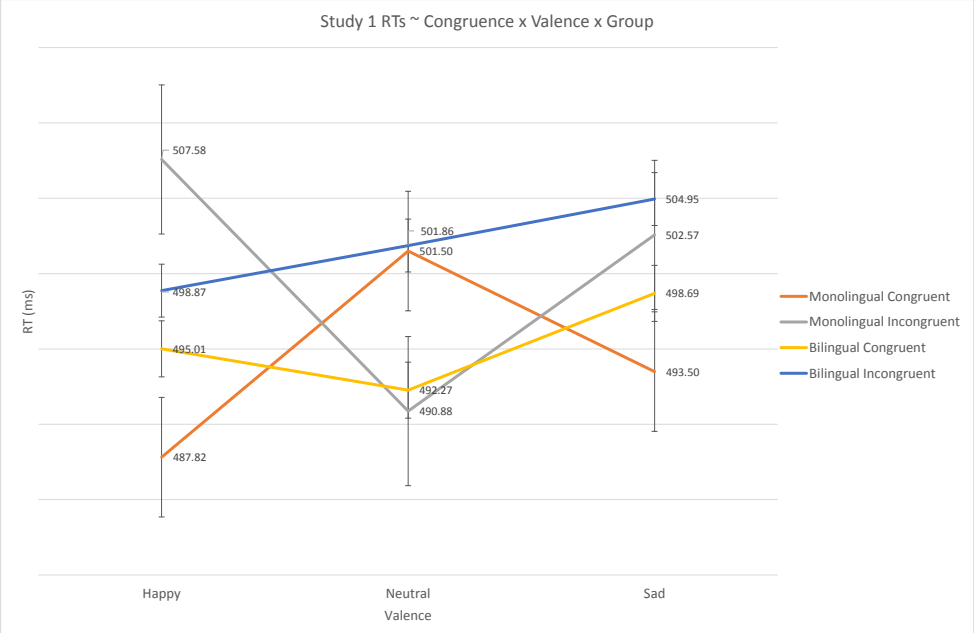

Figure 3.3: Visualisation of the relationships between Group × Congruence × Valence in Experiment 1, with experimental mean RTs in milliseconds. Error bars represent standard error of the mean.

The three-way interaction was driven by a significant Group × Congruence interaction in the positive emotional flanker condition, $t(6023) = -4.58, p = 0.005$. The interaction was not significant in the negative condition [$t(5862) = -0.34$]. Full model estimates are shown in Table A.6, and the ANOVA summary is shown in Table 3.5.

Simple main effects were analysed by changing the factor contrast coding for the LME, and *post-hoc* contrasts were analysed by changing the reference (intercept)

Table 3.4: Mean go accuracy, false alarm (FA) rates, and hit RTs in milliseconds from main experiment.

Group	Congruence	FValence	Hit Rate	SD	FA Rate	SD	Hit RT(ms)	SD
Monolingual	Congruent	Happy	0.98	0.14	0.07	0.26	486.05	94.41
		Neutral	0.99	0.12	0.08	0.28	499.49	98.58
		Sad	0.99	0.11	0.05	0.22	492.35	100.99
	Incongruent	Happy	0.96	0.20	0.08	0.27	505.41	103.21
		Neutral	0.97	0.17	0.10	0.30	490.47	91.43
		Sad	0.98	0.14	0.10	0.30	501.48	94.86
Bilingual	Congruent	Happy	0.97	0.18	0.05	0.21	493.20	97.82
		Neutral	0.97	0.17	0.05	0.22	491.10	93.54
		Sad	0.98	0.13	0.09	0.28	496.36	94.57
	Incongruent	Happy	0.96	0.19	0.11	0.31	495.92	93.60
		Neutral	0.97	0.18	0.11	0.31	500.90	99.25
		Sad	0.96	0.20	0.06	0.24	504.95	93.45

Table 3.5: ANOVA (Type III) table for model terms. FValence = flanker valence.

Term	FStat	DF1	DF2	pValue
(Intercept)	4543.04	1	82.65	0.000
Group	0.16	1	46.49	0.689
FValence	0.98	2	93.16	0.377
Congruence	14.49	1	5867.11	0.0001
PartGender	7.47	1	44.65	0.009
Group:FValence	0.24	2	6034.80	0.790
Group:Congruence	0.09	1	6157.40	0.758
FValence:Congruence	1.39	2	5830.25	0.248
FValence:PartGender	1.05	2	6057.63	0.351
Group:FValence:Congruence	6.19	2	6031.82	0.002

level of Valence (see Appendix A.1.5 for full details). Simple main effect contrasts are summarised in Figure 3.4.

Bilinguals had a significantly smaller flanker interference effect (the difference between incongruent and congruent trials) than did monolinguals in the positive condition. The interference effect in the negative condition was also significant for monolinguals, but not for bilinguals, though the groups did not significantly differ from each other in negative conditions. In neutral conditions, the magnitude of the interference effect was significant for both groups, but interestingly, monolinguals were faster in incongruent conditions than congruent ones: in other words,

monolingual RTs in incongruent neutral trials were comparable to bilingual RTs in congruent neutral trials, and vice versa.

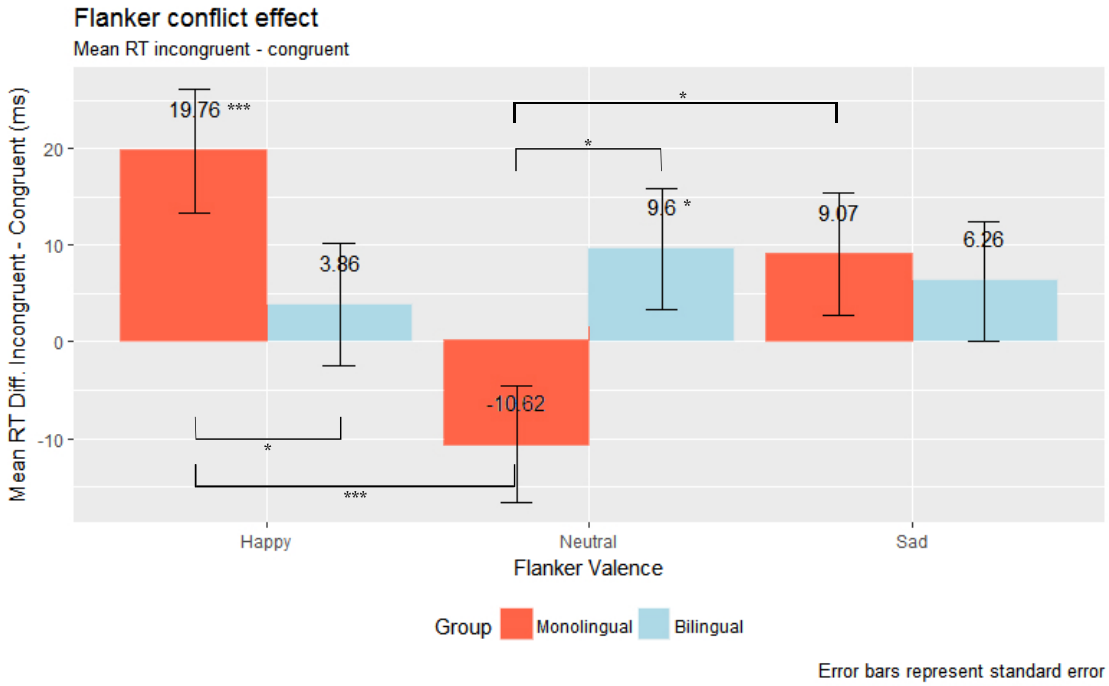

Figure 3.4: Mean reaction time differences in milliseconds between incongruent and congruent trials. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The mean overall hit rates (correct go responses) and false alarm rates (incorrect responses to no-go trials) across conditions are also shown in Table 3.4. Participants were slightly more accurate in congruent compared to incongruent conditions overall. However, performance was close to ceiling in all conditions (lowest go accuracy = 95.6% in bilingual incongruent sad; highest false alarm rate = 10.8% in bilingual incongruent happy). Six participants had 100% accuracy in go conditions and one participant had 100% in no-go conditions. These ceiling effects led to a violation of the assumptions for a generalised linear mixed effects model (namely, normally-distributed residuals and heteroscedasticity), and so the accuracy data in go compared to no-go trials could not be modelled for inference. Instead, by-participant loglinear d' scores (see Stanislaw & Todorov, 1999) were calculated and an LME was fit with d' as the response variable and Group, Congruence, and Flanker Valence as predictors.[2] Random intercepts were included for each

[2] $d' = 4.65$ is considered an effective ceiling value (Macmillan & Creelman, 2004). However, at a hit rate of 100% and a false alarm rate of 0%, d' is an infinite value. Stanislaw and Todorov

participant, but the aggregated dataset was not large enough to include other random variance factors. (See Appendix A for full modelling details.) The d' scores for each condition are shown in figure 3.6 and the three-way relationship between the independent variables is visualised in Figure 3.5.

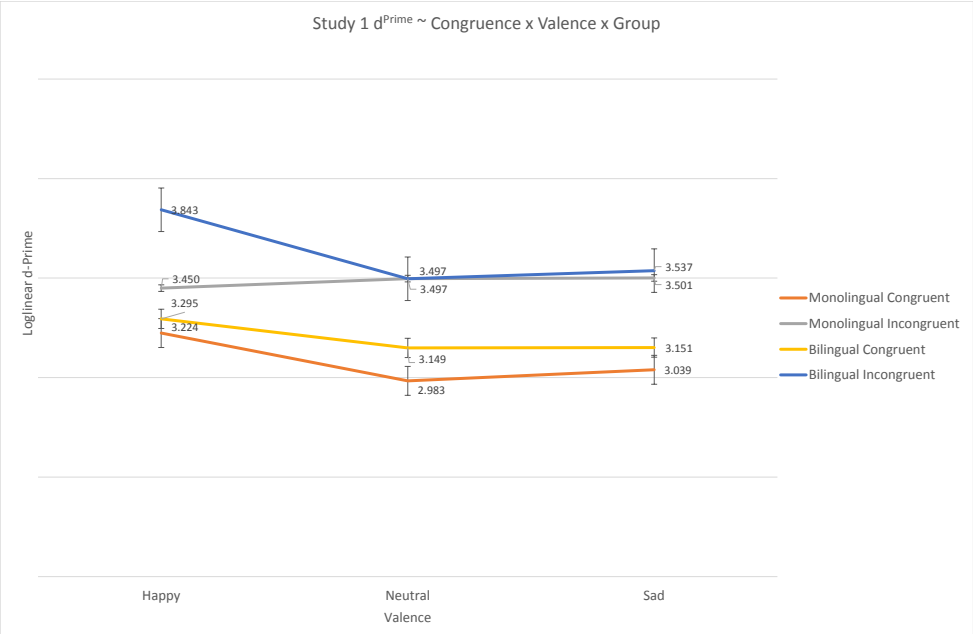

Figure 3.5: Visualisation of the relationships between Group × Congruence × Valence in Experiment 1, with loglinear d' for each condition. Error bars represent standard error of the mean.

There was a significant main effect of congruence on d' ($F(1, 230) = 38.6, p < 0.001$), with lower accuracy in incongruent conditions. None of the other terms or interactions reached significance. Full Type–III ANOVA statistics for the model are listed in Table 3.6.

In summary, the speed of bilinguals' go responses was not significantly affected by positive or negative emotional flankers. An overall significant effect of congruence on both RT and accuracy was observed across all emotional conditions, but was not modulated by valence, consistent with our hypothesis that bilinguals

(1999) recommend the loglinear approach for calculating d' values for data with floor or ceiling effects: the formula for this transformation is $Z((CountFalseAlarms + 0.5) \div NumNogoTrials) - Z((CountHits + 0.5) \div NumGoTrials)$.

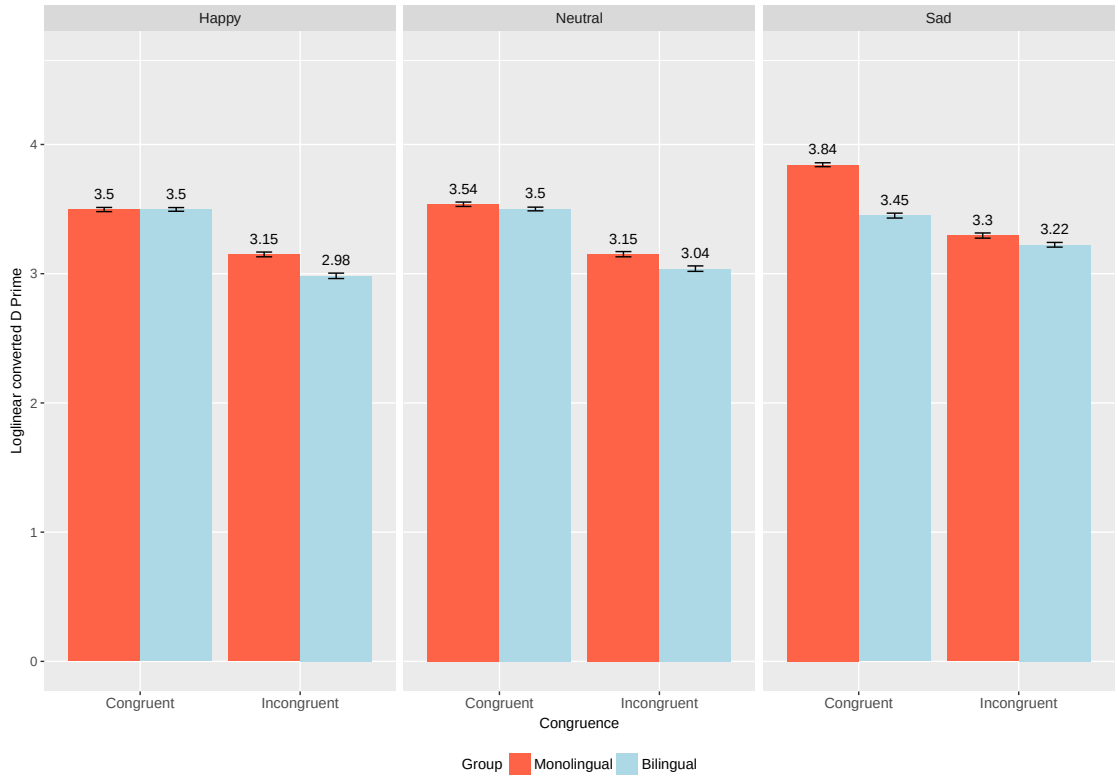

Figure 3.6: Overall loglinear d' values. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.6: ANOVA (Type III) table from LME of accuracy d's.

Term	FStat	DF1	DF2	pValue
(Intercept)	2998.09	1	46	$1.57E-43$
Group	0.36	1	46	0.550
Congruence	38.59	1	230	$2.42E-09$***
FValence	0.23	2	230	0.795
Group:Congruence	0.17	1	230	0.678
Group:FValence	0.05	2	230	0.955
Congruence:FValence	0.38	2	230	0.687
Group:Congruence:FValence	1.78	2	230	0.170

would be able to selectively attend to the task-relevant dimension, gender. This contrasted with the monolinguals' RTs, which showed significant interference from both positive and negative emotional flankers in the gender-incongruent conditions and significant facilitation effects in the gender-congruent conditions, indicating that the irrelevant emotional dimension differentially affected the monolinguals. Furthermore, the congruence effect within the monolingual group was reversed in

the neutral conditions, where gender-incongruent flankers appeared to facilitate RTs while gender-congruent flankers interfered. A possible interpretation of this unusual finding is discussed below. In response accuracy, there was a highly significant overall effect of congruence on d's for both groups, as hypothesised. However, contrary to our hypotheses, there were no effects of group or valence on accuracy. This finding is qualified by a ceiling effect that made the simple main effects of these predictors impossible to analyse separately for go and no-go conditions.

3.5 Discussion

The results indicate that the bilingual group was less influenced by emotional flankers—regardless of valence—than the monolingual group, in both congruent and incongruent conditions. A simple main effect of gender-congruence was observed in the bilingual group whereby gender-congruent flankers facilitated faster RTs, which indicates that the flankers were still processed. This supports our hypotheses that bilinguals have an advantage in selectively attending to task-relevant information (here, gender) and ignoring the distracting effect of valence (Hypothesis 4). Because there is no evidence that emotional valence was processed by the bilingual group, we tentatively conclude that the bilingual advantage in emotion regulation in this task appears to arise at the level of selective attention.

In contrast to the bilinguals, the monolinguals' RTs were influenced by emotional flankers. Both positive and negative flankers caused longer response latencies in incongruent conditions, and shorter response latencies in congruent conditions, relative to neutral flankers. This finding supports Hypothesis 3. In monolinguals, interference suppression of both gender and emotion conflict leads of a behavioural cost. However, the hypothesised difference between positive and negative flankers at the level of interference suppression cannot be confirmed from these data. Only the simple main effect of positive stimuli was significant, but there was no significant difference between positive and negative stimuli in either the congruent or incongruent conditions. Furthermore, we could not support our accuracy predictions due to the ceiling effect.

The findings in monolinguals, however complex, align with Pessoa's (2009) dual-process model, which proposes that task-relevant emotional stimuli enhance

cognitive performance, while task-irrelevant stimuli interfere. Pessoa notes that the spatial locus of emotional stimuli is particularly privileged, which he suggests may be due to systems such as the amygdala that detect the biological relevance of a stimulus (Whalen, 1998). The stimuli used in this study are ambiguous compared to, e.g., angry faces, which signal a clear threat. This may result in additional attentional resources being devoted to the stimuli to resolve this ambiguity (Pessoa, 2009). Thus, the emotional flankers may have captured more attentional resources than the neutral flankers, facilitating conflict resolution in the congruent trials (when the flankers were task-relevant) but enhancing the conflict effect in the incongruent trials (where the flankers were distracting). Of note, our data suggest that this model applies to monolinguals, but not necessarily to bilinguals, who are capable of controlling interference from task-irrelevant emotional stimuli.

Interestingly, in the neutral flanker conditions, the monolinguals were significantly faster on incongruent trials than on congruent trials, which is opposite to the expected response pattern. The faster responses on incongruent trials are consistent with the above account: in the absence of a distracting emotional signal, additional processing resources are available to facilitate conflict resolution. However, this does not fully explain the increase in reaction times observed in the congruent neutral condition relative to congruent emotional and incongruent neutral trials: the slower reaction times suggest that the congruent neutral stimuli *interfered* with responses, rather than simply the absence of the facilitation observed in emotional conditions. This may be due to the interacting interference suppression and response inhibition demands imposed by the combined FNG task design.

Within the flanker component of the task, there is a spatial conflict between the target and the flankers. However, as Posner and Rothbart (1998) point out, this conflict exists in both congruent and incongruent conditions; the behavioural differences between congruent and incongruent trials arise from the conflicting response tendencies induced by incongruent flankers. In addition to resolving conflict between these response tendencies, the participants must determine whether inhibition of the prepotent 'go' response is necessary to avoid making false alarm errors. In the incongruent condition, the 'go' response is selected once disambiguation between flankers and target occurs (which, as noted above, may be faster due to the absence of distracting emotional flankers), while response selection may occur later in the

congruent condition, causing a greater response latency. However, this had no effect on d's, which were lower in incongruent trials across all conditions.

In contrast to previous findings with flanker tasks and counter to our Hypothesis 2, the bilinguals did not have an overall speed advantage across conditions. In fact, the monolinguals were marginally faster in every condition besides incongruent positive (reflecting the significant interference of positive flankers in the monolingual group, discussed above) and congruent neutral (reflecting the simple main effect of congruence in the bilingual group). None of these differences reached significance. However, this finding may also reflect group differences in response strategies to the go/no-go component of the task (see Section 3.1.2.2). A more difficult task with a larger sample may provide further insights into the findings from Costumero et al. (2015) implicating salience detection mechanisms in bilingual response inhibition.

Finally, while the differences between positive and negative stimuli were not significant, the positive stimuli were the most disruptive in incongruent conditions and facilitated the fastest reaction times in congruent conditions, suggesting that the happy faces drew more attentional resources than the sad faces. This supports our Hypothesis 3 and is consistent with the theory proposed by Lang et al. (1997) that positive affect biases approach behaviour, while negative affect biases avoidance behaviour.

The combination of the flanker and go/no-go tasks may also have contributed to the ambiguous findings in emotional trials for the monolinguals. Previous research has shown that happy faces are identified faster than sad or fearful faces in go/no-go trials, likely due to the prepotent tendency to approach positive stimuli (Keightley et al., 2006). In positive congruent trials, this would lead to faster responses, while in positive incongruent trials, faster processing of the flankers due to the positive valence effect increases conflict in the gender dimension, amplifying the flanker effect. In negative trials, by contrast, participants must override the tendency to avoid a negative stimulus. Therefore, while negative faces are still identified more quickly than neutral, the congruence effect is reduced due to this conflicting response tendency. The ceiling effect in go accuracy makes it impossible to determine whether this hypothesis is supported by corresponding differences in hit rates for happy and sad go trials: there was no significant difference between happy and sad d's in any condition, though this may be due to opposite effects of these flankers in

no-go trials. Finally, the stronger effect of positive flankers may also be due to the positive faces being rated as more different from neutral than the negative faces (see Section 3.4.1), or the current sample size may not be sufficient to detect a difference between positive and negative flankers in this design.

3.6 Conclusion

This study investigated whether bilinguals have an advantage over monolinguals in regulating emotional distractions at the level of attention, executive control, or behaviour. The bilingual sample, unlike the monolingual sample, was not influenced by the emotional facial expressions of distracting stimuli during a combined FNG test. These findings indicate for the first time that bilingual adaptations in selective attention to relevant cues (as a result of the social contexts they must navigate) may also have implications for emotion regulation circuits.

An effect of bilingual experience on selective attention in the presence of emotional distractors may help explain compelling recent findings that moral judgements (Geipel et al., 2015) and decision biases such as loss aversion (Keysar et al., 2012) are moderated by language: salience detection (determining the relevance of emotional responses) is linked to both of these processes (Decety et al., 2011; Sokol-Hessner et al., 2012).

However, the distinction between attentional, cognitive, and behavioural processes is not explicitly clear from the results obtained, and future studies are necessary to disentangle the effects of emotion at these respective levels. It was impossible to determine whether the amplification of the flanker effect in the monolingual but not the bilingual group was due to selective attention differences, interference suppression differences, or both. Accuracy data, which indexes response inhibition and could provide evidence of differences between positive and negative flankers at this level, was too skewed to analyse at a by-trial level.

The generalisability of these results is also limited because we endeavoured to recruit samples that were as homogeneous as possible in terms of age, language, and educational background in order to maximise effect size. It also excluded simultaneous bilinguals who learned two languages from birth and highly-proficient trilinguals. Further studies should include more diverse groups of bilinguals to

determine whether the effects observed in this study generalise across cultures, proficiency levels, and social status. (This limitation is addressed in Studies 2 and 3.)

There are several avenues for further research based on the implications of these findings. The first is to explore developmental changes in relevance detection. Hwang et al. (2014) demonstrate that children and adolescents have less efficient neural circuits for salience detection (meaning the ability to differentiate goal-relevant stimuli) compared to adults. Any language-related adaptations in selective attention may therefore be more pronounced in younger participants. As children are also less accurate than adults in go/no-go tasks, replicating the current study in children may also provide analysable accuracy data to help separate the influences of valence on response inhibition and interference suppression, thus helping to explain the ambiguous monolingual findings in this study. Correlations of both language ability and cognitive control with academic outcomes and social integration in adolescents have been reported (Beck et al., 2012; Rueda et al., 2010). Therefore, establishing a link between bilingualism and cognitive control of emotion in children has significant implications for language education and legislation of programmes for children with non-majority home languages. The second avenue for further research is through separate flanker and go/no-go tasks with the same stimuli and conditions. These may make it possible to determine how valence affects the executive processes involved separately. Removing the response inhibition component and conducting a flanker-only task would, for example, remove the possible influence of valence-modulated response tendencies on RTs, which may have masked differences between negative and positive flankers in this task. Using the same design as our FNG task, a flanker-only task which requires participants to respond on every trial would also generate enough observations to analyse longer-lasting conflict adaptation effects with sufficient statistical power. The present task did not include enough trials to account for conflict adaptation due to the limited number of stimuli available (19 per go condition). However, reducing the task-complexity may make it impossible to detect valence effects (Pessoa, 2009), and so a comparison across multiple tasks may be needed.

We expand on both of the above options in Studies 2 and 3.

wenn ich ruf: er wars, nicht ich, der log
rollt er mir grollend einen holzklotz auf die zunge.
junge, brumm ich, du bist eine last, dann er: und du
das letzte.

when i cry: it was him, not me, that lied
he crossly rolls a wooden log onto my tongue.
young man, i grumble, you are quite the load.
he says: and you're the last.

— Uljana Wolf,
'last—lied—list—lump—log', pp. 21.

4 Study 2

Contents

4.1 Background 77
4.1.1 Replicating Study 1 with Mixed Bilinguals 78
4.1.2 Comparing the Flanker–Go/No-go and Flanker-only Tasks 81
4.2 The Present Study 84
4.2.1 Research Questions and Hypotheses 85
4.3 Methods 87
4.3.1 Experiment 1: Flanker–Go/no-go Task with Mixed Bilinguals 87
4.3.2 Experiment 2: Flanker-only task 90
4.3.3 Data Analysis Plan 91
4.4 Results 92
4.4.1 Sample Demographics 92
4.4.2 Experiment 1 92
4.4.3 Experiment 2 96
4.5 Discussion 103
4.5.1 Experiment 1 Results 103
4.5.2 Experiment 2 Results 105
4.6 Conclusion 108

4.1 Background

Bilingualism studies are often criticised for their use of small, highly-selective samples, lack of clearly defined constructs, and resulting lack of generalisability (e.g. Paap & Greenberg, 2013). In the present study, we addressed these criticisms by

attempting to replicate the Study 1 findings in a different population of bilinguals and across two different measures of interference suppression.

In Chapter 3 of this dissertation, we reported the findings of a combined flanker–go/no-go (FNG) task from Study 1. Monolingual and German–English bilingual young adults responded to target pictures of neutral faces, while flanking pictures varied in both their gender-congruence (male/female) and emotional valence (neutral/happy/sad expression). Emotional conflict (between neutral target and happy/sad flanker faces) amplified the effect of gender-congruence in the monolingual sample, but had no significant effect in the bilingual sample. However, a ceiling effect in response accuracy prevented us from ruling out response inhibition-related effects, such as speed-accuracy trade-offs, that may impact our interpretation of the reaction time (RT) data. In addition, the bilingual sample was culturally and linguistically homogeneous, so our findings may have been due to characteristics other than bilingual language proficiency.

Our design of Study 2 was thus guided by two objectives:

1. Attempt to replicate the results of Study 1 in a different population, thereby addressing common criticisms of 'bilingual advantages' studies.

2. Separate the effects of valence on interference suppression from those on response inhibition in the FNG reaction times by comparing them to FO.

4.1.1 Replicating Study 1 with Mixed Bilinguals

The bilingual sample in Study 1 was drawn from a limited group of German–English sequential bilinguals with similar social and linguistic backgrounds in order to introduce as little noise as possible from cultural, social, or educational differences. However, cultural background has an influence on emotion regulation strategies as well, with even northern vs. southern European participants demonstrating differences in "cognitive emotion regulation" (Potthoff et al., 2016). To determine whether the Study 1 results generalise from the "restricted and undetermined circumstances" (Paap et al., 2015) of the chosen task and sample, the present study sought to replicate them with more diverse participants whose primary commonality was a high level of fluency in two or more languages.

4.1.1.1 Addressing common pitfalls in bilingualism research

Based on a small-scale systematic review of bilingualism studies focused on sample size and the language background of participants[1], the majority of studies reporting effects of bilingualism (or lack thereof) on cognitive control fall into four different categories:

1. A specific L1-L2 pairing and one or more control groups of monolingual L1/L2 speakers, or both, $N \approx 30$/Group, complete an experiment measuring particular domain-general executive functions hypothesised to be subject to bilingual adaptation (hereafter 'EF measure').

2. A sample of L1a-L1b simultaneous bilinguals (most frequently Spanish-English and Spanish-Catalan) and a sample of L1a monolinguals complete an EF measure.

3. L1-L2 sequential bilinguals with varying L2 experience complete an EF measure and L2 proficiency is modelled as a predictor of EF ability.

4. L1a-L1b or L1-L2 early sequential children (often EAL or equivalent) are compared to their monolingual peers on a battery of linguistic, academic, and general EF measures.

As noted in Chapter 2, the validity and reliability of these studies has been subject to significant debate in recent years. (For example, the *Bilingualism forum*, 2015, to which Cortex dedicated a special section). Led by a highly critical review by Paap et al. (2015), critics of bilingual advantages research point to insufficient sample sizes, incorrect or non-transparent statistical analyses, failure to replicate seminal findings, lack of convergent validity in myriad tasks ostensibly measuring EF, and ambiguous neuroimaging results that neither explain the effects of increased or decreased brain

[1]This review was conducted by searching PsycINFO for journal articles that included '(bilingual* or multilingual*)' AND '(executive and (control or function*))' in the abstract or title. The 556 returned results were filtered based on the following criteria: 1) Must be an original empirical study (not a review, meta-analysis, or opinion paper), 2) with a primary focus on bilingualism and domain-general cognitive control (i.e. not exclusively about language acquisition/processing), and 3) not exclusively about aging/Alzheimer's Disease/dementia (although studies which reported results of EF measures in aging populations were included). This left 309 articles for review, which were categorised by participants' L1(s), L2(s), control/comparison groups, and age ranges and their sample sizes averaged.

activation nor map onto observed behavioural results, compounded by publication bias favouring positive results (c.f. de Bruin et al., 2015; García-Pentón et al., 2016a; R. M. Klein, 2015a, 2015b; Paap, 2014; Paap & Greenberg, 2013).

However, proponents of the hypothesis that practice in multiple languages leads to beneficial extralinguistic adaptations in cognitive control ability suggest a more positive approach focused on addressing or correcting these criticisms. Luk and Bialystok (2013) caution against treating bilingualism as an undifferentiated construct, pointing out the variety of bi-/multilingual experiences. Fricke et al. (2018) add that this variety, far from making bilingualism research too "hazy" (García-Pentón et al., 2016a), means bilinguals provide a unique perspective on how language processing integrates social and communicative contexts (see also Kroll et al., 2014). A systematic review and meta-analysis by Adesope et al. (2010) refutes the contention of publication bias, although Lehtonen et al. (2018) attest the opposite in a more recent review. On the methodological side, van Heuven and Coderre (2015) as well as Friedman (2016) propose more refined statistical approaches, such as mixed effects regression modelling and latent variable analysis, along with increased use of neuroimaging.

Friedman (2016) offers four concrete recommendations for testing the bilingual advantages hypothesis in light of existing criticisms:

1. Conduct studies with larger sample sizes, in order to account for the impurity and unreliability of many common EF tasks.
2. Use multiple tasks that measure the same EFs to rule out task-specific variance.
3. Consider other sources of individual differences in EF ability besides bilingualism.
4. Draw on literature that discusses the structure of EFs (e.g. unity and diversity, as discussed by Miyake et al., 2000) to interpret findings.

Friedman's recommendations build upon the methodology used in formulating the unity and diversity model of executive functions (see Chapter 2.3.3.1 and Friedman & Miyake, 2004; Miyake & Friedman, 2012; Miyake et al., 2000). These studies are based on latent variable analysis that maps individual EFs' correlations with an underlying common EF factor: a methodology that requires sample sizes orders of magnitude larger than conventional. For example, Friedman notes that the

flanker task has an approximate loading of .40 on an interference suppression latent variable (Friedman & Miyake, 2004), meaning that a sample size of approximately 800 would be required to detect a moderate effect size (0.5 SD) at 80% power (Friedman, 2016, p.4).

Correlations between bilingualism and EF ability have been reported in cohort studies at these sample sizes (Mukadam et al., 2018; Swanson et al., 2017; Zahodne et al., 2014), but such a large-scale study is beyond the scope of most studies whose primary focus is on bilingualism. Friedman's other recommendations are more actionable (Friedman, 2016): even without a latent variable approach, comparing diverse samples across multiple related EF tasks can provide insights into the mechanisms common to them. The bilinguals recruited for the present study were therefore intentionally culturally and linguistically heterogeneous to contrast with the homogeneous German–English sample from Study 1. While it is impossible to rule out all individual differences factors, a replication of the Study 1 results in Experiment 1 would suggest that the group differences observed are more likely due to bilingualism than some other confounding characteristic of the sample.

4.1.2 Comparing the Flanker–Go/No-go and Flanker-only Tasks

Negative valence has been shown to increase flanker interference relative to neutral or positive valence in previous studies (Fenske & Eastwood, 2003; Grose-Fifer et al., 2013). However, the bilinguals' RTs in Study 1 did not differ significantly between happy, sad, or neutral conditions. We thus interpreted the results as evidence for a bilingual advantage in selective attention, meaning the emotional dimension of the flankers was ignored at the attentional level by bilinguals, as opposed to interference suppression, meaning that emotional valence was processed, but successfully inhibited at the executive control level. However, go/no-go accuracy was at ceiling, preventing a statistical analysis of emotional valence effects on hits and false alarm rates. We could therefore not reliably conclude whether response inhibition was affected by emotional valence, as predicted by go/no-go studies (Hare et al., 2005), and whether this impacted the RTs, possibly though speed-accuracy trade-offs.

Brydges, Clunies-Ross et al. (2012) demonstrated that the neural signature of response inhibition in no-go trials is functionally and temporally distinct from interference suppression in incongruent trials contrasted with congruent ones. However, emotion can have competing effects on the two processes. Positive affect activates the appetitive system, creating a bias toward approach behaviour (Lang et al., 1997). In a go/no-go task with infrequent no-go trials, the 'go' response is already prepotent: theoretically, positive stimuli could amplify this response tendency, leading not only to more false-alarm responses (Hare et al., 2005, e.g.) but also to faster RTs in go trials (Keightley et al., 2006). Negative stimuli, meanwhile, activate the aversive system, biasing avoidance behaviour (Lang et al., 1997). In a negative no-go trial, this avoidance tendency would be expected to facilitate response inhibition, but potentially at the expense of RTs in negative go trials. Such a speed-accuracy trade-off has been observed in younger participants (~15 years of age, Tottenham et al., 2011).

However, faster reaction times for negative stimuli have been reported for both go/no-go tasks (Zhang & Lu, 2012) and face gender-identification tasks (Keightley et al., 2006) in adults. Thus, for adult participants, negative stimuli have opposite effects from positive stimuli on false alarm rates, but a similar, facilitatory effect on reaction times. It is therefore possible that the negative stimuli's effects on go responses, leading to faster RTs, are cancelled out by their effects on interference suppression, which produce slower RTs, particularly in incongruent trials. This was one possible explanation for the unexpected findings in Study 1, where positive flankers amplified the effect of gender congruence in monolinguals.

In the present study we included the Study 1 FNG task replication in Experiment 1 alongside a flanker-only (FO) task, i.e. a forced-choice gender identification with no response inhibition demands, in Experiment 2. The FO task in Experiment 2 was intended to disentangle effects of response inhibition and interference suppression on any observed group differences. Based on a hypothesised bilingual advantage in selective attention, we would predict a similar pattern of results in Experiments 1 and 2. Distinct results would point to an effect of response inhibition in Experiment 1. The Experiment 2 results would then provide evidence of group differences in selective attention or interference suppression, should any exist.

4.1.2.1 Response selection and conflict monitoring in the flanker task

As discussed in Chapter 3, Section 3.1.2.1, the flanker task still involves a component of response inhibition even when responses are required on each trial. Tillman and Wiens (2011) found that ERP-indicators of response inhibition were present in a classic flanker task with a low proportion of incongruent trials, but decreased when the proportion was higher, while indicators of interference suppression increased. In order to minimise response inhibition demands in the FO task, and maximise interference suppression, we used an equal proportion of congruent to incongruent trials.

In addition, because the forced-choice flanker task records RTs on every trial, producing up to 216 observations per participant, our design had sufficient statistical power to analyse conflict adaptation across trials. Conflict monitoring was another potential confound that could not be analysed in Study 1 due to the insufficient number of observations, but which provides a possible explanation for the reversal of the congruence effect observed in monolinguals on neutral trials. The conflict adaptation (CA), or 'Gratton' (Gratton et al., 1992), effect occurs on successive trials based on the level of conflict encountered in the first trial. The flanker effect, which is the RT difference between incongruent and congruent trials, is usually smaller when the previous trial was incongruent, and larger when it was congruent (Botvinick et al., 2001).

As discussed in Chapter 2, Section 2.3.3.3, conflict monitoring is a theory of adaptive control that posits the ACC as a 'monitor' that detects both cognitive and emotional conflict (Etkin et al., 2011; Shackman et al., 2011) and signals the dlPFC to engage cognitive control mechanisms that resolve it. Incongruent trials increase the level of control allocated to interference suppression, and thus reduce the flanker effect on succeeding trials (see Egner, 2007, for a review). Grundy et al. (2017) demonstrated that this effect is larger in monolingual compared to bilingual young adults, moderated by stimulus timing: the shorter the response-to-stimulus interval (RSI), the larger the group difference. Furthermore, they provide evidence of post-conflict N2 and P3 ERPs in monolinguals, but not in bilinguals, at RSIs where the behavioural differences disappeared. This, the authors suggest, is due to a bilingual advantage in disengaging attention from previous trials (Grundy et al., 2017).

Inzlicht et al. (2015) suggest that conflict is an aversive experience, and negative affect thus increases perceived conflict. Supporting this theory, induced negative mood has been shown to increase conflict adaptation effects (Van Steenbergen et al., 2010): participants who underwent a negative mood-induction procedure prior to completing a flanker task had a larger reduction in congruence effect when preceding trials were incongruent compared to those with induced positive or neutral moods.

The effects of stimulus valence on this effect are less explored, although there is evidence to suggest that conflict adaptation may only occur when all other features of the succeeding trials are identical (Funes et al., 2010). Funes et al. (2010) observed conflict adaptation on successive incongruent trials when the type of conflict remained the same, but not when it switched from spatial Stroop to flanker conflict. The design of the present study allowed us to compare emotional valence and gender conflict adaptation effects in a similar way. If emotional facial expressions are processed as task-irrelevant, they should represent an additional source of conflict, producing larger conflict adaptation effects. If, however, they are processed as information about the task, sad flankers might be expected to amplify conflict adaptation and happy flankers to attenuate it, as observed in the mood-induction study by (Van Steenbergen et al., 2010)). The absence of previous trial valence effects in the bilingual group would provide further evidence for the tentative conclusion that the bilinguals in Study 1 had an advantage in selective attention to task-relevant stimulus features, and thus did not process emotional valence.

4.2 The Present Study

Experiment 1 sought to generalise the findings from Study 1 to a mixed group of simultaneous and sequential bi- and multilinguals in order to provide evidence that the observed group difference was indeed due to bilingualism. In Experiment 2, the same groups completed a FO task. We measured interference suppression, as indexed by RT differences between congruent and incongruent flankers, across these two variants of the flanker task in an attempt to differentiate it from response inhibition.

If the group differences observed in Study 1 were due to a bilingual advantage in selective attention, the same patterns of results would be expected in both tasks: that is, emotional flankers, regardless of valence, should amplify the effect

of gender-congruence relative to neutral flankers in the monolingual group, but not differ from neutral flankers in the bilingual group.

If, however, the group differences in Study 1 arose due to an advantage in response inhibition, we would anticipate no group differences in the FO task. We address the potential confound of remaining response inhibition demands in the FO by adjusting the ratio of congruent to incongruent trials to minimise this effect (Tillman & Wiens, 2011).

Finally, if the groups differed at the level of interference suppression, then valence-specific differences should be observed in the FO task, with sad flankers creating more interference than happy flankers in incongruent conditions (Fenske & Eastwood, 2003; Grose-Fifer et al., 2013).

The results were interpreted based on the comprehensive framework laid out in Chapter 2, which unites various theories on the structure of EF, conflict monitoring, and attention, as well as emotion and emotion regulation.

4.2.1 Research Questions and Hypotheses

1. Are diverse sequential and simultaneous multilinguals with various L1s better at regulating task-irrelevant affective information relative to monolinguals? How do they compare to the sequential German–English bilinguals from Study 1?
2. Can differences between monolinguals and multilinguals in the emotional FNG task be replicated with a FO task?
3. Comparing both tasks, can group differences be attributed to selective attention, interference suppression, or response inhibition?

4.2.1.1 Experiment 1

Experiment 1 was expected to replicate the findings from Study 1. The bilingual adults were expected to be more effective at diverting their attention away from the distracting emotional features of the flankers. Although differences between simultaneous and sequential bilinguals have been observed anatomically (D. Klein et al., 2006) and linguistically (Gathercole & Moawad, 2010), L2 proficiency and amount of exposure are considered stronger predictors of bilingual advantages than age of acquisition. Furthermore, findings of a bilingual advantage in the flanker task have been reported across numerous L1-L2 combinations.

Hypothesis 1a: We expected a congruence effect unmodulated by valence within the bilingual group for both RT and accuracy.

Hypothesis 1b: The monolinguals were expected to show a Congruence × Valence interaction on both RTs and accuracy, indicating more attentional capture by emotional flankers and a resulting amplification of the congruence effect.

4.2.1.2 Experiment 2

In the absence of response inhibition demands in Experiment 2, differences are expected between positive and negative stimuli for the monolinguals. Grose-Fifer et al. (2013) demonstrated that negative flankers resulted in longer RTs in incongruent trials relative to positive flankers, indicating that negative valence interfered more strongly with conflict resolution. Based on the theory that emotional stimuli are treated as information about the task (Inzlicht et al., 2015; Pessoa, 2009), this may be due to increased conflict in negative relative to positive trials.

Hypothesis 2a: We predicted that monolinguals would have slower RTs in incongruent sad vs. incongruent happy conditions, and faster RTs in congruent positive vs. congruent negative conditions, compared to bilinguals.

However, other theories also suggest that positive emotions broaden the scope of attention, which would be expected to create more interference in incongruent positive trials (Fredrickson, 2004). This is one possible interpretation of the Study 1 result; a replication of this effect in the FO task would therefore provide additional evidence that positive valence interfered with monolinguals' interference suppression more strongly than negative valence. We expected to observe an effect of conflict adaptation and attentional disengagement in the trial sequence effects from Experiment 2.

Hypothesis 3a: An interaction between Current Trial Congruence and Previous Trial Congruence was predicted. Specifically, both groups were expected to show a smaller flanker effect in trials succeeding incongruent trials. Monolinguals were also expected to show facilitation from sequential congruent trials.

The predicted sequence effects would reflect slowed disengagement of attention from the gender-congruent flankers in the monolingual group, as reported by Grundy

et al. (2017). Links between emotion and conflict adaptation are tenuous, but based on the affect-as-conflict signal interpretation put forward by Inzlicht et al. (2015), negative flankers were predicted to result in increased conflict adaptation.

> Hypothesis 3b: Negative flankers were predicted to result in faster RTs in subsequent incongruent trials for both groups.

The effect of previous-trial valence on bilinguals' conflict adaptation effects provided an additional means of testing whether the bilingual Study 1 results were due to selective attention, in which case valence would not be expected to have an effect, or due to interference suppression advantages, in which case previous valence should have similar effects to previous congruence.

4.3 Methods

4.3.1 Experiment 1: Flanker–Go/no-go Task with Mixed Bilinguals

4.3.1.1 Design and Materials

Experiment 1 used the same design and stimulus list as Study 1 (see Chapter 3.3). One stimulus image whose gender was identified below 70% accuracy in Study 1 was removed from the presentation lists to make the experiment slightly shorter. This left 204 total trials, 42 per valence level in the congruent and 30 per valence level in the incongruent condition.

Before starting the experiment, participants were asked to report all of their languages and self-rate their proficiency in each on a sliding scale from 'Beginner' (0) to 'Native' (1). They also completed computer-based vocabulary tests for languages in which a variant of the yes/no-style LexTALE test was available: Dutch, English, and German (Lemhöfer & Broersma, 2012), as well as French (Brysbaert, 2013) and Spanish (Izura et al., 2014).[2]

[2]It is important to note that only the English version of LexTALE has been validated on this population, and thus the results may not be valid measures of bilingual proficiency. However, all tests are based on the same design, making it possible to compare scores in different languages. It was not possible to administer more valid or thorough proficiency tests within the time constraints of the experiment.

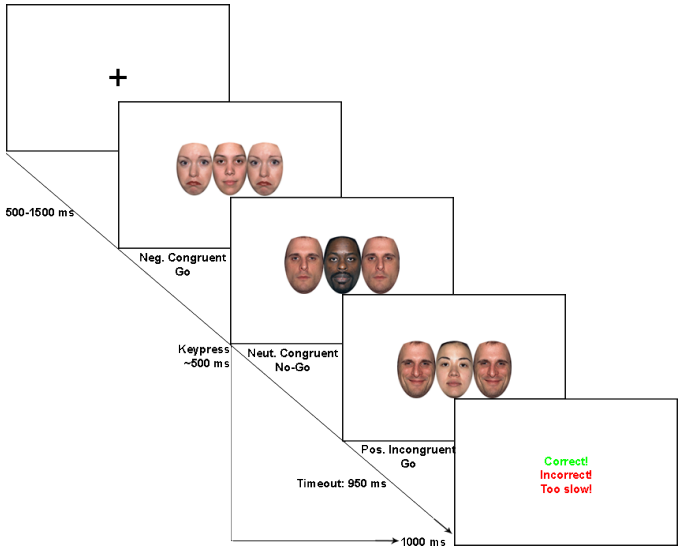

Figure 4.1: Experimental trial sequence. (Go condition = female).

