# The Effects of Language Use on Responses in English/East Asian Bilingual Individuals: A Meta-Analysis

by

Jordan Schriver

Bachelor of Arts (Honours, Psychology), University of New Brunswick

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy

in the Graduate Academic Unit of Psychology

**Supervisor:** W.Q. Elaine Perunovic, PhD, Psychology

**Examining Board:** Biljana Stevanovski, PhD, Psychology
Veronica Whitford, PhD, Psychology
Guohua Yan, PhD, Mathematics and Statistics

**External Examiner:** Rui Zhang, PhD, Psychology, Dickinson College

This dissertation is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

October, 2024

©Jordan Schriver, 2024

## ABSTRACT

According to the linguistic relativity hypothesis, language use can influence the way individuals interact with the world (Whorf, 1956). Compelling evidence for this claim comes from research on bilingual individuals. Specifically, experimental studies in which bilingual individuals have been randomly assigned to communicate in one language or another have shown that language use can affect the way that bilingual individuals think (e.g., Lechuga & Wiebe, 2011), feel (e.g., Alvarado & Jameson, 2011), and behave (e.g., So, 2010). An attempt has not yet been made to integrate these findings, and establish a unified understanding of how the language used by bilingual language users affects their behaviours. Therefore, with this dissertation I addressed this gap by conducting a systematic review and meta-analysis, wherein I attempted to gather all available published and unpublished research up until 2022 in which bilingual individuals proficient in English and an East Asian language were assigned to complete a culturally relevant task in either language, and in which outcomes from the task in each language were compared. Through multilevel meta-analysis including 269 effect sizes drawn from 63 independent samples reflecting the responses of 6693 participants, I found an overall significant effect ($g$ = .27), indicating that when language is manipulated, responses are generally consistent with the cultural norms associated with the language being spoken. The magnitude of this language effect, however, was moderated by the East Asian language being used, and the Task Type under examination. The language effect was not moderated by the Design of the study, Language Acquisition Style, the Country or the City in which the study took place. This meta-analysis offers a synthesis of an existing research program and poses intriguing questions that warrant investigation in future studies.

## ACKNOWLEDGEMENTS

Thank you to my supervisor, Dr. Elaine Perunovic, for her many years of wise and patient guidance. Thank you as well to the rest of my supervisory committee, Dr. Biljana Stevanovski and Dr. Veronica Whitford, for their invaluable insights and helpful comments. A special thank you to Dr. Daniel Voyer, whose guidance in the field of meta-analysis, and R statistical software made this project possible. Thank you as well to Dr. Guohua Yan and Dr. Rui Zhang for serving as my internal and external examiners: your comments and revisions were insightful and greatly appreciated. A big thank-you as well to my dutiful research assistants, Rachel Sharpe, Brooke Tracy and Mallory Murphy. Thank You to the Department of Psychology at the University of New Brunswick for your patience and understanding, as well as to Crandall University for giving me an opportunity to work while finishing this project. I would like to thank my loving family (Candace, Rosie, Joe, and Ellie) for their patience, support, and encouragement throughout this process, and finally I would like to express gratitude to God for this life, and the opportunities therein.

## Table of Contents

ABSTRACT ..... ii
ACKNOWLEDGEMENTS ..... iii
Table of Contents ..... iv
List of Tables ..... viii
List of Figures ..... ix
1.0 Introduction ..... 1
1.1 The Sapir-Whorf Hypothesis ..... 2
1.2 Bilingual Language Users ..... 4
1.3 Language Use and Psychological Processes ..... 6
1.4 Understanding Patterns in Research on Bilingual Task Performance ..... 9
1.5 Examining the Language Effect in Bilinguals ..... 12
1.6 Possible Explanations for Language Effects in Bilinguals ..... 13
*1.6.1 Impression Management* ..... 13
*1.6.2 The Dynamic Constructivist Approach to Culture and Cognition* ..... 15
*1.6.3 Cultural Accommodation and Ethnic Affirmation* ..... 17
1.7 Culturally Relevant Phenomena ..... 19
2.0 The Current Research ..... 22
2.1 The Purpose of the Current Research ..... 22
2.2 The Case of English/East Asian Bilinguals ..... 22
2.3 Objectives ..... 27
2.4 Identifying Relevant Moderators ..... 28

*2.4.1 Task Type (Dependent Variables)* ..... 29

*2.4.2 Language* ..... 30

*2.4.3 Nationality* ..... 30

*2.4.4 City* ..... 31

*2.4.5 Design* ..... 31

*2.4.6 Language Acquisition Style* ..... 33

2.5 Application of a Meta-Analytic Approach ..... 34

3.0 Method ..... 36

3.1 Study Selection ..... 36

3.2 Selection Criteria ..... 44

*3.2.1 The First Screening: Examining Abstracts for Relevance* ..... 45

*3.2.2 The Second Screening: Examining Full Documents for Relevance* ..... 46

*3.2.3 The Third Screening: Excluding Articles from Predatory Journals* ..... 47

3.3 Coding of Variables ..... 48

*3.3.1 Sample Level Variables* ..... 49

*3.3.2 Measurement Level Variables* ..... 52

3.4 Coding Validity ..... 54

3.5 Measure of Effect Size ..... 56

*3.5.1 Direction of Effect* ..... 56

*3.5.2 Adjustment for Comparison of Effects Across Design* ..... 57

3.6 Data Analysis ..... 58

4.0 Results ..... 60
4.1 Preliminary Analyses ..... 60

4.2 Overall Effect Size ..... 61

4.3 Moderator Analysis ..... 62

*4.3.1 Effect of Design* ..... 62

*4.3.2 East Asian Language* ..... 63

*4.3.3 Task Type* ..... 63

*4.3.4 Non-significant Moderators* ..... 64

4.4 Adjustment for Multiple Moderators ..... 64

4.5 Publication Bias ..... 65

5.0 Discussion ..... 69
5.1 Summary of the Research ..... 69

5.2 Summary of Results ..... 70

5.3 Interpretation of Overall Effects ..... 71

5.4 Significant Moderators ..... 72

*5.4.1 Language* ..... 72

*5.4.2 Task Type* ..... 74

*5.4.3 Publication Status* ..... 77

5.5 Non-significant Moderators ..... 77

*5.5.1 Design* ..... 78

*5.5.1 Continent* ..... 79

*5.5.2 Language Acquisition Style* ..... 79

*5.5.3 City* ..... 83

5.6 The Contribution of This Analysis ..... 84

5.7 Limitations and Future Directions ..... 90

5.8 Conclusion ..... 99

References ..... 100

Appendix A: Characteristics of Studies Included in the Analysis ..... 120

Appendix B: Coding Guide ..... 135

Appendix C: Analyses Conducted with Results Presented by Design ..... 141

Curriculum Vitae

## List of Tables

Table 1: Search Terms Used for Each Database ..... 37

Table 2: Results for Task Type as a Moderator, with outliers excluded ..... 63

Table 3: Holm Test of Moderators ..... 64

Table 4: Characteristics of Studies Included in the Analysis ..... 120

Table 5: Results for Task Type as a Moderator, with outliers excluded ..... 144

Table 6: Results for Task Type as a Moderator, in the Between-Subjects Design with outliers excluded ..... 145

Table 7: Holm Test of Moderators for all effects ..... 146

## List of Figures

Figure 1: PRISMA Diagram Involving Both the 2019 and 2022 Searches ..... 43

Figure 2: Funnel Plots for Determining Publication Bias Within the Samples ..... 67

## 1.0 Introduction

How does the language that we use affect the way that we understand and interact with the world? The Linguistic Society of America estimates that there are approximately 6,909 different languages spoken in the world today (Anderson, 2010). If each of those languages has the potential to exert a unique effect on the way that users of that language understand and interact with the world, then understanding how language use affects human psychology is a worthwhile and necessary endeavour. In the first chapter of this dissertation, I review some of the literature in which language use has been shown to affect the responses of bilingual language users. Unfortunately, our current understanding of the pattern formed by these results is incomplete and inconsistent. This dissertation serves as an attempt to help clarify the pattern of results in the literature and to reconcile these inconsistent findings via a meta-analysis that investigates the effect of language on task performance in English/East Asian bilinguals.

To understand the question fully, however, a bit of context is valuable. The study of psycholinguistics (the psychology of language) dates back over 100 years to the very beginning of psychology as a science. Wilhelm Wundt himself was keenly interested in the study of language use as observable human behaviour (Nerlich & Clarke, 1998). Advancements in computers and neuroscience have allowed the field to become much more sophisticated since then, but the central goal of psycholinguistics, to understand the role that language use plays in human psychology, is still being actively pursued more than a century later (De Bot, 2000).

### 1.1 The Sapir-Whorf Hypothesis

One of the best known conceptualizations of the relationship between language use and human psychology is through the Sapir-Whorf hypothesis, or the theory of linguistic relativity (Comrie, 2012). According to Edwin Sapir and his student Benjamin Whorf, the way that a language divides and organizes concepts affects the way that the users of that language understand the world (Whorf, 1956). For example, Sera et al. (2002) had French speakers and Spanish speakers assign voices to unnamed pictures of common objects (e.g. knife, spoon, bed, star), and found that the gender of the voice participants assigned to each picture (masculine or feminine) was consistent with the grammatical gender of the words used for those pictures in French and Spanish. According to the theory of linguistic relativity, if languages differ in the way that they divide and categorize the world, then the way that users of those languages understand the world must also differ (Whorf, 1956).

The linguistic relativity hypothesis has traditionally taken two forms. The strong hypothesis posits that human experience is *determined* by language. Particularly in some earlier versions of the linguistic relativity hypothesis, Edward Sapir argued that all human experience is the product of language, and wrote that, “the ‘real world’ is to a large extent unconsciously built up on the language habits of the group” (Sapir, 1956, p. 209). Researchers have had great difficulty finding solid evidence to support the strong hypothesis, given the wide range of other influences on behaviour that appear to have a stronger effect than language (Chiu et al., 2007; Davies, 1998; Tohidian, 2009). The weak hypothesis, on the other hand, posits that whereas language may not fully determine

behaviour, it does have a *meaningful effect*. Researchers have had much more success finding support for this version of the hypothesis.

Consistent with the latter version of the linguistic relativity hypothesis, when Boroditsky et al. (2011) assigned English-Mandarin bilinguals, who learned Mandarin first, and English later in life, to conceptualize time in either Mandarin or English, they found that participants were more likely to conceptualize time on a vertical plane (up and down) when doing so in Mandarin than in English. This finding was explained as a product of language: in the English language, time is conceptualized on a horizontal plane (e.g., forward in time, back in time), whereas in Mandarin, time is also conceptualized on a vertical plane (above or below). It is this familiarity with perceiving time as a vertical concept, the researchers argued, that made Mandarin-English bilingual language users more likely to respond to time cues on a vertical plane when doing so in Mandarin.

Language use has been found to be associated with perceptual processing in other ways as well. For example, researchers have demonstrated that differences in language use are linked to the ways that speakers of different languages understand and interact with space, such as their typical patterns of intrinsic frames of reference (left, right, front, back) or absolute frames of reference (North, South) as well as their spatial reasoning, wayfinding, and object perception (Levinson, 1996). Researchers have also found that the language a person uses to identify colours and odours is associated with their ability to distinguish between and categorize those colours and odours (Majid & Burenhult, 2014; Winawer et al., 2017). Thus, even if the language that a person uses does not *determine*

how they perceive the world, research findings suggest that there is some association between language and the way that people divide, categorize, and interact with the world.

To what extent, however, are these differences actually due to the unique effect of language? Although the researchers responsible for the work cited above attributed the observed effect to language, the issue with trying to answer this question by comparing speakers of one language to speakers of another language is that observed differences could instead be due to demographic, geographic, or cultural differences between users of each language. To truly understand the effect of language on its own, it is necessary to control for these extraneous variables, which is possible through experimental research conducted with bilingual language users.

### 1.2 Bilingual Language Users

Some of the most interesting research examining the influence of language on behaviour comes from research on bilingual language users - individuals who have the capability of understanding and reacting to the world through at least two distinct languages (Grosjean, 2010). Different categories of bilingualism have been proposed, based on the way that one's second language is acquired. A *compound* bilingual is born into a bilingual household and grows up learning one language in tandem with another. A *coordinate* bilingual, on the other hand, is born into a monolingual household and does all of their primary learning in one language; then, later in life, they are given an opportunity to learn a second language. This process sometimes occurs by choice through education, but often occurs as a necessity of adapting to a new culture (Diller, 1970). In other words, compound bilinguals learn their two languages *concurrently*, creating one

language system, whereas coordinate bilinguals learn their two languages *sequentially*, creating two separate language systems (Nguyen & Ahmadpanah, 2014). For the rest of this literature review (when the information is available to do so), I will clarify whether the sample in the reported research was composed of compound bilinguals, coordinate bilinguals, or a mix of the two.

Research into how the brains of bilingual language users process information has provided some support for this distinction between compound and coordinate bilinguals, demonstrating that whereas the two languages of compound bilinguals activate adjacent parts of the brain's language center (Broca's area in the frontal-lobe), the two languages of coordinate bilinguals activate more distinct parts of the brain's language center. Specifically, when Kim and colleagues (1997) had compound and coordinate bilinguals perform language tasks, they discovered that the distance between the brain areas activated by the two languages in coordinate bilinguals was significantly greater than the distance between the same areas activated in compound bilinguals.. More recent research in this area has revealed that the relationship between age of language acquisition and brain development is more complex, implicating not only differences in regions of activation, but also differences in brain structure (Fedeli et al., 2021; Klein et al., 2014; Tu et al., 2021) and brain functioning (Berken et al., 2015; Li et al., 2024; Liu et al., 2020), demonstrating meaningful neurological differences in the brain structure of compound and coordinate bilinguals (Berken et al., 2017).

The point at which a second-language learner becomes bilingual is a matter for discussion. Certainly, an individual who has equal fluency in two languages would be considered bilingual by most people whereas someone who only knows two or three

words in a second language would not be considered bilingual. To determine one's place between those two points, some studies of bilingualism include a measure of first language (L1) and second language (L2) proficiency (e.g., Bialystok et al., 2005; Verhagen et al., 2017). Generally speaking, a person who has functional proficiency in reading, writing, or speaking two or more languages would most likely be considered bilingual. Because the definition of bilingualism is complicated and nuanced, I will defer to the judgment of my fellow researchers and consider participants bilingual if the authors of the studies included in this meta-analysis judge their participants to be functionally bilingual.

### 1.3 Language Use and Psychological Processes

Researchers have found a great deal of evidence suggesting that bilingual individuals demonstrate differences in cognitions, emotions, self-concept, values, and overt behaviour depending on the language that they are using. Language use has been associated with differences in how bilingual individuals process information (cognition). For example, Lechuga and Wiebe (2011) found that monolingual Spanish-speaking Mexicans tended to be more overconfident in their answers to general knowledge questions than monolingual Americans. Similarly, Spanish-English bilingual individuals of Mexican descent who were randomly assigned to complete a multiple-choice general knowledge task in Spanish were more likely to be overconfident in their answers than participants who completed the same task in English. In other words, Spanish-English bilinguals differed in their level of confidence based on the language they used to complete the task, displaying the confidence typical of monolingual Spanish speakers

when speaking Spanish (compared to when they were speaking English). In another example, Cheng and Broadhurst (2005) observed that Chinese-English bilingual Hong Kong criminology students were more accurate at detecting deception in tasks that were conducted in English, but were more accurate at detecting truthfulness in tasks that were conducted in Chinese. Although the researchers' focus was on comparing the ability to detect deception in a first and second language, the differences in patterns by language were suggestive of different cognitive processes depending on the language being used.

Language use has also been associated with how bilingual language users process emotions. For example, Alvarado and Jameson (2011) observed that Vietnamese-English bilingual language users conceptualize shame differently depending on the language they are using - they conceptualize shame in the same way as monolingual Vietnamese users when conceptualizing it in Vietnamese (similar to other high arousal negative terms such as "fear", "hate", and "anger"), but in the same way as monolingual English users when conceptualizing it in English (comparatively less arousing - closer to concepts such as "anguish", "sad", and "lonely").

Additionally, language use has been associated with differences in self-concept (how bilinguals see themselves), and values (how bilinguals determine morality and merit). For example, Ross et al. (2002) found that Chinese-Canadian bilinguals who completed measures in Chinese tended to report lower levels of self-esteem, greater endorsement of Chinese cultural values, and a more even balance of positive and negative moods than participants who completed the same measures in English. Thus, Chinese-Canadian bilinguals thought of themselves and the world differently when they were speaking Chinese than when they were speaking English. In another study, Chiou and

Mercado (2015) had Chinese-American bilinguals watch a simulated soccer match between a Chinese team and an American team, accompanied by either an English commentary or a Mandarin commentary. After watching the soccer game, participants indicated how likely they would be to root for the Chinese team and the American team. When the commentary was in English, participants were equally likely to say that they would root for the American and Chinese teams, however when the commentary was in Chinese, the participants were significantly more likely to say that they would root for the Chinese team than the American team. This pattern suggests that listening to Chinese commentary reminded the participants of their Chinese identity, causing them to be more supportive of the Chinese team. Thus, the language being used affected the participants' expressed loyalty to (and valuing of) their home and heritage cultures.

Language is also associated with the way that bilingual language users make decisions. For example, Briley et al. (2005) demonstrated that Hong Kong bilingual participants who read hypothetical buying scenarios in Chinese consistently made more conservative consumer decisions than participants who read the same scenarios in English. Although the scenarios were the same, the language affected the participants' approach to those scenarios. Ji, Zhang, and Nisbett (2004) found that Chinese bilinguals were more likely to make relational groupings when they were doing so in Chinese than in English. Again, the options were the same aside from the language of presentation, but the decisions made were significantly different depending on the language being used.

Finally, language use has also been associated with how bilingual language users overtly behave. For example, Sechrest et al. (1968) found that bilingual English-Tagalog Filipinos produced longer responses to questions from superiors when those questions

were posed in English than when they were posed in Tagalog (speaking more in English than in Tagalog). Similarly, Bond and Lai (1986) found that Hong Kong bilingual students (a mix of compound and coordinate bilinguals) who were interviewed in English (their second language) gave longer answers to embarrassing questions than participants who were interviewed in Cantonese (their first language). The authors reasoned that, because second languages are often learned in more neutral settings than first languages, discussing embarrassing topics in one's second language should be less emotionally arousing than discussing them in one's first language (leading to longer answers in English than in Cantonese). In another example of the association between language and behaviour, So (2010) observed the gesturing behaviour of Chinese-English coordinate bilingual Singaporean participants. So noted that participants produced a similar number of gestures as monolingual English users when speaking English, but far fewer nonrepresentational gestures (moving hands along with speech pattern, signalling messages such as "ok" with well-known hand gestures, pointing to specific objects) when speaking Chinese. In other words, the non-verbal behaviour of Chinese-English bilinguals changed along with the verbal behaviour. Taken together, the findings from these studies suggest that the language a bilingual individual uses may have some relationship to the way that they act and make decisions.

### 1.4 Understanding Patterns in Research on Bilingual Task Performance

The research presented in the previous section shows how bilingual individuals can sometimes respond to similar stimuli in markedly different ways, depending on the language being used. Could this be evidence for the linguistic relativity hypothesis, which

posits that language use can affect the way that people understand and interact with the world? One way of testing this hypothesis, while controlling for extraneous variables, is by studying how the responses of bilingual language users differ, depending on the language they are using. Over several decades, researchers have demonstrated associations between language use and responses in bilingual language users across a broad range of measures, languages, cultures, nationalities, and designs. This has led to a general understanding that language use can affect the responses of bilingual language users, but the limits of this effect, and the broad patterns formed by these disparate findings are yet to be explored.

Some important first steps have been made toward a broader understanding of the effect. In a meta-analysis, Oyserman and Lee (2008) sought to investigate the effect of individualism and collectivism priming on participants' values, self-concept, relationality, and well-being. Priming involves "...the nonconscious carryover of a previously stored mental process to a subsequent task" (Oyserman & Lee, 2008, p. 313). The authors examined studies where a first task (a prime) was used to activate knowledge, concepts, and goals associated with individualism or collectivism, and then the effect of that activation was examined on another variable. Ten of the 67 studies included in their analysis used language to prime individualism and collectivism, with English used to prime individualism, and any other language used to prime collectivism. Compared to the other forms of prime (e.g., focusing on similarities and differences, sentence circling tasks), language was found to have a relatively small and heterogeneous effect ($d$ = .10), with the strongest effects found for cognition ($d$ = .77) and the weakest effects found for values ($d$ = .01). However, across all studies surveyed, the language of

the studies themselves also had an effect, with the greatest mean weighted effect size in the five studies conducted in Dutch ($d$ = 1.00), and the smallest mean weighted effect sizes for the study conducted in Chinese ($d$ = -.16). Note that Cohen's $d$ is the difference between the means of each condition (in this case, English and an East Asian language) divided by the pooled standard deviation (Borenstein, 2009).

Oyserman and Lee (2008) took an important step toward a broader understanding of the general effect of priming on the responses of bilingual language users in the context of individualism and collectivism, and included a few studies in which language was used as a prime. In the current research, I examined exclusively the effect of language across a wider range of contexts. Oyserman and Lee (2008) found that across 10 studies where language was deliberately used to prime individualism and collectivism, the effect size on outcomes was small ($d$ = .10); however, with a wider range of outcomes under examination, I reasoned that the effect size might vary.

In the current meta-analysis, I sought to understand the pattern wherein responses were more or less likely to differ as a function of language. Oyserman and Lee's (2008) findings suggest that the effect of language is different depending on the outcome being examined; however again, this is based on studies focused exclusively on individualism and collectivism. Examining the patterns of results with a wide range of theoretically relevant variables in mind, I reasoned, may uncover other interesting patterns. I was also interested in examining the directionality of change as a function of language. Across the 10 studies included in Oyserman and Lee's (2008) analysis, six effect sizes were positive (that is, the effects supported the authors' hypotheses) and four effect sizes were negative (that is, the effects did not support the authors' hypotheses). I aimed to uncover if a

similar pattern emerged across a wider range of contexts. Thus, through this research I sought to describe the size, moderators, and directionality of the effect of language on outcomes.

### 1.5 Examining the Language Effect in Bilinguals

One way of investigating how bilingual language users' outcomes will vary as a function of language use is by comparing the responses of bilinguals to the patterns of monolingual users of each language. For example, Lechuga and Weibe (2011) predicted that bilingual Mexican-Americans would respond to general knowledge questions with more overconfidence when they were randomly assigned to do so in Spanish than in English because they had previously found that monolingual Mexicans tended to be more overconfident in their answers to general knowledge questions than monolingual Americans. Similarly, So (2010) predicted that Singaporean undergraduates who spoke both English and Chinese (coordinate bilinguals) would gesture more with their hands when they were speaking in English than when they were speaking in Chinese, because they first found that monolingual English users from the United States tend to gesture more than monolingual Chinese users in Singapore. The findings of these researchers supported their predictions. Additionally, when Briley and colleagues (2005) predicted that Hong Kong bilinguals would make more conservative consumer decisions during hypothetical buying tasks completed in Chinese than in English, it was because they were aware of existing research, such as that by Hong and Chiu (2001), demonstrating that Chinese individuals tend to value decisions that prevent loss, whereas American individuals tend to value decisions that pursue gains. The researchers' findings supported

their prediction. Taken together, the research presented so far provides evidence for the weak linguistic relativity hypothesis by demonstrating how bilingual language users may understand and interact with the world differently depending on the language they are using. According to this hypothesis, the differences discussed so far are a product of the language itself, and how different language components (such as vocabulary, grammar, and syntax) affect the way that users of that language categorize and experience the world (Whorf, 1956). However, supplementary hypotheses have since been presented that expand the scope of consideration beyond vocabulary, grammar, and syntax, to the cultural meaning behind them, providing alternative explanations for the effect of language on the outcomes of bilingual language users.

## 1.6 Possible Explanations for Language Effects in Bilinguals

### *1.6.1 Impression Management*

One proposed explanation for the effect of language on task performance observed in bilinguals is impression management. In a 2005 study, Briley and colleagues argued that frame switching, defined by Hong et al. (2000) as occurring when "...the individual shifts between interpretive frames rooted in different cultures in response to cues in the social environment" (Hong et al., p.709), is a conscious process based on cultural expectations. Furthermore, languages may present demand characteristics that lead individuals to consciously adjust their behaviour to conform to the social and cultural expectations implied by the social setting in which the language is used (Briley et al., 2005). In support of this idea, Briley and colleagues demonstrated that the effect of language on the buying decisions of Hong Kong bilingual language users disappeared

when participants were led to believe that their responses would be viewed by a Chinese audience in both language conditions. The researchers interpreted this observation as impression management, such that participants became equally conservative in Chinese and English because they wanted to appear socially desirable to a presumably conservative Chinese audience.

Although deliberate impression management strategies may contribute to the effect of language on some outcomes, it is unlikely that impression management is the sole explanation for the effect. In some cases, observations are not just made on mean level differences between language conditions, but also on the correlations between variables as a function of language use. For example, Perunovic et al. (2007) found in a diary study across ten days that, after using a Western language, bilingual bicultural East Asian Canadians showed an inverse relationship between positive affect and negative affect (that is, the more positive affect they reported, the less negative affect they reported). However, after using an East Asian Language, positive affect and negative affect were not associated (reports of positive affect were independent of reports of negative affect), which corresponds with cross-cultural findings of differences in the structure of emotion between Western and East Asian cultures (e.g., Bagozzi et al., 1999). Although it is possible that impression management may be responsible for the effect of language on outcomes when there is a socially desirable way of responding (such as self-descriptions and values), it is unlikely that the relationships between two variables across time are so easily and consciously affected by impression management.

### *1.6.2 The Dynamic Constructivist Approach to Culture and Cognition*

Another explanation for the effect of language use on task performance in bilingual language users is that different languages can activate different sets of cultural knowledge and that cultural knowledge can then affect behaviour. According to the *Dynamic Constructivist Approach to Culture and Cognition* (Hong et al., 2000), the language that bilingual individuals use to communicate activates the cultural knowledge system associated with the context in which that language is typically used, and may, therefore, influence the way people think, feel, and act. This is particularly the case with bilingual individuals who are bicultural (LaFromboise et al., 1993). Through the process of socialization, people learn to internalize the culture around them. This internalized culture then affects their self-concept, and the way that they think, feel, and behave throughout their lives. Bicultural individuals may have internalized two distinct cultures, and could then use knowledge from both of those cultures to understand and navigate the world around them (Ramirez-Esparza et al., 2006). Many (although not all) bilingual language users are biculturals if they developed their different languages in different cultural contexts. For example, a Chinese-Canadian bicultural will likely acquire their Chinese cultural knowledge through interactions with other Chinese-speaking individuals; however, they will likely acquire their Western cultural knowledge through interactions with other Canadians in English. As a result, speaking a different language may prime different sets of cultural knowledge. For example, Ross, Xun, and Wilson (2002) randomly assigned Chinese-Canadian bilingual participants to complete a set of questions in either Chinese or English. Their responses to those questions followed the patterns of Chinese monolinguals when responding in Chinese but were more similar to

the patterns of English monolinguals when responding in English. Ross et al. argued that responding in Chinese activated the cultural knowledge associated with that language, resulting in stronger agreement with typically Chinese values.

Although both the dynamic constructivist approach (Hong et al., 2000), and the linguistic relativity hypothesis (Whorf, 1956) posit that the language one uses to complete a task may affect one's performance on the task, there are important differences in the assumed mechanisms behind this effect. The linguistic relativity hypothesis posits that language affects responses because of differences in the structures of the languages themselves, such as vocabulary, grammar and syntax. For example, when Boroditsky and colleagues (2011) found differences in the way that Mandarin/English bilinguals conceptualize time, they pointed to differences in the structure of language, specifically the vocabulary used to describe time (i.e., use of up and down expressions in Mandarin that do not exist in English). The Dynamic constructivist approach, however, posits that language affects responses because language activates cultural knowledge. For example, Ross and colleagues (2002) argued that bilinguals reported lower self-esteem in English, not because the vocabulary, grammar or syntax of the Chinese language caused individuals to report lower self-esteem, but because speaking the Chinese language brought to mind the Chinese cultural knowledge system, that encompasses the value of self-effacement.

These two mechanisms can be easily reconciled. According to the linguistic relativity hypothesis, culture is a product of language – the distinct culture of a people will be shaped by their distinct language - and so the two are necessarily connected. Therefore, differences between languages in vocabulary, grammar and syntax would

necessarily be translated into differences in culture. In other words, differences in language and culture should work in parallel toward the same pattern of difference. Hence, one cannot separate language from cultural knowledge, and so the former may serve as a cue to activate the latter.

### *1.6.3 Cultural Accommodation and Ethnic Affirmation*

In some cases, when language activates cultural knowledge, it results in behaviour that is more consistent with the language being spoken. This was the case of the participants in the study by Ross and colleagues (2002) – participants speaking Chinese responded in ways that were more consistent with the typical response of Chinese monolinguals than participants speaking English. This is consistent with the *cultural accommodation hypothesis* (Bond & Yang, 1982), which suggests that when participants have access to language A and language B, speaking language A activates the cultural knowledge associated with language A, resulting in behaviour that is more consistent with the culture associated with language A than the culture associated with language B. The cultural accommodation hypothesis assumes that this pattern is due to the activation of cultural knowledge; however, for the sake of a simple description going forward, I will describe this pattern, where responses are consistent with cultural knowledge associated with the language being used, as *linguistic consistency*.

In other cases, the opposite pattern emerges, and speaking one's second language appears to make the cultural identity associated with one's first language more salient. For example, previous research (e.g., Heine et al., 1999) demonstrated that Asian (Japanese) individuals report lower self-esteem than Western (Canadian) individuals.

When Chen and Bond (2007) randomly assigned bilingual Hong Kong University students to complete Rosenberg's Self-Esteem Scale in English or Chinese, they found that participants responding in English reported lower self-esteem than participants responding in Chinese. Chen and Bond interpreted this effect as *ethnic affirmation* - an individual's cultural identity becoming more salient when using a second language associated with a different language. This salience in turn directs that individual to behave in ways that are more *consistent* with their *first* language culture when speaking the *second* language and rejecting the culture associated with that second language (Bond & Cheung, 1984; Bond & Yang, 1982; Yang & Bond, 1980**)**. For the sake of simple description going forward, however, I will refer to this pattern – where responses are inconsistent with the cultural knowledge associated with the language being used as *linguistic inconsistency*.

Thus, according to the linguistic relativity hypothesis, language can affect the way that we interact with the world. The dynamic constructivist approach provides one explanation for the mechanism through which language affects the behaviour of bilinguals – through the activation of different sets of cultural knowledge associated with the language being used (Hong et al., 2000). In some cases, this activation results in linguistic consistency – the responses of bilinguals are consistent with the typical responses of monolingual speakers of the language they are using. In other cases, this activation results in linguistic inconsistency – the responses of bilinguals are inconsistent with the typical responses of monolingual speakers of the language they are using. Regardless of whether it be linguistic consistency or linguistic inconsistency, an

understanding of how culture affects behaviour is necessary for interpretation. This is the topic I will explore next.

### 1.7 Culturally Relevant Phenomena

To the extent that many variables may differ as a function of language, to understand this relationship between language and culture, it is important to narrow our focus to *culturally relevant phenomena*, or phenomena with established differences between cultural groups. In the presentation of their dynamic constructivist approach, Hong and colleagues argue that interpretive tasks must be "influenced by cultural knowledge in a well-understood manner" (Hong et al., 2000, p. 711). For example, Ervin (1964) content-analyzed stories written by French-English coordinate bilingual language users in French and English and observed distinct themes of "verbal aggression toward peers," "withdrawal as a response to anger" and "achievement motivation," which corresponded with observed differences between American and French culture at the time. In another example of value differences as a function of language, Ralston et al. (1995) demonstrated that, whereas bilingual managers from Hong Kong who were randomly assigned to complete the Schwartz Values survey in English tended to place greater emphasis on typically American values (such as achievement and hedonism), managers who completed the same questionnaire in Chinese tended to place greater emphasis on typically Chinese values (such as tradition and security). Ross et al. (2002) demonstrated a similar effect in a Canadian sample, finding that Chinese-Canadian bilingual participants agreed more strongly with Chinese cultural beliefs when they were randomly assigned to respond to a measure of those beliefs in Chinese than when they

were assigned to respond in English. These observations are all examples of linguistic consistency, and are consistent with the cultural accommodation account: by this account, asking bilingual participants to complete tasks in different languages activates different sets of cultural knowledge, and those differences in cultural knowledge are reflected as differences in culture-specific values.

As the studies above demonstrate, when the responses of a bilingual language user change as a function of language, they often change to reflect the culture associated with the language being spoken (linguistic consistency). This is not always the case, however – sometimes when the responses of a bilingual language user change as a function of language, they change to reject the culture associated with the language being spoken (linguistic inconsistency). This inconsistency is also reflected in Oyserman and Lee's (2008) findings that whereas six out of 10 studies in which language was used as a tool to prime individualism and collectivism were associated with changes in the predicted direction, the remaining four studies showed changes in the direction contrary to what was predicted. What are the conditions and circumstances that lead a bilingual language user to reflect or reject the culture associated with the language being spoken?

I reasoned that the best way of reconciling these conflicting findings would be a systematic review and meta-analysis. The meta-analysis is the statistical synthesis of results from a series of studies, accumulated as the result of a systematic review of the literature (Borenstein et al., 2009). In recent years, meta-analyses have become an important part of psycholinguistics, allowing for a greater understanding of subjects such as the efficacy of computer games and simulations for language learning (Peterson, 2010), the role of complexity in artificial grammar learning (Schiff & Katan, 2014), and

the role of bilingualism in executive functioning (Donnelly, 2016), and phonological awareness (Branum-Martin et al., 2012). Indeed, it has been argued that meta-analysis is the ideal tool for resolving conflicts in psycholinguistic research (Rákosi, 2018), such as the one being discussed here.

# 2.0 The Current Research

## 2.1 The Purpose of the Current Research

A great deal of research has demonstrated that speakers of different languages can engage with the world in different ways. However, when comparing the behaviour of individuals who speak different languages, it can be challenging to disentangle the influence of language from other extraneous variables. One effective approach to address this challenge is to examine the effect of language in bilingual individuals while controlling for extraneous variables. Past research has extensively demonstrated that the language in which a bilingual person operates can affect the way that they understand and interact with the world. For years, researchers have been demonstrating associations between language use and responses in bilingual language users across a broad range of responses, languages, nationalities, and study designs. Despite this vast body of research, there is currently no clear understanding of how these findings fit together to form a coherent pattern. To date, an attempt has yet to be made to integrate these findings to establish a unified understanding of how the language used by bilingual language users affects their behaviours. Therefore, the purpose of this research was to work toward a unified understanding through a systematic review and meta-analysis of the effect of language on culturally relevant phenomena for English/East Asian bilinguals.

## 2.2 The Case of English/East Asian Bilinguals

Although Cultural psychology examines the effects of culture on the human experience through a wide range of cultural comparisons, one of the most common pairs of cultures to be compared are Western culture and East Asian culture. These

comparisons are made by examining differences between participants in Western countries and East Asian countries. Henrich, Heine, and Norenzayan (2010) define Western countries as those "clustered in the north-west of Europe (the U.K., France, Germany, Switzerland, the Netherlands, etc.), and British-descent societies such as the U.S., Canada, New Zealand and Australia" (Henrich et al., 2010, p.67). East Asian countries are those heavily influenced by Confucian tradition, primarily North of Indonesia and East of India (i.e., China, Macau, Taiwan, Japan, Mongolia, North Korea, and South Korea, Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, Timor-Leste or Vietnam).

Differences between Western culture and East Asian culture have been a topic of interest for cultural psychologists for decades (Dian & Meijer, 2020; Heine, 2005; Hofstede & Bond, 1984; Hofstede & Bond, 1988). When examining differences between these cultural groups, researchers have uncovered meaningful differences in patterns of response in a variety of areas. However, it is possible to examine differences between Western and East Asian culture without having to make international comparisons. This can be done by examining the self-concept, values, cognitions, and behaviours of individuals who have access to both East Asian and Western ways of thinking through their bilingual use of both English and East Asian languages. A body of research also exists demonstrating that English/East Asian language bilingual participants will often behave in ways that are consistent with Western culture when using English, but East Asian culture when using an East Asian language.

For example, differences have been observed in terms of self-concept. In a meta-analysis published in 2007, Heine and Hamamura examined self-enhancement, which

they defined as “tendencies to dwell on and elaborate positive information about the self relative to negative information” (Heine & Hamamura, 2007, p. 4), among East Asians and Westerners. In 88 of the 91 studies they collected, East Asians self-enhanced less than Westerners, suggesting a difference in self-concept between the two cultural groups. Lee et al. (2010) later found a similar difference as a function of language that paralleled the culture effect observed by Heine and Hamamura; they found higher levels of self-enhancement among Chinese-English bilinguals across three studies by manipulating language: participants self-enhanced more when operating in English than in Chinese. Thus, when participants were operating in English, they thought of themselves like Westerners tend to think of themselves, but in Chinese they thought of themselves like East Asians tend to think of themselves.

Values have also been found to differ between cultural groups, and between language conditions. Guo and Spina (2016) found that Chinese participants were significantly more loss-averse than their British counterparts in a financial decision game. This finding was mirrored by other research wherein Pattaratanakun and Mak (2015) found that bilinguals were more loss-averse when completing the game in Thai than in English, suggesting that values regarding finances were meaningfully changed because of language, and that there are differences in what East Asian participants thought was important (avoiding losses) and what British participants thought was important (pursuing gains).

Chen and colleagues (2023) demonstrated cultural differences in cognition by having American and Chinese participants evaluate emotions through examining ethnically Chinese and European faces, and listening to Chinese and English pseudo

sentences. The authors found that American participants were more responsive to facial expressions than Chinese participants, who were instead more influenced by auditory stimuli, suggesting a difference in how people from the two cultures processed information. Interestingly, this tended to be the pattern when processing sentences and voices from their own culture, however when processing sentences and voices from the other culture, both American and Chinese participants tended to be influenced more by visual stimuli. When Chung-Fat-Yim and colleagues (2022) recruited English-Mandarin bilinguals (a mix of compound and coordinate bilinguals) from Singapore and studied their processing of Asian or Caucasian faces and Mandarin or English sentences, they found that bicultural bilinguals were more sensitive to visual stimuli than audio stimuli when randomly assigned to respond to the Caucasian face/English sentence condition; however, there was no difference in sensitivity to visual over auditory stimuli in the Asian face/Mandarin sentence condition. When faces and voices were mixed, Asian faces with English sentences tended to result in more task interference than Caucasian faces with Mandarin sentences. This suggests that the participants' cognitive processes differed based on the language to which they were exposed – they used a more typically Western style when processing information presented in English, and a more typically East Asian style when processing information presented in Mandarin.

Even overt behaviours have been shown to differ by language group, and those differences have also been observed in bilinguals as a function of language. For example, So (2010) found that English monolingual Americans produced more gestures while speaking than Chinese monolinguals, suggesting Western culture as a relatively high-gesture culture and Chinese culture as a relatively low-gesture culture. So also observed

that Chinese-English coordinate bilinguals from Singapore gestured at a similar rate as English monolinguals when operating in English, and only slightly more than the monolingual Chinese when operating in Chinese. Similarly, in a 2018 study, Nicoladis and colleagues observed a higher gesture rate when coordinate bilingual participants were telling their story in English compared to their native language (i.e., Chinese/Mandarin). Both findings suggest that bilingual participants behaved differently depending on the language – they moved like a typical Westerner when talking in English, but they moved more like a typical East Asian when talking in Chinese.

The decision to narrow the focus of this study from bilinguals *generally* to English East Asian bilinguals *specifically* was not taken lightly. Examining all combinations of languages in the world, or even all English/other language bilinguals would have been too great a task for a single study, so some limiting was necessary. Certainly, a fruitful comparison could be made by looking at English/Spanish bilinguals, or English/French bilinguals, however the decision to focus on English/East Asian bilinguals was motivated by two important considerations. First, I wanted to make sure that I would be able to collect a sufficient number of effects. Although cultural psychology as a field examines a wide range of cultural comparisons, there is a particularly strong emphasis on comparisons between East Asian and Western cultures (Kitayama & Huff, 2015), and so an examination of English/East Asian bilinguals was most likely of all similar comparisons to yield the most effects. Second, I wanted to make sure the effects could be adequately interpreted as culturally relevant tasks. The many studies comparing East Asian and Western cultures have revealed a wide range of differences between those cultural groups. This body of literature enables researchers to

make a wide range of predictions about how the language a bilingual uses might affect their responses in ways that are consistent with existing differences between cultures, and thus, making them culturally relevant tasks. These were my primary reasons for narrowing my focus from bilinguals generally to English East Asian bilinguals specifically.

Thus, there is a body of research wherein East Asian bilinguals are found to behave in more Eastern ways when operating in their East Asian language, and in more Western ways when operating in English. But how strong is this effect? And how pervasive is the effect? The current meta-analysis will shed light on answering these questions.

### 2.3 Objectives

Through this systematic review and meta-analysis, I aimed to address several crucial inquiries. First, I sought to better understand the overall effect of language on the outcomes of East Asian/English bilinguals in the literature. Oyserman and Lee (2008) found that the effect of language priming on their outcome variables was small, but their analysis that involved language effects was limited to only 10 studies from a variety of different ethnic groups. By conducting a systematic review and meta-analysis of all available studies in which an English/East Asian bilingual language user was randomly assigned to complete a culturally relevant task in either English or an East Asian language, I aimed to obtain a more representative estimate of the magnitude of the language effect in these bilingual populations in the literature.

Second, I sought to understand how the effect of language on responses may differ depending on relevant moderators. Oyserman and Lee (2008) found that the effect of language was not uniform across all outcome categories, with a stronger effect on cognition and a weaker effect on values. By considering relevant moderators (discussed below), I sought to identify factors that would modulate the magnitude of the effect of language.

Third, I sought to reconcile conflicting findings in the literature regarding the direction of the effect resulting from language use. Much of the research surveyed so far suggests that outcome differences observed as a function of language use reflect the norms and standards implied by the language being used, leading to linguistic consistency. Other research, however, demonstrates the opposite pattern, with outcomes more typical of one's L1 culture when operating in their L2, leading to linguistic inconsistency. With the current meta-analysis, I aimed to reconcile the varying results found in previous research and determine when performance on culturally relevant tasks would differ in ways that aligned with or opposed the culture associated with the language being used.

### 2.4 Identifying Relevant Moderators

As research on the effect of language on the task responses of bilinguals takes on various forms and formats, I thought it unlikely that the effect would be homogenous across all studies. As such, a systematic review and meta-analysis of this area required careful consideration of variables that may moderate the effect of language. To minimize the risk of Type 1 error, the present study adheres to Lipsey and Wilson's (2001)

recommendation of only considering moderators for which there is empirical or theoretical support. My meta-analysis therefore considered *Task Type*, *Language*, *Nationality*, *City*, *Design*, and *Language Acquisition Style* as possible moderating variables.

### *2.4.1 Task Type (Dependent Variables)*

It is possible that the strength and direction of the effect of language may be affected by the specific type of dependent variables or tasks being measured. For example, Bond and Yang (1982) suggest that overt language manipulation and outcomes directly related to cultural identity are more likely to result in ethnic affirmation (linguistic inconsistency), whereas relatively subtle manipulations and outcomes less associated with one's culture are more likely to elicit cultural accommodation (linguistic consistency). Thus, the nature of the dependent variables examined in the research may be important to the direction of the effect.

The literature reveals substantial variability in the dependent variables examined when investigating differences between languages. Some studies have investigated differences in values, beliefs, and personal opinions (e.g., Ralston et al., 1995), whereas others have examined differences in self-concept, identity, and affiliation with a larger cultural group (e.g., Gocken, 2014). Additionally, some studies have focused on differences in cognition, such as performance on mental tasks (e.g., Lyczak et al., 1976), whereas others have examined differences in overt behaviours, such as engagement in physical activity (e.g., So, 2010). Given the substantial variation in the outcomes examined in this field, clarifying the specific variables being measured was crucial for

comprehending the findings. In their meta-analysis, Oyserman and Lee (2008) grouped outcome variables into four categories: values, self-concept, relationality and well-being and found that the strength of the effect of language was not uniform across the categories. Therefore, I included the nature of responses as a moderator in my analysis, and hereafter refer to it as *Task Type.*

### *2.4.2 Language*

It is important to acknowledge that there is a significant amount of diversity among East Asian languages, and therefore, it would have been misguided to assume that all East Asian languages would affect behaviour in a uniform way. The effect of language on behaviour has been observed, even when languages are very similar, such as between Finnish and Swedish (Lonnqvist et al., 2014). Failing to consider how different East Asian languages may affect behaviour would have also neglected the distinct cultural and historical contexts of each language. Thus, the type of East Asian *Language* was included as a moderator in my analysis.