4.3.1.2 Participants

The target bilingual population was adults aged 18–35 with no psychiatric history. 'Bilingual' was defined as fluency above C1 level and at least 6 months' immersion experience in each language. The target comparison population was adult monolinguals, with the same inclusion criteria. Convenience samples were drawn from students and staff at Oxford and Cambridge Universities: participants were recruited via personal contacts, social media, and posters in college and department common rooms.

All bilinguals spoke English to a high level, had lived and worked/studied in the UK for at least one year. For non-native English speakers, the average LexTALE English score was $87.87\% \pm 10.7$ and the average English self-rating was $0.8\% \pm 0.13$. Their language backgrounds varied by age of acquisition: 'Simultaneous' bilinguals had learned two languages from early childhood, and 'Sequential' bilinguals began learning their L2(s) in school or later. They also varied in the number and type of languages they spoke. Full details of the participants' language backgrounds are given in Table 4.1.

28 mixed bilinguals (8 male) and 28 monolinguals (13 male) from Oxford and Cambridge completed Experiment 1. The mean age of the bilinguals was 25.5 ± 3.8, for the monolinguals 24.6 ± 4.3.

Table 4.1: Summary of participants' language backgrounds for Experiments 1 and 2. N = number of participants completing each task [FNG = Flanker–go/no-go (Experiment 1) only, FO = Flanker (Experiment 2) only, both = Experiments 1 and 2]. For monolinguals, LexTALE scores (as percents) for any familiar languages are reported where available, otherwise self-ratings are reported (continuous 0–1 scale, where 1 is 'Native'). For bi-/multilinguals, only languages with self-ratings above 50% and/or (where available) vocabulary scores above 60% are listed.

Group	L1(s)	L2(s) in addition to English	N
Monolingual	English	—	4 FNG, 4 FO,
		Additional reported lang. exposure (< B2 level):	20 both
		+ French (55%)	1 both
		+ German (48%)	1 both
		+ Hindi (0.65), Kannada (0.65), Konkani (0.55), Tamil (0.1)	1 both
		+ Russian (0.39)	1 FO
		+ Spanish (66%)	1 both
		+ Spanish (46%) and Arabic (0.55)	1 both
Bilingual	Czech	+ Slovak	1 both
	Danish	—	1 FNG
	Dutch	—	1 both
		+ Italian	1 both
		+ Swedish	1 both
	English	+ French (70%)	1 both
		+ French (53%) and Dutch (64%)	1 both
		+ German (76%)	1 both
	German	—	3 FO, 3 both
		+ French and Turkish	1 both
	Mandarin	+ Cantonese	1 both
	Portuguese	—	1 both
	Serbo-Croatian	—	2 both
	Czech–Russian	+ French	1 FNG
	English–Dutch	+ German	1 both
	English–Finnish	+ German	1 both
	English–French	—	1 both
		+ Spanish	1 both
	English–Italian	—	1 both
	English–Mandarin	—	1 both
	English–Punjabi	—	1 both
	English–Turkish	+ German	1 both
	French–Italian	—	1 both
	French–Mandarin	—	1 both
	Spanish–Catalan	—	1 both
		+ Portuguese	1 both

4.3.1.3 Procedure

Before beginning the experiment, participants completed a short computer-based questionnaire indicating their age, gender, languages spoken, and subjective profi-

ciency in each. Following the task instructions, participants completed eight practice trials and were given the opportunity to ask questions afterwards.

The trial sequence was nearly identical to that of Study 1, and is shown in Figure 4.1. Each trial started with a randomised inter-trial interval that varied between 500–1500 ms. A black fixation cross (+) was then presented in the centre of the screen for 500 ms before target and flankers appeared. Participants were instructed to press the Space key, which was marked with a blue sticker, if the centrally presented face appeared to be the same gender as the go condition. The go condition alternated between participants. Feedback, in the form of a green 'Correct' or red 'Incorrect' or 'Too Slow!', was displayed below the stimuli for one second before the start of the next trial. In Study 1, this feedback always appeared after the maximum trial time of 950 ms. However, some participants reported finding this confusing and occasionally pressing the space key repeatedly, so the feedback was altered to appear immediately after keypress. On no-go trials, 'Correct' appeared after the 950 ms timeout if no response was made.

As in Study 1, the order of the blocks and of the 12 stimuli within these blocks was randomised. Participants were instructed to take a short break every 24 trials to prevent fatigue, and the number of remaining trials was displayed on the screen during these breaks.

Upon completing the FNG task, participants were administered the LexTALE yes/no test in English as well as any other language for which they had reported proficiency above 0.25 that had a test version available. For participants who completed both Experiments 1 and 2, these tests were administered between the two experiments, and the order of the experiments alternated.

4.3.2 Experiment 2: Flanker-only task

4.3.2.1 Participants

29 bilinguals (18 female) and 29 monolinguals (15 female) completed Experiment 2. Three of the bilinguals and one of the monolinguals had participated in Study 1, four of the monolinguals completed only Experiment 2, and the rest completed both experiments. See Section 4.3.1.2 and Table 4.1 for full details.

4.3.2.2 Design and Procedure

The design, materials, and trial sequence were identical to Experiment 1, except that participants saw an equal number of male and female target faces and the ratio of congruent/incongruent flankers was 50/50. There were an equal number of happy, sad, and neutral flankers in both congruence conditions. Participants were instructed to press the "A" and "L" keys on the keyboard to indicate the gender of the central face, requiring them to respond on each trial. The gender–key assignment was counterbalanced.

4.3.2.3 Counterbalancing order of tasks

Half the participants completed Experiment 1 first, the other half Experiment 2, with the vocabulary tests in-between to reduce repetition effects. Presentation timing, feedback, and inter-trial interval variation were identical between the two experiments.

4.3.3 Data Analysis Plan

Data analysis methods were the same as in Study 1. Normal distribution of RTs was tested and, where necessary, transformed via a Box-Cox power transformation. Accuracy data with a ceiling effect was aggregated into loglinear d-prime d' scores for each participant (Stanislaw & Todorov, 1999, see Appendix A). Linear mixed-effects models were fit for parametric data (RT, d') and generalised linear mixed effects models for non-parametric data (answers), with the hypothesised interactions and their lower-order terms and with random intercepts for participants and items. Additional random effects terms were added via an information-theoretic forward-model selection approach using AIC, and the inclusion of fixed covariates (e.g. participant gender) was tested using simulated likelihood ratio tests. Significance of all predictors was computed using Type–III ANOVA with a Satterthwaite approximation for degrees of freedom, and planned comparisons conducted using factor contrasts. Full rationale and technical details for these analyses are provided in Appendix A, as are regression equations, model summaries, and power analyses. Scripts for all studies, including model selection steps with the random effect terms tested, are included in Appendix B.

4.4 Results

4.4.1 Sample Demographics

During the model selection process, random effects terms for L1, vocabulary test scores, age, bilingual type (bilingual/multilingual), and sequence of acquisition (sequential or simultaneous) were included to account for additional variance in this diverse sample. A random intercept term for bilingual type × bilingual sequence improved the fit of the RT models for both studies, as indexed by lower AICs compared to models without the term. This implies that both the number of languages spoken and the age of acquisition accounted for significant variance within the bilingual sample. None of the other demographic characteristics improved model fits.

4.4.2 Experiment 1

Table 4.2 shows hit-RT and overall accuracy/false alarm condition means for Experiment 1.

Table 4.2: Experiment 1 group means. Hit rates represent accuracy in go trials, false alarm rates inaccuracy in no-go trials. RTs are given for hits (i.e. correct go responses).

Group	Congruence	Valence	Hit Rate	SD	FA Rate	SD	RT (hits)	SD
Monolingual	Congruent	Neutral	0.978	0.148	0.059	0.236	473.075	94.623
		Happy	0.980	0.141	0.057	0.231	469.378	91.753
		Sad	0.982	0.132	0.127	0.333	468.763	93.157
	Incongruent	Neutral	0.981	0.136	0.125	0.331	468.595	88.492
		Happy	0.991	0.096	0.111	0.315	470.393	90.783
		Sad	0.979	0.143	0.081	0.274	473.579	94.626
Bilingual	Congruent	Neutral	0.984	0.127	0.064	0.245	460.882	89.952
		Happy	0.980	0.141	0.075	0.264	460.809	86.933
		Sad	0.987	0.111	0.111	0.315	455.906	84.658
	Incongruent	Neutral	0.979	0.144	0.083	0.277	466.541	87.330
		Happy	0.986	0.117	0.087	0.282	470.779	96.288
		Sad	0.979	0.142	0.111	0.314	475.748	95.997

4.4.2.1 RT Data

The hypothesised three-way Group × Congruence × Valence interaction was not significant ($F(2, 1237.1) = 0.7, p = 0.49$) for hit-RTs (see Figure 4.2). There was a main effect of Congruence ($F(1, 154.3) = 4.0115, p < 0.05$) and a significant

Group × Congruence interaction ($F(1, 258.7) = 5.43, p < 0.01$), which planned comparisons showed was due to a significant congruence effect in the bilingual but not the monolingual group. The latter was driven by a significant difference between groups in the sad condition: the bilinguals showed a congruence effect in sad flanker trials ($T(163) = -2.631, p < 0.01$), and this effect differed significantly from that of the monolinguals ($T(900.7) = 2.328, p < 0.05$), where congruence was not significant ($T(159) = -0.755, p = 0.45$).

Table 4.3: ANOVA (Type III) table for RT accuracy LME, Experiment 1.

Term	FStat	DF1	DF2	pValue
(Intercept)	6416.61	1	86.46	0.000
Group	1.82E-05	1	55.36	0.997
FValence	0.15	2	181.39	0.859
Congruence	6.10	1	37.42	0.018
PartGender	1.82	1	55.27	0.183
Group:FValence	0.19	2	397.35	0.830
Group:Congruence	5.13	1	214.79	0.025
FValence:Congruence	1.52	2	181.58	0.222
Group:FValence:Congruence	0.77	2	397.42	0.463

4.4.2.2 Accuracy

As in Study 1, accuracy was near-ceiling across conditions; five participants were 100% accurate in no-go conditions, while 15 participants had no misses in go conditions. A generalised-linear model with answers as the response variable was fitted, but the residuals did not meet the prerequisite assumptions of homoscedasticity and showed large negative dispersion as fitted values approached 1, indicating that the ceiling effect created bias in the model. Alternative fit methods and link functions did not improve this significantly. Therefore, no statistical inference could be performed on by-trial accuracy. The descriptive statistics presented in Table 4.2 show increased false alarms by monolinguals in the incongruent neutral and congruent sad conditions, suggesting a slight speed-accuracy trade-off when compared with monolingual RTs in these conditions.

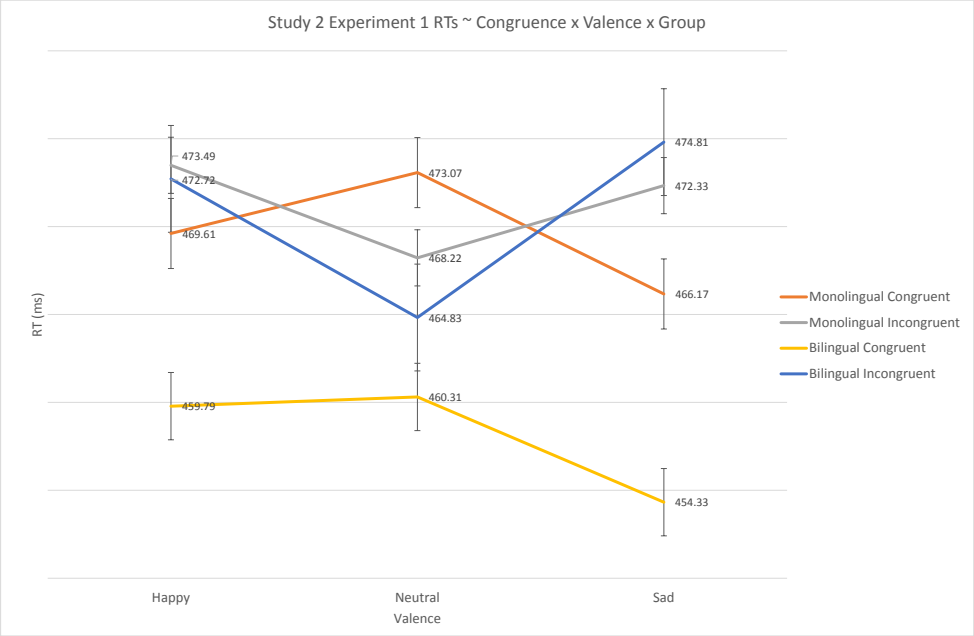

Figure 4.2: Visualisation of the Group × Congruence × Valence interaction, with experimental mean RTs in milliseconds. Error bars represent standard error of the mean.

Following the same procedure as in Study 1, loglinear d' values were calculated for each participant and an LME was fit with d' as the response variable and Group × Congruence × Valence as predictors, with random intercepts for each participant. The ANOVA for this model is shown in Table 4.4.

There were significant main effects of congruence ($F(1, 280) = 22, p < 0.001$) and flanker valence ($F(2, 280) = 4.59, p < 0.05$), with a marginally significant Congruence × Valence interaction ($F(2, 280) = 2.56, p = 0.08$). Planned contrasts revealed the effect of congruence was significant in the happy ($T(280) = -2.46, p < 0.05$) and neutral ($T(280) = -3.47, p < 0.001$), but not in the sad conditions for the monolingual group, with gender-incongruent flankers reducing d' accuracy, as expected. The monolinguals were also more accurate in congruent happy ($T(280) = -3.03, p < 0.01$) and congruent neutral ($T(280) = -2.4, p < 0.05$) compared to congruent sad conditions. In sad flanker conditions, they were actually more accurate in incongruent than in congruent trials, although the difference did not reach significance. For bilinguals, the effect of congruence was significant only

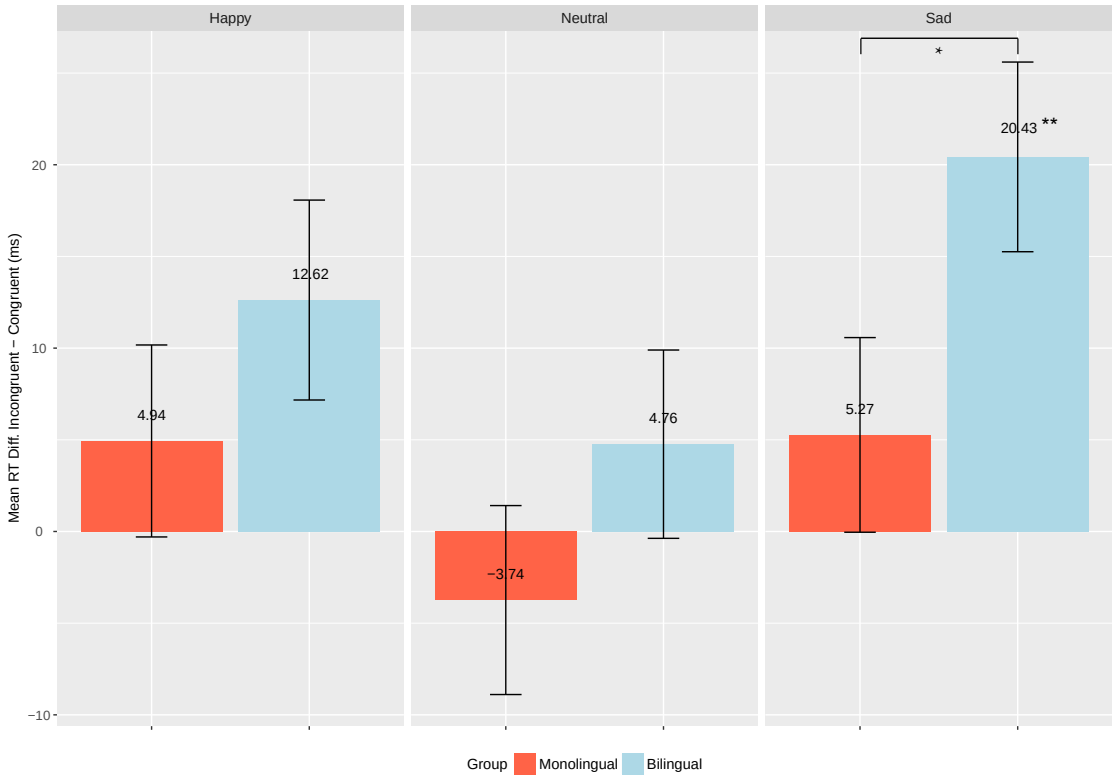

Figure 4.3: Mean reaction time differences in milliseconds between incongruent and congruent trials for Experiment 1. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

in neutral conditions ($T(280) = -2.58, p < 0.05$). These findings are summarised in Figures 4.4 and 4.5.

Table 4.4: ANOVA (Type III) table for d' accuracy LME, Experiment 1.

Term	FStat	DF1	DF2	pValue
(Intercept)	3090.08	1	56	$1.10E-50$
Group	0.09	1	56	0.764
Congruence	22.00	1	280	$4.26E-06$ ***
FValence	4.59	2	280	0.011 *
Group:Congruence	0.00	1	280	0.949
Group:FValence	1.02	2	280	0.363
Congruence:FValence	2.56	2	280	0.079 .
Group:Congruence:FValence	1.50	2	280	0.225

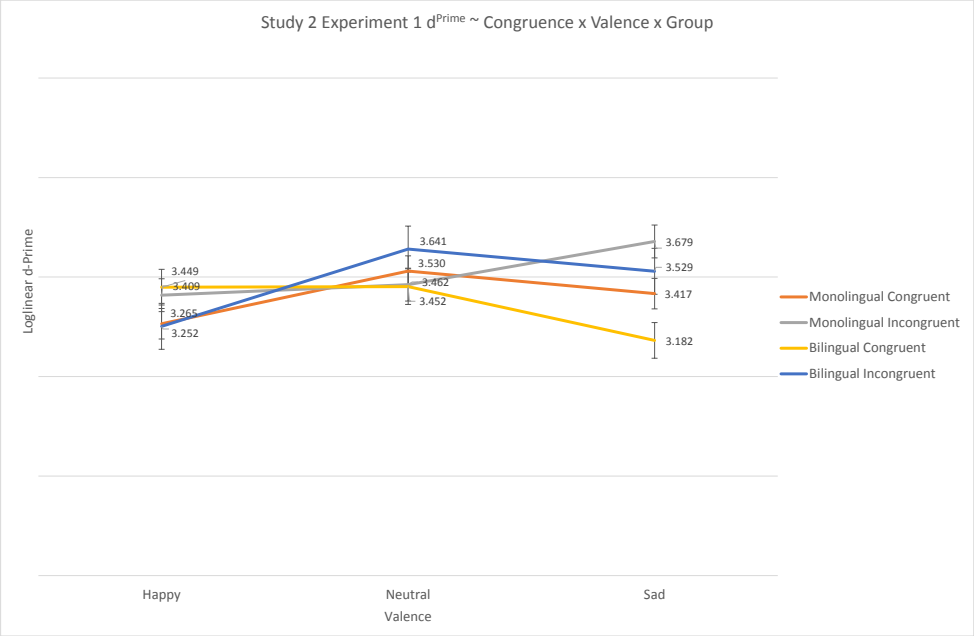

Figure 4.4: Visualisation of the relationship between independent variables on loglinear d' sensitivity thresholds for Experiment 1. Error bars represent the standard error of the difference between hit rate and false alarm rate means.

4.4.3 Experiment 2

The accuracy and RT means from Experiment 2 are listed in Table 4.5.

Table 4.5: Experiment 2 group means. RT means are given for correct responses only.

Group	Congruence	Valence	Accuracy	SD	RT (correct)	SD
Monolingual	Congruent	Neutral	0.927	0.260	510.560	96.155
		Happy	0.933	0.250	514.086	97.414
		Sad	0.926	0.263	512.500	107.029
	Incongruent	Neutral	0.928	0.258	514.570	99.014
		Happy	0.924	0.266	514.900	102.393
		Sad	0.896	0.306	518.599	95.641
Bilingual	Congruent	Neutral	0.951	0.217	514.479	96.488
		Happy	0.968	0.175	517.958	98.915
		Sad	0.952	0.214	511.531	93.664
	Incongruent	Neutral	0.939	0.240	517.802	90.760
		Happy	0.935	0.247	516.616	94.850
		Sad	0.945	0.228	525.987	96.808

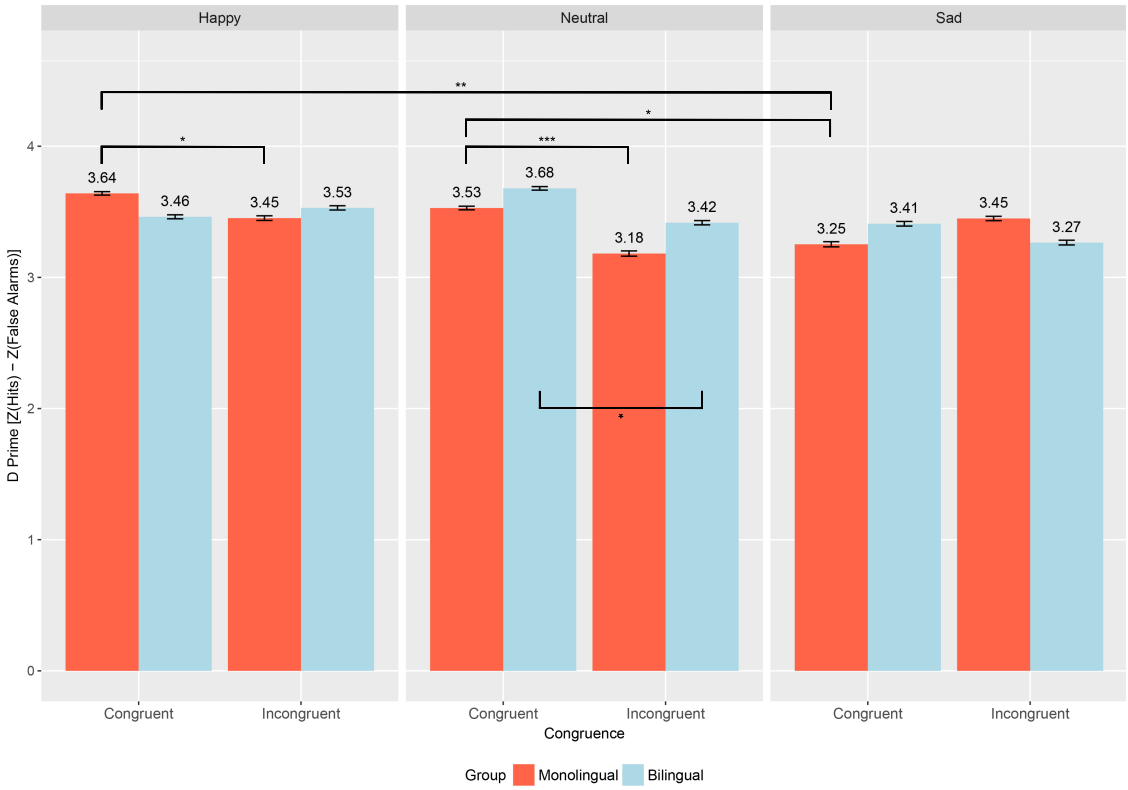

Figure 4.5: Overall loglinear d' rates for Experiment 1. Error bars represent the standard error of the difference between hit rate and false alarm rate means. . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

4.4.3.1 RT Data

There was a highly significant main effect of Congruence ($F(1, 10416) = 21.69, p < 0.0001$) with an underlying Congruence × Valence interaction ($F(2, 10445) = 3.54, p < 0.05$). Sad flankers increased the effect of congruence, while happy flankers decreased it. Planned contrasts revealed that this was driven by a significant simple main effect of congruence in the sad, but not the happy or neutral conditions for both groups, although the bilingual neutral congruence effect was marginally significant, $T(10781) = 5.84, p < 0.1$. See Figures 4.6 and 4.7 for a summary. Table 4.6 shows the Type–III ANOVA statistics for each predictor in the LME model of the RTs.

4.4.3.2 Accuracy

Accuracy was slightly lower compared to Experiment 1, with only one participant achieving 100% accuracy, and thus the accuracy data could be analysed using a generalised linear mixed model (GLMM). Analysis details are listed in Appendix A.

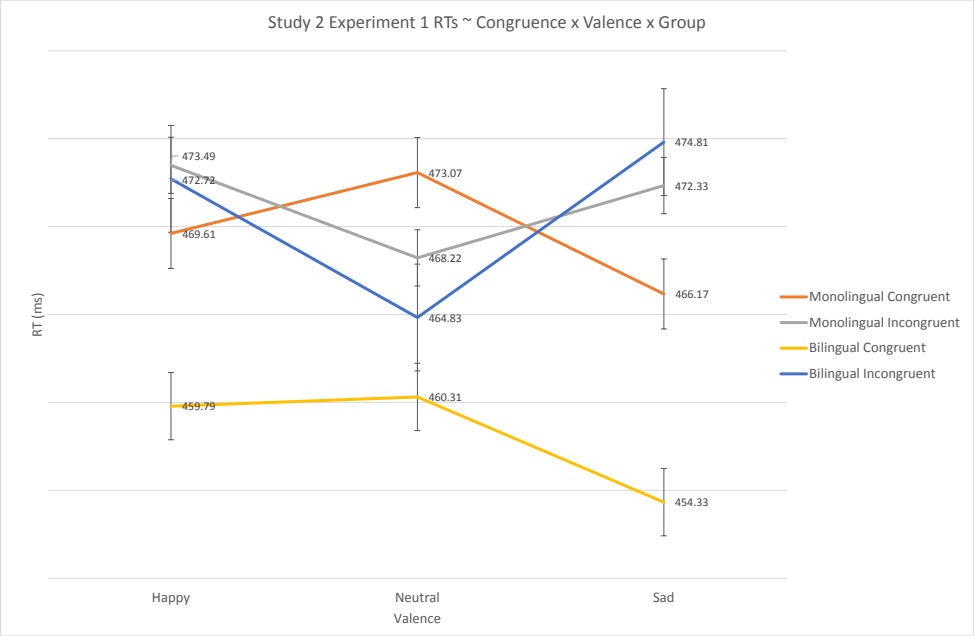

Figure 4.6: Visualisation of the relationships between Group × Congruence × Valence in Experiment 2, with experimental mean RTs in milliseconds. Error bars represent standard error of the mean.

Table 4.6: ANOVA table for Experiment 2 RT model.

Term	FStat	DF1	DF2	pValue
(Intercept)	24737.97	1.00	93.94	$1.41E-115$
Group	0.67	1.00	57.03	0.416
Congruence	21.69	1.00	10416.95	$3.24E-06$ ***
FValence	0.58	2.00	69.36	0.561
PartGender	5.65	1.00	57.01	0.021
Group:Congruence	0.36	1.00	10512.43	0.550
Group:FValence	0.28	2.00	10513.66	0.757
Congruence:FValence	3.54	2.00	10445.21	0.029 *
FValence:PartGender	0.70	2.00	10643.42	0.496
Group:Congruence:FValence	0.01	2.00	10503.31	0.990

There were significant main effects of Group ($F(1, 11622) = 6.81, p < 0.05$) and Congruence ($F(1, 11622) = 13.55, p < 0.01$) with no interactions. As expected, bilinguals were more accurate than monolinguals across all conditions, and

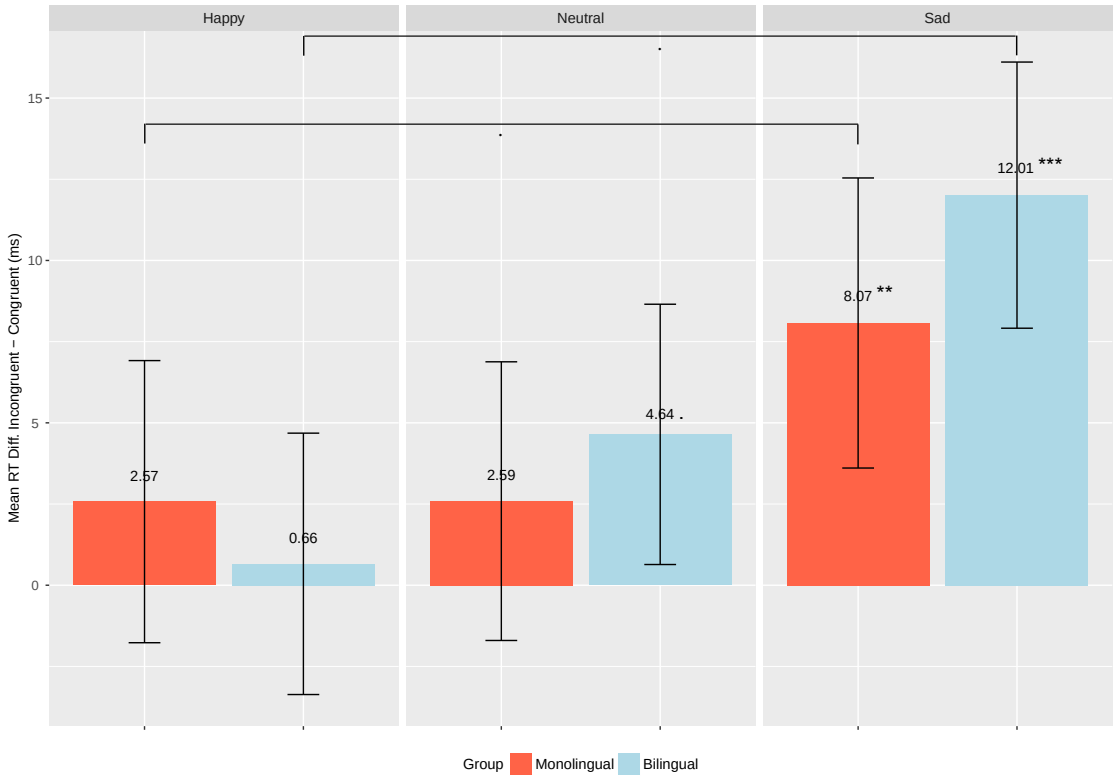

Figure 4.7: Mean reaction time differences in milliseconds between incongruent and congruent trials from Experiment 2. Error bars represent standard errors of the mean differences. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

incongruent trials had lower accuracy than congruent trials. Planned comparisons revealed a simple main effect of congruence within happy trials for the bilinguals; their accuracy was significantly higher in congruent happy than in incongruent happy trials ($Z = 3.56, p < 0.01$). Monolinguals were marginally more accurate in incongruent neutral compared to incongruent sad trials ($Z = 2.56, p < 0.1$). Bilinguals were significantly more accurate than monolinguals in congruent happy ($Z = 2.84, p < 0.05$) and incongruent sad ($Z = 2.71, p < 0.05$) conditions.

4.4.3.3 Conflict adaptation

Trial sequence effects are depicted in Figure 4.10. Because the present study involved both cognitive (gender) and emotional (facial expression) conflict, the influence of both dimensions was modelled with a Previous Congruence × Previous Flanker Valence interaction. RTs on incorrect trials and on trials following errors were removed from analysis. A Group × Valence × Congruence × Previous Congruence × Previous Valence LME model was fit to the RT data, with Participant

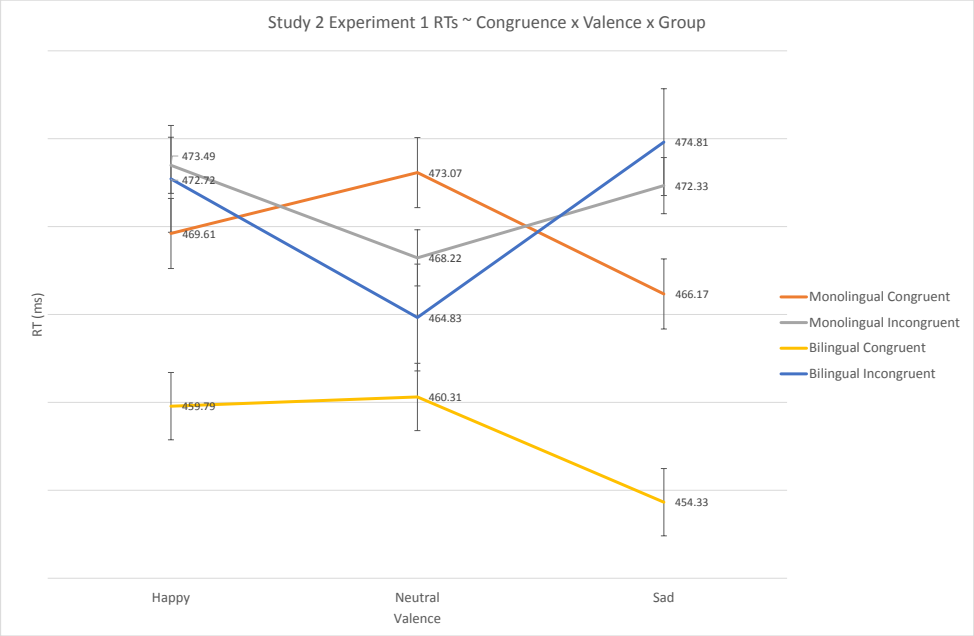

Figure 4.8: Visualisation of the relationships between Group × Congruence × Valence in Experiment 2, with experimental mean accuracy proportions. Error bars represent standard error of the mean.

Table 4.7: Experiment 2 accuracy GLMM ANOVA table. P-values are adjusted with a Bonferroni-Holm correction.* $p < 0.05$, ** $p < 0.01$.

Term	FStat	DF1	DF2	Corr. pValue
(Intercept)	1116.00	1	11622	$1.0E-233$
Group	6.81	1	11622	0.03181**
Congruence	13.55	1	11622	0.0016**
FValence	1.44	2	11622	0.2776
PartGender	0.08	1	11622	0.7721
Group:Congruence	3.24	1	11622	0.1680
Group:FValence	0.96	2	11622	0.3813
Congruence:FValence	1.46	2	11622	0.2776
Group:Congruence:FValence	1.46	2	11622	0.2776

Gender as a fixed covariate and with Target and Participant as crossed random intercepts. Random slopes were not included as the model did not converge with additional random effects parameters.

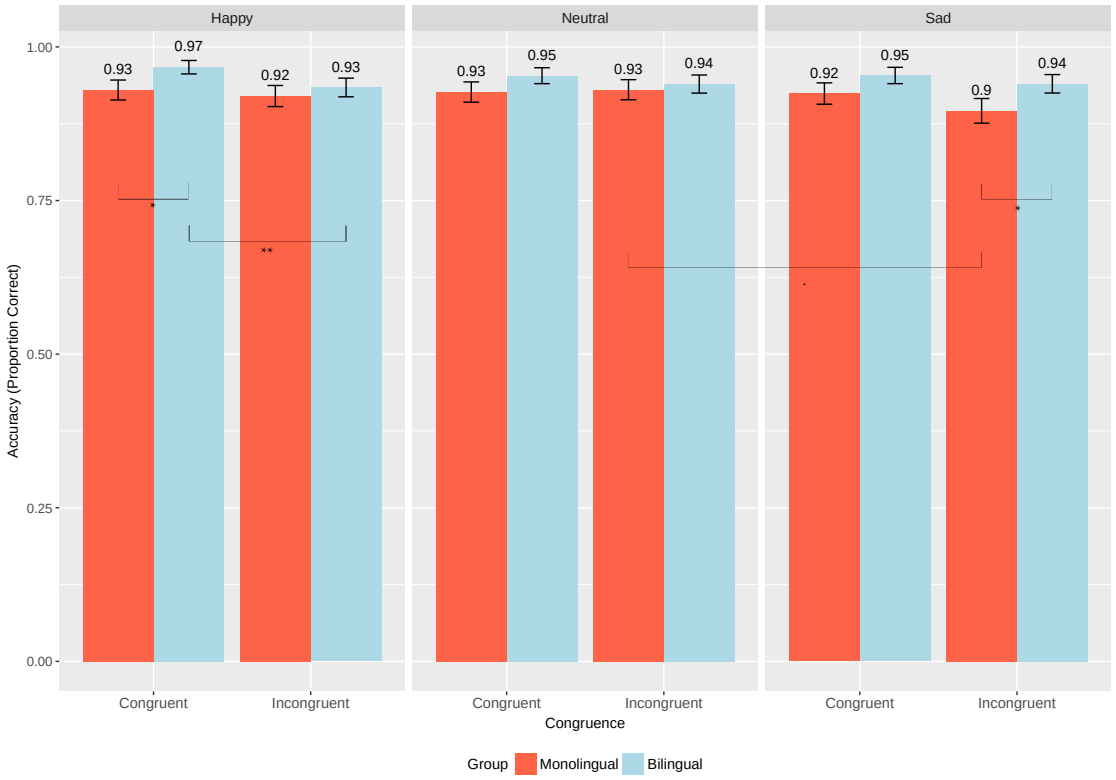

Figure 4.9: Mean proportion correct across conditions in Experiment 2. Error bars represent 95% confidence intervals. . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

An ANOVA of the model indicated a significant main effect of Previous Valence ($F(2, 9946.8) = 4.3246, p < 0.05$) and the expected Congruence × Previous Congruence × Previous Valence interaction ($F(2, 9948.4) = 5.4338, p < 0.01$). Responses in congruent trials were significantly faster for both groups when the previous trial was gender-congruent with happy flankers, and the magnitude of this flanker effect was significantly reduced if the preceding trial instead had happy gender-incongruent flankers. Preceding congruence had no significant effect in any other conditions when the valence of the preceding flankers was neutral or sad. To investigate the direction of these sequence effects further, separate LMEs were fit for valence-switch and valence-nonswitch trials, i.e. for trials preceded by flankers with the same emotional valence (N=2172) and those preceded by a different emotional valence (N=7930).

It was hypothesised that the Congruence × Previous Congruence × Previous Valence interaction was expected to remain significant in the valence-switch, but not the valence non-switch condition.

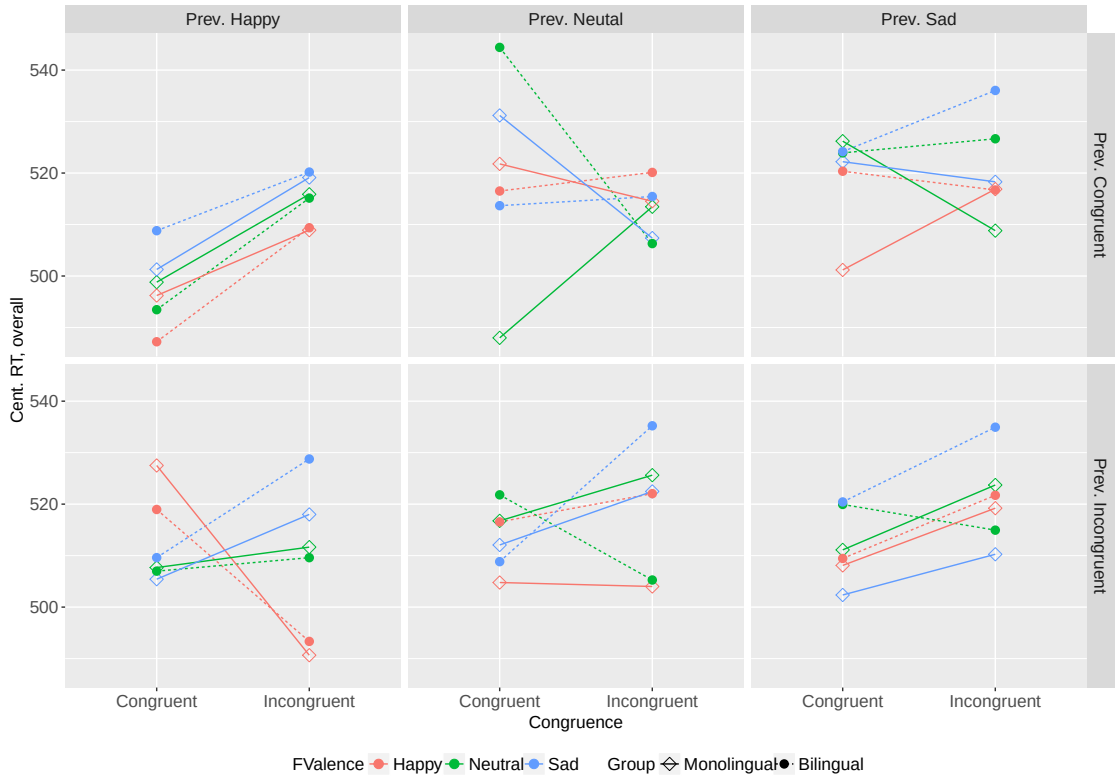

Figure 4.10: Successive trial sequence effects of Experiment 2. The horizontal panels represent previous trial valence, and the vertical panels represent previous trial congruence. Line colours indicate current trial valence (red=happy, green=neutral, blue=sad). Current trial congruence is represented on the x-axis. Solid lines with open diamonds represent the monolingual RTs, and dashed lines with solid circles the bilingual RTs. The slope of each line indicates flanker interference, so a steeper positive slope shows more interference from incongruent compared to congruent flankers. Based on our hypotheses, the slopes should be flatter following incongruent trials (the bottom panels) than following congruent trials (the top panel). The interaction between Congruence, Previous Congruence, and Previous Valence is thus represented by the differences in slopes in the top and bottom panels for each previous valence condition, and the line colours show the simple main effects of current trial valence.

This hypothesis was confirmed, with the critical interaction significant in the valence non-switch condition ($F(2, 2089) = 4.19, p < 0.05$), driven by faster RTs in two subsequent trials of congruent happy or incongruent happy flankers, but not the valence-switch condition ($F(2, 7826.9) = 1.64, p > 0.1$). There was, however, a significant Congruence × Previous Valence interaction in the latter condition, which post-hoc comparisons indicated was due to significantly faster RTs in congruent trials following happy trials compared with any other condition.

No significant between-group contrasts were found in the present model. However,

descriptive statistics revealed an intriguing difference between the groups in one specific condition, namely neutral trials followed by congruent neutral trials. Here, the monolinguals showed a classic flanker effect, with RTs in incongruent conditions approximately 25 ms slower than in congruent conditions. The bilinguals, however, had the opposite effect, with RTs that were on average 38 ms *faster* in incongruent than in congruent conditions.

4.5 Discussion

4.5.1 Experiment 1 Results

Experiment 1 failed to replicate the findings from Study 1. Contrary to Hypotheses 1a and 1b, bilinguals demonstrated a significant effect of emotional valence in this study, with sad flankers amplifying the effect of gender-congruence, and monolinguals had no effect of either congruence or valence in any condition. This contrasts with the Study 1 result, where monolinguals showed a highly significant congruence effect for happy flankers.

Flanker valence did, meanwhile, influence monolinguals' accuracy in this study, which was also not observed in Study 1. The d' results suggest that the monolinguals exhibited a larger speed–accuracy trade-off in sad conditions: while their RTs were fastest to gender-congruent sad flankers, their accuracy was lowest in these congruent sad conditions. This was driven by a high proportion of false alarms (13%), which suggests that the sad flankers interfered more with monolinguals' response inhibition. In happy and neutral trials, monolinguals exhibited the expected significant congruence effect. For bilinguals, congruence significantly impacted accuracy only in neutral conditions.

Returning to Hare et al.'s (2008) findings that negative targets interfered with younger participants' ability to detect the salience of a no-go cue in a go/no-go–only task, it is possible that the negative flankers in the present task had the same effect on monolinguals. It is unclear, however, why this effect was not observed in Study 1, where the main effect of congruence did not vary between groups or conditions. In fact, the monolinguals in Study 1 had the *highest* d' values in sad congruent trials. This may be due to individual differences between participants which could not be controlled or measured. Due to the complexity of the tasks, which required

participants to resolve both emotional and face-gender conflict, it is not possible to draw firm conclusions on this matter from the present data.

These findings refute the hypothesis that bilinguals are less distracted by emotional flankers when generalising across language background, number of languages spoken, ages and modes of acquisition. There are many tentative explanations as to why. Closer examination of the descriptive statistics broken down by number of languages and mode/age of acquisition reveals that multilinguals who spoke three or more languages fluently were, on average, faster than bilinguals. Within these groups, simultaneous multilinguals who had learned two languages from birth, *and* were fluent in at least one additional language, were faster than sequential multilinguals. Bilinguals, in contrast, were slower overall than both multilinguals and monolinguals. While the latter finding is unexpected, the sample sizes for these different groups were not large enough to draw reliable inferences. These findings do underscore the importance of including measures of language proficiency, sequence and age of acquisition, and number of languages spoken in bilingualism research, rather than treating 'bilingual' as a two-level categorical variable. Recent research on trilinguals, particularly those in so-called 'dense code-switching' contexts (Green & Abutalebi, 2013), suggests that proficiency and dominance in all three languages influences performance on executive control tasks (Zantout & Murphy, 2018). The differences between bilinguals in Study 1 and the present study may also be due to cultural differences, especially as emotion and culture are closely related (Markus & Kitayama, 1991; Potthoff et al., 2016). However, comparison across reported L1s showed very slight variation. When L1 was included in statistical models as a random effect, none of the individual levels reached significance, nor did the term improve the overall fit of the model as indexed by AIC.

The differences between Study 1 and the present study may also reflect a minor change in stimulus presentation timings. In Study 1, feedback was always presented at approx. $900ms$ after stimulus onset, regardless of whether or not a key was pressed. Some of the participants made multiple keypresses during a trial as a result and reported that they found the delay in feedback onset confusing. Therefore, the trial sequence was adjusted to occur immediately after keypress, which was, on average, approx. $450ms$ after stimulus onset. Although the time between feedback presentation and onset of the next trial was not changed in any way, the total time

from one stimulus presentation to the next (the inter-stimulus interval, or ISI) was reduced by an average of $344ms$. This slightly faster pace may have led participants to change their response criteria. Several studies have shown that ISI has an effect on N2 and P3 latencies in a go/no-go task (Zamorano et al., 2014). The overall RTs were an average of $40ms$ faster in the present study than in Study 1, suggesting that there were indeed differences between the two tasks.