### *2.4.3 Nationality*

Studies comparing task performance in English to task performance in an East Asian language have been conducted in multiple Western and East Asian countries. Although there are commonalities among East Asian countries, it is important to acknowledge that there are also differences among them. For example, Harzing and collaborators from 26 different countries (2006) examined international differences in survey response style (e.g., acquiescence, extreme response style and middle response style) and found notable differences in typical survey response patterns among

neighbouring East Asian Countries, such as China, South Korea, and Japan. To give consideration to these differences among the various East Asian nations, I used the *Nationality* of the sample as a moderator in my analysis.

### *2.4.4 City*

In the same way that the studies of interest come from several Western and East Asian countries, they also come from many different cities. Although it may seem superfluous to examine the effect of city, I believed that it was necessary due to the large proportion of studies of interest that had been conducted in one particular city: Hong Kong. Hong Kong has been a fertile ground for research on Western and East Asian bilingual/biculturals for decades. In Hong Kong, East Asian cultural influences from Chinese culture blend with Western cultural influences from the many years wherein Hong Kong was a British colony. This cultural dynamic provides ideal conditions for research on bicultural individuals (Hong et al., 2000; Kemmelmeir & Cheng, 2004). Indeed, almost 50% of the included effect sizes came from Hong Kong samples. Given the overrepresentation of effects from such a small geographic area, it was crucial to consider the generalizability of these effects – that is, are the findings from within Hong Kong comparable to the effects outside of Hong Kong? To answer this question, I included *City*, or more specifically, whether the studies originated from Hong Kong, as a moderator.

### *2.4.5 Design*

Although the current meta-analysis only included studies with experimental designs (as opposed to quasi-experimental or correlational designs) that compared

responses when randomly assigned to respond in English or an East Asian language, the experimental designs varied in whether they were between-subjects or within-subjects. There are advantages and disadvantages to either design. On the one hand, between-subjects design studies are less time-consuming to carry out and are not subject to carry-over effects because participants are only exposed to one level of an independent variable. On the other hand, between-subjects designs may be compromised by extraneous variables, or pre-existing differences between groups, which is less of a concern for within-subjects designs that use the same set of participants for all levels of the independent variable. Furthermore, within-subjects designs typically require fewer participants than between-subjects designs, and have less noise and more power (Price et al., 2015).

In many studies that examined the effect of language in English/East Asian bilinguals, researchers used a between-subjects design (e.g., Bond & Yang, 1982; Briley et al., 2005; Lechuga & Weibe, 2011), whereas in other studies researchers used a within-subjects design (e.g., Alvarado & Jameson, 2011; Cheng & Broadhurst, 2005; So, 2010). While it is possible to compare within- and between-subject designs in the same meta-analysis, it requires correlations for the within-subjects designs (Morris & DeShon, 2002), which are rarely available. In such cases, statistical adjustments are recommended to ensure comparability between designs (Borenstein et al., 2009). Should the adjustment be insufficient to establish equivalency between within-subjects and between-subject designs, then including both in the same analysis would be unwise, as including studies with different designs in the same analysis is only reasonable when they are comparable. If there are meaningful differences, then that would suggest a lack of comparability, and

so combined analyses would not be the best course of action (Borenstein, 2009). If that was found to be the case, I determined beforehand that I would conduct separate analyses for studies that measured the effect of language between-subjects and those that measured the effect of language within-subjects.

Conceptualizing the moderator in this way is perhaps an oversimplification of what could have been a much more informative *design* moderator. The purpose of this moderator, however, was to determine if the effects gathered from within-subjects designs were comparable in size to effects gathered from between-subjects designs. Fortunately, all of the gathered effects were easily coded using this dichotomy, allowing for a comparison. Thus, this moderator served its intended purpose. It is indeed possible that other elements of design could be worth examining for their potential effect on the relationship between language and task performance. I explore some of those possibilities in section 5.7.

### *2.4.6 Language Acquisition Style*

As mentioned earlier, an important distinction has been made between two kinds of bilinguals: compound bilinguals who acquire their two languages concurrently, and coordinate bilinguals who learned their two languages sequentially. This difference in language acquisition has been implicated in different ways of approaching the cultures of two different languages. For example, Nguyen and Ahmadpanah (2014) demonstrated that compound English-Spanish bilingual Mexican Americans tended to see a greater overlap between their Mexican and American identities than coordinate English-Spanish bilingual Mexican Americans. In an example using East Asian bilinguals, Ji and

colleagues (2004) examined the tendency of participants to group items based on the relationships between the items (e.g., grouping postman and uniform, because a postman must wear a uniform) rather than grouping items into more abstract categories (e.g., postman and policeman, because both are professions) among Chinese-English bilinguals in Hong Kong. Although participants showed a general preference for relational grouping in Chinese than in English, the difference was stronger for coordinate bilinguals than compound bilinguals.

Furthermore, one's level of exposure to the culture associated with a language has also been shown to play an important role. For example, coordinate bilingual Dutch-English students from Holland demonstrated more prototypically American responses to a prisoner's dilemma game when they participated in English than in Dutch; however, the effect of language on task performance was stronger for participants who had spent time in the United States than for those who had not (Akkermans et al., 2010). Thus, if the style of language acquisition has a bearing on bilinguals' cultural identities and the way they respond to language use, then it would be prudent to consider style of language acquisition as a potential moderator in my meta-analysis.

### 2.5 Application of a Meta-Analytic Approach

My research involved a comprehensive review of the relevant literature, followed by a meta-analysis, to determine the overall effect of language on bilingual English/East Asian language users' responses in the literature, accounting for task type, language, nationality/city, design, and language acquisition style as moderators. This resulted in 63 independent samples from 45 studies, reflecting the responses of 6,693 participants.

Two common models used in a meta-analysis are the fixed-effect model and the random-effects model. In a fixed-effect model, each effect is regarded as a parameter (fixed number) or an estimate of the parameter, and so the model gives us an estimate of what that population parameter might be (with differences between effects assumed to be due to sampling error). A random-effects model, however, appreciates that there may not be a single true population parameter, and that differences in effects are likely due to more than just sampling error (e.g., differences in study design). In a random-effects model, the effects are regarded as random arising from some distribution. The model, therefore, represents the typical effect among a range of heterogeneous effects (Harrer et al., 2021). Both models work on the assumption of independence, which means that each effect is assumed to come from a unique sample. In my data collection, however, many of the effects came from the same sample of participants; this meant that the effects were not truly independent and that neither a fixed-effect nor a random-effects model would be appropriate for my analysis. Instead, I used a multilevel meta-analysis, which allowed for a consideration of the interdependence between effects taken from the same sample, by generating a model at the level of the effect, and a separate model at the level of the sample (Marsh et al., 2009).

# 3.0 Method

## 3.1 Study Selection

In order to ensure sufficient breadth in my search, I incorporated studies from three sources. First, I conducted an exhaustive search through online article databases for all research published in peer-reviewed scientific journals that met my inclusion criteria from the first available articles until August 2019 (see section 3.2, Selection Criteria, for selection criteria). Search terms were created following consultation with a librarian with expertise and experience in systematic review. The databases used included PsycINFO (3,562 files), PubMed (872 files) and ERIC (3,998 matches). I also used the ProQuest Dissertations and Theses Global database to search for unpublished theses and dissertations that met my criteria and added them to the review (11,100 files). I loaded all the obtained files into the open-access research software Zotero (https://www.zotero.org).

Due to the length of time between the initial search and analysis, a search update was deemed appropriate. Therefore, I conducted another search using the same databases and the same search terms in April 2022, limiting my search to only files published between 2019 and 2022. Results from PsycINFO (682 files), PubMed (390 files), ERIC (779 files) and Proquest Dissertations Theses Global (629 files) were again uploaded to Zotero. See Table 1 for the search terms used with each database. The number of files from each database in 2019 is in the left column. The number of files from 2022 is in the right column.

**Table 1**

*Search Terms Used for Each Database*

| Database | Search terms | 2019 Files | 2022 Files |
|---|---|---|---|
| PsycINFO | (multicultural* OR bicultural* OR national* OR cultural* OR sociolinguis* OR "cross cultural" OR cross-cultural*) N2 (dynamic* OR cue* OR identi* OR priming OR prime* OR scheme* OR schema* OR frame* OR framing OR switch* OR shift* OR mobil* OR outcome* OR task* OR "frame switch" OR "frame switching" OR exercise* OR function*) | 3,562 | 682 |
| | S2 "Code Switching" | | |
| | S3 language* | | |
| | S4 bilingual* | | |
| | S5 multilingual* | | |
| | S6 "Foreign languages" OR "English as a Second Language" | | |
| | S7 "Multilingualism" OR "Bilingualism" | | |

| | | | | |
|---|---|---|---|---|
| | S8 | S1 OR S2 | | |
| | S9 | S3 OR S4 OR S5 OR S6 OR S7 | | |
| | S10 | S8 AND S9 | | |
| PubMed | #1 | ((multicultural* OR bicultural* OR national* OR cultural* OR sociolinguis* OR "cross cultural" OR cross-cultural*) AND (dynamic* OR cue* OR identi* OR priming OR prime* OR scheme* OR schema* OR frame* OR framing OR switch* OR shift* OR mobil* OR outcome* OR task* OR "frame switch" OR "frame switching" OR exercise* OR function*)) | 872 | 390 |
| | #2 | Search (Bilingual* OR Multilingual*) | | |
| | #3 | Search Language* | | |
| | #4 | Search (Language*) AND ((Bilingual* OR Multilingual*)) | | |

| | | | | |
|---|---|---|---|---|
| | #5 | ((((multicultural* OR bicultural* OR national* OR cultural* OR sociolinguis* OR “cross cultural” OR cross-cultural*) AND (dynamic* OR cue* OR identi* OR priming OR prime* OR scheme* OR schema* OR frame* OR framing OR switch* OR shift* OR mobil* OR outcome* OR task* OR “frame switch” OR “frame switching” OR exercise* OR function*)))) AND ((Language*) AND ((Bilingual* OR Multilingual*))) | | |
| ERIC | S1 | (multicultural* OR bicultural* OR national* OR cultural* OR sociolinguis* OR “cross cultural” OR cross-cultural*) N2 (dynamic* OR cue* OR identi* OR priming OR prime* OR scheme* OR schema* OR frame* OR framing | 3,998 | 779 |

| | | | | |
|---|---|---|---|---|
| | | OR switch* OR shift* OR mobil* OR outcome* OR task* OR "frame switch" OR "frame switching" OR exercise* OR function*) | | |
| | S2 | "Code Switching (Language)" | | |
| | S3 | language* | | |
| | S4 | bilingual* | | |
| | S5 | multilingual* | | |
| | S6 | "Foreign languages" OR "English as a Second Language" | | |
| | S7 | "Multilingualism" OR "Bilingualism" | | |
| | S8 | S1 OR S2 | | |
| | S9 | S3 OR S4 OR S5 OR S6 OR S7 | | |
| | S10 | S8 AND S9 | | |
| ProQuest | S1 | (multicultural* OR bicultural* OR national* OR cultural* OR sociolinguis* OR "cross cultural" OR cross-cultural*) NEAR/2 (dynamic* OR cue* OR identi* OR priming OR prime* OR scheme* | 111,000 | 629 |

| | |
|---|---|
| | OR schema* OR frame* OR framing OR switch* OR shift* OR mobil* OR outcome* OR task* OR "frame switch" OR "frame switching" OR exercise* OR function*) |
| S2 | language* |
| S3 | bilingual* |
| **S4** | **Multilingual*** |
| **S5** | **S3 OR S4 OR S5** |
| **S6** | **S1 AND S6** |

Second, I actively sought out research that may not have been picked up in the systematic review, but met my search criteria. To do this, I identified six key articles in the field (Bond & Yang, 1982; Ervine, 1964; Hong et al., 2001; Lee & Oyserman, 2010; Oyserman & Lee, 2008; Yang & Bond, 1982) and cross-referenced the works that they cited with my own search results. This search yielded an additional 100 files. I also consulted the conference proceedings from the 2016 and 2018 international congresses of the International Association for Cross-Cultural Psychology (IACCP) in pursuit of current research that was relevant to my findings but had not yet been published, which yielded an additional 61 files (42 files from IACCP 2016 and 19 files from IACCP 2018). During the update in 2022, I consulted conference proceedings from "IACCP 2020+" which took place in 2021 (the "+" symbolizing the fact that the 2020 conference had to

be postponed to 2021 due to the COVID-19 Pandemic), which resulted in an additional seven files. I also advertised through list-serves from the International Association for Cross-Cultural Psychology, and posted on the Facebook group of the Linguistic Society of America. In the post I introduced myself and this project, and asked members to send me unpublished research that met the first three criteria.[1] I also directly contacted the corresponding authors of key articles asking about relevant unpublished research. Unfortunately, these latter two approaches did not yield any new files.

For the purposes of this research, I did not restrict publication language to English. Abstracts written in a foreign language were translated using Google Translate to determine eligibility. I encountered three full articles written in languages in which I lacked proficiency that required additional consideration. Two of the articles were in Chinese and one was in French. I sent the Chinese articles to my supervisor (who is proficient in Chinese) and the French article to a family member (who is proficient in

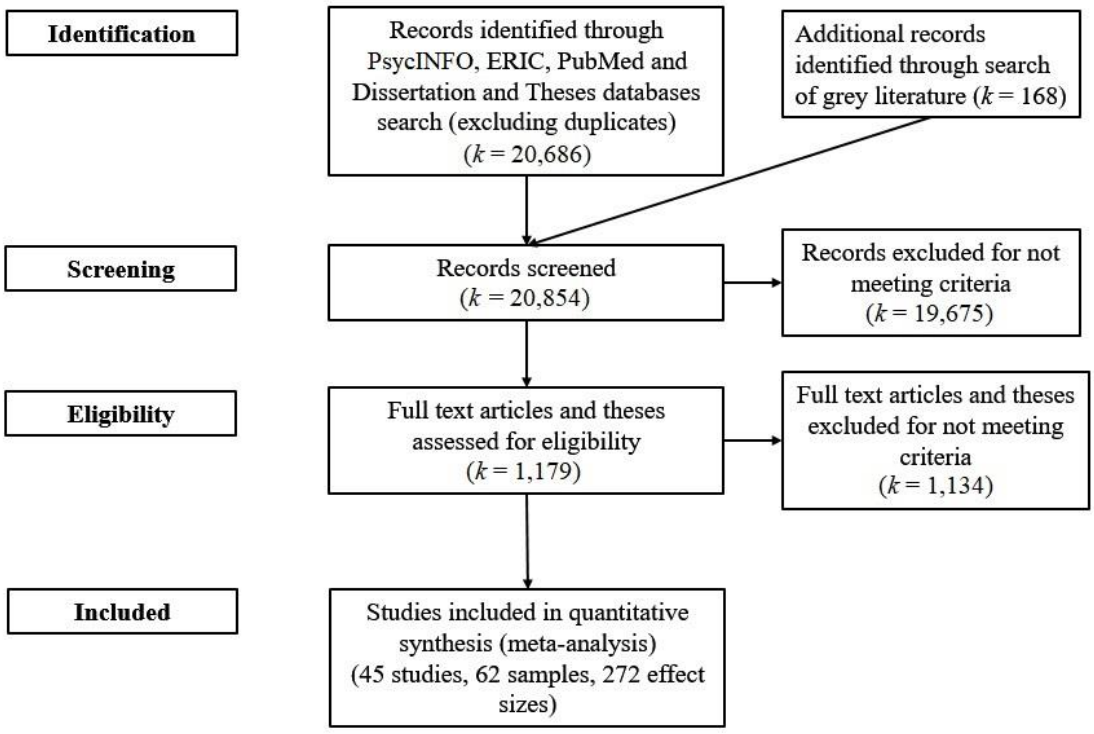
[1] The full post read as follows: My name is Jordan Schriver I am a PhD student from the University of New Brunswick. I am conducting a meta-analysis for my doctoral dissertation, in which I am examining the effect of language on the task performance of bilingual language users, and I could use your help. Specifically, I am looking for unpublished research that meet these criteria:
(1) The research was conducted with bilingual participants who were proficient in both English and at least one East Asian language.
(2) The bilingual participants were assigned to complete the exact same task in either English, or their East Asian language (or, in the case of within-subjects designs, they were assigned to complete the same task in both English and their East Asian language).
(3) Statistics are available to compare outcomes in English to outcomes in an East Asian Language.
If you have any information leading to unpublished research of this nature, please send an email to jrschriver@gmail.com. Thank you very much.

French) to ensure my interpretations were correct. After consultation, all three of the articles were determined to be irrelevant to the analysis.

Once duplicates were removed, the resulting database (from the 2019 search) included 18,507 non-overlapping files in 2019 and an additional 2,347 non-overlapping files, resulting in a total of 20,854 files. See Figure 1 for a graphical display of the search.

**Figure 1**

*PRISMA Diagram Involving Both the 2019 and 2022 Searches.*




*Note.* Each box within the table represents a level of screening. The box on the left represents the level of screening, the center block indicates the number of files included in the screening and the box on the right represents the number of files excluded before moving on to the next level of screening.

### 3.2 Selection Criteria

Once all the files were loaded into Zotero, I reviewed the titles and abstracts of all the files for their relevance. Any file deemed irrelevant was sorted into a separate file, which assistants then reviewed. Due to the large volume of files, a team of three assistants reviewed the excluded files: my supervisor, with extensive knowledge of the subject area and over 20 years of experience in research; a trained undergraduate volunteer research assistant who was helping me in exchange for a course credit (A2), and a paid trained undergraduate research assistant (A3). My supervisor was judged sufficiently aware of the subject matter and process not to need additional training. Each of the undergraduate research assistants was provided with the introduction section of this dissertation to familiarize themselves with the intention behind the study. I also met with the assistants in person (A2) or online (A3) to explain the intention of the study, go over each of the selection criteria in detail, and screen through several abstracts as an example of the process. In 2022, a fourth trained undergraduate research assistant (A4), who received the same reading and one-on-one training as A2 and A3, reviewed the files. Files were deemed relevant only if they met all the following criteria:

1) The files must have contained quantitative data.

2) The files must have included research conducted using bilingual participants who were proficient in both English and at least one East Asian Language.

3) Bilingual participants must have been randomly assigned to complete the same task in either English or the East Asian Language that they can speak (or, in the case of within-subjects designs, they must have been assigned to complete the same task in both English and the East Asian Language that they can speak).

4) Statistics had to be available to compare means on outcome variables in English to means on the same outcome variables in the East Asian Language.

5) The authors must have made hypotheses about the differences between groups based on previously established cultural differences (making those variables culturally-relevant phenomena).

### *3.2.1 The First Screening: Examining Abstracts for Relevance*

Based on the examination of titles and abstracts, I sorted the files into inclusion or exclusion folders.[2] My supervisor, as well as research assistants A2 and A3 looked through both of the exclusion folders for files that were relevant, but were sent to one of the exclusion folders erroneously. The assistants then put those files into a separate folder for me to reconsider. Of the 17,542 files deemed irrelevant, the research assistants

---

[2] Four different folders were used for the first screening. A folder wherein it was clear from the abstract that the file was relevant to the research ("Yes"), a folder wherein it seemed possible that the files might be relevant ("Maybe Yes"), a folder wherein the files had nothing to do with the topic area of the research ("No"), and a folder wherein files were topically related to the area of research but clear from the abstract that they violated one or more of the exclusion criteria ("Maybe No"). Files that I sorted into either of the inclusion folders ("Yes" and "Maybe Yes") were moved along to the second screening. When a file did not meet one or more of the inclusion criteria, I noted which criteria that the file did not meet (1, 2, 3 or 4), and then I sorted that file into one of the two exclusion folders ("No" and "Maybe No").

identified five as potentially relevant and set them aside for me to reconsider. I retrieved the full articles for each of those files and reviewed those for relevance. In total I moved 1,105 files to the second screening.

I followed the same process in 2022, but streamlined the process, such that first, I sorted the files into either a single inclusion folder ("Yes") or a single exclusion folder ("No"). Research Assistant A4 reviewed all the files in the exclusion folder, and of the 2,133 files deemed irrelevant, only identified two files as potentially relevant, and worthy of being reconsidered. I retrieved each of these articles and reviewed them for relevance. In total, 214 files were moved to the second screening in 2022.

### *3.2.2 The Second Screening: Examining Full Documents for Relevance*

The second screening involved only the files that were sorted into the inclusion folders. I retrieved the full articles for each file. Most articles were available online through the initial searched databases. I requested any articles that could not be retrieved through the original searched databases through document delivery. When an article was unavailable through document delivery, A3 and I attempted to contact the authors. A total of 43 emails (36 in 2021 and seven in 2022) were sent to authors asking for copies of their manuscripts. This approach yielded 17 additional articles (12 in 2021, five in 2022). If the authors did not respond to the request, or if contact information was not available then I decided that the article would not be included in the analysis. In the end, I included 965 full articles in the second screening. The same process was followed in 2022, resulting in all 214 full articles being retrieved.

I reviewed the full articles for relevance based on my inclusion criteria.[3] This resulted in 42 articles being considered for the third screening. The same process was followed in 2022, and three additional articles were included in the third screening.

### *3.2.3 The Third Screening: Excluding Articles from Predatory Journals*

The third and final screening included only articles deemed relevant for inclusion. I cross-referenced the journals in which those articles were published with a list of predatory journals available at https://beallslist.net/#update. None of the articles that passed to the third screening came from an identified predatory journal, and so none of the articles were excluded from analysis. The same process was followed in 2022, and none of the articles were found to come from a predatory journal.