The response-to-stimulus interval (RSI) may also have influenced the effect of congruence in monolinguals. Grundy et al. (2017) demonstrated that at shorter RSIs, conflict adaptation, indexed by a reduction in the magnitude of the flanker effect in trials following incongruent compared to those following congruent trials, was stronger in monolinguals than in bilinguals. At longer RSIs, conflict adaptation decreased in monolinguals so that they no longer differed from bilinguals. Neither Study 1 nor the present study had enough observations per participant to analyse conflict adaptation, and so it was not possible to investigate whether Grundy et al's (2017) findings were replicated in our design. However, conflict adaptation could explain the absence of a significant congruence effect in the monolingual group.

In sum, Experiment 1 failed to confirm our hypothesis that bilinguals have an advantage over monolinguals in attending only to task-relevant information. The larger RT flanker effect of bilinguals relative to monolinguals in sad flanker conditions rather indicated a bilingual disadvantage in flanker interference suppression when presented with emotional information. The d' results tell a somewhat different story. While happy and neutral gender-incongruent flankers decreased monolinguals' accuracy, the bilinguals had a significant simple main effect of congruence only in neutral trials. The descriptive statistics indicate that this was driven by increased false-alarm rates in no-go trials. Considered alongside the RT group means, the d' data points to different response strategies between the two groups. Bilinguals may have chosen a more cautious response threshold across all trials while monolinguals were faster, but therefore made more false alarms, in sad trials.

4.5.2 Experiment 2 Results

The FO task was included alongside the FNG task in this study to compare the effects of emotional flankers with and without a response inhibition component. A replication of the Experiment 1 results would be interpreted as evidence that

the RT findings in both experiments were due to the selective attention and/or interference suppression components of the task and not influenced by response inhibition demands. This was true for the bilingual group: as in Experiment 1, they showed an increased effect of gender-congruence in sad conditions. In the absence of response inhibition demands, however, the monolinguals showed the same effect.

The significant Congruence × (Sad) Valence effect in both groups refutes Hypothesis 2a and contradicts the interpretation that the differences between monolinguals and bilinguals reported in Study 1 were due to a bilingual advantage in selective attention. Rather, our results raise the possibility that the response inhibition component of the FNG task appears to have cancelled out the effect of negative valence on interference suppression.

This is further supported by the accuracy data. Positive gender-congruent flankers enhanced bilinguals' accuracy, while negative gender-incongruent flankers impeded monolinguals' accuracy, independently of their effects on reaction times. Group differences in accuracy, but not in RTs, support the interpretation that the groups used different response strategies. As previously discussed, response conflict cannot be reliably measured in a flanker task without concurrent EEG. Nevertheless, the conflict adaptation analysis lends additional credence to this interpretation.

Conflict adaptation effects were observed following happy, but not sad or neutral trials, contradicting Hypothesis 3b. The inter-trial effect does show that emotion had a valence-specific effect on conflict monitoring that cannot be simply attributed to interference suppression demands. The findings counter-intuitively suggest that positive stimulus valence aids conflict resolution, refuting previous findings based on emotional state (Van Steenbergen et al., 2010). Furthermore, incongruent happy trials appeared to result in 'too much' conflict adaptation when followed by another happy trial. A reversal of the flanker effect was observed whereby, following a previous incongruent happy (ih) trial, current congruent happy trials (ihCH) were significantly *slower* than incongruent happy trials (ihIH), with flanker effects of $-37ms$ and $-26ms$ for monolinguals and bilinguals, respectively. This may reflect conflict adaptation across several different control processes, rather than just interference suppression.

Based on the integrative framework of cognitive control and emotion introduced in Chapter 2, Figure 2.6, which accounts for conflict monitoring demands at the

attentional, executive, and response levels, it can be argued that happy flankers generate more conflict than sad flankers, especially in incongruent trials.

1. Going into each trial, the task goal is to identify the central target; gender-incongruent flankers create visual interference which conflicts with this goal.
    - Emotional flankers are more salient than neutral flankers and receive prioritised attention, thus increasing selective attention demands to maintain the goal.
    - Positive emotional states are argued to broaden the scope of attention and increase reliance on heuristic processing; successive happy trials may generate similar effects.
2. Identification of flanker gender interferes with labelling of the target gender, requiring interference suppression.
    - Appraisal of emotional facial expressions draws additional processing resources that interfere with target identification.
    - Negative flankers amplify this perceived conflict.
3. The target and flankers activate competing response tendencies (e.g. 'Respond M/Left' vs. 'Respond F/Right').
    - The tendency to approach appetitive cues amplifies the incorrect response tendency, increasing response conflict. Aversive cues from negative flankers have the opposite effect, facilitating the correct response.

Thus, happy incongruent trials generate additional response conflict that is not present in sad incongruent trials. Because the positive flankers enhance the incorrect response tendency, additional resources are recruited to inhibit this tendency. This adaptation is beneficial in the next trial when the flankers are again incongruent (ihIH), but creates an artificial conflict signal in ihCH trials. Additionally, participants may be slower to disengage attention from the spatial location of the flankers following a happy trial, as suggested by faster RTs in all incongruent conditions following congruent happy trials. This may facilitate

disambiguation and thus faster suppression of gender conflict in the subsequent trial, at the expense of 'over-inhibition' of congruent flankers.

Another intriguing result of the trial sequence analysis is the significant difference between groups on neutral trials following each other. While monolinguals showed a classic flanker effect with conflict adaptation, bilinguals again have a reverse flanker effect. Their mean RTs were *slower* in cnCN trials than in any other condition, with a flanker effect of -38ms relative to cnIN trials. Preceding gender conflict reduced this response latency by 22ms (inCN trials had a flanker effect of -17ms), indicating that preceding gender conflict facilitated the bilinguals' responses even in the absence of any sources of conflict in the present trial.

Counter-intuitively, this finding actually supports the conflict monitoring hypothesis: when neither the current nor the preceding trial contained gender or emotional conflict, bilinguals' monitoring levels were low, resulting in slower RTs. The monolinguals, meanwhile, had longer-lasting conflict adaptation effects (as demonstrated by Grundy et al., 2017). Congruent flankers thus facilitated faster RTs on subsequent congruent trials, supporting our hypothesis. Furthermore, when viewed alongside the RT and accuracy results discussed above, this result hints at an overall higher level of monitoring across all conditions by the bilinguals, with congruence and valence processed as additive sources of conflict, whereas the monolinguals' reduced accuracy in congruent happy and incongruent sad trials suggests more trial-to-trial variation in response tendencies. However, as task monitoring was not the primary focus of our study, this interpretation should be viewed with caution. See Zantout and Murphy (2018) and Costa et al. (2009) for evidence of monitoring under specific task manipulations.

4.6 Conclusion

The bilingual advantage we reported in Study 1 for a homogeneous group of sequential late German–English bilinguals could not be replicated in a more heterogeneous group of bi/multilinguals. While bilingual participants showed less interference from task-irrelevant emotional valence during a combined flanker–go/no-go task relative to monolingual controls in Study 1, the opposite was observed in Experiment 1 of the present study. Unlike in Study 1, effects of emotional valence on

response accuracy were also observed. These results indicated that negative valence, in particular, led to different response strategies in the monolingual compared to the bilingual group.

When the go/no-go component was removed in Experiment 2, there were no differences between monolinguals' and bilinguals' RTs in any condition. For the bilinguals, the results of both experiments were identical (significantly larger flanker effects in sad conditions). The monolinguals had a significant congruence effect in the sad condition only in Experiment 2. This is interpreted as a detrimental influence of negative valence on interference suppression for both groups. Both groups also showed comparable conflict adaptation effects in successive incongruent happy trials, but had opposite results in successive congruent neutral trials. Monolinguals showed facilitation from repeated congruence in the absence of valence conflict, but the bilinguals showed the opposite. In addition, bilinguals had an advantage in response accuracy compared to monolinguals in congruent happy and incongruent sad trials in the FO task. Together, these results are evidence for differing response strategies between groups that have no influence on RTs due to the simplicity of the task.

Overall, the fact that monolinguals were significantly affected by negative flankers in the flanker-only but not the FNG task supports the conclusion that the group differences observed in both Experiment 1 and Study 1 were due to differing response strategies, and under certain conditions a potential bilingual advantage in response inhibition. Negative valence possibly had a detrimental effect on monolinguals' response inhibition in Experiment 1, resulting in faster RTs in sad trials, which may have masked its influence on interference suppression. The fact that bilinguals did not differ between the FNG and FO tasks could be interpreted as a bilingual advantage in controlling the influence of negative emotion on response tendencies and/or coordinating multiple inhibitory processes in the more complex task. This is supported by previous EEG findings where response inhibition differences were observed between bilinguals and monolinguals at a functional level during a go/no-go task, but did not result in different behavioural measures (Costumero et al., 2015). As pointed out by Bialystok (2017), bilingual advantages in behavioural tasks—particularly in young adults who are at peak cognitive capacity—may only emerge in complex tasks, a finding that this phenomenon has in common with emotional influences on cognitive control (Pessoa, 2009).

Because accuracy remained at ceiling in the FNG tasks, it was not possible to analyse no-go trials in isolation and thus determine whether positive and negative flankers had opposite effects on false alarm rates, as predicted by previous studies (see Section 4.1.2). A more difficult version of the task, with smaller images or with faces that are artificially manipulated to make them harder to identify, may provide more conclusive results in this domain.

Additionally, as previously noted in Study 1, Section 3.6, developmental research can provide additional insight into the relationship between bilingualism, emotion, and response inhibition. While interference suppression mechanisms are relatively early to mature, leading to behavioural results similar to adults by around age 10 even when emotional stimuli are involved (Grose-Fifer et al., 2013), response inhibition mechanisms mature later in adolescence (Hwang et al., 2014; Tottenham et al., 2011). A replication of the experiments used in the present study with 11–12 year-old children may therefore result in measurable group differences.

The failure to replicate the results from Study 1 underscores the difficulty in separating possible bilingual advantages from other sources of individual differences such as cultural background, number of languages spoken, context of language use, and AoA. Two influential papers that criticise research on bilingual advantages, Paap and Greenberg (2013) and Paap and Sawi (2014), claim that no effects of bilingualism could be replicated across a battery of executive function tests with large sample sizes. However, while the authors differentiate between balanced and unbalanced bilinguals and/or degree of proficiency, their samples have very mixed language backgrounds, and they fail to differentiate between AoA and number of languages spoken when conducting their statistical analyses. Their participants are thus comparable to those in the present study. On the one hand, the effect of bilingualism in Study 1 may indeed be due to an unidentified feature of the homogeneous group other than bilingualism: culture, for example, predicts differences in cognitive emotional regulation strategies even between European countries (Potthoff et al., 2016). On the other hand, the large amount of individual variation in the heterogeneous Study 2 sample and in Paap et al.'s studies (Paap & Greenberg, 2013; Paap & Sawi, 2014) may prevent detection of a genuine advantage. As discussed in Section 4.1.1.1, with cautious study design—informed by interdisciplinary theoretical frameworks and backed up by neuroimaging studies—bilingual participants provide a unique

means for understanding the way language and environment mediate cognitive control (Fricke et al., 2018). Nevertheless, the results of the present study reflect the criticism of Paap et al. (2015) that bilingual advantages are disproportionately dependent on specific circumstances and nuances of task design and sample.

Les grandes personnes ne comprennent jamais rien toutes seules, et c'est fatigant, pour les enfants, de toujours et toujours leur donner des explications.

Adults never comprehend anything by themselves, and it's exhausting for children to have to explain things to them day in and day out.

— Antoine de Saint-Exupéry,
Le Petit Prince, pp. 6.

5 Study 3

Contents

5.1 Background 113
5.1.1 Maturation of Inhibition and Emotion Regulation Mechanisms 115
5.1.2 Bilingual Advantages in Childhood 117
5.2 The Present Study 118
5.2.1 Research Questions and Hypotheses 118
5.3 Methods 120
5.3.1 Experiment 1: Flanker–go/no-go task 120
5.3.2 Experiment 2: Flanker-only task 122
5.3.3 Data Analysis Plan 122
5.4 Results 123
5.4.1 Sample Demographics 123
5.4.2 Experiment 1 123
5.4.3 Experiment 2 126
5.5 Discussion 128
5.5.1 Experiment 1 Results 128
5.5.2 Experiment 2 Results 133
5.6 Conclusion 135

5.1 Background

The goal of the present study was to test the effect of irrelevant emotional distractors in a population whose executive control circuits are still developing, and thus more susceptible to emotional interference than adults'. We report the results of the

flanker–go/no-go (FNG) and flanker-only (FO) tasks introduced in Studies 1 and 2 with three different groups of 10–11 year-old children: English–German bilinguals, English–French bilinguals, and English monolinguals.

Comparing children's response patterns in FNG and FO tasks to those of the adults in Studies 1 and 2 may help separate interference suppression from response inhibition processes. Developmental research is often enlightening when attempting to disentangle executive processes, as these mature at different rates (Astle & Scerif, 2009). Neuroimaging studies provide solid evidence that children's response inhibition is less mature than adults', and indicate that designated neural networks for interference suppression have not yet developed in this age group (Brydges et al., 2013; Vuillier et al., 2016). Furthermore, children show greater activation in emotion processing circuits during executive control involving emotion, even when they are task-irrelevant (Hwang et al., 2014; Somerville et al., 2011; Tottenham et al., 2011). This means that in inhibition tasks which include emotional stimuli, children are less accurate (Tottenham et al., 2011) and have larger flanker interference effects (Grose-Fifer et al., 2013) than adults.

Study 3 had two objectives:

1. Link cognitive advantages related to bilingualism to socio-emotional advantages related to executive control in children.

2. Determine whether the group differences in response strategies observed in adults translate to a bilingual advantage in response inhibition in children.

Establishing a relationship between bilingualism and emotion regulation in children would have important implications for educational research in this area. Emotion regulation ability is associated with a host of positive academic and social outcomes (Checa et al., 2008). However, only a handful of studies have directly investigated bilingualism as a factor in these associations. The findings so far link bilingual language ability to better social and communicative competence outcomes (Beck et al., 2012; Liberman et al., 2017). To the authors' knowledge, no studies to date have linked bilingualism directly to children's performance on executive control tasks with incorporated emotion regulation demands.

5.1.1 Maturation of Inhibition and Emotion Regulation Mechanisms

Comparison of the FNG tasks in Studies 1 and 2.1 and the FO task in Study 2.2 pointed to differences in the monolingual and bilingual adults' response strategies depending on flanker valence, previous trial conflict, and response-to-stimulus interval (RSI). First of all we observed effects of emotional valence for both groups in Study 2, refuting our hypothesis that bilinguals would selectively attend to the response dimension of the task (face gender) and not process the emotion dimension. Secondly, there were differences between groups in the FNG task, but not the FO task. Critically, there was a negative valence-specific speed-accuracy trade-off in the Study 2 FNG task whereby monolinguals responded more quickly at the expense of higher error rates, while bilinguals responded more slowly but made fewer errors. The same group difference in error rates to sad trials was observed in the FO task, but without accompanying RT differences. This indicated that monolinguals and bilinguals differed at the response inhibition rather than the interference suppression level. Negative valence interfered with monolinguals' response inhibition in the FNG task, and may have masked effects of flanker interference by decreasing RTs in incongruent trials. The interference suppression demands of the task may also simply have been too low to reveal reliable group differences in young adults at ceiling performance.

In order to further disentangle the effects of positive and negative distractors on interference suppression and response inhibition, we collected developmental data. Compared with adults, 8–11 year-old children show less efficiency and functional specialisation of response inhibition as measured by N2 ERPs on no-go trials (Brydges et al., 2013; Vuillier et al., 2016). The same studies observed interference suppression-related N2s in the adults, but not in the children, during go trials with incongruent flankers (Brydges et al., 2013) or following a spatially incongruent pre-stimulus cue (Vuillier et al., 2016). This is consistent with other findings which suggest that children rely on more distributed executive attention mechanisms to resolve conflict (see e.g. discussion in Rueda et al., 2010). Thus, although children's behavioural performance on emotionally neutral flanker tasks is comparable to that of adults by approximately age seven (Rueda et al., 2004), the neural mechanisms they rely on to complete these tasks are less mature and efficient.

Given this relative functional immaturity, children's performance in executive function tasks is more susceptible to emotional interference (e.g. Hwang et al., 2014). Both children and adolescents are less accurate and make more false alarms in emotional go/no-go tasks than do adults (Hare et al., 2008; Somerville et al., 2011; Tottenham et al., 2011). They also exhibit reliable differences between positive and negative stimuli. In the flanker task, negative flankers produce more conflict and thus lead to slower RTs than do positive flankers, and this effect is more pronounced in younger children (<14 years) compared to older adolescents (Grose-Fifer et al., 2013). In the go/no-go task, positive valence increases false alarm rates and decreases RTs relative to neutral and negative valence, while the effect of negative valence varies more depending on age, response condition, and strategies such as speed-accuracy trade-off (Hare et al., 2008; Somerville et al., 2011; Tottenham et al., 2011).

Developmental neuroimaging studies indicate that negative stimuli interfere with children's ability to attend to task-relevant information in conflict tasks: 11–12 year-old children showed significantly increased reactivity to task-irrelevant emotional background images in emotion-related brain circuits during an affective Stroop task, although their behavioural performance on the interference suppression aspect of the task did not differ from adults (Hwang et al., 2014). Thus, negative stimuli may interfere with children's detection of the no-go component of the stimulus during go/no-go tasks, explaining the variation in findings. Hwang et al. (2014) also observed greater activation of frontal and temporal executive control networks in children and adolescents relative to adults in negative trials, which they interpret as an indication that these younger participants exerted greater effort to down-regulate emotional responses. Such increased effort could also be predicted to result in speed-accuracy trade-offs, although none were observed in the behavioural results from their study.

The RSIs were shorter in Study 2 compared to Study 1, which increases monitoring demands and produces more conflict adaptation effects (Grundy et al., 2017). When task demands are sufficiently complex, even adults show interference both behaviourally and neurally from irrelevant negative stimuli (D. Mitchell et al., 2007). It is thus possible that our FNG task was complex enough to produce a behavioural cost of this interference in Study 2, while Study 1 only revealed the more reliable effect of positive valence on response inhibition. In the present study

we increased RSIs to remain consistent with other studies utilising flanker–go/no-go tasks in this age group (Brydges et al., 2013; Checa & Rueda, 2011). If so, we would expect the children in the present study to have higher false alarm rates in both positive and negative trials of the FNG task, while only negative flankers should generate slower RTs in the FO task.

5.1.2 Bilingual Advantages in Childhood

Adaptations in cognitive control due to bilingual experience are also more apparent in younger children. For example, Yang and Yang (2016) reported bilingual advantages in 5–6 year-old children and in young adults who were bilingual in Korean and English. Both groups had an advantage in local and global measures of attention as measured by the child/adult ANT tasks (Fan et al., 2002; Rueda et al., 2004), respectively. Notably, the magnitude of the bilingual children's performance advantage over their monolingual peers on the global measures (overall RT and accuracy) was larger than the equivalent group contrast in adults. In addition, the bilingual children's accuracy in incongruent flanker trials was comparable to the adults', suggesting that experience using multiple languages leads to earlier maturation of interference suppression.

The latter finding is critical as interference suppression, in particular, is associated with positive academic and social outcomes in children (Checa & Rueda, 2011; Rueda et al., 2010). Bilingualism is thus a potential protective factor in populations at risk of adverse outcomes, for example, those with low SES. White and Greenfield (2017) found advantages in bilingual children's performance on a battery of executive function tests in a large cohort of low-income Spanish–English bilingual, Spanish-dominant emerging bilingual, and English monolingual preschoolers. The emerging bilinguals' outcomes were between those of the monolinguals and bilinguals, suggesting that this group was already developing enhanced EF due to learning English. Bilingualism is also associated with increased cortical surface area, an indicator of brain development, in adolescents from low SES backgrounds relative to their monolingual peers (Brito & Noble, 2018). Furthermore, Bialystok and Barac (2012) demonstrate that bilingual advantages in executive functions are dissociable from linguistic ability: among two different groups of emerging bilingual children in language immersion programmes, length of time spent in the programme was

correlated with performance on non-verbal executive control tasks, while proficiency in the immersion language correlated with several measures of metalinguistic awareness. This suggests a causal link between bilingualism and executive control, rather than better innate executive control increasing a child's propensity for learning multiple languages.

While ample behavioural evidence supports the hypothesis that bilingual children have an executive control advantage in interference suppression, such an advantage has not thus far been observed in response inhibition tasks such as the go/no-go task (see Bialystok, 2017; Martin-Rhee & Bialystok, 2008). Neither has any study to the author's knowledge included emotional stimuli in measures of bilingual children's executive control. It is therefore difficult to predict how bilingual children may differ from their monolingual peers on these dimensions when emotions are involved. However, as discussed in Chapter 2.2.3, bilingualism is positively associated with children's social and emotional skills (Beck et al., 2012; Liberman et al., 2017; Yow & Markman, 2011). These same skills, indexed by measures such as temperamental effortful control and peers' sociability ratings, are in turn correlated with adolescents' performance on ANT tasks with and without response inhibition (Checa et al., 2008; Checa & Rueda, 2011). Although tenuous, we hypothesise that these links predict bilingual advantages for both response inhibition and interference suppression in the presence of emotionally distracting stimuli. This would be a significant finding as response inhibition, particularly to positive cues, is associated with decreased risk-taking behaviour in adolescents (Somerville et al., 2011).

5.2 The Present Study

In the present study, 10–11 year old children completed the same FNG and FO tasks as the adults in Studies 1 and 2, although instructions and trial timing were adjusted to make them more child-friendly.

5.2.1 Research Questions and Hypotheses

1. Are bilingual 10–11 year-old children better at regulating task-irrelevant affective information than their monolingual peers?

2. How do bilingual and monolingual children compare to the respective adult samples from Studies 1 and 2 on a) the FNG and b) the FO task?
3. Comparing all three studies, can group differences be attributed to selective attention, interference suppression, or response inhibition?

The first experiment of the present study sought to compare 10–11 year-old children's performance on the FNG task to that of adults in Study 1 and Study 2.1. Experiment 2 compared the same sample's performance on the FO task to the adults' performance in Study 2.2. Comparison of the Experiment 1 and 2 outcomes allowed us to separate effects of response inhibition, due to the go/no-go component, from effects of flanker interference suppression. As in Study 2.2, response inhibition demands in the FO task were minimised by including an equal ratio of congruent to incongruent trials (see Section 4.1.2).

5.2.1.1 Experiment 1

Hypothesis 1a: Both monolingual and bilingual children were expected to show a Congruence × Valence interaction on their RTs in this task, with both positive and negative flankers increasing RTs in incongruent trials and decreasing them in congruent trials relative to neutral flankers. This should mirror the results from the monolingual adults in Study 1.

Hypothesis 1b: The bilingual children were expected to outperform their monolingual peers at the level of interference suppression: that is, they were expected to show less interference from positive and negative flankers in incongruent conditions, and a smaller effect of congruence overall. In line with previous findings, negative flankers were expected to produce a larger flanker effect than positive flankers, particularly for the monolinguals.

Hypothesis 1c: Critically, both groups were expected to show an effect of valence on accuracy, with positive flankers in particular interfering with response inhibition: the children were expected to have more false alarms and fewer misses in happy than in sad trials. The monolinguals were expected to show a Congruence × Valence interaction, with more misses in incongruent sad trials and more false alarms in incongruent happy trials.

5.2.1.2 Experiment 2

Hypothesis 2a: Bilingual children were expected to show an interference suppression advantage over monolingual children, resulting in faster RTs in incongruent trials. As observed in Study 2, negative flankers were expected to cause more flanker interference than positive.

Hypothesis 2b: While the task did not involve response inhibition, an effect of valence on response selection was still expected, resulting in higher accuracy and faster RTs in positive congruent trials. The children were expected to show less evidence of cross-trial response strategies such as speed-accuracy trade-offs or valence-dependent conflict adaptation than the adults in Study 2. A simple Previous Congruence × Current Congruence interaction was expected that may be amplified following emotional trials regardless of valence.

5.3 Methods

5.3.1 Experiment 1: Flanker–go/no-go task

5.3.1.1 Design and Procedure

Experiment 1 replicated the design of Studies 1 and 2.1 (see Sections 3.3 and 4.3.1), except that the instructions were modified to be more child-friendly by adding a story. Children were told that they were attending a party with their friend and that they needed to help find their friend's brother/sister in a crowd of people, but since they didn't know what (s)he looked like, they should respond as quickly as possible every time they saw a boy/girl so that they could get their friend's attention. The maximum response time allowed for each trial was increased from 950 to 1200 ms, and feedback after each trial was changed to "Yay, that was right", "Whoops, that was a boy/girl!", or "Whoops, too slow, try to be quicker!". Figure 5.1 demonstrates the experimental trial sequence. As in Study 1, feedback always appeared after timeout.

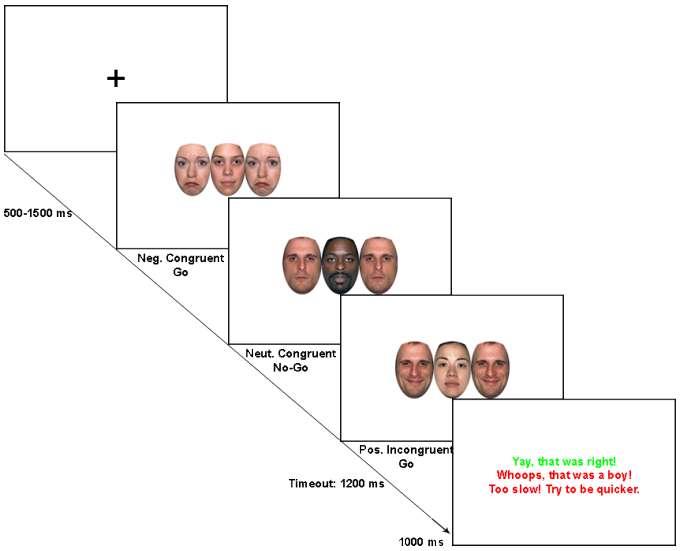

Figure 5.1: Experimental trial sequence. (Go condition = female).

5.3.1.2 Equipment

The experiment was administered using Psychophysics Toolbox (Version 3.0.12) extensions (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli, 1997) for MATLAB R2016b. Seven testing laptops were used for the data collection session. All laptops were synchronised to a screen refresh rate of 60hz, and stimulus flip time-stamping was manually checked on each machine to ensure that the recorded SOTs and RTs were reliable. Drift was <1ms on each laptop. Stimulus size in mm and approx. subtended visual angle were identical on each machine.

5.3.1.3 Participants

20 English–German bilingual, 16 English–French bilingual, and 20 English monolingual children aged 10–11 completed Experiment 1. All of the bilingual students attended a partial-immersion school where they were taught in either German or French for half of the week and in English for the other half. Approximately 50% of these students spoke German or French natively with one or both parents, while the other half had been enrolled in the school for at least four years. Some children also spoke a third language (Arabic, Chinese, Czech, Dutch, Farsi, Greek, Italian, Polish, or Spanish) at varying levels at home, and some had previously attended school in another language (e.g. Arabic, Dutch, Spanish) before moving to the UK[1]. At the time of testing, the bilingual students had completed approximately half of Year 6 (mean age=10.7 years). In addition to the 36 included participants, two students opted out of the study, three students with reported learning disabilities were excluded, and one student requested to end the experiment early.

The monolingual children attended a school in the same geographic region whose catchment area approximately matched the SES and demographic backgrounds of the bilingual school. 22 participants were selected by their teacher[2] from a group of 90 students enrolled in beginner language classes at the start of Year 7.

[1]One student had moved very recently and hardly spoke English, but was included in the study as he was bilingual in French and his home language. Another was a French–English bilingual who had just moved from France, but joined the German class and was not yet proficient in German. She was included in the French group.

[2]Teachers were asked to select the participants in order to protect the childrens' confidentiality: the study necessitated the exclusion of students with learning disabilities or other psychiatric conditions, but the researcher was not authorised to collect this information from the children, so teachers were asked not to send students who met these exclusion criteria to the experiment session. However, allowing teachers to select the participants could have introduced a bias.

The monolinguals were approximately six months older than the bilingual groups (mean age=11.2 years) and were slightly more racially/ethnically diverse. Those who reported fluency in a language other than English and those with learning disabilities were excluded, and one opted out of the study. Data from two participants was excluded due to reported fluency in a second language.

5.3.2 Experiment 2: Flanker-only task

5.3.2.1 Design and Procedure

Study 3.2 was run concurrently to Study 3.1. Participants completed a flanker-only task with the same design, instructions, and trial sequence as the combined task (see Figure 5.1), except that there was an equal number of female and male target faces and the ratio of congruent to incongruent trials was 50/50. The students were instructed to press the "A" and "L" keys on the laptop keyboard (marked with different-coloured stickers) to indicate whether the central face was male or female. This key assignment alternated between participants. The maximum response window was increased from 950 to 1200 seconds. For participants who completed both experiments (the German–English group), the order of the tasks was counterbalanced.

Testing equipment was as described in Experiment 1 above.

5.3.2.2 Participants

Participants were from the same groups and schools as those in Experiment 1. The 20 German–English bilingual students and one French–English bilingual student completed both experiments; the order of experiments was counter-balanced for this group. 14 English–French bilingual students and 23 English monolingual students completed only Experiment 2. In addition to the 23 monolinguals included, one participant's data was removed due to reported fluency in another language.

5.3.3 Data Analysis Plan

The data analysis methods used in this study were the same as those reported in Studies 1 and 2 and are explained in detail in Appendix A. Additional random effects terms were added during model selection to account for the additional variance in the samples due to individual differences in development, as well as

varying schools, classes, SES, language backgrounds, and order of experiments: see the scripts in Appendix B. Participant age and an age × congruence interaction (accounting for more mature interference suppression mechanisms in older children) were included as fixed covariates alongside gender and gender × valence, and all four combinations were tested via simulated LRTs.

5.4 Results

5.4.1 Sample Demographics

There were no significant differences between the German and French children in either RTs or accuracy, and so their data was combined into one bilingual group for all analyses. Random effects terms for L1, language group, bilingualism type (simultaneous or sequential), school, and class were included during model selection, but did not significantly improve the fit of the statistical models. Unfortunately not every class' teacher returned the vocabulary tests, and so LexTALE scores could not be included in the models. The fixed covariate terms for age and age × congruence were not significant in any task. This suggests that these factors did not account for significant additional variance in the data and the samples were thus sufficiently matched on all available measures.

5.4.2 Experiment 1

Table 5.1 lists the mean go accuracy and false alarm rates and the median RTs from Experiment 1. RTs and accuracy were analysed separately with (G)LME models as described in Section 5.3.3 and Appendix A. Fixed covariates for participant age and for gender × valence were not significant based on LRTs and are thus not reported.

5.4.2.1 RT Data

Within hit RTs, there was a significant main effect of Congruence ($F(1, 35.17) = 4.55, p = 0.04$), as well as a Group × FValence interaction ($F(2, 5947.5) = 3.27, p = 0.038$). Planned contrasts indicated that the interaction was driven by a significant simple main effect of valence for the monolingual, but not the bilingual children, in line with our hypotheses. The bilingual children had a significant effect of congruence ($T(43) = 2.16, p = 0.04$) across all conditions of

Table 5.1: Mean go (% hits) and no-go (% false alarms) accuracy and median RTs in ms for Experiment 1.

Group	Congruence	FValence	Hit Rate	SD	FA Rate	SD	Med. Hit RT	SD
Monolingual	Congruent	Sad	0.91	0.29	0.30	0.46	553.57	130.23
		Happy	0.91	0.28	0.29	0.45	561.05	128.45
		Neutral	0.90	0.31	0.32	0.47	565.21	128.76
	Incongruent	Sad	0.91	0.28	0.31	0.46	566.98	132.27
		Happy	0.86	0.34	0.31	0.47	576.16	119.32
		Neutral	0.91	0.29	0.38	0.49	552.05	129.76
Bilingual	Congruent	Sad	0.93	0.25	0.19	0.39	544.60	115.94
		Happy	0.92	0.27	0.17	0.38	540.14	118.12
		Neutral	0.91	0.28	0.16	0.37	542.37	115.16
	Incongruent	Sad	0.92	0.27	0.23	0.42	555.48	113.74
		Happy	0.90	0.30	0.19	0.40	553.02	120.74
		Neutral	0.90	0.30	0.26	0.44	550.81	122.07

valence. Planned comparisons revealed marginally significant simple main effects in neutral ($T(4409.6) = -1.76, p = 0.078$) and sad ($T(4409.6) = -1.82, p = 0.068$) but not in happy trials. The monolinguals had a marginally significant simple main effect of congruence only in happy trials, ($T(94.13) = -1.69, p = 0.09$). In neutral conditions, they had a reverse flanker effect, with faster RTs in incongruent trials. These results are summarised in Figures 5.2 and 5.3.

Table 5.2: Type–III ANOVA table for hit RTs in Experiment 1.

Term	FStat	DF1	DF2	pValue
(Intercept)	16058.999	1	86.739	0.000
Group	0.153	1	53.858	0.697
FValence	0.310	2	5829.088	0.734
Congruence	4.545	1	35.170	0.040
Gender	2.469	1	54.549	0.122
Group:FValence	3.275	2	5947.518	0.038
Group:Congruence	0.190	1	52.161	0.665
FValence:Congruence	0.528	2	5865.414	0.590
Group:FValence:Congruence	0.797	2	5981.053	0.451

5.4.2.2 Accuracy Data

As expected, both groups were significantly more accurate in go compared to no-go trials, $F(1, 11151) = 452.1, p < 0.0001$. There were also significant main effects of

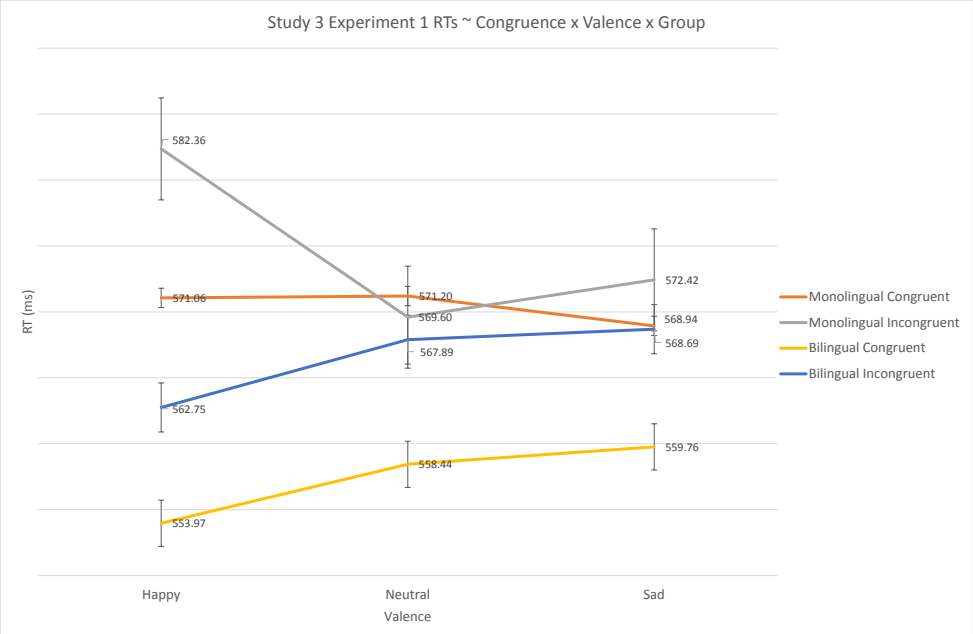

Figure 5.2: Visualisation of the relationships between Group × Congruence × Valence in Experiment 2, with experimental mean RTs in milliseconds. Error bars represent standard error of the mean.

Group $F(1, 11151) = 6.02, p = 0.014$, flanker Valence $F(1, 11151) = 3.32, p = 0.036$, and Congruence $F(1, 11151) = 11.239, p = 0.0001$. Crucially, there was a significant Group × GoStatus interaction, $F(1, 11151) = 9.574, p = 0.002$. The bilingual children committed significantly fewer false alarms than the monolinguals in all conditions, but did not differ from monolingual children in their miss rates. Planned comparisons showed that the bilinguals were also significantly more accurate in congruent compared to incongruent trials across all conditions.

A GoStatus × FValence × Congruence interaction was also observed, $F(2, 11151) = 3.26, p = 0.04$. Specifically, within the bilingual group there were significant simple main effects of congruence in neutral and sad no-go conditions. This was due to higher false alarm rates in incongruent trials. For the monolingual group, there was a significant simple main effect of congruence in positive go trials. The monolingual children had more misses in positive incongruent trials. There were no differences between happy and sad trials, neither in hit (go) nor in false alarm

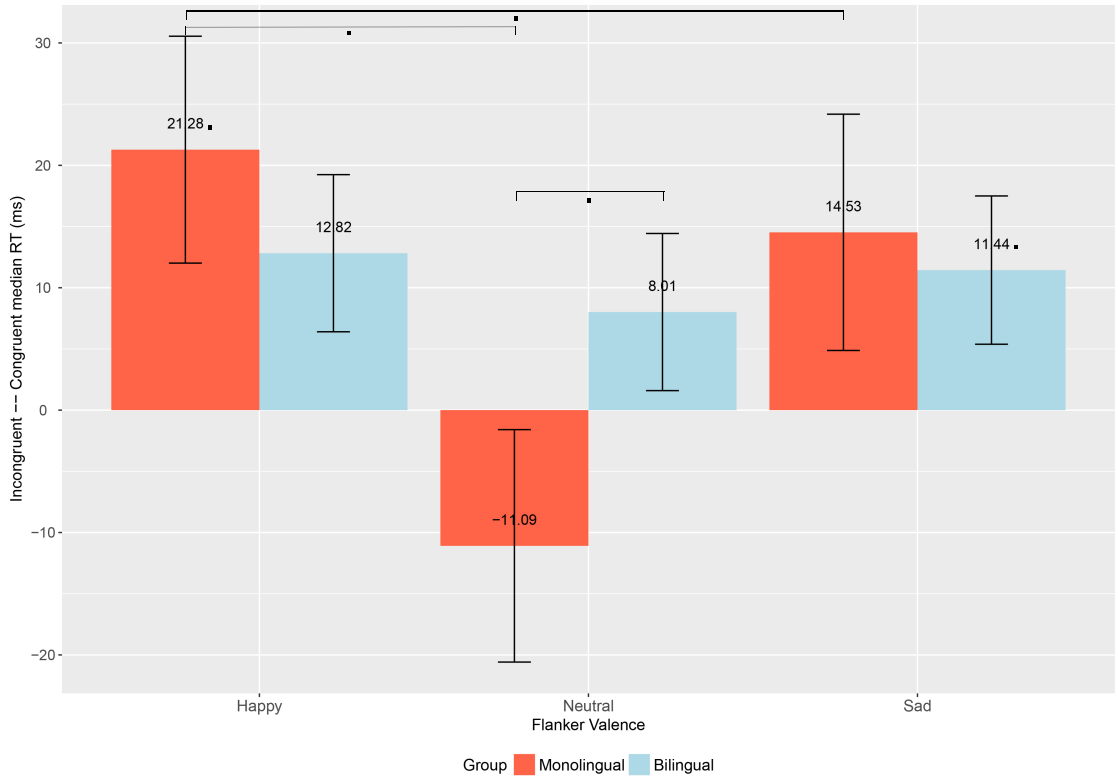

Figure 5.3: Mean reaction time differences in milliseconds between incongruent and congruent trials from Experiment 1. Error bars represent standard errors of the mean differences. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

(no-go) rates, in either congruence condition. The accuracy results are summarised in Figures 5.4, 5.6, 5.5 and 5.7.

5.4.3 Experiment 2

Table 5.4 lists the descriptive statistics for Experiment 2. Median RTs are reported due to large numbers of outliers.

5.4.3.1 RT Data

None of the terms in the overall RT model reached significance. Planned comparisons revealed a marginally significant Congruence × FValence interaction in the bilingual group, where there was a simple main effect of congruence in the happy condition, $T(337.82) = 2.09, p = 0.036$. This effect differed significantly from sad trials ($T(5910) = 2.26, p = 0.027$). As shown in Figure 5.9, this was due to a reverse flanker effect in the sad condition ($-8.26ms$). The monolingual children also showed a reverse flanker effect in neutral trials ($-6.32ms$), but this effect was

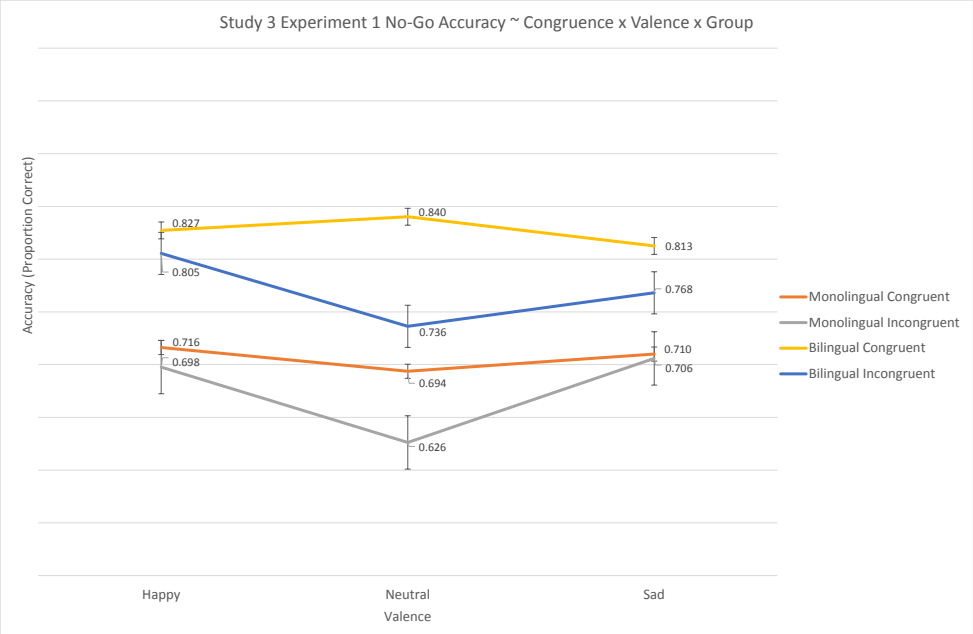

Figure 5.4: Visualisation of the relationships between Group × Congruence × Valence on no-go accuracy in Experiment 1, with experimental mean proportion correct. Error bars represent standard error of the mean.

not significant, nor did the groups differ significantly from each other in any condition. See Figures 5.8 and 5.9.

5.4.3.2 Accuracy Data

For accuracy in the FO task, there was a significant main effect of Congruence, $F(1, 10566) = 4.65, p = 0.031$. There was also a marginally significant Group × Congruence × FValence interaction, $F(2, 10566) = 2.37, p = 0.094$. This was driven by the monolingual children's higher accuracy in congruent sad conditions than in any other condition, $z = -2.27, p = 0.023$, and the bilingual children's higher accuracy in congruent neutral, $z = -2.54, p = 0.011$. In both cases the congruent trials produced significantly fewer errors than the incongruent trials. Post-hoc contrasts between the groups revealed a significant difference only in the bilinguals' higher accuracy relative to monolinguals in congruent neutral, $z = 2.02, p = 0.043$. See Figures 5.10 and 5.11.

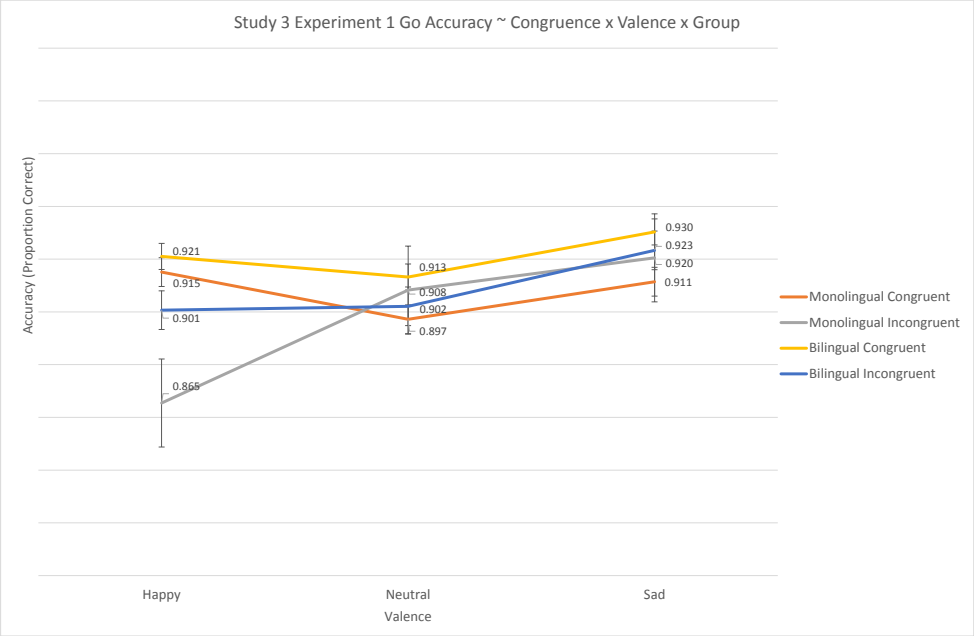

Figure 5.5: Visualisation of the relationships between Group × Congruence × Valence on go accuracy in Experiment 1, with experimental mean proportion correct. Error bars represent standard error of the mean.

5.5 Discussion

5.5.1 Experiment 1 Results

The most significant finding from Experiment 1 was a bilingual advantage in response inhibition. The German and French bilingual children committed fewer false alarms than their monolingual counterparts in all conditions, supporting Hypothesis 1c. This is, to our knowledge, the first study to show a behavioural difference between monolingual and bilingual children on a go/no-go task. The groups did not differ in their accuracy in go conditions, which supports the interpretation that this in indeed an advantage in response inhibition rather than a difference in overall response strategies, as observed, for example, in Study 2. However, further study is required to verify this, as the complexity of the present tasks adds too many confounds.

The RT results replicated our findings from Study 1 almost exactly: the bilinguals had a main effect of congruence unmodulated by flanker valence, while the monolinguals had a significant flanker effect only in positive trials and were faster

Table 5.3: Type–III ANOVA table for binary responses (correct/incorrect) in Experiment 1.

Term	FStat	DF1	DF2	pValue
(Intercept)	216.471	1	11151	0.000
Group	6.020	1	11151	0.014
GoStatus	452.098	1	11151	2E-98
FValence	3.323	2	11151	0.036
Congruence	11.239	1	11151	0.001
Gender	0.003	1	11151	0.957
Group:GoStatus	9.574	1	11151	0.002
Group:FValence	0.441	2	11151	0.643
GoStatus:FValence	0.531	2	11151	0.588
Group:Congruence	1.312	1	11151	0.252
GoStatus:Congruence	2.080	1	11151	0.149
FValence:Congruence	0.378	2	11151	0.685
Group:GoStatus:FValence	0.687	2	11151	0.503
Group:GoStatus:Congruence	0.417	1	11151	0.519
Group:FValence:Congruence	1.244	2	11151	0.288
GoStatus:FValence:Congruence	3.260	2	11151	0.038
Group:GoStatus:FValence:Congruence	0.125	2	11151	0.883

Table 5.4: Mean accuracy and median RTs in ms for Experiment 2.

Group	Congruence	FValence	Accuracy	SD	Med. Correct RT	SD
Monolingual	Congruent	Sad	0.91	0.28	670.08	178.55
		Happy	0.90	0.30	679.39	173.32
		Neutral	0.87	0.33	687.20	176.48
	Incongruent	Sad	0.87	0.33	671.66	167.12
		Happy	0.89	0.31	689.57	175.77
		Neutral	0.88	0.32	681.78	164.09
Bilingual	Congruent	Sad	0.89	0.31	688.12	159.09
		Happy	0.87	0.33	681.59	161.79
		Neutral	0.89	0.31	681.03	156.66
	Incongruent	Sad	0.87	0.33	683.66	160.05
		Happy	0.89	0.32	701.67	156.49
		Neutral	0.86	0.35	686.90	158.33

on incongruent than on congruent neutral trials. These findings support Hypothesis 1b as well as our prediction for monolinguals from Hypothesis 1a. Contrary to Hypothesis 1a, however, the bilingual children had no effect of valence in their

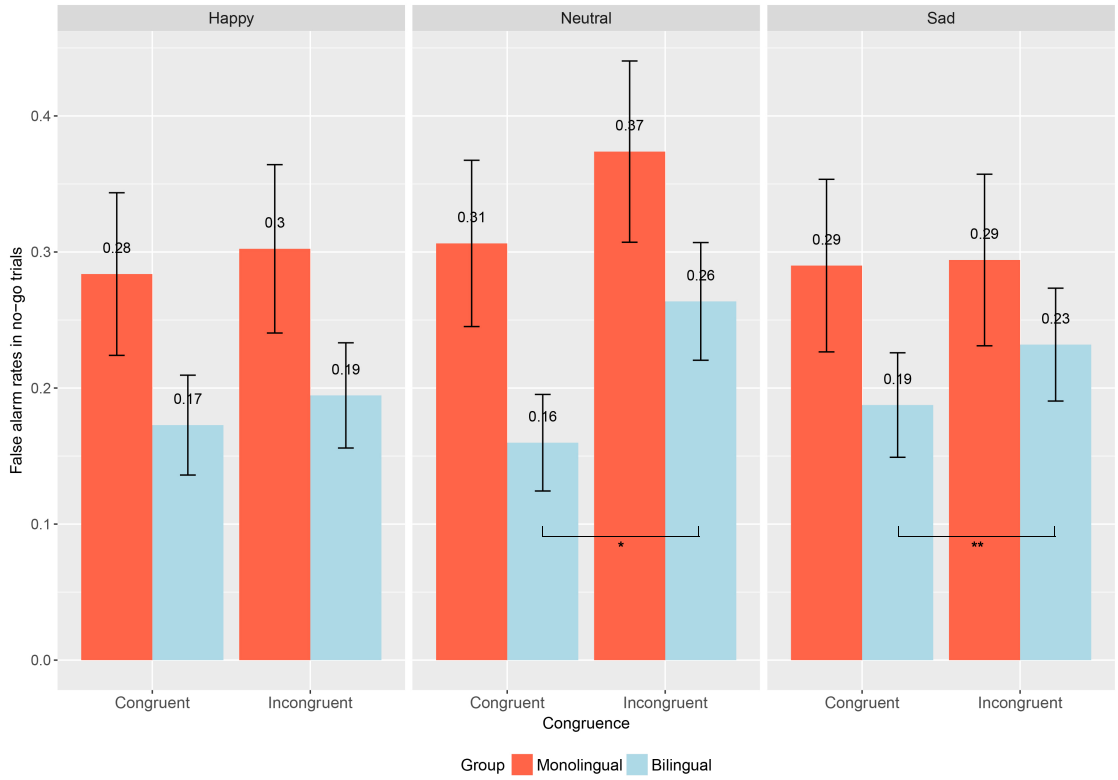

Figure 5.6: False alarm rates in no-go trials for Experiment 1. Error bars represent 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.5: Type–III ANOVA table for hit RTs in Experiment 2.

Term	FStat	DF1	DF2	pValue
(Intercept)	29323.60	1	84.75	0.00
Group	0.25	1	55.52	0.62
FValence	0.16	2	4724.55	0.85
Congruence	1.78	1	33.53	0.19
Gender	0.12	1	55.41	0.73
Group:FValence	0.27	2	4713.35	0.76
Group:Congruence	0.13	1	4546.71	0.72
FValence:Congruence	1.19	2	4531.54	0.30
Group:FValence:Congruence	1.23	2	4536.25	0.29

RTs. This implies that, though slower and less accurate overall, the children were nearing adult-like behavioural performance on this task. This is supported by neuroimaging studies where ERP or fMRI measures reveal greater activation in emotion processing circuits, but comparable behavioural performance, for children

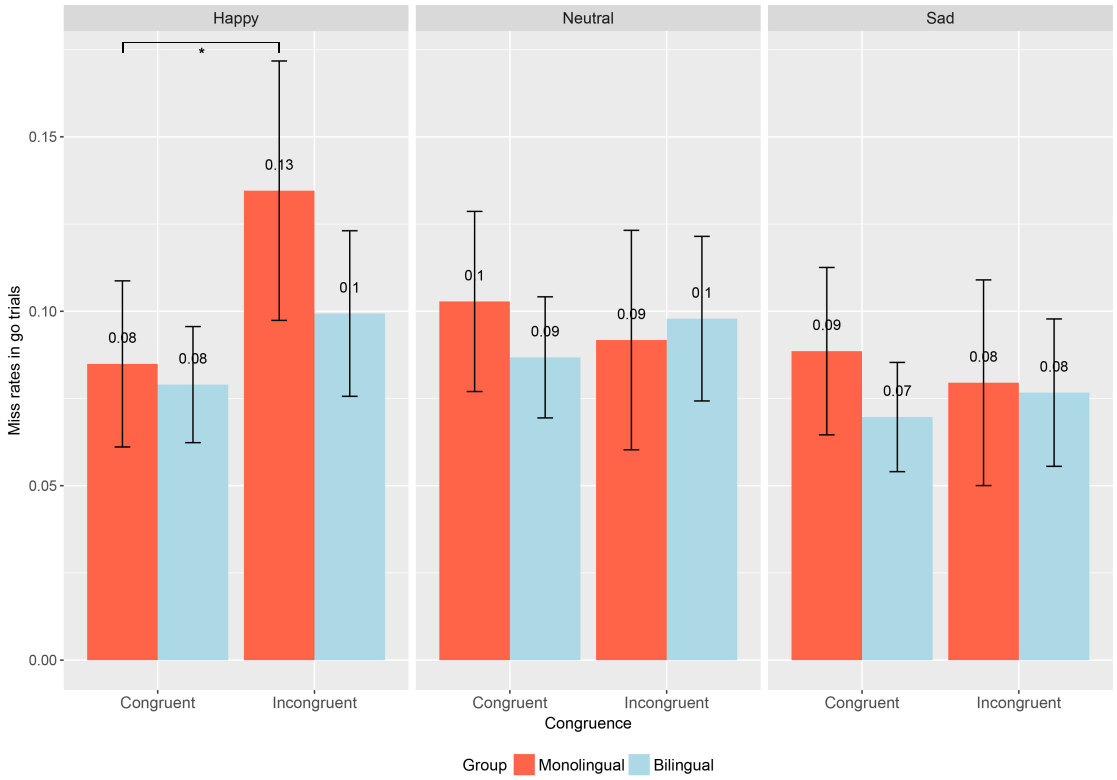

Figure 5.7: Miss rates in go trials for Experiment 1. Error bars represent 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.6: Type–III ANOVA table for binary response accuracy (correct/incorrect) in Experiment 2.

Term	FStat	DF1	DF2	pValue
(Intercept)	245.55	1	10566	0.000
Group	1.11	1	10566	0.293
FValence	0.97	2	10566	0.379
Congruence	4.65	1	10566	0.031
Gender	19.21	1	10566	1E-05
Group:FValence	0.58	2	10566	0.560
Group:Congruence	0.20	1	10566	0.656
FValence:Congruence	2.14	2	10566	0.118
Group:FValence:Congruence	2.37	2	10566	0.094

compared to adults (e.g. Hwang et al., 2014).

In addition to the larger flanker effect in RTs to happy trials, the monolingual

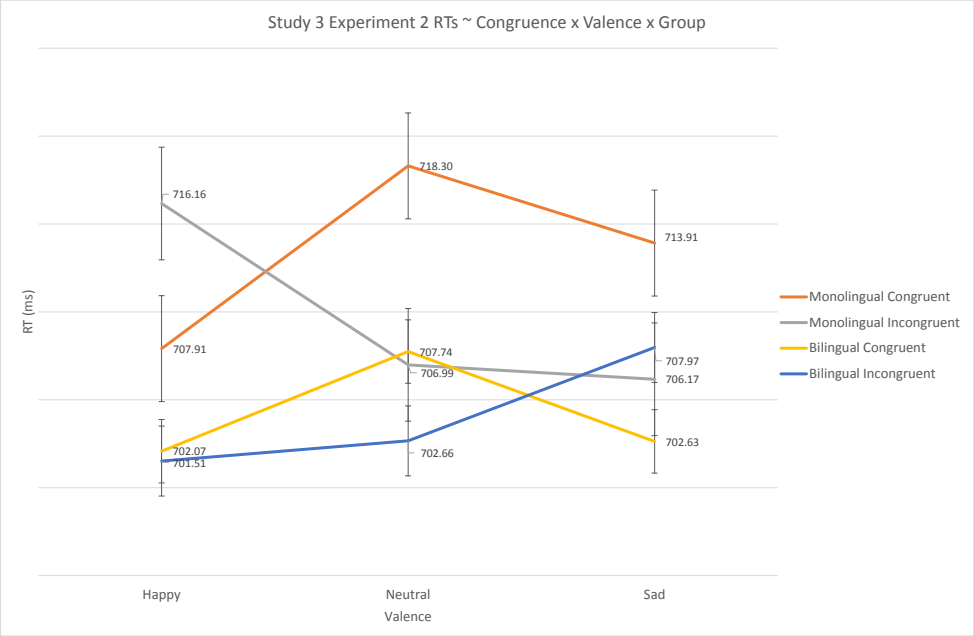

Figure 5.8: Visualisation of the relationships between Group × Congruence × Valence in Experiment 2, with experimental mean RTs in milliseconds. Error bars represent standard error of the mean.

children also made more omission errors (misses) when flankers were positive and gender-incongruent. There was no concurrent increase in false alarms; in fact, there was no effect of congruence in either happy or sad no-go trials for the monolinguals. Positive valence thus appears to increase response conflict in incongruent trials. This supports our interpretation of the conflict adaptation effects reported in Study 2.2, which were only significant following happy trials. It also implies that positive valence is detrimental to monolinguals' interference suppression in this task.

Negative valence, meanwhile, did not have the hypothesised detrimental impact on interference suppression for either group, although it did amplify the effect of congruence on false alarm rates for the bilingual participants. This may once again reflect competing effects of interference suppression and response inhibition that can be resolved through a comparison to the FO data.

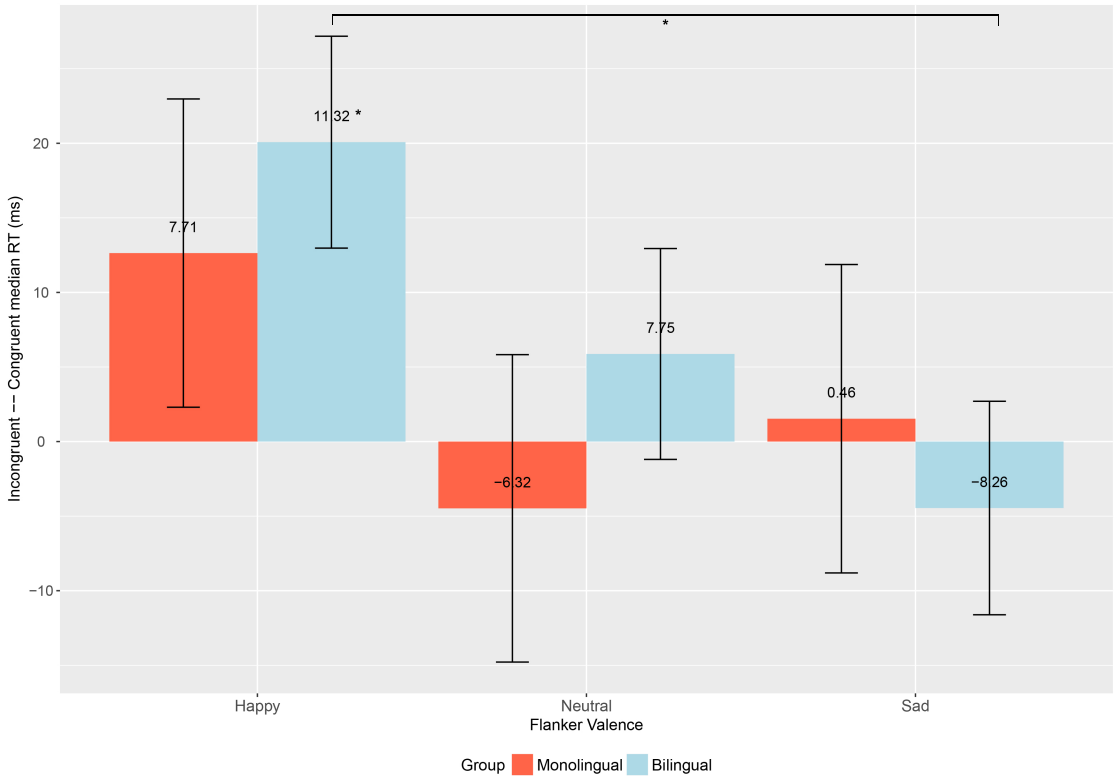

Figure 5.9: Mean reaction time differences in milliseconds between incongruent and congruent trials from Experiment 2. Error bars represent standard errors of the mean differences. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.5.2 Experiment 2 Results

The absence of significant differences between monolingual and bilingual children in the FO task in all but accuracy to congruent neutral trials refutes our Hypothesis 2a, but further supports our interpretation of the Experiment 1 data as a bilingual advantage in response inhibition. The larger flanker effect in happy trials, meanwhile, provides additional evidence that positive valence was more detrimental to interference suppression, which contradicts findings from previous emotional flanker tasks in this age group when emotional facial expression was the response dimension (Grose-Fifer et al., 2013). It is also contrary to our Study 2.2 findings, where only negative flankers increased interference. This may reflect the fact that while children's behavioural performance on flanker tasks is adult-like by age seven (Rueda et al., 2004), they rely on more distributed executive control networks than do adults (Brydges et al., 2013), and may thus be differentially affected by emotion. Although the direction of the positive valence effect is contrary

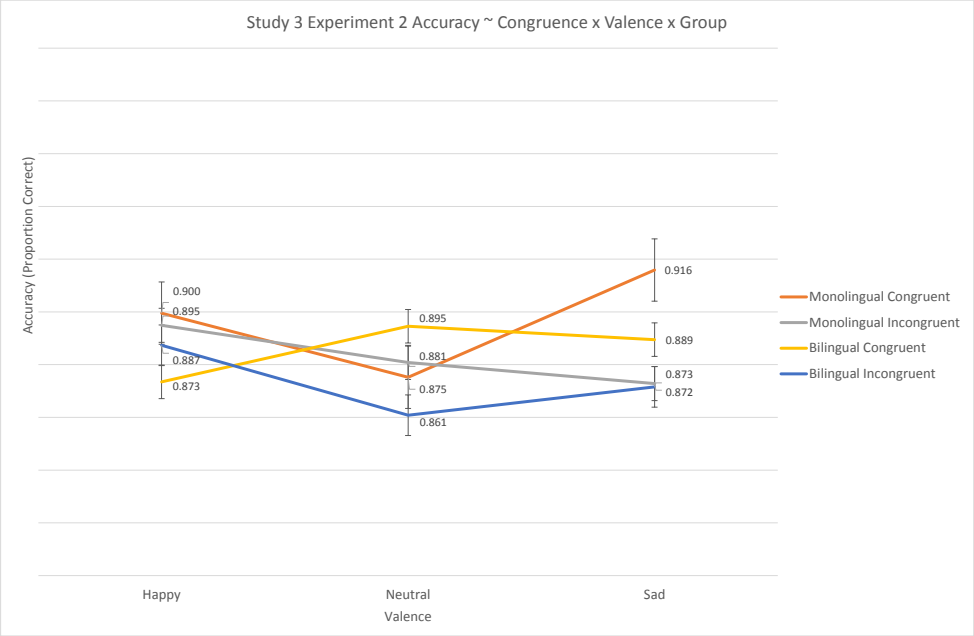

Figure 5.10: Visualisation of the relationships between Group × Congruence × Valence in Experiment 2, with experimental mean proportion correct. Error bars represent standard error of the mean.

to our Hypothesis 2b, this finding may also potentially reflect response tendencies influenced by emotion which our study design could not detect.

As the bilingual children's RTs in sad incongruent trials were faster than in sad congruent trials, another interpretation is that this group read facial expression as a task-relevant conflict signal. Negative incongruent flankers thus aided response selection, while positive incongruent flankers privileged the incorrect response tendency. The fact that negative congruent flankers increased monolingual children's response accuracy (with no equivalent effects on RT) points once again to differences in response strategies between the two groups. Possibly, the monolinguals found the flankers overall more distracting than their bilingual counterparts and thus processed both happy and sad valence as additional conflict. The reverse flanker effect in neutral trials may also be explained by this effect. We did not have sufficient statistical power to conduct a conflict adaptation analysis on the children's RTs. However, our conflict adaptation results from Study 2.2 support this interpretation.

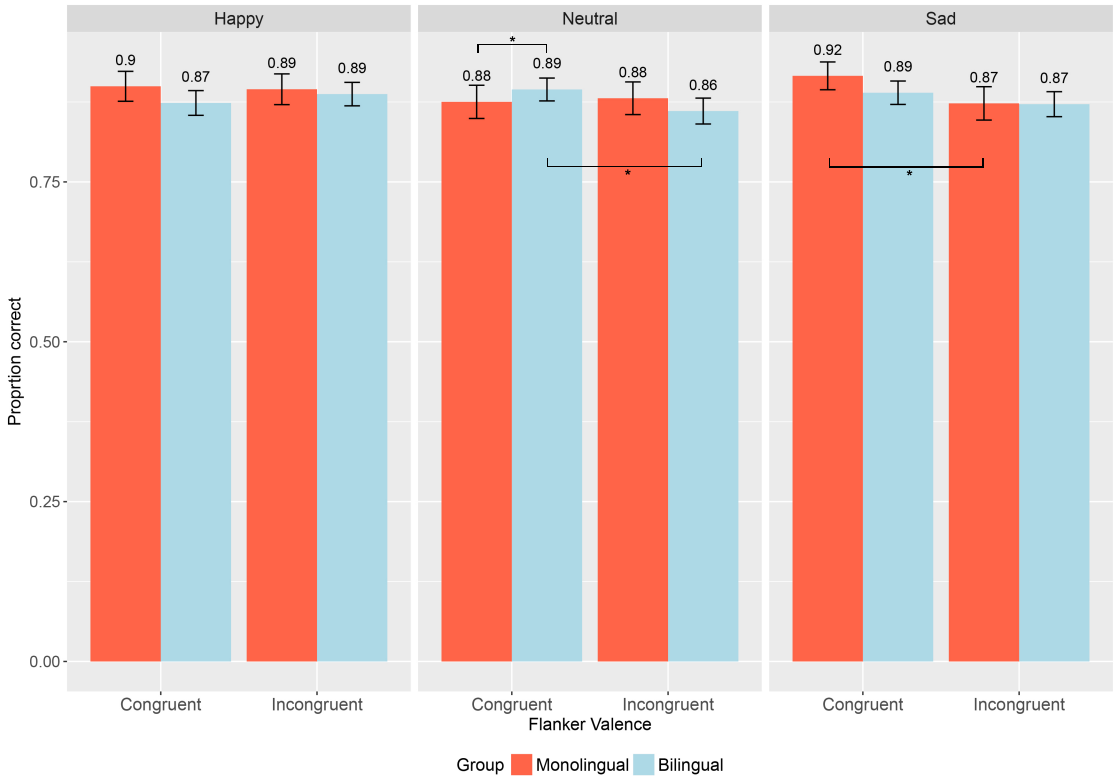

Figure 5.11: Accuracy rates in Experiment 2. Error bars represent standard errors of the mean differences. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The monolingual adults' RTs were most facilitated by successive congruent neutral trials, implying that these trials generated less conflict than congruent happy or sad.

5.6 Conclusion

The present study produced a novel finding of a bilingual advantage in response inhibition in 10–11 year old children during an emotional flanker–go/no-go task. Our finding that bilingualism confers an advantage in children's ability to inhibit both conflicting response tendencies and prepotent responses in the presence of emotional distractions has implications for educational research on executive control. Due to the host of benefits associated with improved executive control ability, researchers have begun exploring the use of 'cognitive training' to improve children's social and educational outcomes (e.g. Rueda et al., 2012). Language is an activity that spans various situations and cognitive demands in a much more ubiquitous way than designated training activities or, e.g., playing a musical instrument (Bialystok & DePape, 2009). The utility of second language acquisition as a

training tool thus raises exciting implications for executive control and emotion regulation development in children.

Study 3 replicated the RT results from Study 1 in a younger sample of bilinguals who were more diverse than the German adults in Study 1, but more homogeneous than the mixed group of adults in Study 2. Thus, we argue that the large number of confounds in Study 2 may have obscured a possible bilingual advantage. Paap and Sawi's (2014) highly cited failed replication study used a sample similar in composition to our Study 2 sample, although much larger. The fact that we were able to replicate our first study with a more controlled sample even though participants were significantly younger emphasises the importance of sample selection in bilingual advantages research. The findings also highlight Bialystok's (2017) argument that bilingual advantages are often only observed in populations who are not at their peak cognitive capacity (e.g. children, elderly adults, those with disabilities).

However, whether our findings from the present study represent a genuine bilingual advantage in response inhibition, or whether the selectivity of the sample and the complexity of the task introduced uncontrolled confounds, remains to be confirmed by future research.

...We are all talkers
It is true, but underneath the talk lies
The moving and not wanting to be moved, the loose
Meaning, untidy and simple like a threshing floor.

— John Ashbery,
'Soonest Mended', pp. 87.

6
General Discussion

Contents

6.1 Summary and Comparison of Study Results 138
6.2 Limitations 141
6.3 Future Directions 143
6.3.1 Bilingualism Research 143
6.3.2 Cognitive Control Research 144
6.3.3 Emotion and Emotion Regulation Research 145
6.4 Concluding Remarks 145

This dissertation integrates theoretical frameworks from bilingualism, cognition, and emotion research in order to establish a relationship between multiple-language use, executive control, and emotion regulation.

There is ample rationale to investigate whether the well-established links between cognitive control and emotion are impacted by bilingual experience. Linguistic interactions are rarely affectively neutral and require the coordination of complex neural mechanisms that underlie everything from representation of word meanings to grammar rules to social cognition. These demands are only amplified further by the addition of multiple languages and their associated cultural nuances and contexts of use. It is therefore hardly surprising that research across numerous academic disciplines has linked bilingualism to changes in everything from subjective emotional impact to moral reasoning to attention and executive control.

In this dissertation we investigate how bilingual experience impacts the processing of emotional stimuli on cognitive control processes whose relationships to emotion regulation are well established, namely, selective attention, interference suppression, and response inhibition. Drawing on paradigms that are widely used in cognitive affective neuroscience, we demonstrated that bilinguals adopt different response strategies from their monolingual counterparts in the presence of task-irrelevant emotional distractors. In 10–11 year-old children, this gave rise to a bilingual advantage in response inhibition.

6.1 Summary and Comparison of Study Results

We found differences in behavioural measures between monolingual and bilingual participants in a combined flanker–go/no-go (FNG) task with emotional flanker faces and neutral target faces across three different studies with different samples. The monolingual samples were all native speakers of English, undergraduate or graduate students in Studies 1 and 2 and 10–11 year old secondary school students in Study 3, but were otherwise culturally and ethnically diverse. The Study 1 bilingual sample was a culturally homogeneous group of native German speakers who acquired English in school and spoke no additional languages fluently. To contrast this group, the Study 2 sample was both linguistically and culturally diverse (multiple L1s and AoAs). Finally, the bilingual child sample in Study 3 was in between the two adult samples in terms of linguistic, cultural, and ethnic diversity. All of the bilingual children were educated in two languages. However, some were native speakers of one or the other, while others grew up with both, and yet others spoke additional languages natively.

The Study 1 bilingual sample showed no effects of emotional valence in the FNG task, while the monolingual controls showed an increased flanker interference effect when flankers were positively valenced. The result was not replicated in the Study 2 sample. Instead, this second bilingual group showed a significant interference effect for negative flankers in the FNG task (Study 2.1) which was significantly larger than the monolingual controls. However, the bilinguals in Study 2.1 demonstrated an advantage over monolinguals in overall accuracy, particularly in congruent sad and incongruent happy trials, while accuracy in Study 1 was at ceiling for both

groups. In Study 3, we replicated the RT results from Study 1 in bilingual 10–11 year-old children and matched monolingual controls. The bilingual children also had a significant advantage in response inhibition, demonstrated by much lower false alarm rates than their monolingual peers across all conditions.

Previous studies of response inhibition have failed to find an advantage for bilinguals, neither in go/no-go (Bialystok & Viswanathan, 2009) nor in stop-signal tasks (Sorge et al., 2017). It is therefore likely that our results were due specifically to the emotional component of the task. This interpretation is supported by evidence of diverging response strategies, even in the absence of differences in RT.

The fact that we replicated the Study 1 results with the child samples in Study 3, but not in the more diverse adult sample in Study 2, suggests two possible causes. First, the Study 2 sample may simply have been too heterogeneous and introduced too many individual differences to detect an effect of bilingualism. This calls into question certain high-profile failed replications that claim to demonstrate fundamental flaws in bilingualism research, but that base this argument on a similarly mixed bilingual sample (Paap & Greenberg, 2013). Second, the failure to replicate in Study 2 may have been due to a methodological inconsistency rather than deficient task validity. We made a minor adjustment to the trial sequence in Study 2, where feedback was presented immediately after participants responded in a trial, in order to reduce repeated keypresses. This reduced the response-to-stimulus interval (RSI). Faster RSIs have been demonstrated to increase conflict adaptation effects, especially in monolinguals (Grundy et al., 2017). In Study 3, feedback was again presented at the maximum trial time regardless of when the children responded. This large discrepancy in findings from a comparatively minor methodological adjustment underscores the importance of meticulous task design based on a combination of theoretical frameworks that account for the overlapping executive control demands imposed by each paradigm.

In Studies 2 and 3 we also ran a flanker-only (FO) variant of the same task. Removing the response inhibition component of the task meant that the participants had to choose between competing responses on each trial, rather than between approaching (go) or avoiding (no-go). However, the selective attention and interference suppression demands were identical. There were no differences between

monolinguals and bilinguals under these conditions. Thus, the group differences in the FNG task can be attributed to response inhibition.

Both groups of adults in Study 2.2 had larger flanker effects in sad conditions compared to all other conditions on this task. The accuracy data indicated that bilinguals were more accurate than monolinguals in the congruent happy and incongruent sad conditions. This further supports the interpretation that the RT results from the combined flanker–go/no-go task are better explained by response inhibition demands: the happy flankers biased approach tendencies, and thus there was no conflict in positive go trials. In contrast, the monolinguals had to override the avoidance tendency induced by negative flankers, producing conflict even in congruent go conditions and thus reducing the facilitation effect of gender-congruence.

This is further supported by the conflict adaptation effects we observed in Study 2.2. Successive incongruent happy trials produced faster RTs than any other sequence of flanker congruence and valence for both groups. The groups differed in their response patterns in consecutive neutral trials. Monolinguals showed facilitation from preceding trial congruence (faster RTs in congruent neutral following other congruent neutral trials), while bilinguals showed interference in the same condition (leading to markedly slower response latencies). This indicated that there was a difference between groups in overall response strategies, but that these did not translate to differences in RTs between the groups during interference suppression.

As addressed in the individual chapters, the complexity of the task and its methodological limitations make it difficult to draw conclusions about the effect of bilingualism on cognitive control involving emotion per se. (See Section 6.2 below.) The individual points of overlap between emotion, cognitive control, and language control presented in Figure 2.6 are impossible to separate in the present results. However, as discussed at the beginning of this Dissertation, the aim our studies was to establish an empirical basis for testing the predictions made by the combined literature. Though highly variable in their outcomes, all three studies presented here demonstrate that manipulating the emotional valence of stimuli in a cognitive task as well as varying the language background of the participant sample produce differences in two measures of executive function.

6.2 Limitations

Although the intention behind selecting a combined FNG task and manipulating both emotional valence and gender congruence was to control for as many factors as possible, the resulting complexity of our experimental tasks limits the conclusions we can draw from the results, as it is unclear which factor(s) played a role in the observed results. For example, as noted several times, the response inhibition element of the FNG confounds the interference suppression results, as it is impossible to separate response latencies induced by flanker/target disambiguation from those induced by evaluating a response tendency. Likewise, conflict induced by gender mismatch relative to that induced by flanker valence cannot be separated in the flanker effect measures when both elements vary within one task. A more informative approach to testing predictions generated by combining cognition and emotion literature would be to build up from very basic tasks and validate each measure on the target population. While we validated the gender-identification task and subjective face valence for both target groups in Study 1, we did not test the robustness of the flanker conflict effect induced by gender conflict in these populations. Interpretation of the FNG task results would have benefited from prior iterations of a flanker-only and go/no-go only task with the same stimuli, optimally varying only gender or only valence.

Our findings are also qualified by participant variance and minor methodological inconsistencies. While we made every effort to match our samples as closely as possible, there were still age and education-level differences amongst the groups. For the adult studies, monolinguals were predominantly undergraduate students aged 18-22, while bilinguals were overwhelmingly graduate students aged 22-30. Most of the monolinguals were students or alumni of Oxford or Cambridge Universities, while the bilinguals came from a broader range of backgrounds including exchange students, visiting researchers, and non-academic staff.

The monolingual sample was also more ethnically and culturally diverse, with participants from the UK, United States, Canada, Australia, India, and New Zealand, and from diverse ethnic backgrounds. The Study 1 bilingual sample was exclusively German or Austrian and predominantly Caucasian. Study 2 was more diverse culturally, but was still majority-Caucasian and most participants were from European countries. This may have impacted the groups' respective responses to

some of the picture stimuli, as these included faces of Caucasian, Afro-Caribbean, and East Asian origin. We included random effects by item with random slopes for groups in our model selection procedures to account for this confound.

In the adult studies we conducted vocabulary yes/no tests to establish baseline language proficiency scores, although these had no significant effects in the statistical analyses and were thus excluded from the final models. In Study 3 these vocabulary tests were intended to be administered in-class as paper questionnaires due to time constraints. However, several of the teachers did not hand out or did not return the tests, so that vocabulary scores were only available for half of the child participants. Their possible contribution to children's performance on our tasks could thus not be assessed. We could also not collect more detailed data about language proficiency, dominance, literacy, etc. from any of the groups within the time constraints of the experiments, and could thus not fully control for variability in these dimensions. However, the random effects terms for the language background information that the participants did provide rarely improved the fit of statistical models, suggesting that there was not a significant amount of uncontrolled variance captured by these factors.

Although psychiatric illness and learning disabilities were named as an exclusion criterion in all participant recruitment materials, due to ethics restrictions no personal medical information could be collected from participants. We could thus not verify that they did not have conditions such as depression, anxiety, or ADHD that could impact their task performance, particularly their responses to emotional stimuli. We also did not collect information about the children's academic achievement or other individual difference factors known to influence cognitive control. Furthermore, we did not control for participants' intelligence and did not include a measure of the frequency with which the bilinguals switched between their languages. Both of these factors have been shown to influence cognitive control measures. Our inclusion of by-subject random effects terms in all statistical models was intended to mitigate the influence of these uncontrolled factors.

Finally, the change in experimental trial sequence between Studies 1 and 2 limits the comparability of the findings. As discussed in Section 4.5, the adults in Study 2 saw feedback immediately after making a response, which reduced the response-to-stimulus interval relative to Study 1, where the same feedback was always presented after trial timeout. This may have influenced participants'

response strategies, an interpretation supported by the fact that Study 3, where the trial sequence was the same as Study 1, replicated our FNG task findings. Future studies should thus ensure that methodological details such as trial presentation intervals are carefully controlled and matched across experiments.

6.3 Future Directions

6.3.1 Bilingualism Research

Our findings have implications for a number of different research areas. First, we have highlighted the methodological and theoretical inconsistencies in bilingualism research that give rise to often contradictory findings. There are several lessons to be learned:

- Studies should report more detailed language background information than is currently conventional, including any L3s, proficiency, frequency of use, age of acquisition, and mode of acquisition.
- The bilingualism literature would benefit from a meta-analysis based on these background factors to identify specific patterns.
- Replications across different groups and related tasks can provide more detailed insights into the executive control mechanisms underlying findings.
- Inter-disciplinary theoretical frameworks are needed to inform both study design and interpretation of results.

Of particular note in our studies was the potential confound of bi- vs. trilingualism. As discussed in Study 2, our descriptive statistics indicated that multilingual participants who had learned two of their languages from birth and achieved fluency in at lest one additional language has faster RTs overall than any other groups, but we did not have sufficient sample sizes to investigate this effect. A replication study comparing bilinguals and trilinguals and the effects of simultaneous or sequential second language acquisition would provide further insights into this intriguing pattern.

6.3.2 Cognitive Control Research

The pervasiveness of language use in all other aspects of cognition, from social interaction through to meta-cognition, makes it an inseparable element of cognitive control research. However, for an experience as widespread as multilingualism, the language background of research participants receives very little mention. This is despite the unique insights multilingual participants can offer as to the interplay of language and cultural background with cognitive skills (Fricke et al., 2018).

Beyond the simplifications and validations suggested in Section 6.2 above, the paradigms introduced in this dissertation can be further modified and expanded to further explore how bilingual populations may differ from monolinguals when emotional processing is factored into cognitive control. We focused specifically on participants' responses to task-irrelevant emotional stimulus dimensions. Future studies could manipulate task-relevance to determine whether bilinguals differ from monolinguals when emotion is relevant to the task. This may provide additional insights into the way emotions influence both groups' response strategies. Manipulating the arousal level of the stimuli may also provide further insights, as high-arousal stimuli are perceived as more salient and tend to be more disruptive to cognitive performance (Pessoa, 2009).

Our finding of a bilingual advantage within response inhibition also requires further investigation using a go/no-go–only task to determine whether it arises only in interplay with interference suppression, as in the FNG task we used, or whether it generalises to other situations where emotions are involved but the complexity of the task is lower.

Finally, we have demonstrated a more pronounced bilingual advantage in children whose cognitive control mechanisms are less mature relative to adults, but this result also has implications for other populations whose cognitive ability is not at peak levels. For example, bilingualism has been associated with cognitive benefits in elderly populations (e.g. Mukadam et al., 2017) that may also extend to emotion regulation.

6.3.3 Emotion and Emotion Regulation Research

The popularity of 'brain training' and related efforts to develop interventions that improve children's executive control and emotion regulation skills highlight the significance of activities that have a positive impact on these outcomes. Bilingualism has already been shown to adapt the way in which everything from subjective emotionality to moral dilemmas are processed, but our findings provide a direct link between bilingualism and the cognitive skills underlying these outcomes. Response inhibition of emotional responses is associated with better impulse control, self-regulation, and reduced risk-taking behaviour in adolescence (Casey et al., 2008; Somerville et al., 2011). Interference suppression predicts positive academic and social outcomes (Checa & Rueda, 2011). Bilingual children from low-income backgrounds perform better on executive control tasks (White & Greenfield, 2017) and show enhanced brain development (Brito & Noble, 2018) relative to their monolingual peers, indicating the bilingualism is a protective factor against adverse outcomes in at-risk populations.

A more applied route for further research based on our findings is to compare monolinguals' and bilinguals' performance on classic emotion regulation paradigms such as reappraisal, which involves changing an emotional response to a stimulus by paying attention to different features.

A better understanding of the positive impact of bilingualism on both executive control and emotion regulation can help prioritise legislation that favours multilingual education and promotes other factors that predict language proficiency in language minority groups, such as immigrant and refugee children (Beirens et al., 2007; Poarch & Bialystok, 2017).

6.4 Concluding Remarks

This dissertation set out to address a gap in the research on bilingualism from both a theoretical and empirical perspective. By adjoining three disciplines—the emotional, the cognitive, and the linguistic—the project was able to offer new insights into the mechanisms underlying human thought and behaviour. The project overcame the limitations of previous studies by adopting an integrative theoretical framework that encompasses the complexity of cognition and emotion.

Based on this framework, complex tasks were devised than enabled us to tap into the different mechanisms of cognitive behaviour and isolate different facets of executive control. The project also offered comparisons between monolinguals and bilinguals, between different groups of bilinguals, and between adults and children, discussing implications of both task type and linguistic experience on task performance. Just as bilinguals and multilinguals experience and reconcile different emotions, distracting stimuli, and ultimately perform at a higher rate compared to their monolingual counterparts, this project reconciles three distinct yet overlapping disciplines which govern behaviour. Bilinguals themselves may not be fully aware of the underlying mechanisms that influence their behaviour. However, it is imperative that researchers be conscious of how, and attempt to explain why, bilinguals process information differently, and advocate second language learning whose benefits transcend several levels of the human experience.

Appendices

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

— Arthur Conan Doyle,
'A Scandal in Bohemia', pp. 63.

A

Data Analysis Details

Contents

A.1 Model Selection and Hypothesis Testing 149
A.1.1 Model Selection Process 150
A.1.2 Data Distribution and Conversions 153
A.1.3 Data Analysis Scripts 153
A.1.4 Model Equations 157
A.1.5 Planned Contrasts and *Post-Hoc* Tests 163
A.2 Sample Size and Power Calculations 167
A.2.1 *A-Priori* Sample Size Calculation 167
A.2.2 *Post-hoc* Power Calculation 168
A.3 Full Model Estimates 173
A.3.1 Study 1 173
A.3.2 Study 2 175
A.3.3 Study 3 178

This appendix provides details for the data analysis procedures used in this dissertation, including power calculations, model selection and fit statistics, and post-hoc testing.

A.1 Model Selection and Hypothesis Testing

Linear mixed-effects (LME) models and their variant for non-parametric data, generalised linear mixed-effects (GLME) models, were used to fit collected data

and perform hypothesis tests throughout this dissertation. The procedures outlined in this appendix were used across all three studies.

A.1.1 Model Selection Process

There is considerable disagreement within the academic literature regarding best practices for fitting (G)LMEs, especially in terms of how to specify the random effects structure. Random intercepts for participants and items are generally accepted to improve the fit of the model as they account for repeated measures (Baayen et al., 2008), but additional random terms, such as variable slopes for fixed effects and covariance with the random intercepts for each participant and item, may also be justified by the experimental design. Barr et al. (2013) recommend using the maximal random structure that fits the design—in other words, including random slopes for each repeated-measures independent variable. However, with complicated designs this often leads to model over-specification and convergence problems. Model complexity also significantly reduces statistical power (Matuschek et al., 2017).

Examination of the random parameter estimates and confidence intervals in the maximal models across Studies 1–3 indicated that many of the random terms were not significant. The second-order derivatives in some cases could not be computed due to insufficient numbers of observations, leading to a degenerate Hessian matrix, indicating that the model was over-specified. Following the recommendations of Matuschek et al. (2017), Yu (2015), and Burnham and Anderson (2003), an information-theoretic forward model selection approach based on Akaike's Information Criterion (AIC) was used to determine the most parsimonious random effects structure.

While Matuschek et al. (2017) recommend a backward-selection approach—beginning with the maximal random structure and iteratively removing terms until the best model is found—Burnham and Anderson (2003) discuss the pitfalls of specifying models for inference without *a priori* assumptions about the data structure: this can lead to over-fitting a model and compromises its precision. Therefore, rather than using backward selection from a maximal model with no theoretical assumptions about the data, random effects were chosen based on predictions about individual and item differences. For example, since there was significant variability in participants' emotional valence ratings (see Section 3.4.1),

random slopes for valence were added for each participant and item. An example of these model selection steps and the resulting AIC values from Study 1 can be seen in the scripts included in Section A.1.3 below.

The full hypothesised fixed effects structure (a three-way interaction between the independent variables of Group, Congruence, and flanker Valence) was included throughout all steps in this process. The model with the lowest AIC following random effects specification was selected for subsequent hypothesis testing. Finally, Gender and a Gender × Valence interaction were added as covariates to control for possible gender differences in the cognitive control of emotion (McRae et al., 2008), and Age and an Age × Congruence term were also added to the models of the children's data in Study 3. The significance of these covariates was tested by re-fitting the model with the Maximum Likelihood fit method (previous steps used Restricted Maximum Likelihood, or REML) and using a simulated LRT with 100 simulations to compare the models with and without the term. A LRT tests the hypothesis that one model represents the dataset better than a reduced model nested within it. Covariate terms were removed one-by-one, starting with higher-order interactions (Gender × Valence). If these interactions improved the fit of the model, then all lower-order terms in these interactions were retained, in line with the principle of marginality. A sample script for these steps, detailing the model selection and hypothesis testing processes for Study 1, is included in Section A.1.3 below. All other studies followed the same steps.

Following the recommendations of Luke (2017), once the final model including all covariates was selected, it was refit using REML and significance of the independent variables was tested using the $fixedEffects$ function of the MATLAB $LinearMixedModel$ class, applying a Satterthwaite approximation for degrees of freedom. This method was shown to produce the lowest Type I error rate with sample sizes comparable to the ones in this study (Luke, 2017). Significance of main effects was determined using a Type–III ANOVA, also with the more conservative Satterthwaite approximation.

A.1.1.1 Categorical Encoding of Contrasts

Unlike the clear conventions for defining contrasts in ANOVA, LMEs have no designated 'best practice' within the psychology literature. The method used in this

paper was recommended by a colleague (M. VanTieghem, personal communication).[1] The model was first fit using difference coding to test for main effects. (This is analogous to the 'main effects' reported in ANOVAs; see Table A.1). Planned factor contrasts were then analysed by changing the contrast coding, adjusting the reference level, and re-fitting. Bonferroni-Holm corrections were applied where appropriate.

Table A.1: Design matrix, difference coding. Grp = Group (ML = Monolingual, BL = Bilingual), C = Congruence (c = Congruent, IC = Incongruent), FV = Flanker Valence (S = Sad, H = Happy, N = Neutral), Gd = Gender (coded as alternating male/female for illustration). The Intercept (vector of 1*s*) and Gender:FValence (product of Gender×FValH/S code) are not included in this table.

	Gen	Grp	C	FVS	FVH	Grp:C	Grp:FVS	Grp:FVH	C:FVS	C:FVH	Grp:C:FVS	Grp:C:FVH
ML C N	−1	−1	−1	−1	−1	1	1	1	1	1	−1	−1
ML C H	1	−1	−1	0	1	1	0	−1	0	−1	0	1
ML C S	−1	−1	−1	1	0	1	−1	0	−1	0	1	0
ML IC N	1	−1	1	−1	−1	−1	1	1	−1	−1	1	1
ML IC H	−1	−1	1	0	1	−1	0	−1	0	1	0	−1
ML IC S	1	−1	1	1	0	−1	−1	0	1	0	−1	0
BL C N	−1	1	−1	−1	−1	−1	−1	−1	1	1	1	1
BL C H	1	1	−1	0	1	−1	0	1	0	−1	0	−1
BL C S	−1	1	−1	1	0	−1	1	0	−1	0	−1	0
BL IC N	1	1	1	−1	−1	1	−1	−1	−1	−1	−1	−1
BL IC H	−1	1	1	0	1	1	0	1	0	1	0	1
BL IC S	1	1	1	1	0	1	1	0	1	0	1	0

Tables A.6 and **??** in the main text show the model estimates and associated statistics for the selected model. The model output listed in Table A.6 gives estimates obtained using difference coding (see design matrix in Table A.1), allowing a test of each factor's contribution to the overall variance using Type III ANOVA (Table 3.5).