Thus, after three screenings, I had 45 studies that met all the outlined inclusion criteria. Although many of these studies involved one group of participants (or one sample), several of them were multi-study designs that contained multiple groups of participants. As a result, these 45 studies yielded 62 independent samples reflecting the responses of 6693 participants, through which I collected 272 effect sizes for inclusion in the current meta-analysis. All included studies are listed in Appendix A.

---

[3] Articles that I judged relevant were sorted into another "Yes" folder and included in the third screening. When I found an article to be irrelevant in the second screening, I provided a specific explanation as to why the articles were deemed to violate the selection criteria and then I moved the articles to the "No" folder.

### 3.3 Coding of Variables

After completing the third screening, trained research assistants (A3 and A4) and I reviewed and coded all articles that passed through the three screenings. I met with both research assistants to explain the variables that we were going to code, and why we were going to code them. I also provided a coding guide for the research assistants (see Appendix B), in which I explained in detail, with relevant examples, how to code the variables. We also coded an example article together, and I remained in regular contact with the research assistants during the coding, examining their work and answering their questions.

In all of these studies, participants who were able to use English and an East Asian language were randomly assigned (either within-subjects or between-subjects) to complete a culturally-relevant task in English and/or an East Asian language, and the outcomes from these tasks were compared. I defined *language use* as any task requiring a response to a stimulus through the medium of either written or spoken language. The presentation of the stimuli and the recorded response were required to be in the same language. The specific East Asian Language used in the study was coded as one of several moderators (described in section 3.3.1). For a study to be included, participants must have been asked to do equivalent tasks in English and/or their East Asian Language, where the only meaningful difference between tasks was the language being used. Furthermore, statistics must have been available for an effect size to be calculated. When the necessary information was not included in the article, research assistant A3 and I contacted the corresponding author for that information. Five such emails were sent to

request that information, and I received responses from three authors. Due to a lack of necessary information, the remaining two studies were excluded.

### *3.3.1 Sample Level Variables*

**3.3.1.1 Publication Status.** At the sample level, we coded the publication status of each article. Published works were coded as "article" (213 effects) and the journal in which the article appeared was also recorded. Unpublished works were coded as "dissertation" (59 effects) and the institutional affiliation of the authors was also recorded.

**3.3.1.2 Language.** Additionally, we coded the specific East Asian language used in each study (all included studies had English as the alternative language). Of the studies that met the rest of our inclusion criteria, the East Asian languages represented were Cantonese (29 effects), Cebuano (3 effects), Chinese (172 effects), Japanese (31 effects), Korean (7 effects), Malay (2 effects), Mandarin (14 effects), Tagalog (10 effects), and Thai (4 effects).

The presence of many languages with relatively small sample sizes presented an issue for moderator analyses, namely a lack of power to detect effects. Furthermore, when analyses were run with all languages included, the statistical program returned error messages due to several particularly small sample sizes. Based on recommendations by Borenstein and colleagues (2009) to avoid small study sizes in moderator/covariate analyses, and in consultation with several members of my supervisory committee, I decided to group this - and several of the following moderators with similarly small sample sizes - into larger, theoretically linked subgroups. Because Cantonese and

Mandarin are both Chinese languages, we combined Chinese, Mandarin, and Cantonese into one category called “Chinese” (215 effects) or “Other” including all non-Chinese languages (57 effects).

**3.3.1.3 Language Acquisition Style**. We also coded language acquisition style. Bilinguals differ in whether they acquire their two languages concurrently (compound bilingual) or sequentially (coordinate bilingual). When information was available to make the judgement that participants learned both of their languages at the same time, the sample was coded as “Compound”, but if one language was acquired after the other, the sample was coded as “Coordinate”.

Although it may have been preferable to use an age cutoff as my primary method of distinguishing between compound and coordinate bilinguals, that data was often unavailable. It was common for articles to indicate whether the individuals began learning their second language in a schooling context or if they had experienced both languages from birth. Therefore, we coded a compound bilingual as anyone who began learning two languages between the time they were born and the time they entered school, and a coordinate bilingual as anyone who only began learning their second language within a school context, or through a language learning program or experience later in life. If that information was not available, then we used the age of six as a cutoff to maintain equivalence.

In some cases, compound and coordinate bilinguals were included in the same sample, which was coded as “Mixed.” If it was not possible to identify whether the participants were compound or coordinate bilinguals based on the information in the article, the results were coded as a third category, “Unclear.” The final dataset had

representation from compound bilinguals (75 effects), coordinate bilinguals (97 effects), mixed (15 effects), and unclear (85 effects).

**3.3.1.4 Nationality** We also coded the nationality of the sample. Nationality of the sample was considered as the country in which the data were collected. For example, if a Canadian researcher conducted a study using participants in the geographic greater Toronto area, we coded the nationality as Canada; however, if the researcher used participants in Beijing, we coded the nationality as China. Sample nationalities included Belgium (8 effects), Canada (10 effects), China (126 effects), Hong Kong pre-reunification (31 effects), Japan, (2 effects), Malaysia (7 effects), England/the UK (10 effects), the Philippines (13 effects), Singapore (6 effects), Thailand (4 effects), and the United States (55 effects). Due to the small sample sizes for each nationality, we initially grouped the countries by *continent*: "East Asia" (189 effects), "Europe" (18 effects), and "North America" (65 effects). Unfortunately, the "Europe" grouping was still too small for analyses, and so we ultimately ended with "East" (189 effects) and "West" (83 effects). As a result of this grouping, this moderator is referred to as "Continent" in subsequent chapters of this dissertation.

**3.3.1.5 City**. Whenever possible, we also coded the *city* or geographic location within a country in which the data was collected. The purpose of coding city was primarily to identify samples from Hong Kong, which has a unique cultural identity within East Asia. Recall from the previous discussion (in section 2.4.4, City), Hong Kong's unique blend of Eastern and Western culture has provided fertile ground for research of this nature, raising the question of whether the effects observed within Hong Kong are meaningfully different from the effects observed outside of Hong Kong. Effects

were coded as coming from “within Hong Kong” (129 effects) and from “outside of Hong Kong” (143 effects).

Additional variables at the sample level that were recorded, but not intended for analysis included the title of the article, age range of participants (mean, standard deviation and range, whenever available), and the gender or sex division of the sample.

### *3.3.2 Measurement Level Variables*

**3.3.2.1 Design.** To determine if the manner in which participants were assigned to complete the tasks moderated the effect of language use on task performance, we coded each effect based on the method of assignment as either “within-subjects” (109 effects) or “between-subjects” (163 effects).

**3.3.2.2 Task Type.** To determine if the nature of the dependent variable used to calculate effect sizes moderated the effect of language on participants’ responses, we coded the *task type* for each effect size. For each of these studies, language (East Asian or English) was the manipulated independent variable. The measured dependent variables in each study were grouped into one of four general task categories – self-concept, values, cognition, or behaviour, and those categories were included as a moderator. The categories were defined in the following ways:

1) ”Self-concept” was defined as "...a collection of beliefs about one’s own nature, unique qualities, and typical behavior” (McCann, 2022, p.472). Tasks coded as self-concept measured participants’ sense of identity. On some occasions, it was identity in a broad sense, such as affiliation to a group or nationality. On other occasions, it was in reference to identity in a narrow sense, such as personality traits. Responses

demonstrating affiliation and belonging to larger cultural groups, as well as responses demonstrating self-reported traits and non-cultural identity, were also coded as self-concept. Responses demonstrating ethnic identity, national affiliation, and acculturation also fell into this category, along with measures of Big Five or Five-Factor personality and trait intelligence.

2) “Values” were defined as "...internalized cognitive structures that guide choices by evoking a sense of basic principles of right and wrong, a sense of priorities and a willingness to make meaning and see patterns” (Oyserman, 2015, p. 36). Tasks categorized as values required participants to make moral judgments or provide their opinion, which could have been in response to a series of statements or hypothetical scenarios. Tasks demonstrating individualism and collectivism, or tradition and modernity, for example, fell into this category.

3) ”Cognition” was defined as “mental processes involved in acquiring knowledge” (McCann, 2022, p.12). Tasks categorized as cognition demonstrated performance on mental tasks that require evaluation of a stimulus. Tasks demonstrating intelligence, mental ability, and attention, were all categorized as cognition. For example, Ji, Zhang, and Nisbett’s (2004) study wherein they examined participants’ tendency to categorize items into relational groups in Chinese or English was categorized as cognition. This category also included examples of social cognition, including any assessment or evaluation of another person. For example, Lyczak and colleagues (1976) had participants evaluate characters in terms of kindness, trustworthiness, honesty, and other similar qualities in either Chinese or English.

4) We coded the remaining effects as “Behaviour”, which was simply "any overt observable response or activity by an organism” (McCann et al., 2022, p.6). Tasks categorized as behaviour required some physical action in response to a stimulus, which includes gesturing (So, 2010), standing position relative to a superior (Sechrest, 1968), and length of a spoken conversation (Bond & Lai, 1986).

In summary, we coded all the tasks as either self-concept (73 effects), values (58 effects), cognition (99 effects), or behaviour (42 effects).

### 3.4 Coding Validity

To validate the coding, research assistant A3 (a paid research assistant to whom I provided one-on-one coding training) and I coded the collected articles independently. I attempted to recruit A3 to return for 2022, however, because they were unavailable, the second set of articles was coded by A4 (who received the same coding training as A3). A coding book was also prepared and made available for the assistants to consult during the coding (See Appendix B). On separate prepared coding sheets, we coded each article for the following: Authors, year of publication, title, publication status (article or dissertation), specific East Asian language (EAL), Language Acquisition Style (compound, coordinate, mixed, unclear), Nationality, City, gender division, age range, Design (within or between), Task Type (self-concept, values, cognition, behaviour), English sample size, East Asian language sample size, *d*, and direction of the effect.

The main variables of interest were EAL, Language Acquisition Style, Nationality, City, Design, and Task Type. EAL sample size and English sample size were also included for the calculation of effect size (*d).* Although it was not one of my primary

moderators, publication status (article or dissertation) was recorded so that publication bias analyses could be conducted. No matter how rigorous the search, the possibility remained that some relevant research might be missed because of publication bias – a tendency for studies with significant results and large effect sizes to be published more often than studies with non-significant results and small effect sizes, which instead disappear into obscurity (Rosenthal, 1979). One way of testing for this is by looking at the moderating effect of publication (published or unpublished) and so publication status was recorded for that purpose.

As a first step, we coded the effects from every included article. Research assistant A3 independently coded all articles up to August 2019, and research assistant A4 coded all articles from September 2019 to April 2022. The initial interrater agreement for coding of effect sizes was 89.74%, with 28 disagreements out of 272 effects. Inter-rater coding for format and language were 100% (no disagreements). Country (12 disagreements out of 272 observations), city (two disagreements out of 272 observations), and design (5 disagreements out of 272 observations) all had agreement above 95%. Therefore, we had a fair degree of confidence in the moderators of Language, Country, City and Design.

By contrast, the initial agreement for Language Acquisition Style (126 disagreements out of 272 observations) and Task Type (107 disagreements out of 272 observations) were both relatively low. The low agreement for Language Acquisition Style was likely due to an inconsistency in the provision of relevant information within the participants sections of the studies: because the authors rarely referred to participants explicitly as “compound” or “coordinate” bilinguals in the text, the research assistants

and I made coding decisions based on other provided evidence (e.g., the average age at which participants first learned their second language). The low agreement for task type was due to the relatively abstract nature of this moderator, and the diversity of tasks that participants were assigned to complete, which made task type far more difficult to code than more objectively descriptive variables such as language or country. Due to the difficulty in coding these variables, and the low interrater agreement, some consideration was given to the possibility of eliminating them as moderators from the analysis. I ultimately decided, however, that despite the issues, these moderators provided valuable information, and should be retained but interpreted with caution.

To compensate for the large discrepancies in interrater agreement, I made note of all coding discrepancies related to the effect sizes and the moderators. I sent this list of discrepancies to the research assistants, and then we independently re-examined each study, and re-considered my coding thereof to check for error. After independent reconsideration, we then met to discuss each discrepancy, provide the reason for my coding thereof, and decide on appropriate coding. We did this for each discrepancy until we reached 100% agreement on all effects and moderators.

## 3.5 Measure of Effect Size

### *3.5.1 Direction of Effect*

My initial summary measure for the meta-analysis was Cohen's *d,* which is the difference between the means of each condition (in this case, English and an East Asian language) divided by the pooled standard deviation (Borenstein, 2009). The variable created by this equation is a standardized indicator of differences in means as a function

of language, which can then be compared across studies. In investigating effect size, I also identified the direction of the effect. In many cases, the pattern of results demonstrated linguistic consistency; in which speaking a language activates the cultural knowledge associated with that language, leading to behaviour that is consistent with that language. I independently codedeffect sizes for differences that represent linguistic consistencyas positive (176 effects). In other cases, the pattern of results demonstrated linguistic inconsistency, in which speaking a second language makes one response more consistent with the protypical behaviour of their first language. I independently coded effect sizes that represented linguistic inconsistency were as negative (85 effects). Eleven effect sizes were exactly 0 (i.e., participants scored exactly the same in the English and East Asian conditions).

### *3.5.2 Adjustment for Comparison of Effects Across Design*

Effect size and confidence intervals were calculated using David Wilson's effect size calculator https://campbellcollaboration.org/effect-size-calculato.html (A companion to Lipsey & Wilson, 2001). Means and standard deviations were used in the calculations whenever possible, and when it was not possible to do so, other available statistics (i.e., *F* and *t*) were used.

The inclusion of both within-subjects effects and between-subjects effects in the analysis required a minor statistical adjustment for the sake of compatibility. Nearly half of our effect sizes (109 effects) came from within-subject designs, which is problematic because Cohen's *d* can only be calculated for within-subjects designs when provided with the standard deviation of the difference and the correlation between the two scores. This

information was, unfortunately, unavailable from our sources, and hence the effect sizes we calculated from the within-subjects designs were merely approximations. I therefore provided an adjustment of .50 as an estimate of correlation and multiplied the *d*'s of the between-subjects effects by 2*sqrt(1-*r*) as recommended in Borenstein and colleagues (2009). An additional adjustment recommended by Hedges and Becker (1986) to use weighted standard deviations instead of pooled standard deviations when sample sizes were small ($n < 20$) was applied to all effect sizes. As a result, the final effect sizes used in the analyses were not Cohen's *d*, but Hedges' *g*.

### 3.6 Data Analysis

To compare effect sizes from multiple samples from multiple studies, I performed a multilevel meta-analysis using the *metafor* package for R statistical software (Viechtbauer, 2010) version 4.4.0. Most of the included studies had multiple effect sizes from the same group of participants, which violated the assumption of independence, and made it inappropriate for either a fixed-effects model or a random-effects model (Borenstein et al., 2009; Marsh et al., 2009). Therefore, I created a multilevel model with effect size at level 1, and sample at level 2 using robust variance estimates. In a multilevel meta-analysis, several effects may be included from the same sample, meaning that the standard errors of those effects would be correlated/dependent. Robust variance estimates were used, therefore, to account for the dependence between effects within the model (Harrer et al., 2021).

Following the analysis, I tested for publication bias in three ways. First, to determine if there was publication bias against small effect sizes, I compared the effect

sizes of the studies, using publication status ("dissertation" and "article") as a moderator. A lack of publication bias would mean that effects of all magnitudes would be contained in both the published and unpublished manuscripts. It is possible, however, that a bias in favor of large effects would lead to more large effects in the published research than in the unpublished research. Using publication status as a moderator would detect this difference and serve as evidence of publication bias.

Second, I generated funnel plots mapping standard error and effect size to look for effect size asymmetry. A funnel plot mapping effect size and standard error with no publication bias would appear symmetrical, with an equal distribution of large and small effects with large standard errors. It is possible, however, that a bias in favor of large effects of all sizes would result in an asymmetry (with a lack of small effects with large standard errors). Third, I ran a test of publication bias using regression analysis, as recommended by Egger et al. (1997), in which effect size was regressed on standard error to determine if there is a significant slope to the distribution (an unbiased sample would have a horizontal, non-significant slope). This method is a more precise way of examining the asymmetry of the sample using regression instead of a visual analysis.

# 4.0 Results

## 4.1 Preliminary Analyses

As a preliminary measure, I conducted analyses to identify outliers. As Tabachnick and Fidell (2007) recommended, outliers were identified as any effect that was more than 3.29 standard deviations from the grand mean, based on the absolute value of *z*-scores. Through this procedure, three effect sizes were identified as outliers. Each outlier was re-calculated from the source article to ensure that the values were not due to a coding error, and all three had been calculated properly. It is worth noting that several of these outliers were calculated using *F* or *t* instead of the preferred *M* and *SD*, and so this calculation might explain the relatively high values.

To determine the influence of these outliers, I conducted all relevant analyses with the outliers included, and then again with the outliers excluded. There were no meaningful differences in terms of significance of mean effect sizes and significance of moderators within the total sample; however, it did affect the significance of moderators when analyses were conducted within-samples. Despite my confidence that these effects were not coded in error, I determined it misleading to report the analyses here with the outliers included; I reasoned that if a moderator is found to be significant when a large outlier is included, but not significant when that large outlier is excluded, then the significance of that moderator would have more to do with the influence of that moderator rather than a true population parameter. The outliers are identified with asterisks in Appendix A, and are thereafter excluded from the main analyses, but mentioned where relevant in footnotes. The final sample consisted of 269 effect sizes

drawn from 63 independent samples reflecting the responses of 6,693 participants. Characteristics of the studies included in the meta-analysis are summarized in Appendix A.

Next, I compared two versions of the null model: a two-level model (considering effect size as level 1; and sample as level 2) and a three-level model (considering effect size as level 1, sample as level 2, and authors as level 3). The purpose of doing so was to determine if the more complex three-level model provided significantly more explanatory power than the more parsimonious two-level model. This model comparison relied on a comparison of deviance and determined that there was no significant difference in fit between the two models ($p = .31$). Accordingly, the following data analyses relied on the more parsimonious two-level model, which included just effect size and sample.

### 4.2 Overall Effect Size

To determine the overall estimated effect size in the full sample, I used the two-level null model, which I determined was the best fit for these data in the previous analysis. The intercept of this two-level null model revealed a mean estimated effect size of $g = .27$, (95% CI: .16, .38), which was significantly different from zero ($p < .001$). This small but significant positive effect suggests that participants, in general, do provide responses that are more consistent with established East Asian cultural practices when responding in their East Asian language than when responding in English.

Calculation of prediction intervals (Borenstein, 2023) revealed that an estimated 95% of effects would land between $g = -.24$ and $g = .79$. I also examined Cochrane's Q test of heterogeneity to determine if there was significant variation in effects between

samples. This test was significant, $Q$ (268) = 1751.03, $p < .001$, indicating significant heterogeneity between samples, which suggests that the overall effect is not a good representation of the sample, and systematic analyses of moderators are appropriate to account for this heterogeneity.

## 4.3 Moderator Analysis

Each test of the moderators was conducted using the rma.mv command followed with the *robust* procedure from the *metafor* package (Viechtbauer, 2010) in R statistical software (R Core Team, 2019) version 4.4.0. Because the typical standard error fails to account for the possibility of dependence between samples in multilevel analyses, revised (robust) standard errors were necessary for a more accurate test of moderators in multilevel models (Rogers & Pustejovsky, 2020).

### *4.3.1 Effect of Design*

The first moderator tested was Design to ensure that the adjustment I applied to effect sizes from within- and between-subject design resulted in comparable groups. If the groups were found to differ, moderator analyses would be conducted separately for the two types of designs. When design was entered as moderator, the test of moderators was not significant, $F(1, 60) = .89$, $p = .348$. The 95% confidence intervals for the estimated mean effects did not contain zero for within-subjects designs ($k = 106$, $g = .34$, 95% $CI = .16, .52$) or between-subjects designs ($k = 163$, $g = .24$, 95% $CI = .10, .37$).

The finding that this moderator was not significant suggests that effects of within-subjects designs are not significantly different from the effects of between-subjects designs. This finding, along with the application of the adjustment recommended by

Borenstein and colleagues (2009), gave me confidence to proceed with my analyses using both within-subjects and between-subjects effects. Nonetheless, to address any remaining concerns about compatibility, I conducted separate

### *4.3.2 East Asian Language*

East Asian Language was a significant moderator ($p$ = .03). The 95% confidence intervals did not contain zero for Chinese ($k$ = 213, $g$ = .32, 95% *CI* = .20, .43), but did contain zero for Other (k = 56, g = .13, 95% CI = -.04, .30). Thus, the effect appears to be significant only for speakers of Chinese, and not speakers of the other East Asian languages used.

### *4.3.3 Task Type*

Task Type was a significant moderator ($p$ = .01). Mean estimates are included in Table 2. The confidence intervals of the mean estimates did not contain zero for any of the task types, suggesting that each task demonstrates a meaningful difference between language conditions. Tukey multiple comparisons indicated that the estimates for behaviour were significantly larger than the estimates for values ($p$ = .003), which suggests that the effect of language is stronger on behavioural tasks than on value-related tasks.

**Table 2**

*Results for Task Type as a Moderator, with outliers excluded.*

| Category | Effects | *g*(95% CI) |
|---|---|---|
| Self-concept | 72 | .20 (.02, .39) |

| | | |
|---|---|---|
| Values | 57 | .26 (.05, .47) |
| Cognition | 99 | .32 (.13, .51). |
| Behaviour | 41 | .37 (.15, .59) |

*Note.* The table contains the number of effect sizes (Effects) and the mean weighted *g* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with $p < .05$ when the 95% CI for *g* does not contain zero.

#### *4.3.4 Non-significant Moderators*

Language acquisition style was not a significant moderator ($p = .13$), nor was Continent ($p = .83$), or City ($p = .81$).

### 4.4 Adjustment for Multiple Moderators

Using a large number of separate analyses may increase the risk of type-one error. Therefore, for a more conservative perspective on the results, I applied the Holm-Bonferroni adjustment method to each of my moderators. This was done only for the significant moderators in the analysis. The pre-adjustment and post-adjustment *p*-values are presented in Table 3. After the adjustment, both significant moderators (East Asian language and task type) remained statistically significant ($p < .05$).

**Table 3**

*Holm Test of Moderators*

| Moderator | *P* | Rank | Transformed |
|---|---|---|---|
| EAL | .03 | 1 | .03 |

| | | | |
|---|---|---|---|
| OUT | .01 | 2 | .02 |

Note: *p*-values (p) for each significant moderator (moderator) were ranked from largest to smallest (rank) and multiplied by their rank.