Table A.2: Model design matrix with custom factor contrasts to control for gender and test the simple main effects of each predictor. Grp = Group (ML = Monolingual, BL = Bilingual), C = Congruence (c = Congruent, IC = Incongruent), FV = Flanker Valence (S = Sad, H = Happy, N = Neutral), Gd = Gender (coded as alternating male/female for illustration). The Intercept (vector of 1*s*) and Gender:FValence (product of Gender×FValH/S code) are not included in this table.

	Gen	Grp	C	FVS	FVH	Grp:C	Grp:FVS	Grp:FVH	C:FVS	C:FVH	Grp:C:FVS	Grp:C:FVH
ML C N	−0.5	0	0	0	0	0	0	0	0	0	0	0
ML C H	0.5	0	0	0	1	0	0	0	0	0	0	0
ML C S	−0.5	0	0	1	0	0	0	0	0	0	0	0
ML IC N	0.5	0	1	0	0	0	0	0	0	0	0	0
ML IC H	−0.5	0	1	0	1	0	0	0	0	1	0	0
ML IC S	0.5	0	1	1	0	0	0	0	1	0	0	0
BL C N	−0.5	1	0	0	0	0	0	0	0	0	0	0
BL C H	0.5	1	0	0	1	0	0	1	0	0	0	0
BL C S	−0.5	1	0	1	0	0	1	0	0	0	0	0
BL IC N	0.5	1	1	0	0	1	0	0	0	0	0	0
BL IC H	−0.5	1	1	0	1	1	0	1	0	1	0	1
BL IC S	0.5	1	1	1	0	1	1	0	1	0	1	0

[1] The author also sought advice from the StackExchange statistics community (https://stats.stackexchange.com/questions/296250/design-matrix-contrast-coding-for-model-selection-and-main-effects-vs-simple, although the question remains unanswered.)

A.1.2 Data Distribution and Conversions

Inference from (G)LMEs is based on the assumption that the dataset to which models are fit approximately matches a specified distribution. LMEs assume a Gaussian (or normal) distribution, while GLMEs allow for non-parametric data with a specified distribution (binomial, Poisson, gamma, or inverse Gaussian) and link function to covert the data (*e.g.* logit, probit, or complementary log-log) for linear modelling (McCullagh & Nelder, 1989). In addition, residual errors from the models should demonstrate homoscedasticity (homogeneity of variance). The RT and accuracy data collected for Studies 1–3 were checked for these assumptions, and where appropriate, a transformation was applied.

RT data was checked for normality using a Box–Cox normality plot, and the data was power-transformed as appropriate to fit an approximately normal distribution. RTs in milliseconds were raised to the power of the maximum correlation coefficient. This was $RTms^{-}0.9$ in Study 2.1 and $RTms^{-}0.5$ in Study 2.2, $RTms^{0}.4$ in Study 3.1 and $RTms^{0}.6$ in Study 3.2. Study 1 RT data was normally distributed an did not require a transformation.

As reported in the main text, the ceiling effect in the Flanker–Go/No-go task led to a violation of the model assumptions. Accuracy data for this task in Studies 1 and 2 was thus converted to loglinear *d′*, a method recommended by Stanislaw and Todorov (1999) for data with extreme values. The formula for this transformation is as follows:

$$\mathtt{d'}_{loglin.} = NORMINV\left(\frac{\mathtt{Count}_{Hits} + 0.5}{\mathtt{Number}_{GoTrials} + 1}\right) - NORMINV\left(\frac{\mathtt{Count}_{FalseAlarms} + 0.5}{\mathtt{Number}_{NogoTrials} + 1}\right)$$

Loglinear *d′*s were calculated for each participant. Because participants responded to *either* male or female faces in the go/no-go task, by-item *d′*s could not be calculated.

A.1.3 Data Analysis Scripts

Statistical analysis was conducted using both the *lme*4 and *lmerTest* packages (Bates et al., 2014) for *R* (R Core Team, 2017) and the Statistics and Machine Learning Toolbox for MATLAB (specifically the *LinearMixedModel* class). MATLAB

was used for the model selection and LRT testing steps because computation times were up to 66% faster (presumably due to the programme being optimised for matrix-based computation).[2] Planned contrasts were conducted in R using the *contrasts*() function, and post-hoc comparisons were performed using the *emmeans* package.

This section provides the scripts used to perform the analysis steps described in Section A.1 above.

A.1.3.1 Random Effects Model Selection

The following MATLAB script was used to determine the most parsimonious random effects structure for the RT moel in Study 1 based on AIC. Steps are explained in the script comments (green text) and each model's AIC is listed after the respective model formula.

```
%% Script for determining random effects structure of lme for reaction times.
randEffects = {};
%fit initial intercept-only model (allowing intercept to vary for each participant and each item), including the expected three-way fixed effects interaction:
randEffects{1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74427*
%% Because participants' emotional valence ratings showed a lot of variance, a random slope term for Valence was tested, allowing the slope of valence for each participant and each item to vary. The model comparisons also test whether valence correlates with the intercept (i.e., each participant and each item differs in the neutral as well as the happy and sad conditions.)
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1+FValence|Participant) + (1+FValence|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74438
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1+FValence|Participant) + (1|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74436
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+FValence|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74429
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (FValence-1|Participant) + (1|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74433
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1|Target) + (FValence-1|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74427
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (FValence-1|Participant) + (1|Target) + (FValence-1|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74433
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1+FValence|Participant) + (1|Target) + (FValence-1|Target)', 'FitMethod','REML', 'DummyVarCoding', 'effects');
%AIC = 74436
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (FValence-1|Participant) + (1+FValence|Target)', 'FitMethod','REML', 'DummyVarCoding', 'effects');
%AIC = 74435
[lmeBest_step1a, randEffects] = bestLme(randEffects); %The bestLme function compares the above models and returns the model(s) with the lowest AIC.
%Intercept-only model retained at AIC = 74427

%% Step 2: The effect of group was expected to vary for each target, as the stimuli represented multiracial faces and the bilingual group came from more racially homogeneous countries than the monolingual group. A group*valence interaction slope was also included as target race may have amplified the effect of valence. These effects were expected to correlate with the intercept (i.e. the groups have different baseline responses to different targets).
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects'); %**
```

[2] Accuracy of the results was subsequently verified by comparing the selected model fit in both software packages. They produced identical results. However, the computation took 10 minutes in MATLAB, compared to 30 minutes in *R*.

```
%AIC = 74425* this model had the lowest AIC and was retained for the next step of comparisons
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group:FValence|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74435
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group+Group:FValence|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74435
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+FValence+Group:FValence|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74444
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group*FValence|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74447
[lmeBest_step2a, randEffects] = bestLme(randEffects);

%% Step 3: In this step, a random slope for congruence and a congruence:valence interaction for each participant was added to accout for individual differences in conflict monitoring. The term was tested with and without intercept covariance, as it was deemed possible that participants responded similarly for congruent trials but differed in incongruent trials.
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (Congruence-1|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74426
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1+Congruence|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74428
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (Congruence:FValence-1|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74429
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1+Congruence|Participant) + (Congruence:FValence-1|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74432
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1+Congruence:FValence|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74426
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1+Congruence:FValence|Participant) + (Congruence-1|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74428
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1+Congruence+Congruence:FValence|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74431
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (Congruence+Congruence:FValence-1|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74433
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (FValence+Congruence:FValence-1|Participant) + (1+Group|Target)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74437
[lmeBest_step3a, randEffects] = bestLme(randEffects);
% None of these models had a lower AIC than 74425 from the previous step ('RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group|Target)')
%% Step 4: Add random intercept for flanker. In this step, an additional random intercept for flanker, nested within valence, was added to account for different effects of each flanker at each valence. Random slopes and intercepts were added for congruence and valence. These were expected to covary with the intercept (i.e. different baselines for each congruent neutral flanker).
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group|Target) + (1|FlankerItem:FValence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74419* this model had the lowest AIC in this step and was thus retained for all further comparisons.
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group|Target) + (1+Congruence|FlankerItem:FValence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74421
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group|Target) + (1+FValence|FlankerItem:FValence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74429
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group|Target) + (1+Congruence:FValence|FlankerItem:FValence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74427
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group|Target) + (1+Congruence+FValence|FlankerItem:FValence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74433
randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|Participant) + (1+Group|Target) + (1+Congruence*FValence|FlankerItem:FValence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%AIC = 74454
[lmeBest_step4a, randEffects] = bestLme(randEffects);
```

```
74
75 %% Step 5: Add random intercept for block
76 % In this step, a random intercept was added for Block, to account for repetition effects.
77 randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|
       Participant) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', 'FitMethod', 'REML
       ', 'DummyVarCoding', 'effects');
78 %AIC = 74391*
79 %% Step 6: Add random intercept of gender (with participants nested in gender)
80 randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|
       Participant:Gender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', 'FitMethod'
       , 'REML', 'DummyVarCoding', 'effects');
81 %AIC = 72778*** this model was retained for subsequent fixed effects and hypothesis testing.
82 randEffects{length(randEffects)+1, 1} = fitlme(hits, 'RTms ~ Group*Congruence*FValence + (1|
       Participant:Gender) + (1+Group|Target) + (1+Gender|FlankerItem:FValence) + (1|Block)', '
       FitMethod', 'REML', 'DummyVarCoding', 'effects');
83 %AIC = 72782
84 %%
85 bestLme_randEffectsICa = bestLme(randEffects);
86 save bestLme_randEffectsICa bestLme_randEffectsICa; save randEffects randEffects
```

A.1.3.2 Fixed Effects Testing

The random effects selection above was performed with the full hypothesised fixed effects structure. Additional covariate terms were then tested using a simulated LRT, which compares the goodness-of-fit of the full model to a reduced model (with one or more fixed effects terms removed). As some previous studies have reported gender differences in responses to emotional stimuli (McRae et al., 2008), a covariate was added for Gender and its interaction with Valence. If the Gender × FValence interaction was significant (as it was in Study 1, see Table A.3), the lower-order term for Gender was also retained.

Including these covariates did not significantly change the magnitude of the estimates and effects. All model estimates are shown in the tables in Section 3.4.2. The following MATLAB script shows the above steps. Again model formulas are shown in lilac and comments (noting test results and decisions) in green.

```
1 fullModel = fitlme(hits, 'RTms ~ Group*FValence*Congruence + (1|Participant) + (1+Group|Target)
       + (1|FlankerItem:FValence) + (1|Block)', 'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
2 %Remove three-way interaction
3 % reducedModel = fitlme(hits, 'RTms ~ Group*FValence*Congruence-Group:FValence:Congruence + (1|
       Participant) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', 'FitMethod', 'ML',
        'DummyVarCoding', 'Effects');
4 % %% Perform simulated LRT: The LRT tests the hypothesis that model 1 represents the data
       better than model 2
5 % LRT = compare(reducedModel, fullModel, 'NSim', 5000);
6 % export(LRT,'XLSFile','LRTtable.xlsx');
7 % %Three-way interaction is significant. Full model retained.
8
9 %% AddPartGender covariate and compare nested models, beginning with highest-order interaction,
        using LRT.
10 lmeGender = fitlme(hits, 'RTms ~PartGender + Group*FValence*Congruence + (1|Participant:
       PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', 'FitMethod', 'REML'
       , 'DummyVarCoding', 'effects');
11 lmeGender1 = fitlme(hits, 'RTms ~PartGender*FValence + Group*FValence*Congruence + (1|
       Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', '
       FitMethod', 'REML', 'DummyVarCoding', 'effects');
12 lmeGender2 = fitlme(hits, 'RTms ~PartGender*Group*FValence*Congruence + (1|Participant:
       PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', 'FitMethod', 'REML'
       , 'DummyVarCoding', 'effects');
13 genderLRT = compare(lmeGender2, lmeGender1, 'NSim', 100);
14 export(genderLRT,'XLSFile','genderLRTtable.xlsx');
15 %G1 better fit than G2. Reduce model.
16 genderLRT = compare(lmeGender1, lmeGender, 'NSim', 100);
17 export(genderLRT,'XLSFile','genderLRTtable.xlsx', 'Sheet', 2);
18 %G1 better fit than G. Retain interaction term.
19 genderLRT = compare(lmeGender1, fullModel, 'NSim', 100);
20 export(genderLRT,'XLSFile','genderLRTtable.xlsx', 'Sheet', 3);
21 %G1 better fit than null. Retain covariate with interaction.
22
```

```
%% Add face luminance covariate and compare nested models, beginning with highest-order
    interaction, using LRT.
lmeLum = fitlme(hits, 'RTms ~PartGender*FValence+Contrast+Group*FValence*Congruence + (1|
    Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', '
    FitMethod', 'REML', 'DummyVarCoding', 'effects');
lmeLum1 = fitlme(hits, 'RTms ~PartGender*FValence+Contrast*Congruence + Group*FValence*
    Congruence + (1|Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|
    Block)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
lmeLum2 = fitlme(hits, 'RTms ~PartGender*FValence+Contrast*Group*FValence*Congruence + (1|
    Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', '
    FitMethod', 'REML', 'DummyVarCoding', 'effects');
lumLRT = compare(lmeLum2, lmeLum1, 'NSim', 100);
export(lumLRT,'XLSFile','lumLRTtable.xlsx');
%L1 better fit than L2. Simplify model.
lumLRT = compare(lmeLum1, lmeLum, 'NSim', 100);
export(lumLRT,'XLSFile','LumLRTtable.xlsx', 'Sheet', 2);
%L1 better fit than L. Retain interaction. Also Contrast x FVal significant. Test
%additional interaction.
lmeLum1a = fitlme(hits, 'RTms ~PartGender*FValence+Contrast:FValence + Group*FValence*
    Congruence + (1|Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|
    Block)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
lumLRT = compare(lmeLum1a, fullModel, 'NSim', 100);
export(lumLRT,'XLSFile','LumLRTtable.xlsx', 'Sheet', 3);

%% Obtain fixed effects estimates.
lmeDiffCoding = fitlme(hits, 'RTms ~PartGender*FValence + Group*FValence*Congruence + (1|
    Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', '
    FitMethod', 'REML', 'DummyVarCoding', 'effects');
[~, ~, diffCodedFixedEffects] = fixedEffects(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %
    obtain conservative Satterthwaite approximations for DF/p value
export(diffCodedFixedEffects,'XLSFile','ModelEstimatesDiffCoding.xlsx');
modelAnova = anova(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %check significance of terms
export(modelAnova,'XLSFile','ModelEstimatesDiffCoding.xlsx', 'Sheet', 2);
%Change factor level order to define monolingual/congruent/neutral as intercept baseline
d.Congruence = reordercats(d.Congruence, {'Congruent', 'Incongruent'});
d.FValence = reordercats(d.FValence, {'Neutral','Sad','Happy'});
d.Group = reordercats(d.Group, {'Monolingual', 'Bilingual'});
hits = d(d.Response=='Hit', :);
%Refit model using dummy coding (default) for simple main effects.
lmeDummyCoding = fitlme(hits, 'RTms ~PartGender*FValence + Group*FValence*Congruence + (1|
    Participant) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)', 'FitMethod', 'REML
    ');
[~, ~, dummyCodedFixedEffects] = fixedEffects(lmeDummyCoding, 'DFMethod', 'Satterthwaite');
export(dummyCodedFixedEffects,'XLSFile','ModelEstimatesDummyCoding.xlsx');
```

Table A.3: Simulated LRT output showing that the full model with three-way interaction is a better fit to the data than a reduced model without the interaction term.

Model	DF	AIC	BIC	LogLik	LRStat	pValue	LowerCI	UpperCI
redModel	21	74572.02	74713.77	−37265.01	0	0	0	0
fullModel	23	74566.83	74722.07	−37260.41	9.196	0.009	0.007	0.012

A.1.4 Model Equations

A.1.4.1 Studies 1 and 2

The final model selected in the procedure shown above retains the three-way interaction of Group, Congruence, and Valence and includes random intercepts for Participants, Items, Flankers, and Blocks, as well as a random slope for Group within each Item. The regression equation for this final model is shown in Equation A.1 below.

$$\begin{aligned}\mathrm{RT}_{ij} = {} & \beta_0 + \beta_1 \mathrm{Group}_{i_g j} + \beta_2 \mathrm{Congruence}_{ijc} + \beta_3 \mathrm{FValSad}_{ijv} + \beta_4 \mathrm{FValHappy}_{ijv} \\ & + \beta_5(\mathrm{Group} \times \mathrm{Congruence}_{i_g jc}) + \beta_6(\mathrm{Group} \times \mathrm{FValSad}_{i_g jv}) \\ & + \beta_7(\mathrm{Group} \times \mathrm{FValHappy}_{i_g jv}) + \beta_8(\mathrm{Congruence} \times \mathrm{FValSad}_{ijcv}) \\ & + \beta_9(\mathrm{Congruence} \times \mathrm{FValHappy}_{ijcv}) + \beta_{10}(\mathrm{Group} \times \mathrm{Congruence} \times \mathrm{FValSad}_{ijcv}) \\ & + \beta_{11}(\mathrm{Group} \times \mathrm{Congruence} \times \mathrm{FValHappy}_{ijcv}) \\ & + b_{1\mathrm{i}_{Gender}} + b_2 \mathrm{Group}_j + b_{3\mathrm{Flanker}_v} + b_{4\mathrm{Block}} + \epsilon_{ij}\end{aligned}$$

Equation A.1: Regression equation for Study 1 RT LME model $\beta = fixed\ effect\ estimate$, $b = random\ effect\ estimate$, $i = participant$, $j = item$, $g = group$, $c = congruence$, $v = valence$, and $\epsilon = residual\ error$. Subscripts indicate nesting. (*E.g.* the term '$\mathrm{Var}_{i_g j}$' indicates the value of Var for the ith participant (nested within group g) on the jth item.)

The fixed effects design matrix used across all three studies is shown in Table A.1, and the corresponding estimates are shown in Tables A.6–A.9.

Following the same model selection procedure, the formula for the RT model in Study 2.1 was:

$$\begin{aligned}\mathrm{RT}_{t_{ij}} = {} & \beta_0 + \beta_1 \mathrm{Group}_{i_g j} + \beta_2 \mathrm{Congruence}_{ijc} + \beta_3 \mathrm{FValSad}_{ijv} + \beta_4 \mathrm{FValHappy}_{ijv} \\ & + \beta_5(\mathrm{Group} \times \mathrm{Congruence}_{i_g jc}) + \beta_6(\mathrm{Group} \times \mathrm{FValSad}_{i_g jv}) \\ & + \beta_7(\mathrm{Group} \times \mathrm{FValHappy}_{i_g jv}) + \beta_8(\mathrm{Congruence} \times \mathrm{FValSad}_{ijcv}) \\ & + \beta_9(\mathrm{Congruence} \times \mathrm{FValHappy}_{ijcv}) + \beta_{10}(\mathrm{Group} \times \mathrm{Congruence} \times \mathrm{FValSad}_{ijcv}) \\ & + \beta_{11}(\mathrm{Group} \times \mathrm{Congruence} \times \mathrm{FValHappy}_{ijcv}) \\ & + b_{1\mathrm{i}} + b_2 \mathrm{Group}_j + b_{3\mathrm{Flanker}_v} + b_{4\mathrm{Block}} + b_{4\mathrm{LangBGType}} + \epsilon_{ij}\end{aligned}$$

Equation A.2: Regression equation for Study 2, Experiment 1 RT LME model following model selection. Transformed $\mathrm{RT}_t = \mathrm{RT}^{-0.9}$, $\beta = fixed\ effect\ estimate$, $b = random\ effect\ estimate$, $i = participant$, $j = item$, $g = group$, $c = congruence$, $v = valence$, and $\epsilon = residual\ error$. LangBGType=number of languages spoken and AOA; *e.g.* 'Multilingual–Simultaneous' or 'Bilingual–Sequential'.

To analyse accuracy in Study 1 and in Study 2, Experiment 1, an intercept-only LME was fitted with random intercepts for participants. The model did not converge with additional random effects terms. There was also not sufficient statistical power to include gender as a covariate. Thus, the models for accuracy analysis in both Study 1 and Study 2.1 were as follows:

$$
\begin{aligned}
d'_{loglin_i} = {} & \beta_0 + \beta_1 \text{Group}_{i_g} + \beta_2 \text{Congruence}_{ic} + \beta_3 \text{FValSad}_{iv} + \beta_4 \text{FValHappy}_{iv} \\
& + \beta_5(\text{Group} \times \text{Congruence}_{i_gc}) + \beta_6(\text{Group} \times \text{FValSad}_{i_gv}) \\
& + \beta_7(\text{Group} \times \text{FValHappy}_{i_gv}) + \beta_8(\text{Congruence} \times \text{FValSad}_{icv}) \\
& + \beta_9(\text{Congruence} \times \text{FValHappy}_{icv}) + \beta_{10}(\text{Group} \times \text{Congruence} \times \text{FValSad}_{i_gcv}) \\
& + \beta_{11}(\text{Group} \times \text{Congruence} \times \text{FValHappy}_{icv}) \\
& + b_{1\text{i}_{Group}} + \epsilon_i
\end{aligned}
$$

Equation A.3: Regression equation for d' accuracy models used in Study 1 and 2.1. $\beta = fixed\ effect\ estimate$, $b = random\ effect\ estimate$, $i = participant$, $g = group$, $c = congruence$, $v = valence$, and $\epsilon = residual\ error$.

For Study 2.2, the regression equation for the model selected to run RT analysis from the Flanker-only task was:

$$
\begin{aligned}
\text{RT}_{t_{ij}} = {} & \beta_0 + \beta_1 \text{Group}_{i_gj} + \beta_2 \text{Congruence}_{ijc} + \beta_3 \text{FValSad}_{ijv} + \beta_4 \text{FValHappy}_{ijv} \\
& + \beta_5(\text{Group} \times \text{Congruence}_{i_gjc}) + \beta_6(\text{Group} \times \text{FValSad}_{i_gjv}) \\
& + \beta_7(\text{Group} \times \text{FValHappy}_{i_gjv}) + \beta_8(\text{Congruence} \times \text{FValSad}_{ijcv}) \\
& + \beta_9(\text{Congruence} \times \text{FValHappy}_{ijcv}) + \beta_{10}(\text{Group} \times \text{Congruence} \times \text{FValSad}_{ijcv}) \\
& + \beta_{11}(\text{Group} \times \text{Congruence} \times \text{FValHappy}_{ijcv}) \\
& + b_{1\text{i}} + b_2 \text{Group}_j + b_{3\text{Flanker}_v} + b_{4\text{Block}} + b_{4\text{LangBGType}} + \epsilon_{ij}
\end{aligned}
$$

Equation A.4: LME regression equation for Study 2, Experiment 2 RT data. Transformed $\text{RT}_t = \text{RT}^{-0.5}$, $\beta = fixed\ effect\ estimate$, $b = random\ effect\ estimate$, $i = participant$, $j = item$, $g = group$, $c = congruence$, $v = valence$, and $\epsilon = residual\ error$.

The GLME for accuracy in the Flanker-only task did not converge with a random intercept term for items or with any random slope terms: this was presumably due to the high accuracy, with the performances again approaching ceiling. However, the model below, including an intercept term only for participants, met the prerequisite assumptions or homoscedasticity and distribution of residuals. Additionally, the complementary log-log[3] link function provided the best fit to the data, as indexed both by lower AIC and Deviation values and by lower dispersion of residuals compared to both the logit and probit functions.

[3]comploglog: $g(p) = log(-log(1-p))$, where p is the probability of a correct response

$$\begin{aligned}\text{Answer}_i = \beta_0 &+ \beta_1 \text{Group}_{i_g} + \beta_2 \text{Congruence}_{ic} + \beta_3 \text{FValSad}_{ijv} + \beta_4 \text{FValHappy}_{iv} \\ &+ \beta_5(\text{Group} \times \text{Congruence}_{i_gc}) + \beta_6(\text{Group} \times \text{FValSad}_{i_gv}) \\ &+ \beta_7(\text{Group} \times \text{FValHappy}_{i_gv}) + \beta_8(\text{Congruence} \times \text{FValSad}_{icv}) \\ &+ \beta_9(\text{Congruence} \times \text{FValHappy}_{icv}) + \beta_{10}(\text{Group} \times \text{Congruence} \times \text{FValSad}_{icv}) \\ &+ \beta_{11}(\text{Group} \times \text{Congruence} \times \text{FValHappy}_{icv}) \\ &+ \beta_{12}(\text{Gender}_i) + b_{1\text{i}} + \epsilon_i\end{aligned}$$

Equation A.5: GLME regression equation for Experiment 2 accuracy data. $\beta = fixed\ effect\ estimate$, $b = random\ effect\ estimate$, $i = participant$, $j = item$, $g = group$, $c = congruence$, $v = valence$, and $\epsilon = residual\ error$. Subscripts indicate nesting. (*E.g.* the term 'Var_{i_gj}' indicates the value of Var for the ith participant (nested within group g) on the jth item.)

A.1.4.2 Study 3

The formulas for the RT models in Study 3 are shown in Equations A.8and A.9.

$$\begin{aligned}\text{RT}_{t_{ij}} = \beta_0 &+ \beta_1 \text{Group}_{i_gj} + \beta_2 \text{Congruence}_{ijc} + \beta_3 \text{FValSad}_{ijv} + \beta_4 \text{FValHappy}_{ijv} \\ &+ \beta_5(\text{Group} \times \text{Congruence}_{i_gjc}) + \beta_6(\text{Group} \times \text{FValSad}_{i_gjv}) \\ &+ \beta_7(\text{Group} \times \text{FValHappy}_{i_gjv}) + \beta_8(\text{Congruence} \times \text{FValSad}_{ijcv}) \\ &+ \beta_9(\text{Congruence} \times \text{FValHappy}_{ijcv}) + \beta_{10}(\text{Group} \times \text{Congruence} \times \text{FValSad}_{ijcv}) \\ &+ \beta_{11}(\text{Group} \times \text{Congruence} \times \text{FValHappy}_{ijcv}) \\ &+ b_1 \text{Congruence}_i + b_2 \text{Congruence}_j + b_{3\text{i:j}} + b_{4\text{Block}} + \epsilon_{ij}\end{aligned}$$

Equation A.6: LME regression equation for Study 3, Experiment 1 RT data. Transformed $\text{RT}_t = \text{RT}^{0.6}$, $\beta = fixed\ effect\ estimate$, $b = random\ effect\ estimate$, $i = participant$, $j = item$, $g = group$, $c = congruence$, $v = valence$, and $\epsilon = residual\ error$.

$$
\begin{aligned}
\mathrm{RT}_{t_{ij}} = {} & \beta_0 + \beta_1 \mathrm{Group}_{i_g j} + \beta_2 \mathrm{Congruence}_{ijc} + \beta_3 \mathrm{FValSad}_{ijv} + \beta_4 \mathrm{FValHappy}_{ijv} \\
& + \beta_5(\mathrm{Group} \times \mathrm{Congruence}_{i_g jc}) + \beta_6(\mathrm{Group} \times \mathrm{FValSad}_{i_g jv}) \\
& + \beta_7(\mathrm{Group} \times \mathrm{FValHappy}_{i_g jv}) + \beta_8(\mathrm{Congruence} \times \mathrm{FValSad}_{ijcv}) \\
& + \beta_9(\mathrm{Congruence} \times \mathrm{FValHappy}_{ijcv}) + \beta_{10}(\mathrm{Group} \times \mathrm{Congruence} \times \mathrm{FValSad}_{ijcv}) \\
& + \beta_{11}(\mathrm{Group} \times \mathrm{Congruence} \times \mathrm{FValHappy}_{ijcv}) \\
& + b_{1\mathrm{i}} + b_2 \mathrm{Congruence}_j + b_3 \mathrm{Congruence}_{flanker:j} + b_{4\mathrm{i:j}} + \epsilon_{ij}
\end{aligned}
$$

Equation A.7: LME regression equation for Study 3, Experiment 2 RT data. Transformed $\mathrm{RT}_t = \mathrm{RT}^{0.4}$, $\beta = fixed\ effect\ estimate$, $b = random\ effect\ estimate$, $i = participant$, $j = item$, $g = group$, $c = congruence$, $v = valence$, and $\epsilon = residual\ error$.

Accuracy was lower in Study 3, and thus GLMEs could be fit to the unaggregated accuracy data without violating assumptions for inference. Equation A.8 represents the model fit to the Experiment 1 (FNG) accuracy data, and A.9 the Experiment 2 (FO) accuracy data. For both models, the probit link function produced the best model fit as indexed by AIC, log deviance, and dispersion of residuals. The models did not converge with additional random effects terms.

$$
\begin{aligned}
\text{Answer}_i = {} & \beta_0 + \beta_1 \text{Group}_{i_g} + \beta_2 \text{Congruence}_{ic} + \beta_3 \text{FValSad}_{ijv} \\
& + \beta_4 \text{FValHappy}_{iv} + \beta_5 \text{GoStatus}_{is} + \beta_6 (\text{Group} \times \text{Congruence}_{i_gc}) \\
& + \beta_7 (\text{Group} \times \text{FValSad}_{i_gv}) + \beta_8 (\text{Group} \times \text{FValHappy}_{i_gv}) + \beta_9 (\text{Group} \times \text{GoStatus}_{i_gcs}) \\
& + \beta_10 (\text{Congruence} \times \text{FValSad}_{icv}) + \beta_11 (\text{Congruence} \times \text{FValHappy}_{icv}) \\
& + \beta_12 (\text{Congruence} \times \text{GoStatus}_{ics}) + \beta_13 (\text{GoStatus} \times \text{FValHappy}_{isv}) \\
& + \beta_14 (\text{GoStatus} \times \text{FValSad}_{isv}) + \beta_{15} (\text{Group} \times \text{Congruence} \times \text{GoStatus}_{ics}) \\
& + \beta_{16} (\text{Group} \times \text{Congruence} \times \text{FValSad}_{icv}) \\
& + \beta_{17} (\text{Group} \times \text{Congruence} \times \text{FValHappy}_{icv}) \\
& + \beta_{18} (\text{Group} \times \text{GoStatus} \times \text{FValSad}_{isv}) + \beta_{19} (\text{Group} \times \text{GoStatus} \times \text{FValHappy}_{isv}) \\
& + \beta_{20} (\text{GoStatus} \times \text{Congruence} \times \text{FValSad}_{iscv}) \\
& + \beta_{21} (\text{GoStatus} \times \text{Congruence} \times \text{FValHappy}_{iscv}) \\
& + \beta_{22} (\text{Group} \times \text{Congruence} \times \text{FValSad} \times \text{GoStatus}_{icvs}) \\
& + \beta_{23} (\text{Group} \times \text{Congruence} \times \text{FValHappy} \times \text{GoStatus}_{icvs}) \\
& + \beta_{24} (\text{Gender}_i) + b_{1\text{i}} + b_{1\text{j}} + \epsilon_i
\end{aligned}
$$

Equation A.8: GLME regression equation for Study 3.1 accuracy data. β = *fixed effect estimate*, b = *random effect estimate*, i = *participant*, j = *item*, g = *group*, c = *congruence*, v = *valence*, s = *gostatus*, and ϵ = *residual error*. Subscripts indicate nesting. (*E.g.* the term 'Var_{i_gj}' indicates the value of Var for the ith participant (nested within group g) on the jth item.)

$$
\begin{aligned}
\text{Answer}_i = {} & \beta_0 + \beta_1 \text{Group}_{i_g} + \beta_2 \text{Congruence}_{ic} + \beta_3 \text{FValSad}_{ijv} + \beta_4 \text{FValHappy}_{iv} \\
& + \beta_5 (\text{Group} \times \text{Congruence}_{i_gc}) + \beta_6 (\text{Group} \times \text{FValSad}_{i_gv}) \\
& + \beta_7 (\text{Group} \times \text{FValHappy}_{i_gv}) + \beta_8 (\text{Congruence} \times \text{FValSad}_{icv}) \\
& + \beta_9 (\text{Congruence} \times \text{FValHappy}_{icv}) + \beta_{10} (\text{Group} \times \text{Congruence} \times \text{FValSad}_{icv}) \\
& + \beta_{11} (\text{Group} \times \text{Congruence} \times \text{FValHappy}_{icv}) \\
& + \beta_{12} (\text{Gender}_i) + b_{1\text{i}} + b_{1\text{j}} + \epsilon_i
\end{aligned}
$$

Equation A.9: GLME regression equation for Study 3.2 accuracy data. β = *fixed effect estimate*, b = *random effect estimate*, i = *participant*, j = *item*, g = *group*, c = *congruence*, v = *valence*, and ϵ = *residual error*. Subscripts indicate nesting. (*E.g.* the term 'Var_{i_gj}' indicates the value of Var for the ith participant (nested within group g) on the jth item.)

A.1.5 Planned Contrasts and *Post-Hoc* Tests

Planned comparisons to test the significance of the flanker effect at each level of Valence for both groups were performed by changing the contrast coding of the Congruence predictor to $(-0.5, 0.5)$, while retaining treatment coding for Group and Valence (see custom design matrix in Table A.4). This provided estimates of the simple effect of Congruence (Incongruent–Congruent) within each level of Group and Valence. Reference levels were changed using the *relevel* function to obtain the desired contrasts. *P*-values were estimated using a Satterthwaite approximation, and a Holm-Bonferroni correction for multiple comparisons was applied. Where necessary, *post-hoc* contrasts were obtained using the *emmeans* package. The contrast coding, model fitting, and contrast comparison steps for Study 1 are shown in the script below, with each step explained in the comments. Command line outputs (including model summaries) are also included. We followed the same procedure for all studies.

Table A.4: Design matrix with custom contrasts to compare the flanker effect (Incongruent–Congruent) at each level of Group × Valence. (See script in Section A.1.5 below for details). Grp = Group (ML = Monolingual, BL = Bilingual), C = Congruence (c = Congruent, IC = Incongruent), FV = Flanker Valence (S = Sad, H = Happy, N = Neutral), Gd = Gender (coded as alternating male/female for illustration). The Intercept (vector of 1*s*) and Gender:FValence (product of Gender×FValH/S code) are not included in this table.

	Gen	Grp	C	FVS	FVH	Grp:C	Grp:FVS	Grp:FVH	C:FVS	C:FVH	Grp:C:FVS	Grp:C:FVH
ML C N	−0.5	−0.5	−0.5	0	0	0.25	0	0	0	0	0	0
ML C H	0.5	−0.5	−0.5	0	1	0.25	0	−0.5	0	−0.5	0	0.25
ML C S	−0.5	−0.5	−0.5	1	0	0.25	−0.5	0	−0.5	0	0.25	0
ML IC N	0.5	−0.5	0.5	0	0	−0.25	0	0	0	0	0	0
ML IC H	−0.5	−0.5	0.5	0	1	−0.25	0	−0.5	0	0.5	0	−0.25
ML IC S	0.5	−0.5	0.5	1	0	−0.25	−0.5	0	0.5	0	−0.25	0
BL C N	−0.5	0.5	−0.5	0	0	−0.25	0	0	0	0	0	0
BL C H	0.5	0.5	−0.5	0	1	−0.25	0	0.5	0	−0.5	0	−0.25
BL C S	−0.5	0.5	−0.5	1	0	−0.25	0.5	0	−0.5	0	−0.25	0
BL IC N	0.5	0.5	0.5	0	0	0.25	0	0	0	0	0	0
BL IC H	−0.5	0.5	0.5	0	1	0.25	0	0.5	0	0.5	0	0.25
BL IC S	0.5	0.5	0.5	1	0	0.25	0.5	0	0.5	0	0.25	0

```
#####
# This script fits the LME chosen through model comparisons in MATLAB
# and applies custom contrasts to perform post-hoc comparisons.
#####

#Load required libraries
library(MASS)
library(lme4)
library(psycholing)
library(lmerTest)

#read in experiment data from a csv spreadsheet (set na.strings to MATLAB missing value string "NaN")
data <- FNGs1 #read.csv("/~/MATLAB/FlankerGoNogo/FlankerGoNogo.csv", header=TRUE, sep = ",", na.strings="NaN")

#Convert variables to factors and label them
data$Group = relevel(data$Group, "Bilingual")

#change contrast codes for the gender and congruence factors to deviation coding.
contrasts(data$Gender)            <-  c(-0.5, 0.5) #center intercept on mean of both levels ("controlling for").
contrasts(data$Congruence)          <-  c(-1, 1)
colnames(contrasts(data$Congruence)) <- c("Incongruent-Congruent") #congruent is expected to be faster and thus coded as negative.
##This sets up contrasts to evaulate the effect of congruence ("flanker effect")
#within specific levels of group and valence (treatment-coded).

#subset correct go trials for RT analysis
hitdata <- data[data$Response=="Hit", ]

#fit model  (the control=... tells the model to run more iterations because it wouldn't converge with the default number.)
m1 <- lmer(RTms ~ Gender+Gender:FValence + Group*Congruence*FValence +
             (1|Participant) + (1+Group|Target) + (1|Flanker:FValence) + (1|Block),
           control=lmerControl(optCtrl=list(maxfun=100000)), data=hitdata)
m1sum <- lmerTest::summary(m1) #lmerTest provides a summary of the regression model with p-values using the Satterthwaite method.

###OUTPUT###
# Linear mixed model fit by REML
# t-tests use  Satterthwaite approximations to degrees of freedom ['lmerMod']
# Formula: RTms ~ Gender + Gender:FValence + Group * Congruence * FValence +      (1 | Participant) + (1 + Group | Target) + (1 | Flanker:FValence) +
#   (1 | Block)
# Data: hitdata
# Control: lmerControl(optCtrl = list(maxfun = 100000))
#
# REML criterion at convergence: 76644.2
#
# Scaled residuals:
#    Min      1Q  Median      3Q     Max
# -2.8464 -0.6491 -0.1430  0.4811  4.5795
#
# Random effects:
#  Groups            Name             Variance Std.Dev. Corr
# Flanker:FValence (Intercept)        39.09    6.252
# Participant      (Intercept)      1329.26   36.459
# Target           (Intercept)       947.52   30.782
# GroupMonolingual   55.41    7.444   -0.57
# Block            (Intercept)        43.09    6.565
# Residual                          7097.37   84.246
# Number of obs: 6528, groups:  Flanker:FValence, 108; Participant, 47; Target, 36; Block, 18
#
# Fixed effects:
#                                                                  Estimate Std. Error         df
      t value             Pr(>|t|)
# (Intercept)                                                     495.5908     9.5370    85.0000
       51.965 < 0.0000000000000002 ***
# Gender1                                                          32.7672    11.4501    50.0000
        2.862             0.006120 **
# GroupMonolingual                                                  1.0959    11.5947    52.0000
        0.095             0.925063
# CongruenceIncongruent-Congruent                                   6.5049     2.5805  6399.0000
        2.521             0.011732 *
# FValenceSad                                                       6.0012     3.9777   273.0000
        1.509             0.132527
# FValenceHappy                                                     2.6807     3.9795   274.0000
        0.674             0.501116
# Gender1:FValenceSad                                              -1.4696     5.1983  6304.0000
       -0.283             0.777409
# Gender1:FValenceHappy                                            -6.4530     5.2194  6311.0000
       -1.236             0.216374
# GroupMonolingual:CongruenceIncongruent-Congruent                 -13.4324     3.7479  6390.0000
       -3.584             0.000341 ***
# FValenceSad:GroupMonolingual                                     -2.1198     5.3929  6224.0000
       -0.393             0.694273
# FValenceHappy:GroupMonolingual                                   -0.4874     5.4072  6326.0000
       -0.090             0.928183
# FValenceSad:CongruenceIncongruent-Congruent                      -0.3618     3.6590  6409.0000
       -0.099             0.921243
# FValenceHappy:CongruenceIncongruent-Congruent                    -4.3202     3.6619  6421.0000
       -1.180             0.238137
```

```
# FValenceSad:GroupMonolingual:CongruenceIncongruent-Congruent      11.4735      5.3059  6261.0000
       2.162              0.030624 *
# FValenceHappy:GroupMonolingual:CongruenceIncongruent-Congruent    19.2359      5.3193  6390.0000
       3.616              0.000301 ***
#   ---


##Comparisons##
#Change the reference (intercept) level of flanker Valence to perform planned comparisons.
hitdata$FValence = relevel(hitdata$FValence, "Happy")
m2 <- lmer(RTms ~ Gender+Gender:FValence + Group*Congruence*FValence +
             (1|Participant) + (1+Group|Target) + (1|Flanker:FValence) + (1|Block),
           control=lmerControl(optCtrl=list(maxfun=100000)), data=hitdata)
m2sum <-lmerTest::summary(m2)


###OUTPUT###
# (Model fit/random effects are identical)
# Fixed effects:
#                                                                    Estimate Std. Error
      df t value             Pr(>|t|)
# (Intercept)                                                        498.2715      9.5395
      85.0000  52.233 < 0.0000000000000002 ***
# Gender1                                                             26.3142     11.4559
      51.0000   2.297              0.025799 *
# GroupMonolingual                                                     0.6085     11.6028
      52.0000   0.052              0.958375
# CongruenceIncongruent-Congruent                                      2.1847      2.5910
      6418.0000   0.843              0.399150
# FValenceNeutral                                                     -2.6807      3.9795
      274.0000  -0.674              0.501115
# FValenceSad                                                          3.3204      3.9795
      279.0000   0.834              0.404778
# Gender1:FValenceNeutral                                              6.4530      5.2194
      6311.0000   1.236              0.216374
# Gender1:FValenceSad                                                  4.9834      5.2123
      6308.0000   0.956              0.339066
# GroupMonolingual:CongruenceIncongruent-Congruent                     5.8035      3.7720
      6417.0000   1.539              0.123963
# FValenceNeutral:GroupMonolingual                                     0.4874      5.4072
      6326.0000   0.090              0.928183
# FValenceSad:GroupMonolingual                                        -1.6325      5.4063
      6320.0000  -0.302              0.762695
# FValenceNeutral:CongruenceIncongruent-Congruent                      4.3202      3.6619
      6421.0000   1.180              0.238137
# FValenceSad:CongruenceIncongruent-Congruent                          3.9584      3.6635
      6430.0000   1.080              0.279962
# FValenceNeutral:GroupMonolingual:CongruenceIncongruent-Congruent   -19.2359      5.3193
      6390.0000  -3.616              0.000301 ***
# FValenceSad:GroupMonolingual:CongruenceIncongruent-Congruent        -7.7624      5.3198
      6382.0000  -1.459              0.144573
# ---

hitdata$FValence = relevel(hitdata$FValence, "Sad")
m3 <- lmer(RTms ~ Gender+Gender:FValence + Group*Congruence*FValence +
             (1|Participant) + (1+Group|Target) + (1|Flanker:FValence) + (1|Block),
           control=lmerControl(optCtrl=list(maxfun=100000)), data=hitdata)
m3sum <-lmerTest::summary(m3)

# ###OUTPUT###
# Fixed effects:
# Estimate Std. Error         df t value             Pr(>|t|)
# (Intercept)                                                        501.5920      9.5372
      85.0000  52.593 < 0.0000000000000002 ***
# Gender1                                                             31.2976     11.4468
      50.0000   2.734              0.00861 **
# GroupMonolingual                                                    -1.0240     11.5946
      52.0000  -0.088              0.92996
# CongruenceIncongruent-Congruent                                      6.1431      2.5810
      6411.0000   2.380              0.01734 *
# FValenceHappy                                                       -3.3204      3.9795
      279.0000  -0.834              0.40478
# FValenceNeutral                                                     -6.0012      3.9777
      273.0000  -1.509              0.13253
# Gender1:FValenceHappy                                               -4.9834      5.2123
      6300.0000  -0.956              0.33907
# Gender1:FValenceNeutral                                              1.4696      5.1983
      6296.0000   0.283              0.77741
# GroupMonolingual:CongruenceIncongruent-Congruent                    -1.9589      3.7489
      6383.0000  -0.523              0.60132
# FValenceHappy:GroupMonolingual                                       1.6325      5.4063
      6312.0000   0.302              0.76270
# FValenceNeutral:GroupMonolingual                                     2.1198      5.3929
      6216.0000   0.393              0.69427
# FValenceHappy:CongruenceIncongruent-Congruent                       -3.9584      3.6635
      6422.0000  -1.080              0.27996
# FValenceNeutral:CongruenceIncongruent-Congruent                      0.3618      3.6590
      6401.0000   0.099              0.92124
# FValenceHappy:GroupMonolingual:CongruenceIncongruent-Congruent       7.7624      5.3198
      6375.0000   1.459              0.14457
# FValenceNeutral:GroupMonolingual:CongruenceIncongruent-Congruent   -11.4735      5.3059
      6253.0000  -2.162              0.03062 *
#   ---
```

```
# Test the contrasts:
#  1) Bilingual Neutral Congruence vs. Monolingual Happy/Sad Congruence
#  2) Bilingual vs. Monolingual Congruence at Sad/Happy/Neutral
#  3) Bilingual Congruent vs. Incongruent at Sad/Happy/Neutral

# Extract p-values and apply Holm-Bonferroni correction for multiple comparisons (Holm, 1979)
pvals = cbind("neutral"=p.adjust(m1sum$coefficients[c(13, 12, 9, 4), 5], "holm"),
              "happy"=c(0,0,p.adjust(m2sum$coefficients[c(9, 4), 5],"holm")),
              "sad"=c(0,0,p.adjust(m3sum$coefficients[c(9, 4), 5],"holm")))

# Reference Level=BL                                          neutral     happy         sad
# 1a) FValenceHappy:CongruenceIncongruent-Congruent     0.476274848 0.0000000 0.00000000
# 1b) FValenceSad:CongruenceIncongruent-Congruent       0.921243488 0.0000000 0.00000000
# 2)  GroupMonolingual:CongruenceIncongruent-Congruent  0.001363793 0.2479266 0.60131587
# 3)  CongruenceIncongruent-Congruent                   0.035197055 0.3991505 0.03467497

hitdata$Group=relevel(hitdata$Group,"Monolingual")
hitdata$FValence = relevel(hitdata$FValence, "Neutral")
m4 <- lmer(RTms ~ Gender+Gender:FValence + Group*Congruence*FValence +
             (1|Participant) + (1+Group|Target) + (1|Flanker:FValence) + (1|Block),
           control=lmerControl(optCtrl=list(maxfun=100000)), data=hitdata)
m4sum <- lmerTest::summary(m4) #lmerTest provides a summary of the regression model with p-
    values using the Satterthwaite method.

hitdata$FValence = relevel(hitdata$FValence, "Happy")
m5 <- lmer(RTms ~ Gender+Gender:FValence + Group*Congruence*FValence +
             (1|Participant) + (1+Group|Target) + (1|Flanker:FValence) + (1|Block),
           control=lmerControl(optCtrl=list(maxfun=100000)), data=hitdata)
m5sum <-lmerTest::summary(m5)

hitdata$FValence = relevel(hitdata$FValence, "Sad")
m6 <- lmer(RTms ~ Gender+Gender:FValence + Group*Congruence*FValence +
             (1|Participant) + (1+Group|Target) + (1|Flanker:FValence) + (1|Block),
           control=lmerControl(optCtrl=list(maxfun=100000)), data=hitdata)
m6sum <-lmerTest::summary(m6)

# Test the contrasts:
#  1) Monolingual Neutral Congruence vs. Monolingual Happy/Sad Congruence
#  2) Monolingual vs. Bilingual Congruence at Sad/Happy/Neutral
#  3) Monolingual Congruent vs. Incongruent at Sad/Happy/Neutral

# Extract p-values and apply Holm-Bonferroni correction for multiple comparisons (Holm, 1979)
pvals = cbind("neutral"=p.adjust(m4sum$coefficients[c(13, 12, 9, 4), 5], "holm"),
              "happy"=c(0,0,p.adjust(m5sum$coefficients[c(9, 4), 5],"holm")),
              "sad"=c(0,0,p.adjust(m6sum$coefficients[c(9, 4), 5],"holm")))

# ref.level=ML                                          neutral      happy        sad
# 1a) FValenceHappy:CongruenceIncongruent-Congruent  0.0005223183
# 1b) FValenceSad:CongruenceIncongruent-Congruent    0.0085388892
# 2) GroupBilingual:CongruenceIncongruent-Congruent  0.0010228513 0.12396334 0.6013159
# 3) CongruenceIncongruent-Congruent                 0.0115891266 0.00763616 0.2541363
```

A.2 Sample Size and Power Calculations

A.2.1 *A-Priori* Sample Size Calculation

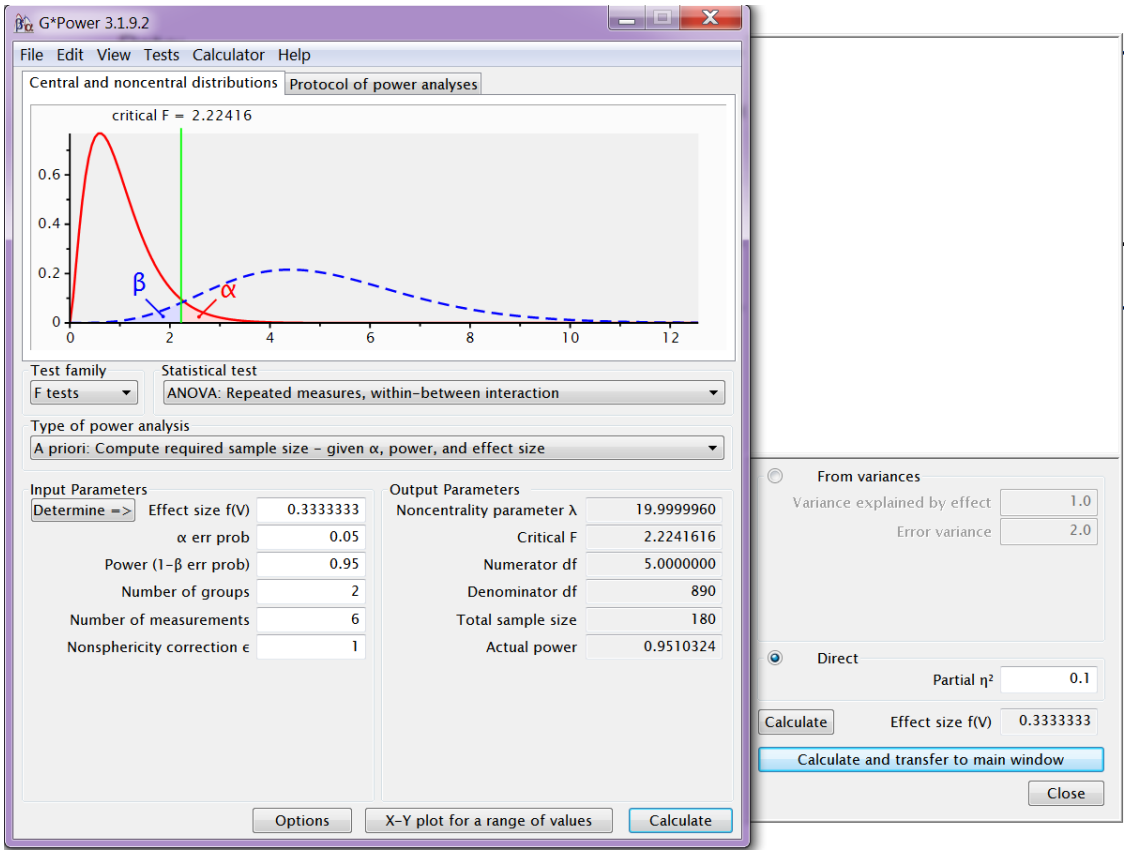

Figure A.10: An *a priori* sample size calculation for the study for a repeated-measures ANOVA with medium effect size in G*Power.