### 4.5 Publication Bias

The current analysis includes effects from both published work ($k$ = 210), and unpublished work ($k$ = 59). Two effects were taken from research which was made available by request to the author, between the point of the research being accepted by a journal and the actual publication of that research, so it was counted as published.

There are a variety of methods for testing publication bias, and each has corresponding strengths and weaknesses. To ensure a thorough test for publication bias, I used three separate methods. First, I sought to determine if the effects taken from published articles were significantly different from the effects taken from unpublished works through moderator analysis. Publication status was a significant moderator when examining all of the effects together ($p$ = .01), with the effects from articles ($g$ = .31, 95%$CI$ = .18, .44) being larger than the effects from dissertations ($g$ = .08, 95%$CI$ = -.02, .19). Publication status was also a significant moderator when just using the between-subjects effects analysis ($p$ = .04), with the effects from articles ($g$ = .26, 95%$CI$ = .11, .40) being larger than effects from dissertations ($g$ = .08, 95%$CI$ = -.01, .17). Publication

status was not a significant moderator for the within-subjects designs ($p$ = .09)[4]. This method is fairly simple and intuitive, but is not as informative as the next two approaches.

Second, I generated funnel plots where each effect was plotted based on size and precision (with standard error plotted as a function of effect size). In an unbiased distribution, these distributions would be essentially symmetrical. Unfortunately, publication bias often occurs when studies with small sample sizes and small effects are less likely to be published than studies with small sample sizes and large effects. On a funnel plot of effect sizes, with effect size on the X axis and standard error on the Y axis, publication bias would appear as an asymmetry, with fewer effects in the lower left quadrant (small effects, and high standard error) than in the lower right quadrant (small effects, low standard error)(Egger et al., 1997). I generated funnel plots for (1) all of the effects together, (2) the between-subjects effects, and (3) the within-subjects effects, which can be seen in Figure 2. Distribution of the effects appears to show a slight asymmetry, consistent with the findings above. Although this provides more information than a test of moderation, it is still imprecise. The third approach involves a more precise examination of this information.

Third, I used an approach recommended by Rodgers and Pustejovsky (2020) to test for publication bias in multilevel analyses. Most traditional forms of publication bias

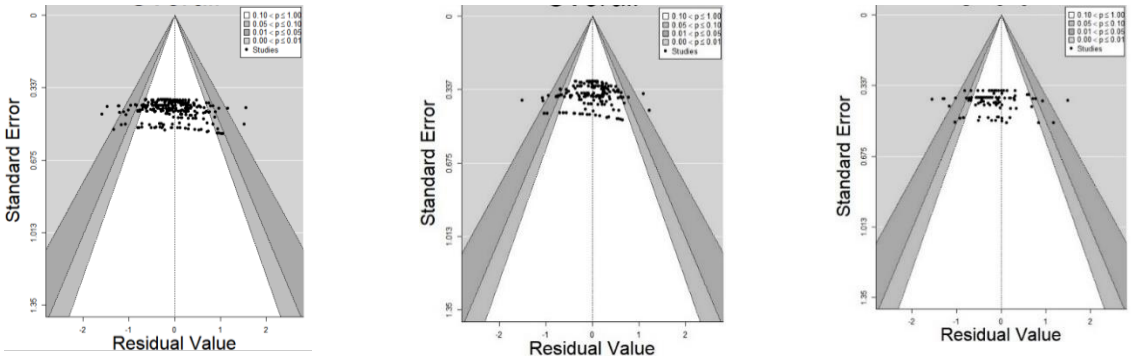
---

[4] When the outliers were included, publication status was a significant moderator for the within-subjects analysis ($p$ = .04), with the effects from articles ($g$ = .46, 95%$CI$ = .21, .71) being larger than the effects from dissertations ($g$ = .1, 95%$CI$ = -.15, .34).

testing use the standard error and sample size to get a sense of the relationship between the two. The issue with multi-level meta-analysis, however, is that I have multiple effect sizes using the same standard error, causing dependence between the effects. To address this issue, standard error needed to be re-calculated as a function of sample size. Therefore, I re-calculated variance as a function of sample and used that to create an adjusted standard error. This standard error was then entered as a moderator in the model. When looking at all of the effects together, the slope was significant ($p$ = .04). The slope was also significant when looking at the between-subjects effects only ($p$ < .001), but not the within-subjects effects ($p$ = .96). Thus, across all three approaches the total sample appeared to demonstrate significant publication bias. Although there also appeared to be evidence of publication bias in the between-subjects effects, there was mixed evidence of publication bias in the within-subjects effects.

**Figure 2**

*Funnel Plots for Determining Publication Bias Within the Samples*



*Note.* Funnel plots map precision through standard error on the y-axis and effect size through observed outcome on the x-axis. Plots represent the total sample (right), only between-subjects effects (middle) and only within-subjects effects (left).

# 5.0 Discussion

## 5.1 Summary of the Research

The purpose of this dissertation was to better understand the effect of language use on task performance, specifically the effect of language use on task performance among participants who speak English and an East Asian language. The current meta-analysis was influenced by the linguistic relativity hypothesis (Whorf, 1956), according to which the language that we use can affect the way that we interact with the world. The dynamic constructivist approach (Hong et al., 2000) extends this idea, suggesting that language affects the way we interact with the world by activating the cultural knowledge associated with that language. The activation of this cultural knowledge may result in linguistic consistency, where behaviour aligns with the cultural knowledge associated with the language being spoken (e.g., behaving like monolingual Chinese when speaking Chinese, but also behaving like monolingual Westerners when speaking English). Activation of cultural knowledge may also produce linguistic inconsistency, where behaviour contradicts the cultural knowledge associated with the language being spoken (e.g., behaving like monolingual Westerners when speaking Chinese, but behaving like monolingual Chinese when speaking English).

Both linguistic consistency and linguistic inconsistency have been demonstrated in studies with English/East Asian bilinguals, who were been randomly assigned to perform a culturally relevant task in either English or their East Asian language. Up until this point, however, the overall size of the difference in performance between language and overall direction of that effect (suggesting a dominant trend toward either linguistic

consistency or inconsistency) had not been examined. Through the current research, I sought to understand the overall size of this difference and direction of the effect by collecting all relevant research in which English/East Asian bilinguals were randomly assigned to complete the same task in either English and their East Asian language (or in both languages). Furthermore, by considering the theoretically relevant moderators, I sought to understand the limits of this effect based on the East Asian language that was used, participants' language acquisition style, the country in which the study was conducted, the city in which the study was conducted, the design of the study, and the task type used for the comparison between languages.

### 5.2 Summary of Results

The results of this meta-analysis demonstrate evidence for the effect of language on culturally relevant phenomena. When considering 269 effect sizes drawn from 63 independent samples, I identified a significant mean effect size of $g = .27$. These findings demonstrate that there is indeed a meaningful difference in performance on culturally relevant tasks, depending on which language is used when the English/East Asian bilingual participant completes the task. These differences correspond with the cultural knowledge associated with the language being spoken – speaking an East Asian language tends to result in responses that are more consistent with observed patterns in East Asian culture (compared to Western culture), which is consistent with the dynamic constructivist approach (Hong et al., 2000). The current meta-analysis provides further boundary details regarding when linguistic consistency is observed. This linguistic consistency pattern appears to be stronger when the East Asian language is Chinese than

when it is a different East Asian language, and when the tasks involve a behavioural response instead of reflection on values. The pattern does not appear to be meaningfully different based on the design of the study, when participants acquired their two languages, or where the study took place.

### 5.3 Interpretation of Overall Effects

There are several points to note concerning the overall effect of language on task performance in the collected studies. First, it is notable that regardless of the design of the studies, and the combination of those designs in the analysis, the overall effect of language was significant. This finding can be interpreted as supporting the linguistic relativity hypothesis – the language being used in the task had a significant effect on responses, resulting in meaningful differences between tasks completed in English and the relevant East Asian language.

Second, it is notable that each of these overall effects were positive, indicating an emerging directionality: when responses in an East Asian language differed from responses in English, the responses in the East Asian language were more likely than those in English to *correspond* with patterns typically observed when examining East Asian culture (i.e., a pattern of linguistic consistency). Therefore, I can conclude that when East Asian bilinguals respond differently depending on the language with which they complete a task, those differences are most likely to be consistent with established patterns of differences between East Asian and Western participants. One explanation for this pattern of linguistic consistency can be provided through cultural accommodation (Bond & Yang, 1982), and the dynamic constructivist approach more broadly (Hong et

al., 2000). By these explanations, when participants in this analysis completed tasks in their East Asian language, it may have activated cultural knowledge associated with that task. That knowledge then would have led them to complete that task in ways that were similar to prototypical speakers of that East Asian language.

Third, I saw heterogeneity in the total model. This pattern suggests that there was sufficient diversity in the total effect that simply examining the overall effect would give us an incomplete picture of the effect and moderator analyses were therefore necessary.

## 5.4 Significant Moderators

Two moderators that were found to explain a significant amount of variance between effects were Language and Task Type, which I will discuss next.

### *5.4.1 Language*

When looking at all of the effects, I observed that Language is a significant moderator. Specifically, there is a greater difference between tasks completed in English and Chinese than between tasks completed in English and other East Asian languages. There are several possible explanations for this difference. There might be something specific about the Chinese language that creates a stronger contrast to Western culture, resulting in larger effects. Indeed, many of the distinctions between Eastern and Western culture are associated with Confucian philosophy, which originated in China (Hofstede & Bond, 1988). It could therefore be that, since the heart of the distinction between Eastern and Western culture lies in Confucian/Chinese philosophy, the strongest Confucian/Chinese values are elicited when speaking the language from that area. Alternatively, it could be that the non-Chinese language groups present a weaker contrast

to Western Culture than the Chinese language group. Two of the most prominent non-Chinese languages within the non-Chinese group were Korean (seven effects) and Japanese (29 effects). Both Korea and Japan are democratic countries, have had strong political and military ties with the United States for over 70 years, and are more Westernised than China. China, however, is a Communist country and spent much of the 20th Century deliberately isolating itself from Western influence (Dian & Meijer, 2020). It is perhaps not surprising, then, that a language coming from China results in a stronger contrast with English – and the associated Western values - than languages coming from relatively more Westernised cultures.

There are alternative explanations for these results. It could also be that there is more diversity within the "other" category, and that diversity is complicating the effect of language use on task performance. Although Chinese represents several dialects (Mandarin and Cantonese), it represents a single language from a single culture. My "Other" group, however, represents six different languages representing six different cultures.

As mentioned in section 3.3.1.2 (Language), distinguishing Chinese from all other East Asian languages was not motivated by purely theoretical reasons, but also the practical and statistical necessities of the meta-analysis. The majority of studies collected within the current meta-analysis used some form of Chinese, and there were comparatively fewer studies wherein any other language was used, so comparing between each individual language would be unsuitable. For that reason, all other languages were grouped together.

With these explanations and caveats taken into consideration, it appears that the effect of language on task performance is stronger when the tasks being compared are conducted in English and Chinese than in English and other East Asian languages.

### *5.4.2 Task Type*

Task Type was a significant moderator (although this should be interpreted with caution, as discussed in the subsequent paragraph). Despite significant differences between English and the East Asian language for all four categories of tasks (behaviour, values, cognition, and self-concept), the effects from behavioural tasks were larger than the effects from tasks that required reflection on values.

One interpretation for this finding is that the shifts in behaviours reflect impression management more than shifts in core values because a person's behaviour is presumably more controllable and easily monitored than their core values. Although arguably people may consciously present how they speak or provide different sets of non-verbal cues in different situations, their core beliefs tend to be more central to who they are and should therefore be less likely to shift from one situation to another.

However, many of the behavioural effects included in my analysis came from observing body language (e.g., Kubota, 1991), dialogue length (e.g., Wang et al., 2010) or eye movements (Lee et al., 2018) that are behaviours performed relatively automatically, that may require less conscious thought than reporting one's beliefs and values. Arguably, reporting one's values is a more deliberate and conscious act than behaviours performed without the same level of thought. Indeed, many of the value effects included in my analysis came from self-reported surveys about values and beliefs

(e.g., Boucher et al., 2011; Earle, 1969; Ralston et al., 1995; Yang & Bond, 1980, Zhang, 2018) or support for a product or group (e.g., Chiou & Mercado, 2016; Moriuchi, 2012; Wang, 2014). It is therefore possible that these kinds of tasks require conscious self-report that may be more susceptible to impression management than behavioural measures. Therefore, a more likely interpretation for the larger effect of behavioural measures is that shifting in response from one language to another is more easily reflected in everyday behaviours where culture is *performed* without much conscious thought, rather than value tasks where culture is consciously *reported* after careful deliberation.

This consideration could lead one to interpret the observed differences in effects consistently with an idea already presented above, that when people complete tasks differently as a function of language, it is more than impression management. In other words, people are not just self-reporting stronger Western values when speaking English and stronger East Asian values when speaking Chinese as a demand characteristic of the study. Certainly, impression management may be a part of the equation, but the finding that behavioural tasks had an overall larger effect than tasks related to self-reported values is consistent with the idea that something beyond conscious impression management is at work; it may suggest that the language one uses to complete a task has a much broader effect on the bilingual individual.

Nevertheless, compared to the effect of Design and Language, this effect of Tasks should be interpreted with a bit more caution. As discussed in section 3.4 (Coding Validity) the coding of Task Type proved to be difficult for both myself and my coders, as the distinction between categories was much less clear than for Design or Language,

and required more interpretation on our part. As a result, initial inter-rater reliability was fairly low, reflecting the higher subjectivity for coding of Task Type. Part of the issue is that this study sought to incorporate a wide range of diverse tasks and group them into fairly general categories. To compensate for this, I made a note of each discrepancy on task type coding and sent the list of discrepancies back to the original coder. We then independently re-examined each relevant code, and met to explain our reasoning for each discrepancy. Based on this discussion, we came to an agreement on a code. One might argue that such an abstract and unsystematic process might not be appropriate for analyses such as these, and there is value in such an argument. Indeed, it could have been beneficial to have a third coder re-examine each effect to ensure that the codes would be reproducible; however, that would simply add a third voice to the conversation without addressing the fact that the moderator is still abstract, diverse, and ultimately subjective. I would argue, however, that the goal of considering Task Type was worthwhile, and our approach to this topic was sufficiently systematic to justify some (admittedly cautious) consideration. Anyone who follows up on this study should take due precautions when considering the coding of Task Type. One option could be to limit the tasks to a specific set of instruments or scales (as was done by Oyserman and Lee in 2008), and pre-sort them into Task Type categories. Another option could be to employ a third coder, so that the Task Type coding could be decided by a simple two-thirds majority. Having either of these procedures planned for Task Type from the beginning could prove beneficial in future research.

### *5.4.3 Publication Status*

Although the moderator of Publication Status was examined within the context of publication bias, the moderator of Publication Status was significant. In other words, the effects taken from published research tended to be much larger than effects taken from unpublished research. This difference, along with the asymmetry of the plotted effects, and the results of the regression of sample size and standard error is evidence of publication bias. Possible explanations for the presence of publication bias in this analysis, as well as implications of this finding are discussed in section 5.7 (Limitations and Future Directions).

## 5.5 Non-significant Moderators

Four moderators were not found to explain a significant amount of variance between effects: Design, Continent, Language Acquisition Style and City. In each of these cases, I failed to find a significant difference in the size of the effects between levels of the moderator. On the one hand, the goal of these kinds of analyses is often to determine the potentially relevant variables that do *not* moderate the overall effect (Borenstein et al., 2009), and so some exploration of the moderators that did not reach significance is valuable. On the other hand, it is dangerous to put too much weight on non-significant findings, as the absence of an effect could simply be due to a lack of power and type II error. With due consideration to both possibilities, in the next section, I interpret what the lack of a significant difference (if it is indeed a true reflection of a lack of an effect) may mean for these moderators.

### *5.5.1 Design*

When looking at the total effect, I observed that Design was not a significant moderator. Specifically, I observed that the effects from within-subjects designs were not overall larger than the effects from between-subjects designs. In other words, there was not a greater difference between English and East Asian language responses when both sets of responses came from the same people than when they came from two separate groups. This finding suggests that, for the sake of this analysis, effects taken from studies using between-subjects designs are not meaningfully different than effects taken from within-subjects designs.

It is possible that there are actually meaningful differences between within-subjects and between-subjects effects, but I was unable to detect them. For example, looking at effect size, or the pattern of significance (in Appendix C), does not help account for the difference in dependence between groups that I see in within-subjects effects compared to between-subjects effects. Although the application of the compatibility coefficient helped to address this difference, any assumptions of absolute equivalence should be done with caution. Nonetheless, if it is indeed the case that there is a similar effect of language when tasks are completed within-subjects and between subjects, then it would permit more freedom when designing language studies. The creation of a study design is a complicated process, and many variables need to be considered; however, the results of this analysis suggest that researchers need not be concerned that using a between-subjects design will result in a smaller effect than using a within-subjects design. Thus, this research will serve as an important reference point for future language research.

### *5.5.1 Continent*

When looking at only within-subjects effects, the moderator of Continent was not significant, which suggests that effects taken from studies conducted in Europe, North America, and Asia did not vary significantly in strength. This failure to reach significance could simply be due to type II error. When this moderator was initially conceptualized, I had intended to compare countries instead of continents. There were many countries, however, that only contributed two or three effects, and so grouping all countries into continents was the same kind of practical and statistical decision as grouping all non-Chinese languages into "other". It is therefore possible that I simply failed to detect an effect of continent due to my decision to aggregate countries. If I had sufficient representation from each country, I could have made more precise distinctions that may have yielded more accurate findings (e.g., comparing the size of the effect in China against the size of the effect in other countries). However, this failure to find a meaningful difference between effects based on continent could also mean that the language an English/East Asian bilingual uses to complete a task will still affect their responses, regardless of where that task is completed.

### *5.5.2 Language Acquisition Style*

The moderator of Language Acquisition Style was not significant. Specifically, effects taken from compound bilingual participants (who learned their two languages simultaneously) did not differ in strength from effects taken from coordinate bilingual participants (who learned one language before the other). There are two methodological explanations for why this moderator may not have been significant, even if Language

Acquisition Style does meaningfully moderate the effect of language use on task performance. First, it is possible that the moderating effect of language acquisition style was not detected due to a small sample size. This is often a concern when examining non-significant effects, and should be considered as a possibility. Second, difficulties in the coding process, which I described in section 3.4 (Coding Validity) and reconsider below, may have compromised the coding to the point that the actual effect of language acquisition style was not detected. Thus, it is still possible that language acquisition style moderates the effect of language on task performance, but due to methodological issues I was unable to demonstrate that relationship.

It is also possible, however, that there is not a meaningful difference in the size of the effect of language on task performance between compound and coordinate bilinguals. One might expect to observe larger effects from coordinate bilinguals than compound bilinguals, as people who learned different languages at different times may be more likely to associate those different languages with different cultural meaning systems. That is, one might expect that the separation in coordinate bilingual language learning would result in two distinct languages and two distinct meaning systems. By contrast, the blending in compound bilingual language learning would result in a more overlap between meaning systems.

Another concern is that I might have conflated language acquisition *style* – that is the sequence in which people learn their language – with language acquisition *context*, the broader set of social and cultural circumstances that are involved with learning a language. The assumptions above presume that when compound bilinguals learn two languages at the same time, they are being learned in the same context, but that may not

be the case: it is entirely possible that someone can learn two languages at the same time, but in very different contexts. For example, a child could spend their days learning Chinese with Eastern-leaning grandparents, and the rest of their time learning English with Western-leaning parents. Alternately, a child could learn Chinese at home with their Eastern-leaning parents, and English out in the world with their Western-leaning friends, and through Western-leaning media. In locations with a history of colonialism, such as Hong Kong, the culture itself becomes a blend of Eastern and Western ideas, and so people will simultaneously learn both sets of ideas as they simultaneously learn both languages. Either way, both languages are learned at the same time, but in different contexts. This speculation is consistent with Gardner's socio-educational model of second language acquisition, which suggests the way we acquire a second language can be affected by the motivation of the language learner (Lalonde & Gardner, 1984). This model was the subject of a meta-analysis by Masgoret and Gardner (2003), which demonstrated that not only motivation, but also attitudes toward the second language learning environment can have an effect on how a second language is learned. Thus, the importance of language learning environment should be appreciated.

These assumptions also presume that all second language learning comes with robust cultural meaning systems; however, it is possible for someone to learn a second language while acquiring very little cultural knowledge alongside it. For example, a 34-year-old graduate student might spend the first three decades of his life learning English, and then take a Chinese course on the popular language-learning application *Duolingo*. He could spend hours every day for years learning vocabulary and grammar rules without ever interacting with Chinese culture. He would certainly be learning another language,

but it would be difficult to argue that he is learning another culture. Research on language instruction suggests that the most effective language education programs should include training in the cultural context of a language, as well as the mechanics of spelling, grammar, and syntax (Hossain, 2024). Indeed, the assumption of all language being associated with cultural knowledge is at the heart of the dynamic constructivist approach, and so if individuals learn a language outside of any kind of cultural context, the dynamic constructivist approach no longer applies. All of this suggests that efforts should have been made to understand each sample's language acquisition context, and not simply language acquisition style.

Unfortunately, language acquisition style was difficult to glean from the collected articles: few articles provided enough information about participants to clearly determine if they were compound or coordinate bilinguals, and even fewer actually used those specific terms. Indeed, most of the effects were coded as either "mixed" (because compound and coordinate bilinguals were analyzed together) or "unclear," because we did not have sufficient information to determine whether they were compound or coordinate bilinguals. This is why initial inter-rater agreement for this variable was quite low – often because one coder would categorize an effect as "unclear" but the other coder would notice some characteristic of the sample that could merit a more precise classification, and that difference had to be discussed. It is highly likely, therefore, that examining language acquisition context (a much more complex phenomenon than language acquisition style) using this meta-analytic approach would be equally frustrating.

It could be that information relevant to language acquisition style was lacking in the collected studies simply because that was not a specific priority for the authors. None of the included studies contained specific hypotheses about differences between compound or coordinate bilinguals. Perhaps sampling was an issue, and a desire to get a sufficient number of participants meant that it was not methodologically feasible to distinguish between compound or coordinate bilinguals in their sample. Nevertheless, the results of this meta-analysis (gleaned from information we were able to collect) suggest the possibility that that there is no meaningful difference in the effect of language on task performance for people who learned their two languages one after the other, and those who learned their two languages at the same time.