An initial target sample of 180 participants (2×90 per group) was determined using an *a priori* power calculation for repeated measures ANOVA (which is based on similar assumptions as an LME) in G*Power (Faul et al., 2007). Input parameters were a moderate effect size ($\eta_p^2 = 0.1$), similar to that reported by Grose-Fifer et al. (2013) for a three-way interaction in a similar study, and the programme's default values for correlation between repeated measures (0.5) and nonsphericity correction (1). This calculation is shown in Figure A.10. It yielded a prospective sample size of 180 participants (2×90 per language group). Due to convenience sampling, this number could not be achieved.

The minimum required sample for an LME was calculated post-hoc (see Section

A.2.2) using the online PANGEA tool (Westfall, 2015).[4] Given the number of items used (18) in the study and the random effects terms included in the model, the calculation for PANGEA showed that 17 per group was sufficient to detect a moderate effect size of $d = 0.3$ (roughly equivalent to $\eta_p^2 = 0.1$).

A.2.2 *Post-hoc* Power Calculation

An *post-hoc* calculation of the minimum required sample size and number of items to detect a significant three-way interaction at a power level above the conventional 0.80 was calculated using fake-data simulation, as recommended for multilevel models by Gelman and Hill (2007).

Previous studies using emotional face flanker and go/no-go tasks reported effect sizes of approximately 20–30 ms for the flanker effect (Grose-Fifer et al., 2013; Zhou & Liu, 2013) and RT differences between happy and negative/neutral faces in go trials (Keightley et al., 2006). Based on the hypotheses listed in Section 3.2.3, expected mean RTs were established for each condition (see Table A.5).

Table A.5: Predicted RTs for each condition based on effect sizes reported in the literature.

Group	Congruence	Valence	Expected Mean (ms)
Monolingual	Congruent	Happy	510
		Neutral	525
		Sad	510
	Incongruent	Happy	550
		Neutral	540
		Sad	555
Bilingual	Congruent	Happy	515
		Neutral	520
		Sad	525
	Incongruent	Happy	535
		Neutral	540
		Sad	545

[4]This tool was not available at the begin of the study.

The standard deviations observed in the data from this study were larger than those reported in studies using a simpler flanker or go/no-go task only (around $100ms$ compared to approximately $40ms$, respectively). Therefore, the standard deviations and variance in mean differences of the collected data were calculated and used to generate the fake data. Variance-covariance matrices were also extracted from LMEs fit from subsets of the collected data.

The following script, adapted from Kerns (2011) and Gelman and Hill (2007), was used to generate fake data from the expected means and a control condition of null-effect means (all set to 500 ± 2). The accompanying comments explain each step.

```
library(MASS)
library(lme4)
library(lmerTest)
library(car)
library(plyr)

## Load experiment data to calculate SDs and variance-covariance matrix
data            <- read.csv("/Users/anna/Dropbox/Oxford/DPhil/DataCollection/FlankerGoNoGo/ParsedData/Study2/FlankerGoNogo2.csv", header = TRUE, sep = ",", na.strings="NaN")
data$Participant<- factor(data$Participant)
data$Group      <- data$Group-1
data$Group      <- factor(data$Group, levels=c(0,1), labels=c("Monolingual", "Bilingual"))
data$FValence   <- factor(data$FValence,levels=c(0,1,-1), labels=c("Neutral", "Happy", "Sad"))
data$Congruence <- factor(data$Congruence, levels=c(1, 0), labels=c("Congruent", "Incongruent"))
data$RT.ms      <- data$RT*1000 #in milliseconds
data$Response   <- factor(data$Response, levels=c(-2, -1, 1, 2), labels = c("False Alarm", "Miss", "Hit", "Correct Reject"))
data$Response   <- relevel(data$Response, "Hit")
hitdata         <- data[data$Response=="Hit", ]
bldata          <- subset(hitdata, hitdata$Group=="Bilingual")
mldata          <- subset(hitdata, hitdata$Group=="Monolingual")

## Calculate group means, SDs, and SDs of condition mean differences
means     <- ddply(hitdata, .(Group, Congruence, FValence), summarize, mean = round(mean(RT.ms)), sd = round(sd(RT.ms)))
monoMeans <- means[1:6,4]
biMeans   <- means[7:12,4]
sds       <- (means[1:6,5] + means[7:12,5])/2
stdiffs   <- sqrt(((means[1:6,5]^2)/nPerGroup)+((means[7:12,5]^2)/nPerGroup))

## Initialise parameters for fake dataframe.
set.seed(1)
nPerGroup   <- 14
nItems      <- 18
nGroups     <- 2
congLevels  <- 2
valLevels   <- 3
nConditions <- congLevels*valLevels
muMonoExp   <- c(525,510,515,540,545,550)   #expected ML means based on previous studies
muBiExp     <- c(520,515,525,530,530,540)   #expected BL means based on previous studies
muMonoNull  <- c(500,501,500,501,500,501)   #means with null effect
muBiNull    <- c(501,500,501,500,501,500)
muMonoObs   <- monoMeans                    #observed monolingual means from experiment data
muBiObs     <- biMeans                      #observed bilingual means from experiment data
# stdevs    <- sds
stdevs      <- c(98.5, 99.5, 101, 96.5, 102.5, 94.5) #SDs of condition means from experiment data
# stdiff    <- mean(stdiffs)              #standard error of the differences between condition means,
stdiff      <- 29                         #assuming equal variances/sphericity (checked in advance!)
nSim        <- 10                         #number of model simulations to run

## Create data frames to generate fake data
SubjectML <- factor(1:nPerGroup)
GroupML <- factor(rep("Mono", each = nPerGroup))
SubjectBL <- factor((nPerGroup+1):(nPerGroup*2))
GroupBL <- factor(rep("Bi", each = nPerGroup))
Item <- factor(1:nItems)
#Conditions <- factor(1: nConditions, labels = c("CNeutral", "CHappy", "CSad", "INeutral" , "IHappy" , "ISad"))
Congruence <- factor(1: congLevels, labels = c("Congruent", "Incongruent"))
Valence <- factor(1: valLevels, labels = c("Neutral", "Happy", "Sad"))

#each subject sees each item in each condition
```

```
dataML <- expand.grid(SubjectML, Item, Congruence, Valence)
names(dataML) <- c("Subject", "Item", "Cong", "Val")
dataBL <- expand.grid(SubjectBL, Item, Congruence, Valence)
names(dataBL) <- c("Subject", "Item", "Cong", "Val")

#assumes equal groups
tmpM <- rep(c("Mono"), each = nPerGroup * nItems * nConditions)
dataML$Group <- factor(tmpM)
tmpB <- rep(c("Bi"), each = nPerGroup * nItems * nConditions)
dataBL$Group <- factor(tmpB)
dataML -> nullML
dataBL -> nullBL

## Create variance-covariance matrices
# set up generic cov. matrix based on equal stdiffs (from http://gjkerns.github.io/R/2012/01/20
    /power-sample-size.html)
ones <- rep(1, nConditions)
A <- stdevs^2 %o% ones
B <- (A + t(A) + (stdiff^2)*(diag(nConditions) - ones %o% ones))/2
C <- B

#Use variance-covariance matrices from LMEs fit to experimental data:
#Fit intercept-only models and extract variance-covariance matrices.
# bllme <- lmer(RT.ms ~ 1 + Congruence*FValence + (1|Participant) + (1|Target), data=bldata)
# mllme <- lmer(RT.ms ~ 1 + Congruence*FValence + (1|Participant) + (1|Target), data=mldata)
# blcov <- as.matrix(vcov(bllme))
# mlcov <- as.matrix(vcov(mllme))
# B     <- mlcov
# C     <- blcov

## Create normally-distributed fake RT data
#Create sample (normal) distributions from expected means (based on previous research or pilots
    )
tmp1       <- mvrnorm(nPerGroup*nItems, mu = muMonoExp, Sigma = B)
tmp2       <- mvrnorm(nPerGroup*nItems, mu = muBiExp, Sigma = C)
dataML$RT <- as.vector(t(c(tmp1[1:nrow(tmp1),1:ncol(tmp1)])))
dataBL$RT <- as.vector(t(c(tmp2[1:nrow(tmp2),1:ncol(tmp2)])))
data       <- rbind(dataML, dataBL)

#Create sample distributions based on null hypothesis (means are all equal)
tmp1       <- mvrnorm(nPerGroup*nItems, mu = muMonoNull, Sigma = B)
tmp2       <- mvrnorm(nPerGroup*nItems, mu = muBiNull, Sigma = C)
nullML$RT <- as.vector(t(c(tmp1[1:nrow(tmp1),1:ncol(tmp1)])))
nullBL$RT <- as.vector(t(c(tmp2[1:nrow(tmp2),1:ncol(tmp2)])))
nulldata  <- rbind(nullML, nullBL)

## Set up functions to simulate model fitting and hypothesis testing based on the expected and
    null hypotheses
# fit and test model based on expected means. Hypothesis test on 3-way interaction terms
# (group x congruence x valence:happy and group x congruence x valence:sad)
runTestHyp  <- function(){
  tmp1       <- mvrnorm(nPerGroup*nItems, mu = muMonoExp, Sigma = B)
  tmp2       <- mvrnorm(nPerGroup*nItems, mu = muBiExp, Sigma = C)
  dataML$RT <- as.vector(t(c(tmp1[1:nrow(tmp1),1:ncol(tmp1)])))
  dataBL$RT <- as.vector(t(c(tmp2[1:nrow(tmp2),1:ncol(tmp2)])))
  data       <- rbind(dataML, dataBL)
  lme        <- lmer(RT ~ Group*Cong*Val + (1|Subject) + (1|Item), control=lmerControl(optCtrl=
      list(maxfun=100000)), data)
  #Option 1: get t-values of interaction terms and retain the larger (absolute value) one
  b          <- base::summary(lme)
  tVal       <- max(abs(b$coefficients[nrow(b$coefficients),ncol(b$coefficients)]),abs(b$
      coefficients[nrow(b$coefficients)-1,ncol(b$coefficients)]))
  tVal > 2     #a t-value >2 is considered significant.
  #Option 2: compare p-values from lmerTest for happy and sad 3x interactions and retain
      smaller one
  # b          <- lmerTest::summary(lme)
  # intPVal    <- min(b$coefficients[nrow(b$coefficients),ncol(b$coefficients)],b$coefficients[
      nrow(b$coefficients)-1,ncol(b$coefficients)])
  #Option 3: obtain p-value for 3-way interaction using anova
  # b        <- car::Anova(lme, type=3)
  # intPVal <- b[nrow(b),3]
  # intPVal < 0.05
}

# fit and test model based on equal (null) means
runTestNull <- function(){
  tmp1       <- mvrnorm(nPerGroup*nItems, mu = muMonoNull, Sigma = B)
  tmp2       <- mvrnorm(nPerGroup*nItems, mu = muBiNull, Sigma = C)
  nullML$RT <- as.vector(t(c(tmp1[1:nrow(tmp1),1:ncol(tmp1)])))
  nullBL$RT <- as.vector(t(c(tmp2[1:nrow(tmp2),1:ncol(tmp2)])))
  nulldata  <- rbind(nullML, nullBL)
  lme        <- lmer(RT ~ Group*Cong*Val + (1|Subject) + (1|Item), control=lmerControl(optCtrl=
      list(maxfun=100000)), nulldata)
  #get t-values of interaction terms and retain the larger (absolute value) one
  b          <- base::summary(lme)
  tVal       <- max(abs(b$coefficients[nrow(b$coefficients),ncol(b$coefficients)]),abs(b$
      coefficients[nrow(b$coefficients)-1,ncol(b$coefficients)]))
  tVal > 2
  #compare p-value for happy and sad 3x interactions and retain smaller one
  # b          <- lmerTest::summary(lme)
  # intPVal    <- min(b$coefficients[nrow(b$coefficients),ncol(b$coefficients)],b$coefficients[
      nrow(b$coefficients)-1,ncol(b$coefficients)])
  #obtain p-value for 3-way interaction using anova
```

```
142   # b        <- car::Anova(lme, type=3)
143   # intPVal <- b[nrow(b),3]
144   # intPVal < 0.05
145 }
146
147 #fit and test model based on observed means (post-hoc)
148 # runTestObs  <- function(){
149 #   tmp1       <- mvrnorm(nPerGroup*nItems, mu = muMonoObs, Sigma = B)
150 #   tmp2       <- mvrnorm(nPerGroup*nItems, mu = muBiObs, Sigma = C)
151 #   dataML$RT <- as.vector(t(c(tmp1[1:nrow(tmp1),1:ncol(tmp1)])))
152 #   dataBL$RT <- as.vector(t(c(tmp2[1:nrow(tmp2),1:ncol(tmp2)])))
153 #   data       <- rbind(dataML, dataBL)
154 #   lme        <- lmer(RT ~ Group*Cong*Val + (1|Subject) + (1|Item), control=lmerControl(optCtrl
      =list(maxfun=100000)), data)
155 #   #Option 1: get t-values of interaction terms and retain the larger (absolute value) one
156 #   b          <- base::summary(lme)
157 #   tVal       <- max(abs(b$coefficients[nrow(b$coefficients),ncol(b$coefficients)]),abs(b$
      coefficients[nrow(b$coefficients)-1,ncol(b$coefficients)]))
158 #   tVal > 2
159 #   #compare p-value for happy and sad 3x interactions and retain smaller one
160 #   # b          <- lmerTest::summary(lme)
161 #   # intPVal    <- min(b$coefficients[nrow(b$coefficients),ncol(b$coefficients)],b$coefficients
      [nrow(b$coefficients)-1,ncol(b$coefficients)])
162 #   #obtain p-value for 3-way interaction using anova
163 #   # b        <- car::Anova(lme, type=3)
164 #   # intPVal <- b[nrow(b),3]
165 #   # intPVal < 0.05
166 # }
167
168 powerExp  <- mean(replicate(nSim, runTestHyp()))
169 powerNull <- mean(replicate(nSim, runTestNull()))
170 #powerObs  <- mean(replicate(nSim, runTestObs()))
```

Number of items was set to 18, as there were 18 items of each gender and only Go-responses were analysed. With participants per group set to 24 and 1000 simulations, the power in the null condition was 0.08. This meant that the t-value of the three-way Group × Congruence × Valence interaction was above the significance level of 2 8% of the time, and so the Type-I error rate of the simulation was slightly above the expected value of 0.05, and the hypothesis test thus slightly unconservative. The same parameters generated a power value of 100% in the hypothesis condition. The simulation models did not converge with smaller sample sizes, even when the random effects structure included only intercepts for Participants and Items.

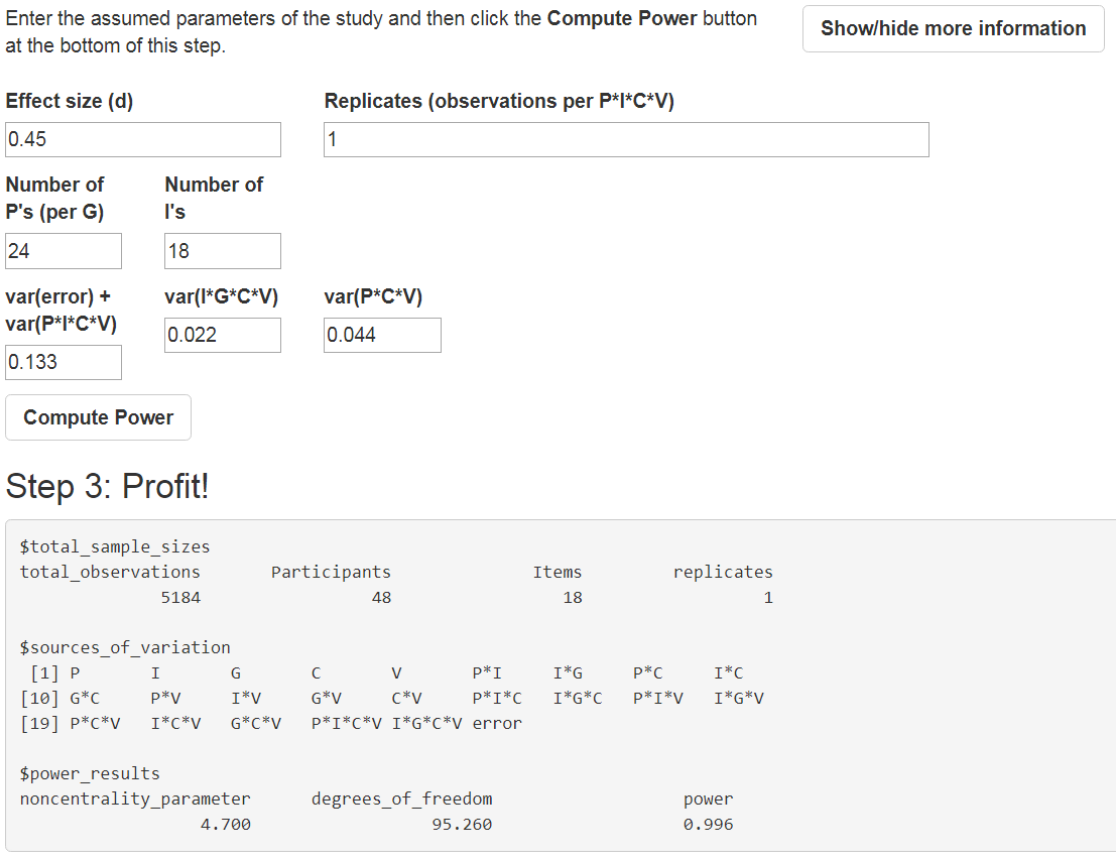

Figure A.11: Input parameters and output calculations from PANGEA.

To account for this limitation, the power and sample size calculations obtained from the simulation were double-checked against the PANGEA online power analysis tool (Westfall, 2015). A screenshot of the tool's output is included in Figure A.11 below. At an effect size of $d = 0.45$, the power of the design with 24 participants per group and 18 items was 0.996. With a reduced effect size of $d = 0.03$, the power was 0.087. A sample of 17 per group was necessary to detect an effect size of $d = 0.3$ (approximately corresponding to the effect size used in *a-priori* calculations for ANOVA) at a power above the conventional > 0.80.

As Studies 2 and 3 had the same design as Study 1, the *post-hoc* calculations from Study 1 were considered *a-priori* calculations for Studies 2 and 3. The Flanker-only task tested the same interactions but had more observations, and the power calculations for Flanker–Go/No-go were thus considered sufficient.

An additional *a-priori* calculation using PANGEA indicated that a sample size of 27 per group was needed at 204 observations per participant in the flanker-only task in order to detect a significant conflict adaptation effect of Previous

Congruence × Previous Valence × Current Congruence × Current Valence at effect size $d = 0.3$ with 0.80 power. However, this assumed 100% accuracy. Although we had 29 participants per group, our conflict adaptation analysis may thus have been slightly under-powered.

A.3 Full Model Estimates

This section lists the estimates for all models used for hypothesis testing in Studies 1–3. Degrees of Freedom are calculated by Satterthwaite's approximation unless otherwise noted (Luke, 2017). Coding schemes are noted in the captions. Where Gender is included as a covariate, it is contrast-coded using weighted Helmert coding [e.g. $(0.42, -0.58)$], thus controlling for the effect of gender and balancing the estimate if the ratios are unequal.

A.3.1 Study 1

Table A.6 shows the model statistics for the LME model fit to the Study 2 reaction time data, based on effects (or difference) coding: thus, estimates represent deviation from the grand mean.

Table A.6: Model estimates for Study 1 reaction time data, Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	t	DF	pValue	LowerCI	UpperCI
(Intercept)	498.36	7.61	65.45	82.67	0.000	483.21	513.50
Grp_BL	−1.96	5.70	−0.34	45.97	0.732	−13.43	9.51
Cong_IC	4.10	1.15	3.56	5800.28	0.000	1.84	6.36
FVal_S	2.09	1.91	1.10	95.00	0.275	−1.69	5.88
FVal_H	0.36	1.88	0.19	95.61	0.848	−3.37	4.09
Gender_M	15.67	5.62	2.79	43.63	0.008	4.33	27.00
Grp_BL:Cong_IC	0.32	1.15	0.28	6035.36	0.777	−1.92	2.57
Grp_BL:FVal_S	1.16	1.66	0.70	5884.51	0.487	−2.10	4.42
Grp_BL:FVal_H	−0.85	1.64	−0.52	5963.52	0.603	−4.07	2.37
Cong_IC:FVal_S	1.13	1.66	0.68	5872.48	0.495	−2.12	4.39
Cong_IC:FVal_H	1.67	1.64	1.02	5829.84	0.309	−1.54	4.88
FVal_S:Gender_M	0.09	1.60	0.06	5939.81	0.954	−3.05	3.24
FVal_H:Gender_M	−1.97	1.59	−1.24	5945.52	0.215	−5.09	1.14
Grp_BL:Cong_IC:FVal_S	−0.34	1.64	−0.21	5861.53	0.834	−3.55	2.87
Grp_BL:Cong_IC:FVal_H	−4.58	1.62	−2.83	6023.10	0.005	−7.75	−1.41

Table A.7 shows the estimates for the Study 1 accuracy model, fit to by-participant loglinear $d\prime$ values.

Table A.7: Model estimates for Study 1 $d\prime$ data with effects coding, Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
(Intercept)	3.068	0.056	54.755	46	0.000	2.955	3.181
GrpBL	−0.034	0.056	−0.602	46	0.550	−0.146	0.079
CongIC	−0.177	0.028	−6.212	230	0.000	−0.233	−0.121
FValS	0.026	0.040	0.644	230	0.520	−0.053	0.105
FValH	−0.006	0.040	−0.141	230	0.888	−0.085	0.074
GrpBL:CongIC	−0.012	0.028	−0.416	230	0.678	−0.068	0.044
GrpBL:FValS	−0.005	0.040	−0.119	230	0.905	−0.084	0.075
GrpBL:FValH	−0.007	0.040	−0.181	230	0.856	−0.087	0.072
CongIC:FValS	0.030	0.040	0.739	230	0.461	−0.050	0.109
CongIC:FValH	0.001	0.040	0.023	230	0.981	−0.078	0.080
GrpBL:CongIC:FValS	0.074	0.040	1.836	230	0.068	−0.005	0.153
GrpBL:CongIC:FValH	−0.053	0.040	−1.303	230	0.194	−0.132	0.027

A.3.2 Study 2

A.3.2.1 Experiment 1

Tables A.8 and A.9 show the fixed effects estimates for Study 2.1 reaction time data and accuracy models based on loglinear $d\prime$s for each participant.

Table A.8: Model estimates for Study 2, Experiment 1 reaction time data with effects coding, Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	*t*Stat	DF	*p*Value	Lower	Upper
(Intercept)	468.18	7.66	61.1	5.77	0	449.25	487.12
GrpBL	1.41	6.22	0.23	2.46	0.84	−21.08	23.9
FValSad	0.08	1.49	0.06	96.82	0.95	−2.87	3.04
FValHappy	0.41	1.48	0.28	97.29	0.78	−2.53	3.35
CongIC	3.62	0.9	4	7151.47	0.0001	1.84	5.39
GenderM	8.09	5.26	1.54	54.16	0.13	−2.45	18.63
GrpBL:FValSad	0.44	1.28	0.35	7436.59	0.73	−2.07	2.96
GrpBL:FValHappy	0.12	1.28	0.1	7586.09	0.92	−2.38	2.63
GrpBL:CongIC	2.18	0.9	2.43	7700.69	0.02	0.42	3.94
FValSad:CongIC	2.38	1.29	1.84	7412.78	0.07	−0.15	4.91
FValHappy:CongIC	0.18	1.28	0.14	7115.81	0.89	−2.34	2.7
FValSad:GenderM	0.87	1.26	0.68	7614.34	0.49	−1.61	3.34
FValHappy:GenderM	−0.83	1.27	−0.65	7612.48	0.51	−3.32	1.66
GrpBL:FValSad:CongIC	1.04	1.28	0.82	7482.84	0.41	−1.46	3.54
GrpBL:FValHappy:CongIC	−0.85	1.27	−0.67	7673.76	0.5	−3.35	1.64

Table A.9: Model estimates for Study 2, Experiment 1 $d\prime$ data with effects coding, Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	*t*Stat	DF	*p*Value	Lower	Upper
(Intercept)	3.14	0.06	55.59	56	0	3.02	3.25
GrpBL	0.02	0.06	0.3	56	0.764	−0.1	0.13
CongIC	−0.11	0.02	−4.69	280	$4.3E-06$	−0.15	−0.06
FValSad	−0.09	0.03	−2.9	280	0.004	−0.16	−0.03
FValHappy	0.07	0.03	2.22	280	0.027	0.01	0.14
GrpBL:CongIC	−0.001	0.02	−0.06	280	0.949	−0.05	0.04
GrpBL:FValHappy	−0.01	0.03	−0.35	280	0.725	−0.08	0.05
GrpBL:FValSad	−0.03	0.03	−1.02	280	0.308	−0.1	0.03
CongIC:FValSad	0.06	0.03	2	280	0.047	0.001	0.13
CongIC:FValHappy	−0.002	0.03	−0.07	280	0.942	−0.07	0.06
GrpBL:GrpBL:FValSad	−0.06	0.03	−1.73	280	0.084	−0.12	0.01
GrpBL:GrpBL:FValHappy	0.03	0.03	0.91	280	0.363	−0.03	0.09

A.3.2.2 Experiment 2

Tables A.10–A.11 show the estimates for the LME fit to Study 2.2 reaction time data and the GLME fit to corresponding accuracy data.

Table A.10: Model estimates for Study 2, Experiment 2 reaction time data with effects coding, Satterthwaite approximation for degrees of freedom. Estimates and standard error have been converted back into milliseconds for interpretability. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
(Intercept)	504.43	6.48	157.28	93.94	0	517.41	491.93
GrpBL	4.13	5.05	−0.82	57.03	0.42	−14.39	5.76
CongIC	3.33	0.71	−4.66	10416.95	0.000003	−4.85	−1.84
FValSad	1.08	1.04	−1.03	70.72	0.31	−3.26	0.97
FValHappy	−0.25	1.02	0.25	69.11	0.81	−1.85	2.21
GenderM	12.3	5.12	−2.38	57.01	0.02	−22.37	−1.83
GrpBL:CongIC	0.42	0.71	−0.6	10512.43	0.55	−1.87	0.93
GrpBL:FValSad	0.17	1.02	−0.17	10348.72	0.86	−2.25	1.76
GrpBL:FValHappy	−0.71	1.002	0.71	10599.68	0.48	−1.29	2.59
FValSad:CongIC	2.68	1.02	−2.61	10279.8	0.009	−4.81	−0.64
FValHappy:CongIC	−1.75	0.998	1.76	10507.87	0.079	−0.21	3.59
FValSad:GenderM	−1.14	1.04	1.11	10643.45	0.27	−0.92	3.07
FValHappy:GenderM	0.2	1.02	−0.2	10643.94	0.84	−2.27	1.73
GrpBL:FValSad:CongIC	0.08	1.01	−0.08	10271.77	0.93	−2.14	1.84
GrpBL:FValHappy:CongIC	−0.14	0.99	0.14	10624.94	0.89	−1.87	2.02

Table A.11: Model estimates for Study 2, Experiment 2 accuracy data with effects coding, Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	*t*Stat	DF	*p*Value	Lower	Upper
(Intercept)	1.01	0.01	76.65	11622	0	0.99	1.04
GrpBL	0.08	0.01	5.93	11622	3.1E-09	0.05	0.1
CongIC	−0.05	0.01	−3.59	11622	0.0003	−0.07	−0.02
FValSad	−0.03	0.02	−1.73	11622	0.08	−0.07	0.004
FValHappy	0.02	0.02	1.14	11622	0.26	−0.02	0.06
GenderM	0.01	0.01	0.51	11622	0.61	−0.02	0.03
GrpBL:CongIC	−0.02	0.01	−1.55	11622	0.12	−0.05	0.01
GrpBL:FValSad	0.02	0.02	1.13	11622	0.26	−0.02	0.06
GrpBL:FValHappy	0.003	0.02	0.19	11622	0.85	−0.03	0.04
CongIC:FValSad	−0.01	0.02	−0.41	11622	0.68	−0.04	0.03
CongIC:FValHappy	−0.02	0.02	−1.22	11622	0.22	−0.06	0.01
GrpBL:CongIC:FValSad	0.03	0.02	1.68	11622	0.09	−0.01	0.07
GrpBL:CongIC:FValHappy	−0.02	0.02	−1.31	11622	0.19	−0.06	0.01

Finally, Table A.12 lists the estimates from the conflict adaptation model.

Table A.12: Model estimates for Study 2, Experiment 2 conflict adaptation analysis. Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
(Intercept)	4E-02	3E-04	150.16	89.37	0.00	4E-02	5E-02
GrpBL	-2E-04	2E-04	-0.70	57.90	0.49	-6E-04	3E-04
CongIC	-2E-04	5E-05	-3.06	9972.74	0.00	-3E-04	-6E-05
FValSad	-1E-04	7E-05	-1.89	9970.95	0.06	-3E-04	5E-06
FValHappy	1E-04	7E-05	2.03	9969.95	0.04	5E-06	3E-04
PrevCong_C	7E-05	5E-05	1.38	9972.37	0.17	-3E-05	2E-04
PartGender	-6E-04	2E-04	-2.55	55.02	0.01	-1E-03	-1E-04
GrpBL:CongIC	1E-05	5E-05	0.22	9966.51	0.83	-9E-05	1E-04
GrpBL:FValSad	-8E-05	7E-05	-1.19	9965.78	0.24	-2E-04	5E-05
GrpBL:FValHappy	9E-05	7E-05	1.31	9966.43	0.19	-5E-05	2E-04
CongIC:FValSad	-1E-04	7E-05	-1.80	9971.44	0.07	-3E-04	1E-05
CongIC:FValHappy	1E-04	7E-05	1.62	9970.98	0.11	-2E-05	3E-04
GrpBL:PrevCong_C	-9E-06	5E-05	-0.19	9967.22	0.85	-1E-04	9E-05
CongIC:PrevCong_C	1E-05	5E-05	0.20	9973.28	0.84	-9E-05	1E-04
FValSad:PrevCong_C	-2E-06	7E-05	-0.03	9971.73	0.98	-1E-04	1E-04
FValHappy:PrevCong_C	-3E-05	7E-05	-0.40	9971.14	0.69	-2E-04	1E-04
GrpBL:CongIC:FValSad	-3E-05	7E-05	-0.45	9966.26	0.65	-2E-04	1E-04
GrpBL:CongIC:FValHappy	-6E-05	7E-05	-0.85	9966.50	0.39	-2E-04	8E-05
GrpBL:CongIC:PrevCong_C	6E-06	5E-05	0.11	9966.99	0.91	-9E-05	1E-04
GrpBL:FValSad:PrevCong_C	8E-06	7E-05	0.12	9966.84	0.91	-1E-04	1E-04
GrpBL:FValHappy:PrevCong_C	3E-05	7E-05	0.38	9968.39	0.70	-1E-04	2E-04
CongIC:FValSad:PrevCong_C	1E-04	7E-05	1.86	9973.53	0.06	-7E-06	3E-04
CongIC:FValHappy:PrevCong_C	-2E-04	7E-05	-2.58	9973.89	0.01	-3E-04	-4E-05
GrpBL:CongIC:FValSad:PrevCong_C	6E-06	7E-05	0.08	9966.76	0.93	-1E-04	1E-04
GrpBL:CongIC:FValHappy:PrevCong_C	-3E-05	7E-05	-0.43	9968.22	0.67	-2E-04	1E-04
FValSad:PrevCong_C:PrevVal_Sad	-7E-05	9E-05	-0.77	9948.45	0.44	-2E-04	1E-04
FValHappy:PrevCong_C:PrevVal_Sad	1E-04	8E-05	1.57	9948.60	0.12	-3E-05	3E-04
FValSad:PrevCong_C:PrevVal_Happy	9E-05	8E-05	1.07	9949.14	0.29	-7E-05	2E-04
FValHappy:PrevCong_C:PrevVal_Happy	-5E-05	9E-05	-0.52	9948.30	0.60	-2E-04	1E-04
GrpBL:CongIC:FValSad:PrevCong_C	6E-06	6E-05	0.10	9942.87	0.92	-1E-04	1E-04
GrpBL:CongIC:FValHappy:PrevCong_C	-4E-06	6E-05	-0.07	9944.83	0.95	-1E-04	1E-04
GrpBL:CongIC:FValSad:PrevVal_Sad	4E-05	9E-05	0.44	9944.98	0.66	-1E-04	2E-04
GrpBL:CongIC:FValHappy:PrevVal_Sad	5E-05	8E-05	0.57	9944.32	0.57	-1E-04	2E-04
GrpBL:CongIC:FValSad:PrevVal_Happy	8E-05	8E-05	1.01	9943.26	0.31	-8E-05	2E-04
GrpBL:CongIC:FValHappy:PrevVal_Happy	-1E-05	9E-05	-0.11	9944.48	0.91	-2E-04	2E-04
GrpBL:CongIC:PrevCong_C:PrevVal_Sad	-3E-06	6E-05	-0.05	9944.65	0.96	-1E-04	1E-04
GrpBL:CongIC:PrevCong_C:PrevVal_Happy	-1E-05	6E-05	-0.18	9943.96	0.86	-1E-04	1E-04
GrpBL:FValSad:PrevCong_C:PrevVal_Sad	-2E-05	9E-05	-0.22	9945.45	0.82	-2E-04	2E-04
GrpBL:FValHappy:PrevCong_C:PrevVal_Sad	6E-06	8E-05	0.08	9944.90	0.94	-2E-04	2E-04
GrpBL:FValSad:PrevCong_C:PrevVal_Happy	-6E-05	8E-05	-0.71	9946.30	0.48	-2E-04	1E-04
GrpBL:FValHappy:PrevCong_C:PrevVal_Happy	-7E-05	9E-05	-0.74	9944.39	0.46	-2E-04	1E-04
CongIC:FValSad:PrevCong_C:PrevVal_Sad	-7E-05	9E-05	-0.75	9948.27	0.46	-2E-04	1E-04
CongIC:FValHappy:PrevCong_C:PrevVal_Sad	7E-05	8E-05	0.83	9946.32	0.40	-9E-05	2E-04
CongIC:FValSad:PrevCong_C:PrevVal_Happy	1E-04	8E-05	1.36	9946.72	0.17	-5E-05	3E-04
CongIC:FValHappy:PrevCong_C:PrevVal_Happy	-1E-04	9E-05	-1.17	9948.52	0.24	-3E-04	7E-05
GrpBL:CongIC:FValSad:PrevCong_C:PrevVal_Sad	5E-05	9E-05	0.60	9943.47	0.55	-1E-04	2E-04
GrpBL:CongIC:FValHappy:PrevCong_C:PrevVal_Sad	1E-04	8E-05	1.19	9943.62	0.23	-6E-05	3E-04
GrpBL:CongIC:FValSad:PrevCong_C:PrevVal_Happy	6E-05	8E-05	0.79	9943.72	0.43	-1E-04	2E-04
GrpBL:CongIC:FValHappy:PrevCong_C:PrevVal_Happy	-8E-05	9E-05	-0.84	9944.05	0.40	-3E-04	1E-04

A.3.3 Study 3

A.3.3.1 Experiment 1

Tables A.13 and A.14 list the model estimates for the RT and accuracy models from Study 3, Experiment 1.

Table A.13: Model estimates for Study 3, Experiment 1 RT data with effects coding, Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
(Intercept)	44.822	0.354	126.724	86.739	0.000	44.119	45.525
Grp_BL	-0.114	0.291	-0.391	53.858	0.697	-0.698	0.470
FValSad	-0.020	0.099	-0.198	5779.761	0.843	-0.213	0.174
FValHappy	0.075	0.099	0.758	5861.416	0.449	-0.119	0.269
CongIC	0.210	0.098	2.132	35.170	0.040	0.010	0.409
GenderM	0.421	0.268	1.571	54.549	0.122	-0.116	0.959
GrpBL:FValSad	0.157	0.098	1.600	5958.236	0.110	-0.035	0.349
GrpBL:FValHappy	-0.249	0.098	-2.532	5927.488	0.011	-0.442	-0.056
GrpBL:CongIC	0.034	0.077	0.435	52.161	0.665	-0.121	0.189
FValSad:CongIC	-0.024	0.099	-0.247	5839.477	0.805	-0.218	0.169
FValHappy:CongIC	0.097	0.099	0.986	5877.396	0.324	-0.096	0.291
GrpBL:FValSad:CongIC	0.046	0.098	0.473	6010.424	0.636	-0.146	0.238
GrpBL:FValHappy:CongIC	-0.123	0.098	-1.250	5930.468	0.211	-0.316	0.070

A.3.3.2 Experiment 2

Model estimates for Experiment 2 of Study 3 are listed in Tables A.15 and A.16.

Table A.14: Model estimates for Study 3, Experiment 1 accuracy data with effects coding, Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
(Intercept)	1.083	0.074	14.713	11151.000	0.000	0.939	1.227
GrpBL	0.111	0.045	2.453	11151.000	0.014	0.022	0.199
GoStatus	0.362	0.017	21.263	11151.000	0.000	0.328	0.395
FValSad	0.065	0.025	2.555	11151.000	0.011	0.015	0.114
FValHappy	-0.027	0.024	-1.094	11151.000	0.274	-0.075	0.021
CongIC	-0.056	0.017	-3.352	11151.000	0.001	-0.089	-0.023
Gender	-0.002	0.043	-0.054	11151.000	0.957	-0.087	0.083
GrpBL:GoStatus	-0.053	0.017	-3.094	11151.000	0.002	-0.086	-0.019
GrpBL:FValSad	-0.022	0.024	-0.890	11151.000	0.374	-0.069	0.026
GrpBL:FValHappy	0.017	0.024	0.712	11151.000	0.477	-0.029	0.063
GoStatus:FValSad	-0.002	0.026	-0.082	11151.000	0.935	-0.053	0.049
GoStatus:FValHappy	-0.020	0.024	-0.815	11151.000	0.415	-0.068	0.028
GrpBL:CongIC	-0.019	0.017	-1.146	11151.000	0.252	-0.052	0.014
GoStatus:CongIC	0.024	0.017	1.442	11151.000	0.149	-0.009	0.057
FValSad:CongIC	0.013	0.025	0.540	11151.000	0.589	-0.035	0.062
FValHappy:CongIC	-0.020	0.024	-0.860	11151.000	0.390	-0.067	0.026
GrpBL:GoStatus:FValSad	0.020	0.024	0.838	11151.000	0.402	-0.027	0.068
GrpBL:GoStatus:FValHappy	0.007	0.024	0.276	11151.000	0.782	-0.040	0.053
GrpBL:GoStatus:CongIC	0.011	0.017	0.646	11151.000	0.519	-0.022	0.044
GrpBL:FValSad:CongIC	-0.012	0.024	-0.505	11151.000	0.613	-0.060	0.035
GrpBL:FValHappy:CongIC	0.036	0.024	1.543	11151.000	0.123	-0.010	0.083
GoStatus:FValSad:CongIC	0.043	0.025	1.710	11151.000	0.087	-0.006	0.091
GoStatus:FValHappy:CongIC	-0.059	0.024	-2.494	11151.000	0.013	-0.105	-0.013
GrpBL:GoStatus:FValSad:CongIC	0.002	0.024	0.075	11151.000	0.940	-0.046	0.049
GrpBL:GoStatus:FValHappy:CongIC	0.009	0.024	0.387	11151.000	0.699	-0.037	0.055

Table A.15: Model estimates for Study 3, Experiment 2 RT data with effects coding, Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
(Intercept)	13.77	0.08	168.33	83.51	0.00	13.60	13.93
GrpBL	0.02	0.07	0.34	54.46	0.73	-0.11	0.16
FValSad	-0.01	0.02	-0.55	4664.36	0.58	-0.04	0.02
FValHappy	0.00	0.02	0.13	4657.33	0.90	-0.03	0.04
CongIC	0.02	0.01	1.36	34.00	0.18	-0.01	0.05
Gender	-0.01	0.07	-0.19	54.33	0.85	-0.15	0.12
GrpBL:FValSad	0.01	0.02	0.49	4651.00	0.62	-0.03	0.04
GrpBL:FValHappy	0.00	0.02	0.27	4646.82	0.79	-0.03	0.04
GrpBL:CongIC	0.00	0.01	0.31	4502.09	0.76	-0.02	0.03
FValSad:CongIC	-0.02	0.02	-1.31	4464.76	0.19	-0.06	0.01
FValHappy:CongIC	0.03	0.02	1.49	4465.34	0.14	-0.01	0.06
GrpBL:FValSad:CongIC	-0.03	0.02	-1.50	4471.57	0.13	-0.06	0.01
GrpBL:FValHappy:CongIC	0.01	0.02	0.36	4466.21	0.72	-0.03	0.04

Table A.16: Model estimates for Study 3, Experiment 2 accuracy data with effects coding, Satterthwaite approximation for degrees of freedom. Grp = group (BL = bilingual), Cong = congruence (IC = incongruent), FVal = flanker valence (S = sad, H = happy).