### *5.5.3 City*

When looking at only the within-subjects effects, the moderator of City was not significant. Specifically, effects taken from Hong Kong were not meaningfully different from effects taken elsewhere in the world. Again, this lack of effect could be a type II error, and there may actually be a difference in the effects observed within Hong Kong and outside of Hong Kong. However, unlike language and continent, which were grouped together out of practical necessity, the decision to compare effects from within Hong Kong to effects outside of Hong Kong was theoretically premeditated. Hong Kong is a very interesting place for cultural psychologists. On the one hand, it has been an important center of trade and culture for centuries, and so one would expect a strong representation of East Asian culture there. On the other hand, Hong Kong experienced a century of British colonial rule, only returning to China in 1997 (Hung, 2022). Therefore,

one would expect a strong representation of Western culture as well. Because of this interesting blend of cultures, it has become a popular location for studies on frame switching, resulting in the inclusion of 17 studies and 126 effects in our study. However, this unique situation may cause some to question the generalizability of the effects found in Hong Kong, suggesting that perhaps there is something particular about Hong Kong that is behind these effects.

Through this meta-analysis, however, I was unable to find a meaningful difference in the size of effects taken from studies conducted in Hong Kong, compared to studies that were conducted outside of Hong Kong. Although this does not completely dispel concerns about the possibility of Hong Kong's unique situation affecting results in some way beyond the scope of this analysis, based on the information I used in this analysis Hong Kong's unique situation appears to yield effects similar in magnitude to effects found outside of Hong Kong. The implication, therefore, is that concerns about the effect of language on task performance being stronger or weaker because the studies took place in Hong Kong can be at least partially lessened by the non-significant finding here.

Taken together, these non-significant moderators provide additional context to the overall effect. The effects in this analysis did not differ based on where the studies were conducted, or the type of bilingual sample in the studies.

### 5.6 The Contribution of This Analysis

The current meta-analysis offers significant contributions to the literature in several ways. First, it addresses a long-standing question in the field that had not been

adequately answered. Although there is extensive evidence documenting the effect of language use in East Asian/English bilingual language users, and it seemed very likely that cultural knowledge played a role, there has not been a clear pattern established in the literature for how these findings fit together. When considering all effects together, the effect was small, but positive ($g$ = .27). The size of this effect will provide useful context to researchers who want to examine the effect of language on the task performance of English/East Asian bilinguals: knowing the size of the overall effect will allow researchers to perform more accurate a priori power analyses, so that they can know with confidence how many participants they will need to recruit to detect this small effect.

Second, this study provides an interesting elaboration upon the results of Oyserman and Lee (2008). When they examined language as a tool for priming individualism and collectivism, they observed that it had a generally small and heterogeneous effect ($d$ = .10). They also observed that it had the smallest effect on values, compared to other types of tasks. Although I found a slightly larger overall effect ($g$ = .27), it is still small and, based on the result of our Q test, heterogenous. I also found that the effect of language on values was smaller than the effect of language on behaviour. Thus, my observations are consistent with Oyserman and Lee's finding that the effect of language as a prime is generally small, heterogenous, and less effective on tasks that require self-reported values.

This study differs from Oyserman and Lee in scope and intention. They examined the effect of languages from around the world on levels of individualism and collectivism. Although I comparatively narrowed my scope to include only East Asian languages, I also widened my scope to examine the effect of language on a broader range

of culturally relevant phenomena. Oyserman and Lee used 10 studies where Chinese was used to prime individualism and collectivism, and found a small, negative effect. After examining over four-times as many studies where an East Asian language was used to activate a wider range of cultural knowledge, I found a small positive effect. Perhaps this difference occurred because tasks which allow for an assessment of individualism and collectivism (the exclusive focus of language studies in Oyserman and Lee) are particularly likely to make one aware of their ethnolinguistic identity, which makes ethnic affirmation more likely (Chen & Bond, 2007). Whatever the reason for the difference, the entirety of this meta-analysis expands upon a small part of Oyserman and Lee's meta-analysis, providing a much broader exploration of the effect of language on task responses in English/East Asian bilinguals.

Third, this finding also adds to our theoretical understanding of the effect by providing supportive evidence for the linguistic relativity hypothesis (Whorf, 1956), demonstrating that people do indeed respond to a task in a meaningfully different way when using English than when using an East Asian language. Furthermore, not only does language affect the way that English/East Asian bilinguals respond to tasks, this meta-analysis has demonstrated that the pattern of difference typically corresponds with the cultural knowledge associated with that language. This latter finding should be considered as supportive evidence for the dynamic constructivist approach (Hong et al., 2000), that when behaviour changes as an effect of language, the cultural knowledge associated with that language plays an important role in the directionality of that change.

As noted in section 1.6.2 (The Dynamic Constructivist Approach to Culture and Cognition), the seeming inconsistencies between the linguistic relativity hypothesis and

dynamic constructivist approach can be reconciled when one recognizes that language, and the culture of that language, work in parallel. Nonetheless, it is worth noting that all of the studies in this analysis were included because they examined culturally relevant tasks – that is, tasks with previously existing differences between *culture groups*. Thus, all of the studies in this analysis can be explained as reflecting differences in culture.

Fourth, this research provides a meaningful contribution to the field as a resource to inform future research. In terms of the literal outcomes, or dependent variables, it is important to note that all task groups showed an overall significant effect of language ($g > .2$). This research is important confirmation that language affects a wide range of tasks, and so researchers examining self-concept should have as much confidence in a language prime as researchers examining cognition, values, or behaviours. Nonetheless, the finding that effects based on behaviours showed greater difference than effects that were based on values suggests that some variables are more influenced by language than others, and that the amount of deliberation required when responding to a task may affect how much it can be influenced by language. Although this finding should be interpreted with caution due to its abstract and subjective nature, it does suggest that when language affects behaviours, particularly the behaviours of English/East Asian bilinguals, some tasks will perhaps be affected more than others. Future research examining the effect of language on task performance should pay particular attention to how much a task is affected by language, relative to how much conscious thought is required for completing the task. Future research could also attempt to understand how deliberation, or cognitive busyness, might affect the effect of language on responses.

Beyond dependent variables, my moderators revealed several elements of studies that seem to play an important role in the effect of language on task performance. Effects taken from studies that compared English and Chinese, for example, were significantly larger than effects taken from studies that compared English and other East Asian languages, suggesting that perhaps this phenomenon is centered more squarely around Chinese culture than I had anticipated. Thus, when considering the possibility that language affects the way we interact with the world (as suggested by the linguistic relativity hypothesis), or that the effect of language is due to the activation of cultural knowledge (as suggested by the dynamic constructivist approach), we must hereafter consider the possibility that not all languages affect the way we interact with the world to the same degree. This is not to say that we should *only* consider the effect when working with the Chinese language, because the effect was observed across a range of languages; however, it does suggest that *particular* attention should be paid to the possibility of this effect when working with the Chinese language. This particular attention should certainly be paid in academic contexts when language is deliberately used to prime cultural knowledge among Chinese-English bilingual participants, but that particular attention should also be paid in non-academic, social and economic contexts as well. Whenever Chinese-English bilinguals are required to use their multiple languages, the influence of culture should be considered.

It is also worth noting that there were several moderators that did not demonstrate a significant effect. Although it is possible that the lack of effect was simply due to a lack of power to detect a real population parameter, there is also some room to speculate that perhaps the effect of language on task performance does not meaningfully differ based on

whether participants are assigned to conditions within-subjects or between subjects, the geographic location of the participants (continent and city), or the style in which those participants learned their two languages. These moderators also provide a contribution to our understanding of the relationship between language and human behaviour.

By outlining and clarifying these patterns of results present in the literature, this research facilitates the creation of more informed hypotheses. Additionally, by compiling a comprehensive collection of all studies immediately relevant to this area, I have created a resource that will allow other researchers to identify which questions have been thoroughly explored, and which questions require further investigation. By establishing a solid foundation, this research facilitates the advancement of future lines of research in the field.

Finally, the application of meta-analytic techniques in this area helps move the science of psychology in the right direction. Although the issue of replicability has been a concern of psychologists for generations, increased concerns over the lack of replicability in recent years, particularly in fields such as social psychology, present an existential crisis for the entire discipline (Braver et al., 2014; Earp & Trafimow, 2015; Rodgers & Shrout, 2018; Wagner, 2022). Meta-analysis offers a potential solution by synthesising results from multiple studies and multiple labs and testing the generalisability and limits of the existing observations. As such, this research is particularly timely, and helps to shore up our understanding of this area. By conducting this meta-analysis, I am not only contributing to our understanding of the relationship between language use and behaviours of East Asian/English bilingual language users, but also providing an example of the importance of meta-analytic techniques in advancing the field of psychology.

### 5.7 Limitations and Future Directions

Admittedly, this research is not without limitations. First, the results of these analyses are shaped by a particular methodology, and that methodology has its limitations. As with all research, the quality of the analysis depends on the quality of the information used in those analyses. Although steps were taken to follow best practices, there is always the possibility that something was missed or omitted. Additionally, the strictness of my search criteria meant that many otherwise relevant studies had to be omitted. For example, I previously mentioned an excellent study by Boroditsky and colleagues (2011), which demonstrated differences in how English-Mandarin bilinguals conceptualize time and space. Although this article was picked up in the initial search for files, it was screened out of the final analysis due to a lack of random assignment. Thus, the results of this meta-analysis should be understood within the context of the specific kinds of studies that were included - laboratory studies wherein English/East Asian bilinguals were randomly assigned to complete a task in English or their East Asian Language. Future research could examine these same effects, but as they occur in more naturalistic settings, although with such analysis one would not be able to make the inferences about the causal effect of language that I am able to do here.

Second, several groupings in this analysis are based on pragmatics more than theory. Specifically, I intended to analyze the moderators of language and country with each unique language and country as independent levels of each moderator. Unfortunately, due to some issues with sample size this was not feasible, and larger groupings were constructed out of necessity. Those groupings affected how the analyses were conducted, and the eventual results, all leading to some rather broad implications

and unanswered questions about the role of language and country on the effect observed. Thus, some of the results of this meta-analysis should be understood as being influenced by statistical necessity as well as theory. Future research could address this through more deliberate and intentional sampling across language groups and countries if more studies with this information become available.

Another decision made for the sake of pragmatics involved the conceptualization of the *design* moderator. The practical reason for differentiating between within-subjects designs and between-subjects designs was to determine if effects from the different designs were sufficiently comparable for combined analysis. There are other elements of design, however, that could have been examined. For example, it could have been worthwhile to measure the level of anonymity in the studies. Recall that Briley and colleagues (2005) observed language differences in purchasing decisions when participants completed a task anonymously; however, when participants were led to believe that their responses were going to be reviewed by peers, the language effect disappeared. It is possible that when there are interlocutors involved in the design of the study, and/or participant responses are not anonymous, an element of impression management is introduced that is not present in more secluded situations. It could have been worthwhile, then, to distinguish between studies where the language manipulation was done privately or in the presence of others.

Beyond level of anonymity, other design elements such as the amount of language switching required, the medium of instruction (auditory or visual) or the medium of response (speech or writing), could all be worth examining for how they affect the relationship between of language and task performance. Although all of this design

information might certainly have been relevant to this analysis, it would have been incredibly difficult to code. For example, properly coding anonymity would require a great deal of knowledge about the amount of contact participants had with research assistants, information that was not readily available in many of the studies included in this analysis. Thus, although there are elements of design that may moderate the effect of language on task performance, the examination of these elements may be better investigated through experimental studies.

The third limitation of my study has to do with a failure to distinguish between studies that demonstrate linguistic consistency and linguistic inconsistency. Through my analysis, I did not identify a moderator that reliably distinguishes between positive and negative effects. Although the overall effect in my analysis was positive, there were many effects that were negative. In other words, although overall using an East Asian language tends to result in a person responding in a way that is more consistent with East Asian culture, there are many cases where the opposite occurs, and speaking an East Asian language results in a person responding in a way that is less consistent with East Asian Culture. Although there could be many reasons behind these negative effects, the pattern is consistent with the concept of *ethnic affirmation* - that when one speaks a language that is not associated with their native culture, they are reminded of their culture and work harder to represent that culture. These negative effects are by no means rare - in fact, of 269 effects in my analysis, 84 were negative. So although the overall effect is positive, it would be misleading to suggest that negative effects do not occur, or occur less often.

It could be that there is a moderator that I did not identify which could help me distinguish between the overall positive and overall negative effects. I had considered examining bicultural identity integration as a moderator. Bicultural identity integration is the extent to which an individual's two identities are considered together or separately (Benet-Martínez & Haritatos, 2005). Indeed, there is evidence suggesting that individuals with lower bicultural identity integration would be more likely to demonstrate linguistic consistency than people with high bicultural identity integration due to the clearer distinction between the cultural groups (e.g., Benet et al., 2002). Unfortunately, there were not a sufficient number of studies that measured bicultural identity integration for inclusion as a moderator. However, future research may be able to better address the difference between the positive and negative effects by considering bicultural identity integration.

Another moderator worth considering could be language entropy (Gullifer & Titone, 2020), which is understood as a combination of age of language acquisition, amount of language exposure as well as the contexts in which language exposure typically takes place. People with low language entropy (also known as compartmentalized bilinguals) have a short list of scenarios when L1 is appropriate, and an equally short list of scenarios when L2 is clearly appropriate. People with high language entropy (also known as integrated bilinguals) have much more diverse and nuanced experiences with their two languages. After demonstrating this in a sample from Montreal, the authors suggested that language entropy may serve to distinguish between cases when languages are seen as distinct and competing (for compartmentalized bilinguals), and cases when languages are treated as complimentary (for integrated

bilinguals). Given my speculation about how my exploration of language acquisition lacks appropriate use of context (see section 5.3.2 Language Acquisition Style), future research in this area using the idea of language entropy could provide valuable insights.

Although language entropy would allow for an understanding of how, where, and with whom bilinguals *use* their languages, it would also be valuable to know how, where, and with whom bilinguals *acquired* their languages. Particularly for coordinate bilinguals, the way that they learned their second language could affect the way that culture meaning systems are connected to, and activated by, that language. For example, coordinate bilingual A, who learned English at home and Chinese through weekly language lessons, might not develop the same level of cultural understanding and corresponding connection with language as coordinate bilingual B, who learned English at home, and Chinese through several years of immersion in Chinese society. As a consequence, one might expect a weaker effect of language on responses to culturally relevant tasks from coordinate bilingual A than coordinate bilingual B.

Similarly, through our examination of language acquisition style, we sought to understand if there were differences between bilinguals who learned their two *languages* concurrently or sequentially, but it could also be worthwhile to look for differences between biculturals who learned their two *cultures* concurrently or sequentially. Specifically, one might expect that biculturals who acquired both of their cultures concurrently might develop less of a distinction between those cultures than someone who acquired their two cultures sequentially, resulting in a weaker effect of language among the former than the latter. Although information about how participants acquired their two languages and two cultures could be informative, to use that information as an

effective moderator in a meta-analysis one would need to limit their analysis to research wherein that contextual information is reported and used to make comparisons. Although that information was not readily available in most of the studies in this meta-analysis, future researchers who are willing to limit themselves to only studies where that information is available could uncover valuable insights into the relationship between language, culture, and task responses in bicultural bilinguals.

It is also possible that by oversimplifying design, I missed an opportunity to examine how methodological elements may have influenced the direction of effects. As I mentioned above (Section 2.4.1: Task Type), Bond and Yang (1982) suggest that overt language manipulation and outcomes directly related to cultural identity are more likely to result in ethnic affirmation (linguistic inconsistency), whereas relatively subtle manipulations and outcomes less associated with one's culture are more likely to elicit cultural accommodation (linguistic consistency). There is certainly some evidence of this assertion when considering our task type results, as there was a smaller overall effect for value tasks (outcomes directly related to cultural identity) than behaviour tasks (outcomes more indirectly related). The design implications of Bond and Yang's assertion, however, have gone unexplored.

If I had been able to distinguish between overt language manipulations and subtle language manipulations, I could have gained a better understanding of how language use affects the direction of the effect. Indeed, this assertion is consistent with the broader literature on priming and persuasion: subtle priming is more likely to produce an assimilation effect (when a persuasive message is followed by a congruent change in attitudes), however blatant priming increases awareness of persuasion, and is more likely

to activate a contrast effect (when a persuasive message is followed by a change in the opposite direction). Interestingly, this difference can be observed at multiple levels of awareness. For example, Vianello and colleagues (2009) exposed participants to both subtle and blatant persuasive messages, and found that participants tended to show an assimilation effect in their explicit attitudes when exposed to subtle advertisements, but a short-term contrast effect in their implicit attitudes when exposed to blatant advertisements. Although it may be difficult to develop a concise and reliable method for coding subtlety in language manipulation studies, or depth of participant awareness, it may prove worthwhile for future study in this area.

It is also possible that I was unable to differentiate between positive and negative effects due to a false equivalency made for the sake of this meta-analysis: I had considered the effect of language on task performance a single continuum, with positive scores indicating linguistic consistency, a zero meaning no effect, and negative scores indicating linguistic inconsistency. It could be the case that effects that demonstrate linguistic consistency and effects that demonstrate linguistic inconsistency actually differ in other qualitatively important ways that were not detectable through my analysis. Future research could benefit by deliberately separating, and independently analyzing the broader contexts behind tasks that result in linguistic consistency and linguistic inconsistency.

Finally, the results of my analyses should be considered in light of publication bias. Given the prevalence of publication bias in the literature (Rodgers & Pustejovsky, 2021), the presence of publication bias here is not surprising. Moreover, despite my efforts to obtain unpublished research (see section 3.1 Study Selection), I had far more

effects from published research ($k$ = 210) than unpublished research ($k$ = 59). The effects from published studies were significantly larger than the effects from unpublished studies, and a plot of the effect size and standard error showed a disproportionate number of studies with large effects and small samples.

This difference in effects could be due to the fact that my unpublished research effects came from theses and dissertations. In the current academic climate, where publication is key to upward mobility, theses and dissertations with significant differences and large effects will typically be repackaged into one or more articles and submitted for publication. It is possible, therefore, that the theses and dissertations in my analysis have only remained as such because they had lower levels of significance or smaller effects. Thus, it makes sense that the theses and dissertations would have smaller effects, as otherwise they would have been published as articles.

One might be less likely to expect that there would be publication bias in this area, however, as most findings would be interesting and publishable. On the one hand, if a researcher randomly assigned participants to complete a task in English or an East Asian language and they found a significant difference between language conditions, then that would be interesting: it would suggest that language affected performance of the task, and they could likely publish that finding. Furthermore, the results would be interpretable and publishable regardless of the direction of the language effect. If participants behaved in ways that were more similar to Chinese monolinguals when speaking Chinese, that would demonstrate linguistic consistency, and could be interpreted as evidence of cultural accommodation (Bond & Yang, 1982). If the participants behaved in ways that were more similar to English monolinguals when speaking Chinese,

however, that would demonstrate linguistic inconsistency, and could be interpreted as evidence for ethnic affirmation (Chen & Bond, 2007). On the other hand, if a researcher did not find a significant difference between language conditions on the task, that would also be interesting, as it would suggest that the task was robust to the effects of language, and that information might also be worth publishing. The issue, however, is that most research courses teach the widely accepted convention that the findings of a research project must be interpreted in light of the researchers' initial hypotheses. If researchers hypothesized that a task was robust to language and would show no difference between conditions but found a meaningful difference, then they would interpret the observed difference as their hypotheses not being supported and their research not worth publishing. Certainly many researchers operate outside of this convention, and through p-hacking or other shady research practices may find a way to publish their results anyway, however for those who are willing to follow the conventions, the file drawer remains a threat to this area of research.

My conclusions must therefore be interpreted with publication bias in mind: it is possible that the effect of language is over-estimated here, due to the suppression of smaller or less-significant effects. The field of psychology generally could benefit by assessing research based on theoretical and methodological merit, rather than effect size or level of significance. Recent trends toward pre-registration of research and the proliferation of open-access journals are certainly steps in the right direction. This area of research specifically could benefit from a more widespread understanding of the dynamic constructivist approach. Asking participants to complete a task in one of their multiple languages may activate specific sets of cultural knowledge associated with those

languages. This may lead to effects demonstrating linguistic consistency or linguistic inconsistency, however both types of effects are interpretable within the dynamic constructivist approach. An awareness of these possibilities while generating hypotheses should lead to a more comprehensive range of effects being published.

### 5.8 Conclusion

According to the weak version of the linguistic relativity hypothesis, language can affect the way that we think about the world. Although there is evidence for this hypothesis across languages, this dissertation has explored this dynamic specifically among speakers of English and East Asian languages. Across numerous studies I found meaningful differences in the way that English/East Asian bilinguals think, act, view themselves and view the world depending on the language in which they were assigned to think, act, view themselves and view the world. The dynamic constructivist approach (Hong et al., 2000) elaborates on this idea by suggesting that language affects the way we think about the world by activating cultural knowledge systems associated with that language. This dissertation provides support for this idea by demonstrating that when English/East Asian bilinguals are asked to complete culturally relevant tasks in English and their East Asian language, not only are the differences meaningful, but they follow patterns that parallel the prototypical cultural knowledge systems of the respective language. This dissertation should be considered as legitimate evidence for both concepts, and I humbly suggest that any discussion of those concepts should take the preceding work into consideration.

## References


Akkermans, D., Harzing, A.W., & van Witteloostuijn, A. (2010). Cultural accommodation and language priming. *Management International Review, 50*(5), 559–583. https://doi.org/10.1007/s11575-010-0053-0

Alvarado, N., & Jameson, K.A. (2011). Shared knowledge about emotion among Vietnamese and English bilingual and monolingual speakers. *Journal of Cross-Cultural Psychology, 42*(6), 963-982. https://doi.org/10.1177/00220221103811

Anderson, S.R. (2010). How many languages are there in the world? *Linguistic Society of America*. Retrieved from https://www.linguisticsociety.org/content/how-many-languages-are-there-world

Bagozzi, R. P., Wong, N., Yi, Y. (1999). The role of culture and gender in the relationship between positive and negative affect. *Cognition and Emotion*, *13*(6), 641 – 672. https://doi.org/10.1080/026999399379023

Benet, V., Leu, J., Lee, F., & Morris, M. (2002). Negotiating biculturalism: Cultural frame switching in biculturals with oppositional versus compatible cultural identities. *Journal of Cross-Cultural Psychology*, *33*, 492–516. https://doi.org/10.1177/0022022102033005005

Benet-Martínez, V., & Haritatos, J. (2005). Bicultural identity integration (BII): Components and psychosocial antecedents. *Journal of Personality*, *73*(4), 1015-1050. https://doi.org/10.1111/j.1467-6494.2005.00337.x

Berken, J. A., Gracco, V. L., Chen, J. K., Watkins, K. E., Baum, S., Callahan, M., & Klein, D. (2015). Neural activation in speech production and reading aloud in

native and non-native languages. *NeuroImage*, *112*, 208–217. https://doi.org/10.1016/j.neuroimage.2015.03.016

Berken, J. A., Gracco, V. L., & Klein, D. (2017). Early bilingualism, language attainment, and brain development. *Neuropsychologia*, *98*, 220–227. https://doi.org/10.1016/j.neuropsychologia.2016.08.031

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology*, *97*(4), 580–590. https://doi.org/10.1037/0022-0663.97.4.580

*Bokhorst-Heng, W. D., & Caleon, I. S. (2009). The language attitudes of bilingual youth in multilingual Singapore. *Journal of Multilingual and Multicultural Development*, *30*(3), 235–251. https://doi.org/10.1080/01434630802510121

Bond, M. H., & Yang, K. S. (1982). Ethnic affirmation versus cross cultural accommodation: The variable impact of questionnaire language on Chinese Bilinguals from Hong Kong. *Journal of Cross-Cultural Psychology*, *13*(2), 169-185. https://doi.org/10.1177/0022002182013002003

*Bond, M. H. (1983). How language variation affects inter-cultural differentiation of values by Hong Kong bilinguals. *Journal of Language and Social Psychology*, *2*(1), 57–66. https://doi.org/10.1177/0261927X8300200104

Bond, M. H., & Cheung, M.K. (1984). Experimenter language choice and ethnic affirmation by Chinese trilinguals in Hong Kong. *International Journal of Intercultural Relations*, *8*(4), 347-356. https://doi.org/10.1016/0147-1767(84)90014-2

*Bond, M. H. (1985). Language as a carrier of ethnic stereotypes in Hong Kong. *The Journal of Social Psychology, 125*(1), 53–62. https://doi.org/10.1080/00224545.1985.9713508

Bond, M. H., & Lai, T. M. (1986). Embarrassment and code switching into a second language. *The Journal of Social Psychology, 126*(2), 179-186.

Borenstein, M., Hedges, L.V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis.* Wiley.

Borenstein, M. (2024). Avoiding common mistakes in meta-analysis: Understanding the distinct roles of , -squared, tau-squared, and the prediction interval in reporting heterogeneity. *Research Synthesis Methods*, *15*(2), 354–368. https://doi.org/10.1002/jrsm.1678

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition, 118*(1), 123-129. https://doi.org/10.1016/j.cognition.2010.09.010

*Boucher, H. C., & O'Dowd, M. C. (2011). Language and the bicultural dialectical self. *Cultural Diversity and Ethnic Minority Psychology, 17*(2), 211–216. https://doi.org/10.1037/a0023686

Branum-Martin, L., Tao, S., Garnaat, S., Bunta, F., & Francis, D. J. (2012). Meta-analysis of bilingual phonological awareness: Language, age, and psycholinguistic grain size. *Journal of Educational Psychology, 104*(4), 932–944. https://doi.org/10.1037/a0027755

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-

analysis and replicability. *Perspectives on Psychological Science*, *9*(3), 333–342. https:/doi.org/10.1177/1745691614529796

*Briley, D. A., Morris, M. W., & Simonson, I. (2005). Cultural chameleons: Biculturals, conformity motives, and decision making. *Journal of Consumer Psychology*, *15*(4), 351–362. https://doi.org/10.1207/s15327663jcp1504_9

*Chang, A. M., Chau, J. P. C., & Holroyd, E. (1999). Translation of questionnaires and issues of equivalence. *Journal of Advanced Nursing*, *29*(2), 316–322. https://doi.org/10.1046/j.1365-2648.1999.00891.x

*Chen, S. X., & Bond, M. H. (2007). Explaining language priming effects: Further evidence for ethnic affirmation among Chinese-English Bilinguals. *Journal of Language and Social Psychology*, *26*(4), 398–406. https://doi.org/10.1177/0261927X07306984

*Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin*, *36*(11), 1514–1528. https://doi.org/10.1177/0146167210385360

*Chen, S. X., Benet-Martínez, V., & Ng, J. C. K. (2014). Does language affect personality perception? A functional approach to testing the Whorfian hypothesis. Journal of Personality, 82(2), 130–143. https://doi.org/10.1111/jopy.12040

Chen, P., Chung-Fat-Yim, A., Guo, T., & Marian, V. (2023). Cultural background and input familiarity influence multisensory emotion perception. *Cultural Diversity and Ethnic Minority Psychology*. https://doi.org/10.1037/cdp0000577

Cheng, K. H. W., & Broadhurst, R. (2011). The detection of deception: The effects of first and second language on lie detection ability. *Psychiatry, Psychology and Law, 12*(1), 107 - 118. https://doi.org/10.1375/pplt.2005.12.1.107

*Cheng, M., Setoh, P., Bornstein, M. H., & Esposito, G. (2020). She thinks in English, but she wants in Mandarin: Differences in Singaporean bilingual English-Mandarin maternal mental-state-talk. *Behavioral Sciences (Basel, Switzerland), 10*(7). https://doi.org/10.3390/bs10070106

*Chiou, A. Y., & Mercado, B. K. (2016). Flexible loyalties: How malleable are bicultural loyalties? *Frontiers in Psychology*, *7*. https://doi-org.proxy.hil.unb.ca/10.3389/fpsyg.2016.01985

Chung, R. H. G., Kim, B. S. K., & Abreu, J. M. (2004). Asian American multidimensional acculturation scale: Development, factor analysis, reliability, and validity. *Cultural Diversity and Ethnic Minority Psychology, 10*(1), 66-80. https://doi.org/10.1037/1099-9809.10.1.66

Chung-Fat-Yim, A., Chen, P., Chan, A. H. D., & Marian, V. (2022). Audio-visual interactions during emotion processing in bicultural bilinguals. *Motivation and Emotion, 46*(5), 719–734. https://doi.org/10.1007/s11031-022-09953-2

Comrie, B. (n.d.). *Language and thought.* Linguistic Society of America. https://www.linguisticsociety.org/resource/language-and-thought.

Davies, I. R. L., (1998). A study of colour grouping in three languages: A test of the linguistic relativity hypothesis. *British Journal of Psychology, 89*(3), 433-452. https://doi.org/10.1111/j.2044-8295.1998.tb02695.x

de Bot, K. (2000). Psycholinguistics in applied linguistics: Trends and perspectives. *Annual Review of Applied Linguistics, 20,* 224-237. https://doi.org/10.1017/S0267190500200147

Dian, M., & Meijer, H., Networking hegemony: alliance dynamics in East Asia. *International Politics, 57(2),* 131-149. DOI: 10.1057/s41311-019-00190-

*Dixon, D. J. (2007). The effects of language priming on independent and interdependent self-construal among Chinese university students currently studying English. *Current Research in Social Psychology, 13*(1), 1–9. ISSN: 1088-7423 (Electronic).

Diller, K. C. (1970). "Compound" and "coordinate" bilingualism: A conceptual artifact. *Word, 26*(2), 254-261. https://doi.org/10.1080/00437956.1970.11435596

Donnelly, S. (2016). *Re-examining the bilingual advantage on interference-control and task-switching tasks: A meta-analysis.* (Publication No. 10009829) [Doctoral dissertation, City University of New York City]. ProQuest Dissertations & Theses Global.

*Earle, M. J. (1969). A cross-cultural and cross-language comparison of dogmatism scores. *The Journal of Social Psychology, 79*(1), 19–24. https://doi.org/10.1080/00224545.1969.9922382

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00621

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis

detected by a simple, graphical test. *British Medical Journal*, *315*, 629-634. https://doi.org/10.1136/bmj.315.7109.629

Ervin, S.M. (1964). Language and TAT content in bilinguals. *Journal of Abnormal and Social Psychology, 68,* 500-507. doi:10.1037/h0044803

Fedeli, D., Del Maschio, N., Sulpizio, S., Rothman, J., & Abutalebi, J. (2021). The Bilingual Structural Connectome: Dual-language Experiential Factors Modulate Distinct Cerebral Networks. *Brain and Language*, *220*, 104978. https://doi.org/10.1016/j.bandl.2021.104978

Gocken, E., Furnham, A., Mavoreli, S., & Petrides, K.V.(2014). A cross-cultural investigation of trait emotional intelligence in Hong Kong and the UK. *Personality and Individual Differences, 65,* 30-35. https://doi.org/10.1016/j.paid.2014.01.053

Grosjean, F. (2010). *Bilingual*. Harvard University Press.

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*(2), 283–294. https://doi.org/10.1017/S1366728919000026

Guo, T., & Spina, R. (2016). Chinese are more loss averse than British. *Asian Journal of Social Psychology*, *19*(2), 154–162. https://doi.org/10.1111/ajsp.12129

Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Chapman & Hall/CRC Press. ISBN 978-0-367-61007-4.

Hedges, L. V., & Becker, B.J. (1986). Statistical methods in the meta-analysis of research

on gender differences. In J. Hyde, & M. C. Linn (Eds.), *The psychology of gender: Advance through meta-analysis* (pp. 14-50). John Hopkins University Press.

Heine, S. J. (2005). Constructing good selves in Japan and North America. In R. M. Sorrentino, D. Cohen, J. M. Olsen, & M. P. Zanna (Eds.), *Culture and social behavior: The tenth Ontario symposium* (pp. 115–143). Lawrence Erlbaum.

Heine, S.J., Lehman, D.R., Markus, H.R., & Kitayama, S. (1999). Is there a universal need for positive regard? *Psychological Review, 106,* 766-794. doi:10.1037/0033-295X.106.4.766

Heine, S. J., & Hamamura, T. (2007). In Search of East Asian Self-Enhancement. *Personality and Social Psychology Review*, *11*(1), 4–27. https://doi.org/10.1177/1088868306294587

Henrich, J., Heine, S., Norenzayan, A. (2010). The weirdest people in the world? *Behavioural and Brain Sciences, 33(2-3), 61 -135* doi:10.1017/S0140525X0999152X

Hong, Y.-y., Morris, M. W., Chiu, C.-y., & Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist, 55,* 709-720. doi:10.1037//0003-066X.55.7.70

Hong, Y.Y., & Chiu, C.Y. (2001). Toward a paradigm shift: From cross cultural differences in social cognition to social-cognitive mediation of cultural differences. *Social Cognition, 19,* 181-196. https://doi.org/10.1521/soco.19.3.181.21471

Hofstede, G., & Bond, M. H. (1984). Hofstede's culture dimensions: An independent

validation using Rokeach's Value Survey. *Journal of Cross-Cultural Psychology*, *15*(4), 417–433. https://doi.org/10.1177/0022002184015004003

Hofstede, G., & Bond, M.H., (1988). The Confucius connection: from cultural roots to economic growth. *Organizational Dynamics, 16 (4),* 5 - 21, https://doi.org/10.1016/0090-2616(88)90009-5.

Hossain, K. I. (2024). Reviewing the role of culture in English language learning: Challenges and opportunities for educators. *Social Sciences & Humanities Open*, *9*, 100781. https://doi.org/10.1016/j.ssaho.2023.100781

Hung, H.T., (2022). Small city drawing great concerns: Hong Kong's historic role in Sino-US relations. *Webster Review of International History, 2 (2),* 61-87. https://websterreview.lse.ac.uk/articles/42

*Imbo, I., & LeFevre, J.-A. (2011). Cultural differences in strategic behavior: A study in computational estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1294–1301. https://doi-org.proxy.hil.unb.ca/10.1037/a0024070

*Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science*, *14*(1), 39–46. https://doi.org/10.1111/1467-9280.01416

*Ji, L.-J., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology*, *87*(1), 57–65. https://doi.org/10.1037/0022-3514.87.1.57

*Kemmelmeier, M., & Cheng, B. Y.-M. (2004). Language and self-construal priming: A replication and extension in a Hong Kong sample. *Journal of Cross-Cultural Psychology*, *35*(6), 705–712. https://doi.org/10.1177/0022022104270112

*Kim, H. K. (2007). *The role of the learner subjectivity and pragmatic transfer in the performance of requests by Korean ESL learners* (Order No. 3270355). Available from ProQuest Dissertations & Theses Global. (304728982). https://login.proxy.hil.unb.ca/login?url=https://www.proquest.com/dissertations-theses/role-learner-subjectivity-pragmatic-transfer/docview/304728982/se-2

Kim, K. H. S., Relkin, N. R., Lee, K.-M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, *388*(6638), 171–174. https://doi.org/10.1038/40623

Kitayama, S., & Huff, S. (2015). Cultural Neuroscience: Connecting Culture, Brain, and Genes. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Scienes* (1st ed., pp. 1–16). Wiley. https://doi.org/10.1002/9781118900772.etrds0062

Klein, D., Mok, K., Chen, J.-K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, *131*, 20–24. https://doi.org/10.1016/j.bandl.2013.05.014

*Kobayashi, C., Glover, G. H., & Temple, E. (2006). Cultural and linguistic influence on neural bases of "Theory of Mind": An fMRI study with Japanese bilinguals. *Brain and Language*, *98*(2), 210–220. https://doi.org/10.1016/j.bandl.2006.04.013

*Kobayashi, C., Glover, G. H., & Temple, E. (2007). Cultural and linguistic effects on neural bases of "Theory of Mind" in American and Japanese children. *Brain Research, 1164*, 95–107. https://doi.org/10.1016/j.brainres.2007.06.022

*Kubota, M. (1991). *The use of back channel behaviors by Japanese and American bilingual persons* (Order No. 9217399). Available from ProQuest Dissertations & Theses Global. (303923186). https://login.proxy.hil.unb.ca/login?url=https://www.proquest.com/dissertations-theses/use-back-channel-behaviors-japanese-american/docview/303923186/se-2

LaFromboise, T., Coleman, H. L. K., & Gerton, J. (1993). Psychological impact of biculturalism: Evidence and theory. *Psychological Bulletin, 114*(3), 395 – 412. https://doi.org/10.1037/0033-2909.114.3.395

Lalonde, R. N., & Gardner, R. C. (1984). Investigating a causal model of second language acquisition: Where does personality fit? *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement, 16*(3), 224–237. https://doi.org/10.1037/h0080844

Lechuga, J., & Wiebe, J. S. (2011). Culture and probability judgment accuracy: The influence of holistic reasoning. *Journal of Cross-Cultural Psychology, 42*(6), 1054–1065. https://doi.org/10.1177/0022022111407914

*Lee, A. (2018). *The role of language and culture in face and scene processing and description strategies* (Order No. 11002103). Available from ProQuest Dissertations & Theses Global. (2116938011). https://login.proxy.hil.unb.ca/login?url=https://www.proquest.com/dissertations-theses/role-language-culture-face-scene-processing/docview/2116938011/se-2

*Lee, E. S. (2014). *A dynamic approach to acculturation: The influence of cultural priming method on acculturation among Asians* (Order No. 3673780). Available from ProQuest Dissertations & Theses Global. (1652534416). https://login.proxy.hil.unb.ca/login?url=https://www.proquest.com/dissertations-theses/dynamic-approach-acculturation-influence-cultural/docview/1652534416/se-2

*Lee, K. S. (2006). Navigating between cultures: A new paradigm for Korean American cultural identification. *Pastoral Psychology*, *54*(4), 289–311. psyh. https://doi.org/10.1007/s11089-005-0001-2

*Lee, S. W. S., Oyserman, D., & Bond, M. H. (2010). Am I doing better than you? That depends on whether you ask me in English or Chinese: Self-enhancement effects of language as a cultural mindset prime. *Journal of Experimental Social Psychology*, *46*(5), 785–791. https://doi.org/10.1016/j.jesp.2010.04.005

Li, J., Yao, C., Li, Y., Liu, X., Zhao, Z., Shang, Y., Yang, J., Yao, Z., Sheng, Y., & Hu, B. (2024). Effects of second language acquisition on brain functional networks at different developmental stages. *Brain Imaging and Behavior*, *18*(4), 808–818. https://doi.org/10.1007/s11682-024-00865-y

*Li, K. K. (2017). How does language affect decision-making in social interactions and decision biases? *Journal of Economic Psychology*, *61*, 15–28. https://doi.org/10.1016/j.joep.2017.03.003

*Li, Y., Yang, J., Suzanne Scherf, K., & Li, P. (2013). Two faces, two languages: An fMRI study of bilingual picture naming. *Brain and Language*, *127*(3), 452–462. https://doi.org/10.1016/j.bandl.2013.09.005

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.

*Liu, C., Li, L., Jiao, L., & Wang, R. (2021). Bilingual language control flexibly adapts to cultural context. *Frontiers in Psychology*, *12*, 744289. https://doi.org/10.3389/fpsyg.2021.744289

Liu, X., Tu, L., Chen, X., Zhong, M., Niu, M., Zhao, L., Lu, Z., & Huang, R. (2020). Dynamic language network in early and late Cantonese–Mandarin bilinguals. *Frontiers in Psychology*, *11*, 1189. https://doi.org/10.3389/fpsyg.2020.01189

Levinson, S. (1996). Language and space. *Annual Review of Anthropology*, *25,* 353-382. https://doi.org/10.1146/annurev.anthro.25.1.353

Lonnqvist, J.E., Konstable, K., Lonnqvist, N., & Verkasalo, M. (2014). Accuracy, consensus, in-group bias and cultural frame shifting in the context of national character stereotypes. *The Journal of Social Psychology, 154*, 40-58. https://doi.org/10.1080/00224545.2013.843500

*Lyczak, R., Fu, G. S., & Ho, A. (1976). Attitudes of Hong Kong bilinguals toward English and Chinese speakers. *Journal of Cross-Cultural Psychology*, *7*(4), 425–438. https://doi.org/10.1177/002202217674004

Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130,* 266-270. https://doi.org/10.1016/j.cognition.2013.11.004

Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O'Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis

comparing traditional and multilevel approaches. *Review of Educational Research*, *79*(3), 1290–1326. https://doi.org/10.3102/0034654309334143

Masgoret, A.-M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, *53*(S1), 167–210. https://doi.org/10.1111/1467-9922.00227

McCann, D., Weiten, W., Matheson, D. (2022). Psychology: Themes and variations, 6th Canadian edition. Cengage.

*Moriuchi, E. (2013). *Who Am I? A Cross-Cultural Study on Japanese-American Biculturals' Consumption Preference Towards Hedonic and Utilitarian Products* (Order No. 10033675). Available from ProQuest Dissertations & Theses Global. (1774246048). https://login.proxy.hil.unb.ca/login?url=https://www.proquest.com/dissertations-theses/who-am-i-cross-cultural-study-on-japanese/docview/1774246048/se-2

Morris, S. B., & Deshon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105-125. Doi: 10.1037/1082-989X.7.1.105

Nerlich, B., & Clarke, D. (1998). The linguistic repudiation of Wundt. *History of Psychology, 1,* 179-204. doi 10.1037/1093-4510.1.3.179

Nguyen, A.-M. D., & Ahmadpanah, S. S. (2014). The interplay between bicultural blending and dual language acquisition. *Journal of Cross-Cultural Psychology*, *45*(8), 1215–1220. https://doi.org/10.1177/0022022114542849

Nicoladis, E., Nagpal, J., Marentette, P., & Hauer, B. (2018). Gesture frequency is linked

to story-telling style: Evidence from bilinguals. *Language and Cognition*, *10*(4), 641–664. https://doi.org/10.1017/langcog.2018.25

Oyserman, D., & Lee, S.W.S. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychological Bulletin, 134,* 311-342. doi 10.1037/0033-2909.134.2.311

Oyserman, D. (2015). Values, Psychology of. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences (Second Edition)* (pp. 36–40). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.24030-0

*Pattaratanakun, J. A., & Mak, V. (2015). Culture moderates biases in search decisions. *Psychological Science*, *26*(8), 1229–1240. https://doi.org/10.1177/0956797615583979

Perunovic, W.Q. E., Heller, D., Rafaeli, E.(2007). Within-person changes in the structure of emotion: The role of cultural identification and language. *Psychological Science, 18,* 607 – 613. doi 10.1111/j.1467-9280.2007.01947.x

Peterson, M. (2010). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation & Gaming*, *41*(1), 72–93. https://doi.org/10.1177/1046878109355684

Price, P., Jhangiani, R., & Chiang, I. (2015). Research Methods in Psychology – 2nd Canadian Edition - Open Textbook Library. Retrieved May 19, 2019, from https://open.umn.edu/opentextbooks/textbooks/431

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Rákosi, C. (2018). Dealing with the conflicting results of psycholinguistic experiments: How to resolve them with the help of statistical meta-analysis. *Journal of Psycholinguistic Research*, 47, 777-801 https://doi.org/10.1007/s10936-018-9558-7

*Ralston, D. A., Cunniff, M. K., & Gustafson, D. J. (1995). Cultural accommodation: The effect of language on the responses of bilingual Hong Kong Chinese managers. *Journal of Cross-Cultural Psychology*, *26*(6), 714–727. https://doi.org/10.1177/002202219502600612

Ramirez-Esparza, N., Gosling, S.D., Benet-Martinez, V., Potter, J.P., & Pennebaker, J.W. (2006). Do bilinguals have two personalities? A special case of cultural frame switching. *Journal of Research in Personality, 40*, 99-120. doi 10.1016/j.jrp.2004.09.001

Rodgers, J. L., & Shrout, P. E. (2018). Psychology's replication crisis as scientific opportunity: A précis for policymakers. *Policy Insights from the Behavioral and Brain Sciences*, *5*(1), 134–141. https://doi.org/10.1177/2372732217749254

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological methods*, *26*(2), 141-160. doi 10.1037/met0000300

*Romney, A. K., Moore, C. C., & Rusch, C. D. (1997). Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 5489–5494. https://doi.org/10.1073/pnas.94.10.5489

*Ross, M., Xun, W. Q. E., & Wilson, A. E. (2002). Language and the bicultural self. *Personality and Social Psychology Bulletin*, *28*(8), 1040–1050. https://doi.org/10.1177/01461672022811003

*Smith, L., Leung, W. G., Crane, B., Parkinson, B., Toulopoulou, T., & Yiend, J. (2018). Bilingual comparison of Mandarin and English cognitive bias tasks. *Behavior Research Methods*, *50*(1), 302–312. https://doi.org/10.3758/s13428-017-0871-0

*Sakamoto, I. (2001). *Negotiating multiple cultural contexts: Flexibility and constraint in the cultural selfways of Japanese academic migrants* (Order No. 3016948). Available from ProQuest Dissertations & Theses Global; ProQuest Dissertations & Theses Global Closed Collection. (275874417). https://login.proxy.hil.unb.ca/login?url=https://www.proquest.com/dissertations-theses/negotiating-multiple-cultural-contexts/docview/275874417/se-2

Sapir, E. (1956*). Selected writings of Edward Sapir in language, culture and personality*. In D.G. Mandelbaum (Ed.). University of California press.