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
(Intercept)	1.25	0.08	15.67	10566	0.000	1.09	1.40
GrpBL	0.05	0.04	1.05	10566	0.293	-0.04	0.13
FValSad	0.01	0.03	0.51	10566	0.610	-0.04	0.06
FValHappy	0.02	0.03	0.85	10566	0.395	-0.03	0.07
CongIC	-0.04	0.02	-2.16	10566	0.031	-0.07	0.00
Gender	-0.19	0.04	-4.38	10566	0.000	-0.28	-0.11
GrpBL:FValSad	-0.01	0.03	-0.56	10566	0.576	-0.06	0.04
GrpBL:FValHappy	-0.01	0.03	-0.50	10566	0.616	-0.06	0.04
GrpBL:CongIC	0.01	0.02	0.45	10566	0.656	-0.03	0.04
FValSad:CongIC	-0.04	0.03	-1.61	10566	0.107	-0.09	0.01
FValHappy:CongIC	0.05	0.03	1.94	10566	0.053	0.00	0.10
GrpBL:FValSad:CongIC	0.03	0.03	1.07	10566	0.283	-0.02	0.08
GrpBL:FValHappy:CongIC	0.03	0.03	1.07	10566	0.283	-0.02	0.08

We are, I am, you are by cowardice or courage the one who find our way back to this scene carrying a knife, a camera a book of myths in which our names do not appear.

— Adrienne Rich,
'Diving Into the Wreck', pp. 55.

B
Supplementary Data Analysis

Contents

B.1 Face Colour Contrast Analysis 181
B.1.1 Study 1 182
B.1.2 Study 2 182
B.1.3 Study 3 184
B.2 Additional Model Selection and Hypothesis Testing Scripts 185
B.2.1 Study 2 186
B.2.2 Study 3 198

B.1 Face Colour Contrast Analysis

As noted in the main text, the face images used as stimuli contained faces of multiple races/ethnicities. A factor overlooked during the main analyses was the role that differences in face colour could play in the flanker congruence effect. Colour has been used as a conflict dimension in multiple studies (Kanske & Kotz, 2010b; Zhou & Liu, 2013). However, Zhou and Liu (2013) reported a modulation of the flanker conflict effect by colour only when it was a task-relevant response dimension, and in all studies known to the author, the colours utilised had a very high contrast (e.g. blue/red or green/red).

In order to analyse the influence of face colour in the studies reported in this

dissertation, the average brightness of each face was determined by calculating an average of each image's mean RGB[1] values.

Using the image brightness scores, the difference between target and flankers in each experimental trial was calculated to obtain a ´Contrast' score. This value (positive if the flanker was brighter than the target and negative if the target was brighter than the flanker) was added as a covariate to each model equation.

The nested models were compared using the LRT method with 100 simulations. This process is illustrated in the script for Study 1 included in Section A.1.3.2 with detailed comments.

In Study 1 and Study 2.1, this covariate could only be included in the RT model, as the accuracy model was fit to aggregate data without trial-level information. An effect of colour contrast on accuracy can thus not be fully ruled out for the FNG tasks in Studies 1 and 2.

B.1.1 Study 1

There was no significant main effect of Contrast in Study 1 RTs. Table B.1 gives the Type–III ANOVA table for the hit RT model including Contrast as a continuous covariate. This can be compared to Table 3.5 in Chapter 3, which shows the model terms without the covariate.

B.1.2 Study 2

B.1.2.1 Experiment 1

Again there was no significant main effect of Contrast. None of the other results reported for Study 2.1 were significantly impacted by the inclusion of this covariate. Table B.2 gives the Type–III ANOVA table for the hit RT model including Contrast as a continuous covariate. This can be compared to Table 4.3 in Chapter 4.

[1]Red, Green, Blue, the standard format for reporting colour information for each pixel of a digital image. Each value in the triplet ranges from 1–255, indicating the saturation of the respective colour. For example, the colour white is represented by the triplet [255, 255, 255], while the colour black is represented by [0, 0, 0]. The average of the three values represents the overall brightness. For the present analysis, the averages of the R, G, and B vectors were averaged to obtain the mean brightness of the whole image.

Table B.1: Type–III ANOVA table for the Study 1 RT model including a continuous covariate for face contrast.

Term	Fstat	DF1	DF2	pValue
(Intercept)	4557.50	1	82.46	0.000
Group	0.17	1	46.48	0.686
FValence	0.98	2	92.26	0.379
Congruence	14.41	1	5865.80	0.000
PartGender	7.46	1	44.63	0.009
Contrast	0.33	1	110.20	0.565
Group:FValence	0.23	2	6035.10	0.791
Group:Congruence	0.08	1	6155.90	0.777
FValence:Congruence	1.38	2	5834.60	0.252
FValence:PartGender	1.05	2	6057.10	0.351
Group:FValence:Congruence	6.19	2	6033.00	0.00

Table B.2: Type–III ANOVA table for the Study 2.1 RT model including a continuous covariate for face contrast.

Term	Fstat	DF1	DF2	pValue
(Intercept)	6429.32	1	86.68	0.000
Group	0.00	1	55.38	0.995
FValence	0.07	2	197.33	0.934
Congruence	6.04	1	37.28	0.019
PartGender	1.85	1	55.25	0.180
Contrast	0.17	1	234.95	0.684
Group:FValence	0.23	2	401.14	0.791
Group:Congruence	4.99	1	218.21	0.027
FValence:Congruence	1.49	2	185.26	0.227
FValence:PartGender	0.45	2	7616.63	0.636
Group:FValence:Congruence	0.72	2	393.98	0.49

Table B.3: Type–III ANOVA table for the Study 2.2 RT model including a continuous covariate for face contrast.

Term	Fstat	DF1	DF2	pValue
(Intercept)	23059.51	1	86.07	0.000
Group	0.46	1	55.00	0.501
Congruence	12.24	1	169.28	0.001
FValence	0.95	2	155.03	0.390
PartGender	6.43	1	55.01	0.014
Contrast	0.40	1	180.52	0.527
Group:Congruence	0.51	1	1666.99	0.477
Group:FValence	0.16	2	8024.10	0.850
Congruence:FValence	1.75	2	155.24	0.178
Group:Congruence:FValence	0.08	2	8029.79	0.927

B.1.2.2 Experiment 2

Again there was no significant main effect of Contrast. None of the other results reported for Study 2.2 were significantly impacted by the inclusion of this covariate. Table B.3 gives the Type–III ANOVA table for the hit RT model including Contrast as a continuous covariate. This can be compared to Table 4.6 in Chapter 4.

B.1.3 Study 3

B.1.3.1 Experiment 1

Again there was no significant main effect of Contrast. None of the other results reported for Study 3.1 were significantly impacted by the inclusion of this covariate. Table B.4 gives the Type–III ANOVA table for the hit RT model including Contrast as a continuous covariate. This can be compared to Table 5.2 in Chapter 5.

B.1.3.2 Experiment 2

Again there was no significant main effect of Contrast. None of the other results reported for Study 3.2 were significantly impacted by the inclusion of this covariate. Table B.5 gives the Type–III ANOVA table for the hit RT model including Contrast as a continuous covariate. This can be compared to Table 5.5 in Chapter 5.

Table B.4: Type–III ANOVA table for the Study 3.1 RT model including a continuous covariate for face contrast.

Term	Fstat	DF1	DF2	pValue
(Intercept)	16133.88	1	86.50	0.000
Group	0.15	1	53.84	0.704
FValence	0.07	2	5680.06	0.931
Congruence	4.39	1	34.81	0.043
Gender	2.56	1	54.49	0.115
Contrast	0.85	1	467.34	0.358
Group:Congruence	0.19	1	52.20	0.664

Table B.5: Type–III ANOVA table for the Study 3.2 RT model including a continuous covariate for face contrast.

Term	Fstat	DF1	DF2	pValue
(Intercept)	28118.42	1	83.17	0.000
Group	0.12	1	54.47	0.730
FValence	0.12	2	4572.68	0.891
Congruence	1.94	1	33.65	0.173
Gender	0.04	1	54.34	0.851
Contrast	0.27	1	187.35	0.605
Group:Congruence	0.10	1	4506.56	0.754

B.2 Additional Model Selection and Hypothesis Testing Scripts

The following scripts were used for the model selection and hypothesis testing (including covariate testing) workflows for Studies 2 and 3. They follow the same steps and structure as those presented in Appendix A and are thus not as extensively commented.

B.2.1 Study 2

B.2.1.1 Experiment 1

The following script presents the model selection workflow for Study 2.1.

```
1
2 FNG2
3 randEffects = {};
4 %% Test random intercepts
5 1
6 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
7 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)',...
8 'FitMethod', 'REML', 'CheckHessian',1,'DummyVarCoding', 'effects');
9 %
10 2
11 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
12 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)',...
13 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
14 %
15 3
16 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
17 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem)',...
18 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
19 %
20 4
21 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
22 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem)+(1|FlankerItem:Participant)',...
23 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
24 %
25 5
26 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
27 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)',...
28 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
29 %
30 6
31 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
32 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:
       Congruence)',...
33 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
34 %
35 7
36 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
37 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:
       Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
38 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
39 %
40 8
41 %
42 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
43 'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|
       FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|
       FlankerItem:FValence:Congruence:Participant)',...
44 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
45 %
46 9
47 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
48 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
       FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|
       FlankerItem:FValence:Congruence:Participant)',...
49 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
50 %
51 10
52 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
53 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
       Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:
       FValence:Congruence:Participant)',...
54 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
55 %
56 11
57 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
58 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|
       FlankerItem:FValence:Congruence:Participant)',...
59 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
60 %
61 12
62 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
63 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:
       FValence:Congruence:Participant)',...
64 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
65 %
66 13
67 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
```

```
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
14
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
%
15
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Target:Participant)+(1|FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
16
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
17
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
18
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
19
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
20
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
%
21
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Target)+(1|FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
22
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
23
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
24
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
25
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
26
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
```

```
135 'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|
        FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)',...
136 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
137 %
138 27
139 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
140 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem:
        Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:
        FValence:Congruence:Participant)',...
141 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
142 %
143 28
144 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
145 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
        FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
        Congruence:Participant)',...
146 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
147 %
148 29
149 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
150 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
        FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
        Congruence:Participant)',...
151 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
152 %
153 30
154 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
155 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
        FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence:
        Participant)',...
156 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
157 %
158 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
159 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
        FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)',...
160 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
161 %
162 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
163 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
        FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:
        Participant)',...
164 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
165 %
166 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
167 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
        Participant)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:
        Participant)',...
168 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
169 %
170 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
171 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
        FValence)+(1|FlankerItem:FValence:Congruence:Participant)',...
172 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
173 %
174 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
175 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
        Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)',...
176 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
177 %
178 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
179 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
        FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence:Participant)',...
180 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
181 %
182 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
183 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
        FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)',...
184 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
185 %(1)AIC -90665.994270
186
187 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
188 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
        FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence:Participant)',...
189 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
190 %
191 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
192 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
        FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
193 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
194 %(1)AIC -90665.994270***
195
196 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
197 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
        FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
198 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
199
200 [lmeBest_step1, randEffects] = bestLme(randEffects);
201 sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
202 sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})
203
204 %% Add random slopes
205
206 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
```

```
207 'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Group|Target)+(1|Target:Participant)+(1|
        FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
208 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
209 %
210 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
211 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(Group-1|Target)+(1|Target:
        Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
212 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
213 %
214 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
215 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
        FlankerItem:Participant)+(1+Group|FlankerItem:FValence:Congruence)',...
216 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
217 %
218 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
219 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
        FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(Group-1|FlankerItem:FValence:
        Congruence)',...
220 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
221 4%
222 %
223 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
224 'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1|Target)+(1|Target:Participant)
        +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
225 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
226 %
227 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
228 'RTconv~Group*Congruence*FValence+(1+FValence|Participant)+(1|Target)+(1|Target:Participant)
        +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
229 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
230 %
231 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
232 'RTconv~Group*Congruence*FValence+(1|Participant)+(Congruence-1|Participant)+(1|Target)+(1|
        Target:Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
233 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
234 %
235 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
236 'RTconv~Group*Congruence*FValence+(1|Participant)+(FValence-1|Participant)+(1|Target)+(1|Target
        :Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
237 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
238 %
239 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
240 'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1|Target:Participant)
        +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
241 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
242 %(1)AIC  -85613.398442****
243 9
244 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
245 'RTconv~Group*Congruence*FValence+(1|Participant)+(Congruence-1|Target)+(1|Target:
        Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
246 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
247 %
248 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
249 'RTconv~Group*Congruence*FValence+(1|Participant)+(1+FValence|Target)+(1|Target:Participant)
        +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
250 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
251 %
252 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
253 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(FValence-1|Target)+(1|Target:
        Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
254 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
255 %
256
257 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
258 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+Congruence|Target:Participant)
        +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
259 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
260 %
261 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
262 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(Congruence-1|Target:Participant)
        +(1|Target:Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)'
        ,...
263 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
264 %
265 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
266 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+FValence|Target:Participant)
        +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
267 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
268 %
269 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
270 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(FValence-1|Target:Participant)
        +(1|Target:Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)'
        ,...
271 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
272 %
273 16
274 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
275 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1+
        Congruence|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
276 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
277 %
278 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
```

```
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(Congruence-1|FlankerItem:Participant)+(1|FlankerItem:FValence:
    Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1+FValence
    |FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(FValence-1|FlankerItem:Participant)+(1|FlankerItem:FValence:
    Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
20
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1+Congruence|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(Congruence-1|FlankerItem:
    FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%

[lmeBest_step2, randEffects] = bestLme(randEffects);
sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})

%% Combine random slopes

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1|Target:Participant)
    +(1|FlankerItem:Participant)+(1+Group|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%(1)AIC -85614.575773*****
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1|Target:Participant)
    +(1|FlankerItem:Participant)+(1+Congruence|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|Target:
    Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1|Target:Participant)
    +(1|FlankerItem:Participant)+(1+Group+Congruence|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|Target:
    Participant)+(1|FlankerItem:Participant)+(1+Group|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|Target:
    Participant)+(1|FlankerItem:Participant)+(1+Congruence|FlankerItem:FValence:Congruence)'
    ,...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%

[lmeBest_step3, randEffects] = bestLme(randEffects);
sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})


%%
```

The following script presents the covariate testing and hypothesis testing workflow for Study 2.1.

```
FNG2

fullModel = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|
    FlankerItem:Participant)+(1+Group|FlankerItem:FValence:Congruence)+(1|FullType)',...
    'FitMethod', 'ML', 'DummyVarCoding', 'effects');
%% Add Gender covariate and compare nested models, beinning with highest-order interaction,
    using LRT.
lmeGender = fitlme(hits, 'RTconv ~ PartGender + Group*FValence*Congruence + (1+Congruence|
    Participant)+(1+Congruence|Target)+(1|FlankerItem:Participant)+(1+Group|FlankerItem:
    FValence:Congruence)+(1|FullType)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
```

```
lmeGender1 = fitlme(hits, 'RTconv ~ PartGender*FValence + Group*FValence*Congruence + (1|Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)+ (1|FullType)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
lmeGender2 = fitlme(hits, 'RTconv ~ PartGender*Group*FValence*Congruence + (1|Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)+ (1|FullType)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
genderLRT = compare(lmeGender2, lmeGender1, 'NSim', 100);
export(genderLRT,'XLSFile','S2E1genderLRT.xlsx');
%G1 better fit than G2. Reduce model.
genderLRT = compare(lmeGender1, lmeGender, 'NSim', 100);
export(genderLRT,'XLSFile','S2E1genderLRT.xlsx', 'Sheet', 2);
%G1 better fit than G. Retain interaction term.
genderLRT = compare(lmeGender, fullModel, 'NSim', 100);
export(genderLRT,'XLSFile','S2E1genderLRT.xlsx', 'Sheet', 3);

%% Add face luminance covariate and compare nested models, beinning with highest-order interaction, using LRT.
lmeLum = fitlme(hits, 'RTconv ~ PartGender*FValence+Contrast + Group*FValence*Congruence + (1+Congruence|Participant)+(1+Congruence|Target)+(1|FlankerItem:Participant)+(1+Group|FlankerItem:FValence:Congruence)+(1|FullType)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
lmeLum1 = fitlme(hits, 'RTconv ~ PartGender*FValence+Contrast*Congruence + Group*FValence*Congruence + (1|Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)+ (1|FullType)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
lmeLum2 = fitlme(hits, 'RTconv ~ PartGender*FValence+ Contrast * Group*FValence*Congruence + (1|Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)+ (1|FullType)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
lumLRT = compare(lmeLum2, lmeLum1, 'NSim', 100);
export(lumLRT,'XLSFile','S2E1lumLRT.xlsx');
%L1 better fit than L2. Reduce model.
lumLRT = compare(lmeLum1, lmeLum, 'NSim', 100);
export(lumLRT,'XLSFile','S2E1lumLRT.xlsx', 'Sheet', 2);
%G1 better fit than G. Retain interaction term.
lumLRT = compare(lmeLum, fullModel, 'NSim', 100);
export(lumLRT,'XLSFile','S2E1lumLRT.xlsx', 'Sheet', 3);

%% Obtain fixed effects estimates.
lmeDiffCoding = fitlme(hits, 'RTconv ~ PartGender*FValence + Group*FValence*Congruence + (1+Congruence|Participant)+(1+Congruence|Target)+(1|FlankerItem:Participant)+(1+Group|FlankerItem:FValence:Congruence)+(1|FullType)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
[~, ~, diffCodedFixedEffects] = fixedEffects(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %obtain conservative Satterthwaite approximations for DF/p value
export(diffCodedFixedEffects,'XLSFile','S2E1ModelEstimates.xlsx');
modelAnova = anova(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %check significance of terms
export(modelAnova,'XLSFile','S2E1ModelEstimates.xlsx', 'Sheet', 2);
%Change factor level order to define monolingual/congruent/neutral as intercept baseline
d.Congruence = reordercats(d.Congruence, {'Congruent', 'Incongruent'});
d.FValence = reordercats(d.FValence, {'Neutral','Sad','Happy'});
d.Group = reordercats(d.Group, {'Monolingual', 'Bilingual'});
hits = d(d.Response=='Hit', :);
%Refit model using dummy coding (default) for simple main effects.
lmeDummyCoding = fitlme(hits, 'RTconv ~ PartGender*FValence + Group*FValence*Congruence + (1+Congruence|Participant)+(1+Congruence|Target)+(1|FlankerItem:Participant)+(1+Group|FlankerItem:FValence:Congruence)+(1|FullType)', 'FitMethod', 'REML');
[~, ~, dummyCodedFixedEffects] = fixedEffects(lmeDummyCoding, 'DFMethod', 'Satterthwaite');
export(dummyCodedFixedEffects,'XLSFile','S2E1ModelEstimates.xlsx','Sheet',3);

%% Calculate d-primes and fit lme.
FNG2
hitRate = grpstats(go, {'Participant','Group', 'Congruence', 'FValence'}, {'mean'}, 'DataVars', 'Answer');
faRate = grpstats(nogo, {'Participant','Group', 'Congruence', 'FValence'}, {'mean'}, 'DataVars', 'Answer');
faRate.faRate(:,1) = 1-faRate.mean_Answer;

respP = hitRate;
respP.Properties.VarNames{6} = 'HitRate';
respP.faGroupCount(:,1) = faRate.GroupCount;
respP.faRate(:,1) = faRate.faRate;
respP.dprime(:,1) = dprime_simple(respP.HitRate,respP.faRate);

%loglinear dprime conversion for biased data
numhits=respP.HitRate.*respP.GroupCount;
numfas=respP.faRate.*respP.faGroupCount;
hrconv=(numhits+0.5)./(respP.GroupCount+1);
faconv=(numfas+0.5)./(respP.faGroupCount+1);
respP.dprimelog(:,1) = norminv(hrconv)-norminv(faconv);
respP.respCriterion(:,1) = -0.5.*(norminv(hrconv)+ norminv(faconv));

%fit model
DPpartLMs2 = fitlme(respP, 'dprimelog ~ Group*FValence*Congruence + (1|Participant:Group)',...
'FitMethod', 'REML','CheckHessian',true,'DummyVarCoding', 'effects');
%does not converge with more random terms.
dprime2anova=anova(DPpartLMs2,'DFMethod','Satterthwaite');
[~,~,s2e1dpfixec]=fixedEffects(DPpartLMs2,'DFMethod','Satterthwaite');
export(s2e1dpfixec,'XLSFile','S2E1DprimeModel.xlsx','Sheet',1);
export(dprime2anova,'XLSFile','S2E1DprimeModel.xlsx','Sheet',2);

respP.Congruence = reordercats(respP.Congruence, {'Congruent', 'Incongruent'});
respP.FValence = reordercats(respP.FValence, {'Neutral','Sad','Happy'});
respP.Group = reordercats(respP.Group, {'Monolingual', 'Bilingual'});
DPpartLMs2DC = fitlme(respP, 'dprimelog ~ Group*FValence*Congruence + (1|Participant:Group)'
```

```
      ,...
'FitMethod', 'REML','CheckHessian',true);
[~,~,s2e1dpfixdc]=fixedEffects(DPpartLMs2DC,'DFMethod','Satterthwaite');
export(s2e1dpfixdc,'XLSFile','S2E1DprimeModel.xlsx','Sheet',3);
```

B.2.1.2 Experiment 2

The following script presents the model selection workflow for Study 2.2.

```

FO
randEffects = {};
%% Test random intercepts

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem)+(1|FlankerItem:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:
    Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:
    Congruence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|
    FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|
    FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
    Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:
    FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|
    FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:
    FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem
    :FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
```

```
60 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
61 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
62 %
63 %
64 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
65 'RTconv~Group*Congruence*FValence+(1|Target:Participant)+(1|FlankerItem)+(1|FlankerItem:
       Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:
       FValence:Congruence:Participant)',...
66 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
67 %
68 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
69 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|FlankerItem)+(1|FlankerItem:Participant)
       +(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
70 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
71 %
72 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
73 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem:Participant)+(1|
       FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
74 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
75 %
76 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
77 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
78 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
79 %
80 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
81 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:
       Participant)',...
82 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
83 %
84 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
85 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
       FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence)',...
86 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
87 %
88 %
89 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
90 'RTconv~Group*Congruence*FValence+(1|Target)+(1|FlankerItem)+(1|FlankerItem:Participant)+(1|
       FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
91 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
92 %
93 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
94 'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem:Participant)
       +(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
95 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
96 %
97 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
98 'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|
       FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
99 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
100 %
101 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
102 'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|
       FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
103 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
104 %
105 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
106 'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|
       FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence:
       Participant)',...
107 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
108 %
109 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
110 'RTconv~Group*Congruence*FValence+(1|Target)+(1|Target:Participant)+(1|FlankerItem)+(1|
       FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)',...
111 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
112 %
113 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
114 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem:
       Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:
       FValence:Congruence:Participant)',...
115 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
116 %
117 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
118 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
       FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
119 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
120 %
121 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
122 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
       FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:
       Congruence:Participant)',...
```

```
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence:
    Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target:Participant)+(1|FlankerItem)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
    FValence)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:
    Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
    Participant)+(1|FlankerItem:FValence:Congruence)+(1|FlankerItem:FValence:Congruence:
    Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
    FValence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|FlankerItem)+(1|FlankerItem:
    Participant)+(1|FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:FValence)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%(1)AIC -90665.994270

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%(1)AIC -90665.994270***

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');

[lmeBest_step1, randEffects] = bestLme(randEffects);
sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})

%% Add random slopes

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Group|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(Group-1|Target)+(1|Target:
    Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1+Group|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(Group-1|FlankerItem:FValence:
    Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1|Target)+(1|Target:Participant)
    +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
```

```
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+FValence|Participant)+(1|Target)+(1|Target:Participant)
    +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(Congruence-1|Participant)+(1|Target)+(1|
    Target:Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(FValence-1|Participant)+(1|Target)+(1|Target
    :Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+Congruence|Target)+(1|Target:
    Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(Congruence-1|Target)+(1|Target:
    Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+FValence|Target)+(1|Target:
    Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(FValence-1|Target)+(1|Target:
    Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+Congruence|Target:Participant)
    +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(Congruence-1|Target:Participant)
    +(1|Target:Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)'
    ,...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+FValence|Target:Participant)
    +(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(FValence-1|Target:Participant)
    +(1|Target:Participant)+(1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)'
    ,...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1+
    Congruence|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%(1)AIC -92506.335369***
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(Congruence-1|FlankerItem:Participant)+(1|FlankerItem:FValence:
    Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1+FValence
    |FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(FValence-1|FlankerItem:Participant)+(1|FlankerItem:FValence:
    Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1+Congruence|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
```

```
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1|
    FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+(Congruence-1|FlankerItem:
    FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%

[lmeBest_step2, randEffects] = bestLme(randEffects);
sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})

%% Combine random slopes

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(Congruence
    -1|Target:Participant)+(1+Congruence|FlankerItem:Participant)+(1|FlankerItem:FValence:
    Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%(1)AIC -92506.701820***

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(Congruence
    -1|Target:Participant)+(1|FlankerItem:Participant)+(Congruence-1|FlankerItem:Participant)
    +(1|FlankerItem:FValence:Congruence)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');
%

[lmeBest_step3, randEffects] = bestLme(randEffects);
sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})


%%
```

The following script presents the covariate testing and hypothesis testing workflow for Study 2.2.

```
FO
fullModel = fitlme(hits, 'RTconv ~ Group*FValence*Congruence + (1|Participant)+(1|Target)+(1|
    Target:Participant)+(1|FlankerItem:Participant)+(Congruence-1|FlankerItem:Participant)+(1|
    FlankerItem:FValence:Congruence)+ (1|BLType:BLSequence)', 'FitMethod', 'ML', '
    DummyVarCoding', 'Effects');

%% Add PartGender covariate and compare nested models, beginning with highest-order interaction
    , using LRT.
lmeGender = fitlme(hits, 'RTconv ~ PartGender + Group*FValence*Congruence + (1|Participant)+(1|
    Target)+(1|Target:Participant)+(1|FlankerItem:Participant)+(Congruence-1|FlankerItem:
    Participant)+(1|FlankerItem:FValence:Congruence)+ (1|BLType:BLSequence)', 'FitMethod', '
    REML', 'DummyVarCoding', 'effects');
lmeGender1 = fitlme(hits, 'RTconv ~ PartGender*FValence + Group*FValence*Congruence + (1|
    Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)+ (1|
    BLType:BLSequence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
lmeGender2 = fitlme(hits, 'RTconv ~ PartGender*Group*FValence*Congruence + (1|Participant:
    PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)+ (1|BLType:BLSequence
    )', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
genderLRT = compare(lmeGender2, lmeGender1, 'NSim', 100);
export(genderLRT,'XLSFile','S2E2genderLRTtable.xlsx');
%G1 better fit than G2. Simplify.
genderLRT = compare(lmeGender1, lmeGender, 'NSim', 100);
export(genderLRT,'XLSFile','S2E2genderLRTtable.xlsx', 'Sheet', 2);
%G1 better fit than G. Retain interaction.
genderLRT = compare(lmeGender, fullModel, 'NSim', 100);
export(genderLRT,'XLSFile','S2E2genderLRTtable.xlsx', 'Sheet', 3);
%G beter fit than null. Retain interaction.

%% Add face colour covariate and compare nested models, beginning with highest-order
    interaction, using LRT.
lmeLum = fitlme(hits, 'RTconv ~ PartGender+Contrast + Group*FValence*Congruence + (1|
    Participant)+(1|Target)+(1|Target:Participant)+(1|FlankerItem:Participant)+(Congruence-1|
    FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+ (1|BLType:BLSequence)', '
    FitMethod', 'REML', 'DummyVarCoding', 'effects');
lmeLum1 = fitlme(hits, 'RTconv ~ PartGender*FValence+Contrast*Congruence + Group*FValence*
    Congruence + (1|Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|
    Block)+ (1|BLType:BLSequence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
lmeLum2 = fitlme(hits, 'RTconv ~ PartGender*Contrast * Group*FValence*Congruence + (1|
    Participant:PartGender) + (1+Group|Target) + (1|FlankerItem:FValence) + (1|Block)+ (1|
    BLType:BLSequence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');

lumLRT = compare(lmeLum2, lmeLum1, 'NSim', 10);
export(lumLRT,'XLSFile','S2E2lumLRTtable.xlsx');
%G1 better fit than G2. Simplify.
lumLRT = compare(lmeLum1, lmeLum, 'NSim', 100);
export(lumLRT,'XLSFile','S2E2lumLRTtable.xlsx', 'Sheet', 2);
%G1 better fit than G. Retain interaction.
lumLRT = compare(lmeLum, fullModel, 'NSim', 100);
export(lumLRT,'XLSFile','S2E2lumLRTtable.xlsx', 'Sheet', 3);
%G beter fit than null. Retain interaction.


%% Obtain fixed effects estimates.
```

```
lmeDiffCoding = fitlme(hits, 'RTconv ~ PartGender*FValence + Group*FValence*Congruence + (1|Participant)+(1|Target)+(1|Target:Participant)+(1|FlankerItem:Participant)+(Congruence-1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+ (1|BLType:BLSequence)', 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
[~, ~, diffCodedFixedEffects] = fixedEffects(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %obtain conservative Satterthwaite approximations for DF/p value
export(diffCodedFixedEffects,'XLSFile','S2E2ModelEstimates.xlsx');
modelAnova = anova(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %check significance of terms
export(modelAnova,'XLSFile','S2E2ModelEstimates.xlsx', 'Sheet', 2);
%Change factor level order to define monolingual/congruent/neutral as intercept baseline
d.Congruence = reordercats(d.Congruence, {'Congruent', 'Incongruent'});
d.FValence = reordercats(d.FValence, {'Neutral','Sad','Happy'});
d.Group = reordercats(d.Group, {'Monolingual', 'Bilingual'});
hits = d(d.Answer==1, :);
%Refit model using dummy coding (default) for simple main effects.
lmeDummyCoding = fitlme(hits, 'RTconv ~ PartGender*FValence + Group*FValence*Congruence + (1|Participant)+(1|Target)+(1|Target:Participant)+(1|FlankerItem:Participant)+(Congruence-1|FlankerItem:Participant)+(1|FlankerItem:FValence:Congruence)+ (1|BLType:BLSequence)', 'FitMethod', 'REML');
[~, ~, dummyCodedFixedEffects] = fixedEffects(lmeDummyCoding, 'DFMethod', 'Satterthwaite');
export(dummyCodedFixedEffects,'XLSFile','S2E2ModelEstimates.xlsx','Sheet',3);

%% Conflict adaptation
FO
lmCA=fitlme(hits(hits.PrevAns==1,:), 'RTconv ~ PartGender+Group*Congruence*FValence*PrevCong*ValType + (1|Participant:Group)+(1|Target)', 'FitMethod', 'REML','DummyVarCoding','effects','CheckHessian',1);
[~, ~, CAFixedEffects] = fixedEffects(lmCA, 'DFMethod', 'Satterthwaite'); %obtain conservative Satterthwaite approximations for DF/p value
export(CAFixedEffects,'XLSFile','S2E2ModelEstimates.xlsx','Sheet', 4);
CAmodelAnova = anova(lmCA, 'DFMethod', 'Satterthwaite'); %check significance of terms
export(CAmodelAnova,'XLSFile','S2E2ModelEstimates.xlsx', 'Sheet', 5);
%Change factor level order to define monolingual/congruent/neutral as intercept baseline
d.Congruence = reordercats(d.Congruence, {'Congruent', 'Incongruent'});
d.FValence = reordercats(d.FValence, {'Happy','Sad','Neutral'});
d.PrevVal = reordercats(d.PrevVal, {'Happy','Sad','Neutral'});
%d.Group = reordercats(d.Group, {'Monolingual', 'Bilingual'});
hits = d(d.Answer==1, :);
%Refit model using dummy coding (default) for simple main effects.
lmCAdc=fitlme(hits, 'RTconv ~ Congruence*PrevCong*PrevVal + (1|Participant:Group)+(1|Target)', 'FitMethod', 'REML','CheckHessian',1);
[~, ~, CAdcFixedEffects] = fixedEffects(lmCAdc, 'DFMethod', 'Satterthwaite'); %obtain conservative Satterthwaite approximations for DF/p value
export(CAdcFixedEffects,'XLSFile','S2E2ModelEstimates.xlsx','Sheet', 6);

%%
%GLMM
FO

glmeDiffCoding = fitglme(d, 'Answer ~ PartGender+Group*FValence*Congruence + (1|Participant)',...
    'Distribution','Binomial','FitMethod', 'Laplace', 'Link', 'Logit','CheckHessian',true,'DummyVarCoding', 'effects');
[~, ~, diffCodedFixedEffectsACC] = fixedEffects(glmeDiffCoding);
export(diffCodedFixedEffectsACC,'XLSFile','S2E2ModelEstimatesAcc.xlsx');
glmemodelAnova = anova(glmeDiffCoding); %check significance of terms
%adjust p-values
glmemodelAnova.pValue(~(strcmp(glmemodelAnova.Term,'PartGender')...
    |strcmp(glmemodelAnova.Term,'(Intercept)')))...
    = pval_adjust(glmemodelAnova.pValue(~(strcmp(glmemodelAnova.Term,'PartGender')...
    |strcmp(glmemodelAnova.Term,'(Intercept)'))),'BH');
export(glmemodelAnova,'XLSFile','S2E2ModelEstimatesAcc.xlsx', 'Sheet', 2);
%Change factor level order to define monolingual/congruent/neutral as intercept baseline
d.Congruence = reordercats(d.Congruence, {'Congruent', 'Incongruent'});
d.FValence = reordercats(d.FValence, {'Happy','Neutral','Sad'});
d.Group = reordercats(d.Group, {'Monolingual', 'Bilingual'});
%Refit model using dummy coding (default) for simple main effects.
glmeDummyCoding = fitglme(d, 'Answer ~ PartGender+Group*FValence*Congruence + (1|Participant)',...
    'Distribution','Binomial','FitMethod', 'Laplace', 'Link', 'Logit');
[~, ~, dummyCodedFixedEffectsACC] = fixedEffects(glmeDummyCoding);
%apply p-value correction to terms being compared. Exclude gender
%(control), intercept, and Group:Congruence:FValence (meaningless in dummy coding
%as not comparing e.g. ML cong neutral to BL incong sad)
dummyCodedFixedEffectsACC.pValue(~(strcmp(dummyCodedFixedEffectsACC.Name,'PartGender')...
    |strcmp(dummyCodedFixedEffectsACC.Name,'(Intercept)')|...
    contains(dummyCodedFixedEffectsACC.Name,'lingual:Congruence_Incongruent:F')))...
    = pval_adjust(dummyCodedFixedEffectsACC.pValue(~(strcmp(dummyCodedFixedEffectsACC.Name,'PartGender')|...
    strcmp(dummyCodedFixedEffectsACC.Name,'(Intercept)')|...
    contains(dummyCodedFixedEffectsACC.Name,'lingual:Congruence_Incongruent:F'))),'BH');
export(dummyCodedFixedEffectsACC,'XLSFile','S2E2ModelEstimatesAcc.xlsx','Sheet',3);
```

B.2.2 Study 3

B.2.2.1 Experiment 1

The following script presents the model selection workflow for Study 3.1.

```
1
2 FNGschool
3 randEffects = {};
4 %% Test random intercepts
5
6 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
7 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)',...
8 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
9 1%
10 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
11 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)',...
12 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
13 2%
14 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
15 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Flanker)',...
16 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
17 3%
18 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
19 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Flanker:Participant)',...
20 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
21 3%
22 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
23 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Block)',...
24 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
25 4%
26 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
27 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|L1)',...
28 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
29 5%
30 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
31 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Language)',...
32 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
33 6%
34 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
35 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Type)',...
36 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
37 7%
38 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
39 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Sequence)',...
40 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
41 8%
42 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
43 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Class)',...
44 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
45 9%
46 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
47 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|School)',...
48 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
49
50
51 [lmeBest_step1, randEffects] = bestLme(randEffects);
52 sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
53 sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})
54
55 %%
56
57 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
58 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Block)+(1|Target:Participant)'
      ,...
59 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
60 1%
61 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
62 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Block)+(1|Flanker)',...
63 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
64 2%
65 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
66 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Block)+(1|Flanker:Participant)'
      ,...
67 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
68 3%
69 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
70 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Block)+(1|Target:Participant)
      +(1|Flanker:Participant)',...
71 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
72 4%
73
74 [lmeBest_step2, randEffects] = bestLme(randEffects);
75 sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
76 sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})
77 %% Add random slopes
78
79 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
80 'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Group|Target)+(1|Target:Participant)+(1|
      Block)',...
```

```
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
1%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1+Group|
    Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
1.5%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1|Target)+(1|Target:Participant)
    +(1|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
2%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1|Target:Participant)
    +(1|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
3%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+Congruence|Target:Participant)
    +(1|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
4%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1+
    Congruence|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
4.5%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+FValence|Participant)+(1|Target)+(1|Target:Participant)
    +(1|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
5%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+FValence|Target)+(1|Target:Participant)
    +(1|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
6%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+FValence|Target:Participant)
    +(1|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
7%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)+(1+FValence
    |Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
8%

[lmeBest_step3, randEffects] = bestLme(randEffects);
sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})

%% Combine random slopes

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|Target:
    Participant)+(1|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
1%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1+Congruence|Target:
    Participant)+(1|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
2%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1|Target:Participant)
    +(1+Congruence|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
2.25%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|Target:
    Participant)+(1+Congruence|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
2.5%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1+
    Congruence|Target:Participant)+(1+Congruence|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
2.75%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1|Target:Participant)
    +(1+Group|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
3%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|Target:
    Participant)+(1+Group|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
4%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|Target:
    Participant)+(1+Group+Congruence|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
```

```
5%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence+Group|Target)+(1+
    Congruence|Target:Participant)+(1+Group+Congruence|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
6%

[lmeBest_step4, randEffects] = bestLme(randEffects);
sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})



%%
selectedFNGkids = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|Target:
    Participant)+(1+Congruence|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
```

The following script presents the covariate testing and hypothesis testing workflow for Study 3.1.

```
FNGschool
%%
fullModel = fitlme(hits, ...
'RTconv~1+Group*Congruence+FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|Target:
    Participant)+(1|Block)',...
'FitMethod', 'ML','CheckHessian',1, 'DummyVarCoding', 'effects');
%% Add Gender covariate and compare nested models, beinning with highest-order interaction,
    using LRT.
lmeGender = fitlme(hits, ...
'RTconv~1+Gender+Group*Congruence+FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|
    Target:Participant)+(1|Block)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
lmeGender1 = fitlme(hits, ...
'RTconv~1+Gender*FValence+Group*Congruence+FValence+(1+Congruence|Participant)+(1+Congruence|
    Target)+(1|Target:Participant)+(1|Block)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
genderLRT = compare(lmeGender1, lmeGender, 'NSim', 100)
export(genderLRT,'XLSFile','FNGkids.xlsx','Sheet',1);
genderLRT = compare(lmeGender, fullModel, 'NSim', 100)
export(genderLRT,'XLSFile','FNGkids.xlsx','Sheet',2);
%Gender significant.
%% Age covariate
lmeAge = fitlme(hits, ...
'RTconv~1+Age+Gender+Group*Congruence+FValence+(1+Congruence|Participant)+(1+Congruence|Target)
    +(1|Target:Participant)+(1|Block)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
lmeAge1 = fitlme(hits, ...
'RTconv~1+Age*Congruence+Gender+Group*Congruence+FValence+(1+Congruence|Participant)+(1+
    Congruence|Target)+(1|Target:Participant)+(1|Block)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
ageLRT = compare(lmeAge1, lmeAge, 'NSim', 100);
export(ageLRT,'XLSFile','FNGkids.xlsx','Sheet',3);
ageLRT = compare(lmeAge, fullModel, 'NSim', 100);
export(ageLRT,'XLSFile','FNGkids.xlsx','Sheet',4);
%age not significant
%% Face colour covariate
lmeLum = fitlme(hits, ...
'RTconv~1+Contrast+Gender+Group*Congruence+FValence+(1+Congruence|Participant)+(1+Congruence|
    Target)+(1|Target:Participant)+(1|Block)',...
'FitMethod', 'REML', 'DummyVarCoding', 'Effects');
lmeLum1 = fitlme(hits, ...
'RTconv~1+Contrast*Congruence+Gender+Group*Congruence+FValence+(1+Congruence|Participant)+(1+
    Congruence|Target)+(1|Target:Participant)+(1|Block)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
ageLRT = compare(lmeLum1, lmeLum, 'NSim', 100);
export(ageLRT,'XLSFile','FNGkids.xlsx','Sheet',5);
ageLRT = compare(lmeLum, fullModel, 'NSim', 100);
export(ageLRT,'XLSFile','FNGkids.xlsx','Sheet',6);
%
%% Obtain fixed effects estimates.
lmeDiffCoding = fitlme(hits, ...
'RTconv~Gender+Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|
    Target:Participant)+(1|Block)',...
    'FitMethod', 'REML', 'DummyVarCoding', 'effects');
[~, ~, diffCodedFixedEffects] = fixedEffects(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %
    obtain conservative Satterthwaite approximations for DF/p value
export(diffCodedFixedEffects,'XLSFile','FNGkids.xlsx','Sheet',5);
modelAnova = anova(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %check significance of terms
export(modelAnova,'XLSFile','FNGkids.xlsx','Sheet',6);

%Change factor level order to define monolingual/congruent/neutral as intercept baseline
hits.Congruence = reordercats(hits.Congruence, {'Congruent', 'Incongruent'});
hits.FValence = reordercats(hits.FValence, {'Neutral', 'Happy', 'Sad'});
hits.Group = reordercats(hits.Group, {'Monolingual', 'Bilingual'});
%Refit model using dummy coding (default) for simple main effects.
lmeDummyCoding = fitlme(hits, ...
'RTconv~Gender+Group*Congruence*FValence+(1+Congruence|Participant)+(1+Congruence|Target)+(1|
    Target:Participant)+(1|Block)',...
```

```
    'FitMethod', 'REML');
[~, ~, dummyCodedFixedEffects] = fixedEffects(lmeDummyCoding, 'DFMethod', 'Satterthwaite');
export(dummyCodedFixedEffects,'XLSFile','FNGkids.xlsx','Sheet',7);



%%
%GLMM
FNGschool

glmeDiffCoding = fitglme(d, 'Answer ~ Gender+GoStatus*Group*FValence*Congruence-GoStatus:Group:
    FValence:Congruence + (1|Participant) + (1|Target)',...
    'Distribution','Binomial','FitMethod', 'Laplace', 'Link', 'Probit','CheckHessian',1,'
        DummyVarCoding', 'effects');
[~, ~, diffCodedFixedEffectsACC] = fixedEffects(glmeDiffCoding);
export(diffCodedFixedEffectsACC,'XLSFile','FNGkids.xlsx','Sheet',8);
glmemodelAnova = anova(glmeDiffCoding); %check significance of terms
%adjust p-values
% glmemodelAnova.pValue(~(strcmp(glmemodelAnova.Term,'Gender')...
%     |strcmp(glmemodelAnova.Term,'(Intercept)'))))...
%     = pval_adjust(glmemodelAnova.pValue(~(strcmp(glmemodelAnova.Term,'Gender')...
%     |strcmp(glmemodelAnova.Term,'(Intercept)'))),'BH');
export(glmemodelAnova,'XLSFile','FNGkids.xlsx','Sheet',9);
%Change factor level order to define monolingual/congruent/neutral as intercept baseline
d.Congruence = reordercats(d.Congruence, {'Congruent', 'Incongruent'});
d.FValence = reordercats(d.FValence, {'Happy','Neutral','Sad'});
d.Group = reordercats(d.Group, {'Monolingual', 'Bilingual'});
%Refit model using dummy coding (default) for simple main effects.
glmeDummyCoding = fitglme(d, 'Answer ~ Gender+GoStatus*Group*FValence*Congruence + (1|
    Participant) + (1|Target)',...
    'Distribution','Binomial','FitMethod', 'Laplace', 'Link', 'Probit');
[~, ~, dummyCodedFixedEffectsACC] = fixedEffects(glmeDummyCoding);
%apply p-value correction to terms being compared. Exclude gender
%(control), intercept, and Group:Congruence:FValence (meaningless in dummy coding
%as not comparing e.g. ML cong neutral to BL incong sad)
% dummyCodedFixedEffectsACC.pValue(~(strcmp(dummyCodedFixedEffectsACC.Name,'PartGender')...
%     |strcmp(dummyCodedFixedEffectsACC.Name,'(Intercept)')|...
%     contains(dummyCodedFixedEffectsACC.Name,'lingual:Congruence_Incongruent:F')))...
%     = pval_adjust(dummyCodedFixedEffectsACC.pValue(~(strcmp(dummyCodedFixedEffectsACC.Name,'
    PartGender')|...
%     strcmp(dummyCodedFixedEffectsACC.Name,'(Intercept)')|...
%     contains(dummyCodedFixedEffectsACC.Name,'lingual:Congruence_Incongruent:F'))),'holm');
export(dummyCodedFixedEffectsACC,'XLSFile','FNGkids.xlsx','Sheet',10);
```

B.2.2.2 Experiment 2

The following script presents the model selection workflow for Study 3.2.

```

FOschool
randEffects = {};
%% Test random intercepts

randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
1%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Target:Participant)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
2%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Flanker)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
3%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Flanker:Participant)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
3%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Block)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
4%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|L1)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
5%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Language)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
6%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Type)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
7%
randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Sequence)',...
'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
```

```
41 8%
42 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
43 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Age)',...
44 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
45 9%
46 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
47 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|School)',...
48 'FitMethod', 'REML','CheckHessian',1, 'DummyVarCoding', 'effects');
49
50 [lmeBest_step1, randEffects] = bestLme(randEffects);
51 sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
52 sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})
53
54 %%
55
56 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
57 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Flanker:Participant)+(1|Target:
       Participant)',...
58 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
59 1%
60
61 [lmeBest_step2, randEffects] = bestLme(randEffects);
62 sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
63 sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})
64 %% Add random slopes
65
66 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
67 'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Group|Target)+(1|Flanker:Participant)+(1|
       Target:Participant)',...
68 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
69 1%
70 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
71 'RTconv~Group*Congruence*FValence+(1+Congruence|Participant)+(1|Target)+(1|Flanker:Participant)
       +(1|Target:Participant)',...
72 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
73 2%
74 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
75 'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1|Flanker:Participant)
       +(1|Target:Participant)',...
76 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
77 3%
78 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
79 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+Congruence|Flanker:Participant)
       +(1|Target:Participant)',...
80 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
81 4%
82 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
83 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1|Flanker:Participant)+(1+
       Congruence|Target:Participant)',...
84 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
85 5%
86 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
87 'RTconv~Group*Congruence*FValence+(1+FValence|Participant)+(1|Target)+(1|Flanker:Participant)
       +(1|Target:Participant)',...
88 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
89 6%
90 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
91 'RTconv~Group*Congruence*FValence+(1|Participant)+(1+FValence|Target)+(1|Flanker:Participant)
       +(1|Target:Participant)',...
92 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
93 7%
94 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
95 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+FValence|Flanker:Participant)
       +(1|Target:Participant)',...
96 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
97 8%
98 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
99 'RTconv~Group*Congruence*FValence+(1+FValence|Participant)+(1|Target)+(1|Flanker:Participant)
       +(1+FValence|Target:Participant)',...
100 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
101 9%
102 [lmeBest_step3, randEffects] = bestLme(randEffects);
103 sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
104 sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})
105
106 %% Combine random slopes
107
108 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
109 'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1+Congruence|Flanker:
       Participant)+(1|Target:Participant)',...
110 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
111 3%
112 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
113 'RTconv~Group*Congruence*FValence+(1|Participant)+(1|Target)+(1+Congruence|Flanker:Participant)
       +(1+Congruence|Target:Participant)',...
114 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
115 4%
116 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
117 'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1|Flanker:Participant)
       +(1+Congruence|Target:Participant)',...
118 'FitMethod', 'REML', 'DummyVarCoding', 'effects');
119 5
120 randEffects{length(randEffects)+1, 1} = fitlme(hits, ...
```

```
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1+Congruence|Flanker:
    Participant)+(1+Congruence|Target:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'effects');

[lmeBest_step4, randEffects] = bestLme(randEffects);
sprintf('%s, (1)AIC %f', randEffects{1,3}, randEffects{1,2})
sprintf('%s, (2)AIC %f', randEffects{2,3}, randEffects{2,2})



%%
```

The following script presents the covariate testing and hypothesis testing workflow for Study 3.2.

```
FOschool %parse data
%%
fullModel = fitlme(hits, ...
'RTconv~Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1+Congruence|Flanker:
    Participant)+(1|Target:Participant)',...
'FitMethod', 'ML','CheckHessian',1, 'DummyVarCoding', 'effects');
%% Add Gender covariate and compare nested models, beinning with highest-order interaction,
    using LRT.
lmeGender = fitlme(hits, ...
'RTconv~1+Gender+Group*Congruence+FValence+(1|Participant)+(1+Congruence|Target)+(1+Congruence|
    Flanker:Participant)+(1|Target:Participant)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
lmeGender1 = fitlme(hits, ...
'RTconv~1+Gender*FValence+Group*Congruence+FValence+(1|Participant)+(1+Congruence|Target)+(1+
    Congruence|Flanker:Participant)+(1|Target:Participant)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
genderLRT = compare(lmeGender1, lmeGender, 'NSim', 100);
export(genderLRT,'XLSFile','FOkids.xlsx','Sheet',1);
genderLRT = compare(lmeGender, fullModel, 'NSim', 100);
export(genderLRT,'XLSFile','FOkids.xlsx','Sheet',2);
%Gender significant.
%% Age covariate
lmeAge = fitlme(hits, ...
'RTconv~1+Age+Gender+Group*Congruence+FValence+(1|Participant)+(1+Congruence|Target)+(1+
    Congruence|Flanker:Participant)+(1|Target:Participant)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
lmeAge1 = fitlme(hits, ...
'RTconv~1+Age*Congruence+Gender+Group*Congruence+FValence+(1|Participant)+(1+Congruence|Target)
    +(1+Congruence|Flanker:Participant)+(1|Target:Participant)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
ageLRT = compare(lmeAge1, lmeAge, 'NSim', 100);
export(ageLRT,'XLSFile','FOkids.xlsx','Sheet',3);
ageLRT = compare(lmeAge, fullModel, 'NSim', 100);
export(ageLRT,'XLSFile','FOkids.xlsx','Sheet',4);
%age not significant
%% Face colour covariate
lmeLum = fitlme(hits, ...
'RTconv~1+Contrast+Gender+Group*Congruence+FValence+(1|Participant)+(1+Congruence|Target)+(1+
    Congruence|Flanker:Participant)+(1|Target:Participant)',...
'FitMethod', 'REML', 'DummyVarCoding', 'Effects');
lmeLum1 = fitlme(hits, ...
'RTconv~1+Contrast*Congruence+Gender+Group*Congruence+FValence+(1|Participant)+(1+Congruence|
    Target)+(1+Congruence|Flanker:Participant)+(1|Target:Participant)',...
'FitMethod', 'ML', 'DummyVarCoding', 'Effects');
lumLRT = compare(lmeLum1, lmeLum, 'NSim', 100);
export(ageLRT,'XLSFile','FOkids.xlsx','Sheet',5);
ageLRT = compare(lmeLum, fullModel, 'NSim', 100);
export(ageLRT,'XLSFile','FOkids.xlsx','Sheet',6);
%
%% Obtain fixed effects estimates.
lmeDiffCoding = fitlme(hits, ...
'RTconv~Gender+Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1+Congruence|
    Flanker:Participant)+(1|Target:Participant)',...
    'FitMethod', 'REML', 'DummyVarCoding', 'effects');
[~, ~, diffCodedFixedEffects] = fixedEffects(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %
    obtain conservative Satterthwaite approximations for DF/p value
export(diffCodedFixedEffects,'XLSFile','FOkids.xlsx','Sheet',5);
modelAnova = anova(lmeDiffCoding, 'DFMethod', 'Satterthwaite'); %check significance of terms
export(modelAnova,'XLSFile','FOkids.xlsx','Sheet',6);

%Change factor level order to define monolingual/congruent/neutral as intercept baseline
hits.Congruence = reordercats(hits.Congruence, {'Congruent', 'Incongruent'});
hits.FValence = reordercats(hits.FValence, {'Neutral', 'Happy', 'Sad'});
hits.Group = reordercats(hits.Group, {'Monolingual', 'Bilingual'});
%Refit model using dummy coding (default) for simple main effects.
lmeDummyCoding = fitlme(hits, ...
'RTconv~Gender+Group*Congruence*FValence+(1|Participant)+(1+Congruence|Target)+(1+Congruence|
    Flanker:Participant)+(1|Target:Participant)',...
    'FitMethod', 'REML');
[~, ~, dummyCodedFixedEffects] = fixedEffects(lmeDummyCoding, 'DFMethod', 'Satterthwaite');
export(dummyCodedFixedEffects,'XLSFile','FOkids.xlsx','Sheet',7);


%%
%GLMM
```

```
FOschool

glmeDiffCoding = fitglme(d, 'Answer ~ Gender+Group*FValence*Congruence + (1|Participant) + (1|Target)',...
    'Distribution','Binomial','FitMethod', 'Laplace', 'Link', 'Probit','CheckHessian',1,'DummyVarCoding', 'effects');
[~, ~, diffCodedFixedEffectsACC] = fixedEffects(glmeDiffCoding);
export(diffCodedFixedEffectsACC,'XLSFile','FOkids.xlsx','Sheet',8);
glmemodelAnova = anova(glmeDiffCoding); %check significance of terms
%adjust p-values
% glmemodelAnova.pValue(~(strcmp(glmemodelAnova.Term,'PartGender')...
%     |strcmp(glmemodelAnova.Term,'(Intercept)')))...
%     = pval_adjust(glmemodelAnova.pValue(~(strcmp(glmemodelAnova.Term,'PartGender')...
%     |strcmp(glmemodelAnova.Term,'(Intercept)'))),'BH');
export(glmemodelAnova,'XLSFile','FOkids.xlsx','Sheet',9);
%Change factor level order to define monolingual/congruent/neutral as intercept baseline
d.Congruence = reordercats(d.Congruence, {'Congruent', 'Incongruent'});
d.FValence = reordercats(d.FValence, {'Happy','Neutral','Sad'});
d.Group = reordercats(d.Group, {'Monolingual', 'Bilingual'});
%Refit model using dummy coding (default) for simple main effects.
glmeDummyCoding = fitglme(d, 'Answer ~ Gender+Group*FValence*Congruence + (1|Participant) + (1|Target)',...
    'Distribution','Binomial','FitMethod', 'Laplace', 'Link', 'Probit');
[~, ~, dummyCodedFixedEffectsACC] = fixedEffects(glmeDummyCoding);
%apply p-value correction to terms being compared. Exclude gender
%(control), intercept, and Group:Congruence:FValence (meaningless in dummy coding
%as not comparing e.g. ML cong neutral to BL incong sad)
% dummyCodedFixedEffectsACC.pValue(~(strcmp(dummyCodedFixedEffectsACC.Name,'PartGender')...
%     |strcmp(dummyCodedFixedEffectsACC.Name,'(Intercept)')|...
%     contains(dummyCodedFixedEffectsACC.Name,'lingual:Congruence_Incongruent:F')))...
%     = pval_adjust(dummyCodedFixedEffectsACC.pValue(~(strcmp(dummyCodedFixedEffectsACC.Name,'PartGender')|...
%     strcmp(dummyCodedFixedEffectsACC.Name,'(Intercept)')|...
%     contains(dummyCodedFixedEffectsACC.Name,'lingual:Congruence_Incongruent:F'))),'holm');
export(dummyCodedFixedEffectsACC,'XLSFile','FOkids.xlsx','Sheet',10);
```

Alles Gescheite ist schon gedacht worden.
Man muss nur versuchen, es noch einmal zu denken.

All intelligent thoughts have already been thought;
one must only try to think them again.

— Johann Wolfgang von Goethe,
Wilhelm Meisters Wanderjahre

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta psychologica*, *128*(3), 466–478.

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., Cappa, S. F. & Costa, A. (2011). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral cortex*, *22*(9), 2076–2086.

Adesope, O. O., Lavin, T., Thompson, T. & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*(2), 207–245.

Albert, J., López-Martín, S. & Carretié, L. (2010). Emotional context modulates response inhibition: Neural and behavioral data. *Neuroimage*, *49*(1), 914–921.

Alguacil, S., Tudela, P. & Ruz, M. (2013). Cognitive and affective control in a flanker word task: Common and dissociable brain mechanisms. *Neuropsychologia*, *51*(9), 1663–1672.

Altarriba, J. & Canary, T. M. (2004). The influence of emotional arousal on affective priming in monolingual and bilingual speakers. *Journal of Multilingual and Multicultural Development*, *25*(2-3), 248–265.

Anzaldúa, G. (1987). How to tame a wild tongue. *Borderlands: La frontera*. Aunt Lute.

Ashbery, J. (1986). Soonest mended. *Selected poems: John ashbery*. Penguin.

Astheimer, L., Janus, M., Moreno, S. & Bialystok, E. (2014). Electrophysiological measures of attention during speech perception predict metalinguistic skills in children. *Developmental cognitive neuroscience*, *7*, 1–12.

Astle, D. E. & Scerif, G. (2009). Using developmental cognitive neuroscience to study behavioral and attentional control. *Developmental Psychobiology*, *51*(2), 107–118.

Ayneto, A. & Sebastián-Gallés, N. (2017). The influence of bilingualism on the preference for the mouth region of dynamic faces. *Developmental science*, *20*(1).

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390–412.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829–839.

Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617–645.

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in cognitive science*, *2*(4), 716–724.

Bates, D., Kliegl, R., Vasishth, S. & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Maechler, M., Bolker, B., Walker, S. et al. (2014). Lme4: Linear mixed-effects models using eigen and s4. *R package version*, *1*(7), 1–23.

Baumeister, S., Hohmann, S., Wolf, I., Plichta, M. M., Rechtsteiner, S., Zangl, M., Ruf, M., Holz, N., Boecker, R., Meyer-Lindenberg, A. et al. (2014). Sequential inhibitory control processes assessed through simultaneous eeg–fmri. *Neuroimage*, *94*, 349–359.

Beck, L., Kumschick, I. R., Eid, M. & Klann-Delius, G. (2012). Relationship between language competence and emotional competence in middle childhood. *Emotion*, *12*(3), 503.

Beirens, H., Hughes, N., Hek, R. & Spicer, N. (2007). Preventing social exclusion of refugee and asylum seeking children: Building new networks. *Social Policy and Society*, *6*(2), 219–229.

Benjamin, W. (1980). *Das kunstwerk im zeitalter seiner technischen reproduzierbarkeit: Drei studien zur kunstsoziologie*. Suhrkamp Verlag.

Berch, D. B., Krikorian, R. & Huha, E. M. (1998). The corsi block-tapping task: Methodological and theoretical considerations. *Brain and cognition*, *38*(3), 317–338.

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

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological bulletin*, *143*(3), 233.

Bialystok, E. & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*(1), 67–73.

Bialystok, E., Craik, F. I. & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in cognitive sciences*, *16*(4), 240–250.

Bialystok, E. & DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 565.

Bialystok, E. & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and language*, *109*(2), 93–100.

Bialystok, E., Hawrylewicz, K., Wiseheart, M. & Toplak, M. (2017). Interaction of bilingualism and attention-deficit/hyperactivity disorder in young adults. *Bilingualism: Language and Cognition*, *20*(3), 588–601.

Bialystok, E. & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, *112*(3), 494–500.

Bilingualism forum (Vol. 73). (2015).

Bloom, L. & Beckwith, R. (1989). Talking with feeling: Integrating affective and linguistic expression in early language development. *Cognition & Emotion*, *3*(4), 313–342.

Bloom, L., Tinker, E. & Scholnick, E. K. (2001). The intentionality model and language acquisition: Engagement, effort, and the essential tension in development. *Monographs of the Society for Research in Child Development*, i–101.

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

Botvinick, M. M., Cohen, J. D. & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in cognitive sciences*, *8*(12), 539–546.

Brainard, D. H. & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436.

Brito, N. & Noble, K. (2018). The independent and interacting effects of socioeconomic status and dual-language use on brain structure and cognition. *Developmental science*, e12688–e12688.

Brosch, T., Pourtois, G. & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cognition and Emotion*, *24*(3), 377–400.

Brydges, C. R., Anderson, M., Reid, C. L. & Fox, A. M. (2013). Maturation of cognitive control: Delineating response inhibition and interference suppression. *PloS one*, *8*(7), e69826.

Brydges, C. R., Clunies-Ross, K., Clohessy, M., Lo, Z. L., Nguyen, A., Rousset, C., Whitelaw, P., Yeap, Y. J. & Fox, A. M. (2012). Dissociable components of cognitive control: An event-related potential (erp) study of response inhibition and interference suppression. *PLoS One*, *7*(3), e34482.

Brydges, C. R., Reid, C. L., Fox, A. M. & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence*, *40*(5), 458–469.

Brysbaert, M. (2013). Lextale_fr a fast, free, and efficient test to measure language proficiency in french. *Psychologica Belgica*, *53*(1), 23–37.

Bull, R. & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental neuropsychology*, *19*(3), 273–293.

Burnham, K. P. & Anderson, D. R. (2003). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer Science & Business Media.

Bush, G., Luu, P. & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences*, *4*(6), 215–222.

Caldwell-Harris, C. L. & Ayçiçegi-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, *71*(3), 193–204.

Casey, B., Getz, S. & Galvan, A. (2008). The adolescent brain. *Developmental Review*, *28*(1), 62–77.

Checa, P., Rodríguez-Bailón, R. & Rueda, M. R. (2008). Neurocognitive and temperamental systems of self-regulation and early adolescents' social and academic outcomes. *Mind, Brain, and Education*, *2*(4), 177–187.

Checa, P. & Rueda, M. R. (2011). Behavioral and brain measures of executive attention and school competence in late childhood. *Developmental Neuropsychology*, *36*(8), 1018–1032.

Chwilla, D. J., Virgillito, D. & Vissers, C. T. W. (2011). The relationship of language and emotion: N400 support for an embodied view of language comprehension. *Journal of cognitive neuroscience*, *23*(9), 2400–2414.

Colzato, L. S., Bajo, M. T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W. & Hommel, B. (2008). How does bilingualism improve executive control? a comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 302.

Conan Doyle, A. (1891). A scandal in bohemia. *The Strand Magazine*, *2*(7), 61–75.

Conboy, B. T., Brooks, R., Meltzoff, A. N. & Kuhl, P. K. (2015). Social interaction in infants' learning of second-language phonetics: An exploration of brain–behavior relations. *Developmental neuropsychology*, *40*(4), 216–229.

Conboy, B. T., Sommerville, J. A. & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental psychology*, *44*(5), 1505.

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J. & Keysar, B. (2014). Your morals depend on language. *PloS one*, *9*(4), e94842.

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., Hernández, M. & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ant task. *Cognition*, *106*(1), 59–86.

Costumero, V., Rodríguez-Pujadas, A., Fuentes-Claramonte, P. & Ávila, C. (2015). How bilingualism shapes the functional architecture of the brain: A study on executive control in early bilinguals and monolinguals. *Human brain mapping*, *36*(12), 5101–5112.

De Zulueta, F. (1984). The implications of bilingualism in the study and treatment of psychiatric disorders: A review. *Psychological Medicine*, *14*(3), 541–557.

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

Decety, J., Michalska, K. J. & Kinzler, K. D. (2011). The contribution of emotion and cognition to moral sensitivity: A neurodevelopmental study. *Cerebral cortex*, *22*(1), 209–220.

de Saint-Exupéry, A. (1946). *Le petit prince*. Gallimard.

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of multilingual and multicultural development*, *25*(2-3), 204–222.

Dewaele, J.-M. (2008). The emotional weight of i love you in multilinguals' languages. *Journal of Pragmatics*, *40*(10), 1753–1780.

Diamond, A. & Doar, B. (1989). The performance of human infants on a measure of frontal cortex function, the delayed response task. *Developmental Psychobiology*, *22*(3), 271–294.

Dijkstra, T. & van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(3), 175–197.

Duckworth, A. L. & Seligman, M. E. (2005). Self-discipline outdoes iq in predicting academic performance of adolescents. *Psychological science*, *16*(12), 939–944.

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(4), 380–390.

Egner, T. & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature neuroscience*, *8*(12), 1784–1790.

Eriksen, B. A. & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention, Perception, & Psychophysics*, *16*(1), 143–149.

Etkin, A., Egner, T. & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in cognitive sciences*, *15*(2), 85–93.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A. & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of cognitive neuroscience*, *14*(3), 340–347.

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*(2), 175–191.

Fenske, M. J. & Eastwood, J. D. (2003). Modulation of focused attention by faces expressing emotion: Evidence from flanker tasks. *Emotion*, *3*(4), 327.

Fernandes, M. A. & Moscovitch, M. (2002). Factors modulating the effect of divided attention during retrieval of words. *Memory & Cognition*, *30*(5), 731–744.

Fernandez, M., Tartar, J. L., Padron, D. & Acosta, J. (2013). Neurophysiological marker of inhibition distinguishes language groups on a non-linguistic executive function test. *Brain and cognition*, *83*(3), 330–336.

Festman, J. & Münte, T. F. (2012). Cognitive control in russian–german bilinguals. *Frontiers in Psychology*, *3*, 115.

Fox, E., Yates, A. & Ashwin, C. (2012). Trait anxiety and perceptual load as determinants of emotion processing in a fear conditioning paradigm. *Emotion*, *12*(2), 236.

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1449), 1367.

Fricke, M., Zirnstein, M., Navarro-Torres, C. & Kroll, J. F. (2018). Bilingualism reveals fundamental variation in language processing. *Bilingualism: Language and Cognition*, 1–8.

Friedman, N. P. (2016). Research on individual differences in executive functions. *Linguistic approaches to bilingualism*, *6*(5), 535–548.

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.

Frischen, A., Eastwood, J. D. & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological bulletin*, *134*(5), 662.

Funes, M. J., Lupiáñez, J. & Humphreys, G. (2010). Analyzing the generality of conflict adaptation effects. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 147.

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodríguez-Pujadas, A., Belloch, V., Hernández, M., Costa, A. & Ávila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *Neuroimage*, *53*(4), 1272–1278.

García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J. A. & Carreiras, M. (2016a). ‘‘hazy” or “jumbled”? putting together the pieces of the bilingual puzzle. *Language, Cognition and Neuroscience*, *31*(3), 353–360.

García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J. A. & Carreiras, M. (2016b). The neuroanatomy of bilingualism: How to turn a hazy view into the full picture. *Language, Cognition and Neuroscience*, *31*(3), 303–327.

Gathercole, V. C. M. & Moawad, R. A. (2010). Semantic interaction in early and late bilinguals: All words are not created equally. *Bilingualism: Language and Cognition*, *13*(4), 385–408.

Geipel, J., Hadjichristidis, C. & Surian, L. (2015). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, *59*, 8–17.

Geipel, J., Hadjichristidis, C. & Surian, L. (2016). Foreign language affects the contribution of intentions and outcomes to moral judgment. *Cognition*, *154*, 34–39.

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevelhierarchical models* (Vol. 1). Cambridge University Press New York, NY, USA.

Goldstein, M., Brendel, G., Tuescher, O., Pan, H., Epstein, J., Beutel, M., Yang, Y., Thomas, K., Levy, K., Silverman, M. et al. (2007). Neural substrates of the

interaction of emotional stimulus processing and motor inhibitory control: An emotional linguistic go/no-go fMRIstudy. *Neuroimage*, *36*(3), 1026–1040.

Gollek, C. & Doherty, M. J. (2016). Metacognitive developments in word learning: Mutual exclusivity and theory of mind. *Journal of experimental child psychology*, *148*, 51–69.

Granger, L. A. (2018). Encouraging the use and activation of heritage languages in the broader educational system. *Handbook of Research and Practice in Heritage Language Education*, 333–354.

Gratton, G., Coles, M. G. & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480.

Gray, J. R. (2004). Integration of emotion and cognitive control. *Current Directions in Psychological Science*, *13*(2), 46–48.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and language*, *27*(2), 210–223.

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

Grose-Fifer, J., Rodrigues, A., Hoover, S. & Zottoli, T. (2013). Attentional capture by emotional faces in adolescence. *Advances in cognitive psychology*, *9*(2), 81.

Grundy, J. G., Chung-Fat-Yim, A., Friesen, D. C., Mak, L. & Bialystok, E. (2017). Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults. *Cognition*, *163*, 42–55.

Guzmán-Veléz, E. & Tranel, D. (2015). Does bilingualism contribute to cognitive reserve? cognitive and neural perspectives. *Neuropsychology*, *29*(1), 139.

Hare, T. A., Tottenham, N., Davidson, M. C., Glover, G. H. & Casey, B. (2005). Contributions of amygdala and striatal activity in emotion regulation. *Biological psychiatry*, *57*(6), 624–632.

Hare, T. A., Tottenham, N., Galván, A., Voss, H. U., Glover, G. H. & Casey, B. (2008). Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. *Biological psychiatry*, *63*(10), 927–934.

Harman, C., Rothbart, M. K. & Posner, M. I. (1997). Distress and attention interactions in early infancy. *Motivation and Emotion*, *21*(1), 27–43.

Hayakawa, S., Costa, A., Foucart, A. & Keysar, B. (2016). Using a foreign language changes our choices. *Trends in cognitive sciences*, *20*(11), 791–793.

Hernández, M., Costa, A. & Humphreys, G. W. (2012). Escaping capture: Bilingualism modulates distraction from working memory. *Cognition*, *122*(1), 37–50.

Hernández, M., Martin, C. D., Barceló, F. & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language*, *69*(3), 257–276.

Hobson, N. M., Saunders, B., Al-Khindi, T. & Inzlicht, M. (2014). Emotion down-regulation diminishes cognitive control: A neurophysiological investigation. *Emotion*, *14*(6), 1014.

Hsu, C.-T., Jacobs, A. M. & Conrad, M. (2015). Can harry potter still put a spell on us in a second language? an fMRIstudy on reading emotion-laden literature in late bilinguals. *Cortex*, *63*, 282–295.

Hwang, S., White, S. F., Nolan, Z. T., Sinclair, S. & Blair, R. (2014). Neurodevelopmental changes in the responsiveness of systems involved in top down attention and emotional responding. *Neuropsychologia*, *62*, 277–285.

Inzlicht, M., Bartholow, B. D. & Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends in cognitive sciences*, *19*(3), 126–132.

Izura, C., Cuetos, F. & Brysbaert, M. (2014). Lextale-esp: A test to rapidly and efficiently assess the spanish vocabulary size. *Psicológica*, *35*(1), 49–66.

Kanske, P. & Kotz, S. A. (2010a). Emotion speeds up conflict resolution: A new role for the ventral anterior cingulate cortex? *Cerebral Cortex*, *21*(4), 911–919.

Kanske, P. & Kotz, S. A. (2010b). Modulation of early conflict processing: N200 responses to emotional words in a flanker task. *Neuropsychologia*, *48*(12), 3661–3664.

Kanske, P. & Kotz, S. A. (2011a). Conflict processing is modulated by positive emotion: ERPdata from a flanker task. *Behavioural brain research*, *219*(2), 382–386.

Kanske, P. & Kotz, S. A. (2011b). Emotion triggers executive attention: Anterior cingulate cortex and amygdala responses to emotional words in a conflict task. *Human brain mapping*, *32*(2), 198–208.

Kanske, P. & Kotz, S. A. (2012). Effortful control, depression, and anxiety correlate with the influence of emotion on executive attentional control. *Biological psychology*, *91*(1), 88–95.

Kanske, P., Plitschka, J. & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and n400 ERPeffects. *Neuropsychologia*, *49*(11), 3121–3129.

Keightley, M. L., Winocur, G., Burianova, H., Hongwanishkul, D. & Grady, C. L. (2006). Age effects on social cognition: Faces tell a different story. *Psychology and aging*, *21*(3), 558.

Kerns, G. J. (2011). *Power and sample size for repeated measures anova with r*. https://www.r-bloggers.com/power-and-sample-size-for-repeated-measures-anova-with-r/

Keysar, B., Hayakawa, S. L. & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological science*, *23*(6), 661–668.

Klein, D., Zatorre, R. J., Chen, J.-K., Milner, B., Crane, J., Belin, P. & Bouffard, M. (2006). Bilingual brain organization: A functional magnetic resonance adaptation study. *Neuroimage*, *31*(1), 366–375.

Klein, R. M. (2015a). Is there a benefit of bilingualism for executive functioning? *Bilingualism: Language and Cognition*, *18*(1), 29–31.

Klein, R. M. (2015b). On the belief that the cognitive exercise associated with the acquisition of a second language enhances extra-linguistic cognitive functions: Is "type-i incompetence" at work here. *Cortex*, *73*, 340–341.

Klein, R. M. (2016). What cognitive processes are likely to be exercised by bilingualism and does this exercise lead to extra-linguistic cognitive benefits? *Linguistic Approaches to Bilingualism*, *6*(5), 549–564.

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., Broussard, C. et al. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1.

Kousta, S.-T., Vinson, D. P. & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481.

Kroll, J. F. & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497–514.

Kroll, J. F., Bobb, S. C. & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science*, *23*(3), 159–163.

Kroll, J. F., Bobb, S. C., Misra, M. & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta psychologica*, *128*(3), 416–430.

Kroll, J. F. & Chiarello, C. (2016). Language experience and the brain: Variability, neuroplasticity, and bilingualism. *Language, Cognition and Neuroscience*, *31*(3), 345–348.

Kroll, J. F. & McClain, R. (2013). What bilinguals tell us about culture, cognition, and language. *Proceedings of the National Academy of Sciences*, *110*(28), 11219–11220.

Lang, P. J., Bradley, M. M., Cuthbert, B. N. et al. (1997). Motivated attention: Affect, activation, and action. *Attention and orienting: Sensory and motivational processes*, *97*, 135.

Larsen, R. J. & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.). Thousand Oaks, CA: Sage Publications.

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.

Lemhöfer, K. & Broersma, M. (2012). Introducing lextale: A quick and valid lexical test for advanced learners of english. *Behavior Research Methods*, *44*(2), 325–343.

Lewis, P. A., Critchley, H., Rotshtein, P. & Dolan, R. (2006). Neural correlates of processing valence and arousal in affective words. *Cerebral cortex*, *17*(3), 742–748.

Li, P., Zhang, F., Tsai, E. & Puls, B. (2014). Language history questionnaire (lhq 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*(3), 673–680.

Liberman, Z., Woodward, A. L., Keysar, B. & Kinzler, K. D. (2017). Exposure to multiple languages enhances communication skills in infancy. *Developmental science*, *20*(1), 1–11.

Lim, S.-L., Padmala, S. & Pessoa, L. (2008). Affective learning modulates spatial competition during low-load attentional conditions. *Neuropsychologia*, *46*(5), 1267–1278.

Lindquist, K. A., Satpute, A. B. & Gendron, M. (2015). Does language do more than communicate emotion? *Current directions in psychological science*, *24*(2), 99–108.

Luk, G. & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in r. *Behavior Research Methods*, *49*(4), 1494–1502.

MacDonald, A. W., Cohen, J. D., Stenger, V. A. & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*(5472), 1835–1838.

Macmillan, N. A. & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Markus, H. R. & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, *98*(2), 224.

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.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McCullagh, P. & Nelder, J. (1989). *Generalized linear models, second edition*. Taylor & Francis. https://books.google.co.uk/books?id=h9kFH2%5C_FfBkC

McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J. & Gross, J. J. (2008). Gender differences in emotion regulation: An fMRIstudy of cognitive reappraisal. *Group processes & intergroup relations*, *11*(2), 143–162.

Meara, P. & Miralpeix, I. (2015). *Tools for vocabulary research* (tech. rep.). Bristol: Multilingual Matters.

Mercier, J., Pivneva, I. & Titone, D. (2014). Individual differences in inhibitory control relate to bilingual spoken word processing. *Bilingualism: Language and Cognition*, *17*(1), 89–117.

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.

Mitchell, D., Nakic, M., Fridberg, D., Kamel, N., Pine, D. & Blair, R. (2007). The impact of processing load on emotion. *Neuroimage*, *34*(3), 1299–1309.

Mitchell, R. L. & Phillips, L. H. (2007). The psychological, neurochemical and functional neuroanatomical mediators of the effects of positive and negative mood on executive functions. *Neuropsychologia*, *45*(4), 617–629.

Miyake, A. & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current directions in psychological science*, *21*(1), 8–14.

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.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S. et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698.

Morales, J., Calvo, A. & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of experimental child psychology*, *114*(2), 187–202.

Moreno, S., Bialystok, E., Wodniecka, Z. & Alain, C. (2010). Conflict resolution in sentence processing by bilinguals. *Journal of Neurolinguistics*, *23*(6), 564–579.

Moreno, S., Wodniecka, Z., Tays, W., Alain, C. & Bialystok, E. (2014). Inhibitory control in bilinguals and musicians: Event related potential (erp) evidence for experience-specific effects. *PloS one*, *9*(4), e94169.

Mukadam, N., Jichi, F., Green, D. & Livingston, G. (2018). The relationship of bilingualism to cognitive decline: The australian longitudinal study of ageing. *International journal of geriatric psychiatry*, *33*(2), e249–e256.

Mukadam, N., Sommerlad, A. & Livingston, G. (2017). The relationship of bilingualism compared to monolingualism to the risk of cognitive decline or dementia: A systematic review and meta-analysis. *Journal of Alzheimer's Disease*, *58*(1), 45–54.

Mulder, H., Verhagen, J., Van der Ven, S. H., Slot, P. L. & Leseman, P. P. (2017). Early executive function at age two predicts emergent mathematics and literacy at age five. *Frontiers in psychology*, *8*, 1706.

Myles, F. (2017). Learning foreign languages in primary schools: Is younger better? *Languages, Society & Policy*.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological bulletin*, *126*(2), 220.

Norman, D. A. & Shallice, T. (1986). Attention to action. *Consciousness and self-regulation* (pp. 1–18). Springer.

Ochsner, K. N. & Gross, J. J. (2005). The cognitive control of emotion. *Trends in cognitive sciences*, *9*(5), 242–249.

Ochsner, K. N., Silvers, J. A. & Buhle, J. T. (2012). Functional imaging studies of emotion regulation: A synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*, *1251*(1), E1–E24.

Opitz, B. & Degner, J. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia*, *50*(8), 1961–1967.

Oquendo, M. A. (1996). Psychiatric evaluation and psychotherapy in the patient's second language. *Psychiatric services*, *47*(6), 614–618.

Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology*, *26*(3), 242–255.

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.

Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R. & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, *29*(2), 89–112.

Paap, K. R. & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in psychology*, *5*, 962.

Paradis, M. (2008). Bilingualism and neuropsychiatric disorders. *Journal of Neurolinguistics*, *21*(3), 199–230.

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and cognition*, *11*(2), 147–164.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. *The Bilingual Mental Lexicon: Interdisciplinary Approaches*, *70*, 125.

Payne, B. K. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of personality and social psychology*, *89*(4), 488.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*(4), 437–442.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature reviews neuroscience*, *9*(2), 148–158.

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in cognitive sciences*, *13*(4), 160–166.

Pessoa, L. (2010). Emotion and cognition and the amygdala: From "what is it?" to "what's to be done?" *Neuropsychologia*, *48*(12), 3416–3429.

Pessoa, L. (2013). *The cognitive-emotional brain: From interactions to integration*. MIT press.

Pessoa, L., McKenna, M., Gutierrez, E. & Ungerleider, L. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, *99*(17), 11458–11463.

Poarch, G. J. & Bialystok, E. (2017). Assessing the implications of migrant multilingualism for language education. *Zeitschrift für Erziehungswissenschaft*, *20*(2), 175–191.

Ponari, M., Rodríguez-Cuadrado, S., Vinson, D., Fox, N., Costa, A. & Vigliocco, G. (2015). Processing advantage for emotional words in bilingual speakers. *Emotion*, *15*(5), 644.

Posner, M. I. & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*(1), 25–42.

Posner, M. I. & Rothbart, M. K. (1998). Attention, self–regulation and consciousness. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *353*(1377), 1915–1927.

Potthoff, S., Garnefski, N., Miklósi, M., Ubbiali, A., Domínguez-Sánchez, F. J., Martins, E. C., Witthöft, M. & Kraaij, V. (2016). Cognitive emotion regulation and psychopathology across cultures: A comparison between six european countries. *Personality and Individual Differences*, *98*, 218–224.

Poulin-Dubois, D., Blaye, A., Coutya, J. & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of experimental child psychology*, *108*(3), 567–579.

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Ren, Y., Wyver, S., Xu Rattanasone, N. & Demuth, K. (2016). Social competence and language skills in mandarin–english bilingual preschoolers: The moderation effect of emotion regulation. *Early Education and Development*, *27*(3), 303–317.

Rich, A. (1993). *Adrienne rich's poetry and prose: Poems, prose, reviews, and criticism* (B. Gelpi & A. Gelpi, Eds.). W.W. Norton.

Rodríguez-Pujadas, A., Sanjuán, A., Fuentes, P., Ventura-Campos, N., Barrós-Loscertales, A. & Ávila, C. (2014). Differential neural control in early bilinguals and monolinguals during response inhibition. *Brain and language*, *132*, 43–51.

Rothbart, M. K. (1989). Temperament in childhood: A framework. *Temperament in childhood*, *59–73*.

Rueda, M. R., Checa, P. & Combita, L. M. (2012). Enhanced efficiency of the executive attention network after training in preschool children: Immediate changes and effects after two months. *Developmental cognitive neuroscience*, *2*, 192–204.

Rueda, M. R., Checa, P. & Rothbart, M. K. (2010). Contributions of attentional control to socioemotional and academic development. *Early Education and Development*, *21*(5), 744–764.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P. & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029–1040.

Rueda, M. R., Posner, M. I. & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental neuropsychology*, *28*(2), 573–594.

Sandoval, T. C., Gollan, T. H., Ferreira, V. S. & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? the dual-task analogy. *Bilingualism: Language and Cognition*, *13*(2), 231–252.

Schmeichel, B. J. & Tang, D. (2014). The relationship between individual differences in executive functioning and emotion regulation: A comprehensive review. *The control within: Motivation and its regulation*, 133–151.

Schwarz, N. & Clore, G. L. (1996). Feelings and phenomenal experiences. *Social psychology: Handbook of basic principles*, *2*, 385–407.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J. & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience*, *12*(3), 154–167.

Simon, J. R. & Berbaum, K. (1990). Effect of conflicting cues on information processing: The 'stroop effect'vs. the 'simon effect'. *Acta psychologica*, *73*(2), 159–170.

Sokol-Hessner, P., Camerer, C. F. & Phelps, E. A. (2012). Emotion regulation reduces loss aversion and decreases amygdala responses to losses. *Social cognitive and affective neuroscience*, *8*(3), 341–350.

Somerville, L. H., Hare, T. & Casey, B. (2011). Frontostriatal maturation predicts cognitive control failure to appetitive cues in adolescents. *Journal of cognitive neuroscience*, *23*(9), 2123–2134.

Somerville, L. H., Jones, R. M. & Casey, B. (2010). A time of change: Behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain and cognition*, *72*(1), 124–133.

Somerville, L. H., Kim, H., Johnstone, T., Alexander, A. L. & Whalen, P. J. (2004). Human amygdala responses during presentation of happy and neutral faces: Correlations with state anxiety. *Biological psychiatry*, *55*(9), 897–903.

Sorge, G. B., Toplak, M. E. & Bialystok, E. (2017). Interactions between levels of attention ability and levels of bilingualism in children's executive functioning. *Developmental science*, *20*(1), 1–16.

Spalek, K., Damian, M. F. & Bölte, J. (2013). Is lexical selection in spoken word production competitive? introduction to the special issue on lexical competition in language production. *Language and Cognitive Processes*, *28*(5), 597–614.

Stanislaw, H. & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, and computers*, *31*(1), 137–149.

Stocco, A., Lebiere, C. & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological review*, *117*(2), 541.

Swanson, H. L., Orosco, M. J. & Kudo, M. (2017). Does growth in the executive system of working memory underlie growth in literacy for bilingual children with and without reading disabilities? *Journal of learning disabilities*, *50*(4), 386–407.

Tillman, C. M. & Wiens, S. (2011). Behavioral and ERPindices of response conflict in stroop and flanker tasks. *Psychophysiology*, *48*(10), 1405–1411.

Tottenham, N., Hare, T. A. & Casey, B. (2011). Behavioral assessment of emotion discrimination, emotion regulation, and cognitive control in childhood, adolescence, and adulthood. *Frontiers in psychology*, *2*, 39.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. & Nelson, C. (2009). The nimstim set of facial expressions: Judgments from untrained research participants. *Psychiatry research*, *168*(3), 242–249.

Van Steenbergen, H., Band, G. P. & Hommel, B. (2010). In the mood for adaptation: How affect regulates conflict-driven control. *Psychological Science*, *21*(11), 1629–1634.

van Heuven, W. & Coderre, E. L. (2015). A call for sophisticated statistical approaches and neuroimaging techniques to study the bilingual advantage. *Cortex; a journal devoted to the study of the nervous system and behavior*, *73*, 330.

Verhees, M. W., Chwilla, D. J., Tromp, J. & Vissers, C. T. (2015). Contributions of emotional state and attention to the processing of syntactic agreement errors: Evidence from p600. *Frontiers in psychology*, *6*, 388.

von Goethe, J. W. (1829). *Wilhelm Meisters Wanderjahre oder die Entsagenden*. Cotta.

Vuilleumier, P., Armony, J. L., Driver, J. & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRIstudy. *Neuron*, *30*(3), 829–841.

Vuillier, L., Bryce, D., Szücs, D. & Whitebread, D. (2016). The maturation of interference suppression and response inhibition: ERPanalysis of a cued go/nogo task. *PloS one*, *11*(11), e0165697.

Weinberg, A. & Hajcak, G. (2010). Beyond good and evil: The time-course of neural activity elicited by specific picture content. *Emotion*, *10*(6), 767.

Westfall, J. (2015). Pangea: Power analysis for general anova designs. *Unpublished manuscript. Available at http://jakewestfall.org/publications/pangea.pdf*.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current directions in psychological science*, *7*(6), 177–188.

White, L. J. & Greenfield, D. B. (2017). Executive functioning in spanish-and english-speaking head start preschoolers. *Developmental science*, *20*(1), 1–14.

Wilde, O. (2003). *The picture of dorian gray*. Penguin.

Wolf, U. (2009). L. *Falsche freunde: Gedichte*. kookbooks.

Wolfe, C. D. & Bell, M. A. (2007a). The integration of cognition and emotion during infancy and early childhood: Regulatory processes associated with the development of working memory. *Brain and Cognition*, *65*(1), 3–13.

Wolfe, C. D. & Bell, M. A. (2007b). Sources of variability in working memory in early childhood: A consideration of age, temperament, language, and brain electrical activity. *Cognitive Development*, *22*(4), 431–455.

Yang, S. & Yang, H. (2016). Bilingual effects on deployment of the attention system in linguistically and culturally homogeneous children and adults. *Journal of experimental child psychology*, *146*, 121–136.

Ye, Z. & Zhou, X. (2009a). Conflict control during sentence comprehension: fMRIevidence. *Neuroimage*, *48*(1), 280–290.

Ye, Z. & Zhou, X. (2009b). Executive control in language processing. *Neuroscience & Biobehavioral Reviews*, *33*(8), 1168–1177.

Yeung, N., Botvinick, M. M. & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological review*, *111*(4), 931.

Yow, W. Q., Li, X., Lam, S., Gliga, T., Chong, Y. S., Kwek, K. & Broekman, B. F. (2017). A bilingual advantage in 54-month-olds' use of referential cues in fast mapping. *Developmental science*, *20*(1), e12482.

Yow, W. Q. & Markman, E. M. (2011). Bilingualism and children's use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition*, *14*(4), 562–569.

Yu, H.-T. (2015). Applying linear mixed effects models with crossed random effects to psycholinguistic data: Multilevel specification and model selection. *The Quantitative Methods for Psychology*, *11*(2), 78–88.

Zahodne, L. B., Schofield, P. W., Farrell, M. T., Stern, Y. & Manly, J. J. (2014). Bilingualism does not alter cognitive decline or dementia risk among spanish-speaking immigrants. *Neuropsychology*, *28*(2), 238.

Zamorano, F., Billeke, P., Hurtado, J. M., López, V., Carrasco, X., Ossandón, T. & Aboitiz, F. (2014). Temporal constraints of behavioral inhibition: Relevance of inter-stimulus interval in a go-nogo task. *PloS one*, *9*(1), e87232.

Zantout, M. & Murphy, V. (2018). *Investigating multilingual adolescents' cognitive and code-switching performance: An exploratory application of language control models* [Child Language Symposium] [Presentation]. https://indico.conference4me.psnc.pl/event/166/session/13/contribution/51

Zhang, W. & Lu, J. (2012). Time course of automatic emotion regulation during a facial go/nogo task. *Biological psychology*, *89*(2), 444–449.

Zhou, P. & Liu, X. (2013). Attentional modulation of emotional conflict processing with flanker tasks. *PloS one*, *8*(3), e60548.