Schiff, R., & Katan, P. (2014). Does complexity matter? Meta-analysis of learner performance in artificial grammar tasks. *Frontiers in Psychology*, *5*. (2015-09869-001). doi: 10.3389/fpsyg.2014.01084

Sechrest, L., Flores, L., Arellano, L (1968). Language and social interaction in a bilingual culture. *The Journal of Social Psychology*, *76*, 155-161. https://doi.org/10.1080/00224545.1968.9933607

So, W.C. (2010). Cross-cultural transfer in gesture frequency in Chinese-English bilinguals. *Language and Cognitive Processes*, *25*, 1335-1353. https://doi.org/10.1080/01690961003694268

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon/Pearson Education.

Tohidian, I. (2009). Examining linguistic relativity hypothesis as one of the main views on the relationship between language and thought. *Journal of Psycholinguistic Research*, 38, 65 - 74. doi: 10.1007/s10936-008-9083-1

Tu, L., Niu, M., Pan, X., Hanakawa, T., Liu, X., Lu, Z., Gao, W., Ouyang, D., Zhang, M., Li, S., Wang, J., Jiang, B., & Huang, R. (2021). Age of acquisition of Mandarin modulates cortical thickness in high-proficient Cantonese–Mandarin bidialectals. *Journal of Psycholinguistic Research*, *50*(4), 723–736. https://doi.org/10.1007/s10936-020-09716-5

Verhagen, J., Grassmann, S., Kuntay, A. (2017). Monolingual and bilingual children's resolution of referential conflicts: Effects of bilingualism and relative language proficiency. *Cognitive Development, 41,* 10 – 18. https://doi.org/10.1016/j.cogdev.2016.10.003

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1-48. URL: http://jsatsoft.org/v36/i03/

Wagner, J. A. (2022). The influence of unpublished studies on results of recent meta-analyses: Publication bias, the file drawer problem, and implications for the replication crisis. *International Journal of Social Research Methodology*, *25*(5), 639–644. https://doi.org/10.1080/13645579.2021.1922805

*Wang, Q., Shao, Y., & Li, Y. J. (2010). "My way or mom's way?" The bilingual and bicultural self in Hong Kong Chinese children and adolescents. *Child Development*, *81*(2), 555–567. https://doi.org/10.1111/j.1467-8624.2009.01415.x

*Wang, W. (2014). *Advertising to Bicultural Consumers: The Role of Dialectical Thinking and Bicultural Identity Integration on Dual-Focused Persuasive Appeals* (1837035937) [Ph.D., The University of Manchester (United Kingdom)]. ProQuest Dissertations & Theses Global. https://login.proxy.hil.unb.ca/login?url=https://search.proquest.com/docview/1837035937?accountid=14611

*Watkins, D., & Gerong, A. (1999). Language of response and the spontaneous self-concept: A test of the cultural accommodation hypothesis. *Journal of Cross-Cultural Psychology*, *30*(1), 115–121. https://doi.org/10.1177/0022022199030001007

Whorf, B. L. (1956). *Language, thought, and reality: selected writings of Benjamin Lee Whorf*. In J.B. Carroll (Ed.). The M.I.T press.

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 7780-7785. https://doi.org/10.1073/pnas.0701644104

*Yang, K.-S., & Bond, M. H. (1980). Ethnic affirmation by Chinese bilinguals. *Journal of Cross-Cultural Psychology*, *11*(4), 411–425. https://doi.org/10.1177/0022022180114002

*Zhang, W. (2018). *Political Implications of Bilingual Cognition* (Order No. 13804659). Available from Politics Collection; ProQuest Dissertations & Theses Global; ProQuest Dissertations & Theses Global Closed Collection. (2168353970).

https://login.proxy.hil.unb.ca/login?url=https://www.proquest.com/dissertations-theses/political-implications-bilingual-cognition/docview/2168353970/se-2

*Zhou, C., Dewaele, J., Ochs, C.M., DeLeersnyder, J. (2021). The role of language and cultural engagement in emotional fit with culture: An experiment comparing Chinese-English bilinguals to British and Chinese monolinguals. *Affective Science, 2,* 128-141. https://doi.org/10.1007/s42761-021-00037-x

**Note: References with * were included in the meta-analysis**

## Appendix A: Characteristics of Studies Included in the Analysis

**Table 4**

*Characteristics of Studies Included in the Analysis*

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | Earle | 1969 | A | CH | CP | AS | HK | WS | VL | 101 | 0.65 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | 0.27 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | 0.20 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | 0.33 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | 0.19 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | 0.22 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | 0.63 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | 0.18 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | 0.11 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | -0.35 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | -0.49 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | -0.35 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | -0.19 |
| 2 | 2.1 | Lyczak et al. | 1976 | A | CH | CR | AS | HK | WS | CO | 210 | -0.08 |
| 3 | 3.1 | Yang & Bond | 1980 | A | CH | CR | AS | HK | BS | VL | 80 | -0.52 |
| 4 | 4.1 | Bond & Yang | 1982 | A | CH | CR | AS | HK | BS | VL | 368 | -0.37 |
| 4 | 4.1 | Bond & Yang | 1982 | A | CH | CR | AS | HK | BS | VL | 368 | -0.38 |
| 5 | 5.1 | Bond | 1983 | A | CH | UN | AS | HK | BS | VL | 30 | 1.16 |
| 6 | 6.1 | Bond | 1984 | A | CH | CR | AS | HK | BS | VL | 142 | 0.72 |
| 6 | 6.1 | Bond | 1984 | A | CH | CR | AS | HK | BS | VL | 142 | 0.72 |
| 6 | 6.1 | Bond | 1984 | A | CH | CR | AS | HK | BS | VL | 142 | 0.61 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 7.1 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 13 | 0.37 |
| 7 | 7.1 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 13 | 0.21 |
| 7 | 7.1 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 13 | 0.13 |
| 7 | 7.1 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 13 | -0.72 |
| 7 | 7.1 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 13 | -0.17 |
| 7 | 7.2 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 16 | 0.29 |
| 7 | 7.2 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 16 | 0.25 |
| 7 | 7.2 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 16 | -0.60 |
| 7 | 7.2 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 16 | -0.15 |
| 7 | 7.2 | Kubota | 1991 | D | JA | CR | NA | OL | WS | BH | 16 | -0.52 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | 0.46 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | 0.57 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | 0.12 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | 0.57 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | 0.08 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | 0.03 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | 0.25 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | -0.14 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | -0.41 |
| 8 | 8.1 | Ralston et al. | 1995 | A | CH | UN | AS | HK | BS | VL | 104 | -0.04 |
| 9 | 9.1 | Romney et al. | 1997 | A | JA | UN | AS | OL | WS | VL | 17 | *2.45 |
| 9 | 9.1 | Romney et al. | 1997 | A | JA | UN | AS | OL | WS | VL | 17 | 1.54 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 10.1 | Chang et al. | 1999 | A | CH | UN | AS | HK | WS | SC | 51 | -0.06 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | CO | 89 | 0.49 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | CO | 89 | 0.53 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | CO | 89 | 0.39 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | CO | 89 | 0.11 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | CO | 89 | -0.53 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | CO | 89 | -0.58 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | SC | 92 | 0.26 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | SC | 89 | 0.11 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | SC | 89 | 0.16 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | SC | 89 | -0.22 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | SC | 89 | -0.05 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | SC | 93 | -0.47 |
| 11 | 11.1 | Sakamoto | 2001 | D | JA | CR | AN | OL | BS | SC | 93 | -0.79 |
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | AN | OL | BS | SC | 52 | 1.86 |
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | AN | OL | BS | SC | 52 | 0.62 |
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | AN | OL | BS | SC | 52 | 0.79 |
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | AN | OL | BS | SC | 52 | 0.78 |
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | AN | OL | BS | SC | 52 | 0.39 |
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | AN | OL | BS | SC | 52 | 0.30 |
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | AN | OL | BS | SC | 52 | 0.47 |
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | AN | OL | BS | SC | 52 | 0.95 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | NA | OL | BS | SC | 52 | 0.46 |
| 12 | 12.1 | Ross et al. | 2002 | A | CH | UN | NA | OL | BS | SC | 52 | 0.72 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | 0.09 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | 0.00 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | 0.87 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | 0.48 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | 0.15 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | 0.35 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | 1.51 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | 0.49 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | -0.31 |
| 13 | 13.1 | Ishii et al. | 2003 | A | SE | UN | AS | OL | BS | CO | 31 | -0.38 |
| 14 | 14.1 | Ji et al. | 2004 | A | CH | MX | NA | OL | BS | CO | 45 | 1.07 |
| 14 | 14.1 | Ji et al. | 2004 | A | CH | MX | NA | OL | BS | CO | 45 | 1.20 |
| 14 | 14.2 | Ji et al. | 2004 | A | CH | MX | NA | OL | BS | CO | 86 | 0.26 |
| 14 | 14.2 | Ji et al. | 2004 | A | CH | MX | NA | OL | BS | CO | 86 | 0.37 |
| 14 | 14.3 | Ji et al. | 2004 | A | CH | MX | AS | OL | BS | CO | 119 | 0.46 |
| 14 | 14.3 | Ji et al. | 2004 | A | CH | MX | AS | OL | BS | CO | 119 | 0.46 |
| 14 | 14.4 | Ji et al. | 2004 | A | CH | MX | AS | HK | WS | CO | 116 | 0.95 |
| 14 | 14.4 | Ji et al. | 2004 | A | CH | MX | AS | HK | WS | CO | 116 | 0.95 |
| 15 | 15.1 | Kemmelmeier & Cheng | 2004 | A | CH | CP | AS | HK | BS | SC | 126 | -0.41 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 15.1 | Kemmelmeier & Cheng | 2004 | A | CH | CP | AS | HK | BS | SC | 126 | -0.41 |
| 16 | 16.1 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | VL | 72 | 0.52 |
| 16 | 16.1 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | VL | 72 | 0.77 |
| 16 | 16.2 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | CO | 61 | 0.63 |
| 16 | 16.2 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | VL | 61 | 0.66 |
| 16 | 16.2 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | VL | 61 | 0.75 |
| 16 | 16.2 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | VL | 61 | 0.66 |
| 16 | 16.3 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | VL | 150 | 0.48 |
| 16 | 16.3 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | VL | 150 | 0.38 |
| 16 | 16.3 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | VL | 150 | -0.66 |
| 16 | 16.4 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | VL | 65 | -0.79 |
| 16 | 16.5 | Briley et al. | 2005 | A | CH | UN | AS | HK | BS | CO | 61 | 1.18 |
| 17 | 17.1 | Kobayashi | 2006 | A | JA | CR | NA | OL | WS | CO | 16 | 0.54 |
| 18 | 18.1 | Lee | 2006 | A | KO | UN | NA | OL | BS | SC | 75 | -0.18 |
| 18 | 18.1 | Lee | 2006 | A | KO | UN | NA | OL | BS | SC | 75 | -1.34 |
| 19 | 19.1 | Chen & Bond | 2007 | A | CH | UN | AS | HK | BS | SC | 452 | -0.52 |
| 19 | 19.1 | Chen & Bond | 2007 | A | CH | UN | AS | HK | BS | SC | 452 | -0.53 |
| 19 | 19.1 | Chen & Bond | 2007 | A | CH | UN | AS | HK | BS | SC | 452 | -0.26 |
| 20 | 20.1 | Dixon | 2007 | A | CH | CR | AS | OL | BS | SC | 153 | 0.32 |
| 20 | 20.1 | Dixon | 2007 | A | CH | CR | AS | OL | BS | SC | 153 | -0.52 |
| 21 | 21.1 | Kim | 2007 | D | KO | CR | NA | OL | WS | SC | 30 | 0.55 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 22 | 22.1 | Kobayashi et al. | 2007 | A | JA | CP | NA | OL | WS | CO | 12 | 0.47 |
| 23 | 23.1 | Bokhorst-Heng & Santos Caleon | 2009 | A | CH | CP | AS | OL | WS | VL | 204 | 0.46 |
| 23 | 23.1 | Bokhorst-Heng & Santos Caleon | 2009 | A | CH | CP | AS | OL | WS | VL | 204 | 0.38 |
| 23 | 23.1 | Bokhorst-Heng & Santos Caleon | 2009 | A | SE | CP | AS | OL | WS | VL | 204 | 0.13 |
| 23 | 23.2 | Bokhorst-Heng & Santos Caleon | 2009 | A | SE | CP | AS | OL | WS | VL | 204 | 0.05 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | 0.16 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | 0.07 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | 0.22 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | 0.15 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | 0.04 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | 0.02 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | 0.09 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | 0.24 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | 0.56 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | -0.43 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | -0.28 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | CO | 213 | -0.25 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | SC | 213 | 0.05 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | SC | 213 | -0.34 |
| 24 | 24.1 | Chen & Bond | 2010 | A | CH | CP | AS | HK | BS | SC | 213 | -0.07 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.00 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.00 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.00 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.00 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.00 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.00 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.00 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.12 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 1.10 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.06 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.01 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.21 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.41 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.54 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | 0.51 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.33 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.01 |
| 24 | 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.05 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.18 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.22 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.38 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.23 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.03 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.16 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.12 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.06 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.20 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.22 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.96 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.46 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | *-1.22 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | CO | 76 | -0.70 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | SC | 76 | 0.04 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | SC | 76 | 0.00 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | SC | 76 | 0.08 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | SC | 76 | 0.12 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | SC | 76 | 0.28 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | SC | 76 | -0.22 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | SC | 76 | -0.09 |
| 24.2 | Chen & Bond | 2010 | A | CH | CP | AS | HK | WS | SC | 76 | -0.98 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 25.1 | Lee et al. | 2010 | A | CH | UN | AS | HK | WS | SC | 47 | 0.21 |
| 25 | 25.1 | Lee et al. | 2010 | A | CH | UN | AS | HK | WS | SC | 47 | 0.30 |
| 25 | 25.2 | Lee et al. | 2010 | A | CH | UN | AS | HK | BS | SC | 37 | 0.70 |
| 25 | 25.2 | Lee et al. | 2010 | A | CH | UN | AS | HK | BS | SC | 39 | 0.46 |
| 25 | 25.3 | Lee et al. | 2010 | A | CH | UN | AS | HK | BS | SC | 89 | 0.18 |
| 25 | 25.3 | Lee et al. | 2010 | A | CH | UN | AS | HK | BS | SC | 85 | 1.01 |
| 26 | 26.1 | Wang et al. | 2010 | A | CH | CP | AS | HK | BS | BH | 125 | 0.70 |
| 26 | 26.1 | Wang et al. | 2010 | A | CH | CP | AS | HK | BS | BH | 125 | 0.84 |
| 26 | 26.1 | Wang et al. | 2010 | A | CH | CP | AS | HK | BS | SC | 125 | 1.36 |
| 26 | 26.1 | Wang et al. | 2010 | A | CH | CP | AS | HK | BS | SC | 125 | 0.77 |
| 26 | 26.1 | Wang et al. | 2010 | A | CH | CP | AS | HK | BS | VL | 125 | 0.57 |
| 28 | 28.1 | Boucher & O'Dowd | 2011 | A | CH | CR | NA | OL | BS | SC | 31 | 1.40 |
| 28 | 28.1 | Boucher & O'Dowd | 2011 | A | CH | CR | NA | OL | BS | SC | 31 | 1.37 |
| 28 | 28.1 | Boucher & O'Dowd | 2011 | A | CH | CR | NA | OL | BS | SC | 31 | 0.54 |
| 28 | 28.1 | Boucher & O'Dowd | 2011 | A | CH | CR | NA | OL | BS | SC | 31 | 1.46 |
| 28 | 28.1 | Boucher & O'Dowd | 2011 | A | CH | CR | NA | OL | BS | SC | 31 | 0.66 |
| 28 | 28.1 | Boucher & | 2011 | A | CH | CR | NA | OL | BS | SC | 31 | 0.34 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 28 | 28.1 | O'Dowd Boucher & O'Dowd | 2011 | A | CH | CR | NA | OL | BS | SC | 31 | 0.72 |
| 28 | 28.1 | Boucher & O'Dowd | 2011 | A | CH | CR | NA | OL | BS | SC | 31 | 0.97 |
| 28 | 28.1 | Boucher & O'Dowd | 2011 | A | CH | CR | NA | OL | BS | VL | 31 | 1.18 |
| 29 | 29.1 | Imbo & LeFevre | 2011 | A | CH | CR | EU | OL | BS | CO | 80 | 0.00 |
| 29 | 29.1 | Imbo & LeFevre | 2011 | A | CH | CR | EU | OL | BS | CO | 80 | 0.47 |
| 29 | 29.1 | Imbo & LeFevre | 2011 | A | CH | CR | EU | OL | BS | CO | 80 | 0.12 |
| 29 | 29.1 | Imbo & LeFevre | 2011 | A | CH | CR | EU | OL | BS | CO | 80 | 1.01 |
| 29 | 29.1 | Imbo & LeFevre | 2011 | A | CH | CR | EU | OL | BS | CO | 80 | 0.93 |
| 29 | 29.1 | Imbo & LeFevre | 2011 | A | CH | CR | EU | OL | BS | CO | 80 | 0.33 |
| 29 | 29.1 | Imbo & LeFevre | 2011 | A | CH | CR | EU | OL | BS | CO | 80 | -0.88 |
| 29 | 29.1 | Imbo & LeFevre | 2011 | A | CH | CR | EU | OL | BS | CO | 80 | -0.35 |
| 30 | 30.1 | Moriuchi | 2012 | D | JA | UN | NA | OL | BS | VL | 298 | 0.19 |
| 30 | 30.1 | Moriuchi | 2012 | D | JA | UN | NA | OL | BS | VL | 298 | 0.27 |
| 30 | 30.1 | Moriuchi | 2012 | D | JA | UN | NA | OL | BS | VL | 298 | -0.06 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | 30.1 | Moriuchi | 2012 | D | JA | UN | NA | OL | BS | VL | 298 | -0.25 |
| 31 | 31.1 | Li et al. | 2013 | A | CH | CR | NA | OL | WS | BH | 15 | 1.20 |
| 31 | 31.1 | Li et al. | 2013 | A | CH | CR | NA | OL | WS | BH | 15 | 0.05 |
| 31 | 31.1 | Li et al. | 2013 | A | CH | CR | NA | OL | WS | BH | 15 | 0.36 |
| 31 | 31.1 | Li et al. | 2013 | A | CH | CR | NA | OL | WS | BH | 15 | *3.65 |
| 32 | 32.1 | Chen et al. | 2014 | A | CH | UN | AS | HK | BS | SC | 198 | 0.33 |
| 32 | 32.1 | Chen et al. | 2014 | A | CH | UN | AS | HK | BS | SC | 198 | 0.28 |
| 32 | 32.2 | Chen et al. | 2014 | A | CH | UN | AS | HK | WS | BH | 76 | 1.05 |
| 32 | 32.2 | Chen et al. | 2014 | A | CH | UN | AS | HK | WS | BH | 76 | 1.83 |
| 32 | 32.2 | Chen et al. | 2014 | A | CH | UN | AS | HK | WS | SC | 76 | 0.61 |
| 32 | 32.2 | Chen et al. | 2014 | A | CH | UN | AS | HK | WS | SC | 76 | *2.44 |
| 33 | 33.1 | Chen et al. | 2014 | A | CH | UN | AS | HK | WS | SC | 129 | 0.35 |
| 33 | 33.1 | Chen et al. | 2014 | A | CH | UN | AS | HK | WS | SC | 129 | 0.16 |
| 33 | 33.2 | Chen et al. | 2014 | A | CH | UN | AS | OL | BS | SC | 229 | 0.44 |
| 33 | 33.2 | Chen et al. | 2014 | A | CH | UN | AS | OL | BS | SC | 229 | 0.38 |
| 33 | 33.3 | Chen et al. | 2014 | A | CH | UN | AS | HK | BS | BH | 106 | 0.52 |
| 34 | 34.1 | Lee | 2014 | D | KO | CR | NA | OL | BS | SC | 89 | 0..94 |
| 34 | 34.1 | Lee | 2014 | D | KO | CR | NA | OL | BS | SC | 89 | 1.12 |
| 34 | 34.1 | Lee | 2014 | D | KO | CR | NA | OL | BS | SC | 89 | -0.51 |
| 34 | 34.1 | Lee | 2014 | D | KO | CR | NA | OL | BS | VL | 89 | -0.25 |
| 35 | 35.1 | Wang | 2014 | D | CH | UN | EU | OL | BS | SC | 82 | 0.54 |
| 35 | 35.1 | Wang | 2014 | D | CH | UN | EU | OL | BS | VL | 82 | -0.56 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 35 | 35.2 | Wang | 2014 | D | CH | UN | AS | OL | BS | SC | 82 | 0.38 |
| 35 | 35.2 | Wang | 2014 | D | CH | UN | AS | OL | BS | VL | 82 | 0.13 |
| 36 | 36.1 | Pattaratanakun | 2015 | A | SE | CP | AS | OL | BS | VL | 84 | 0.69 |
| 36 | 36.1 | Pattaratanakun | 2015 | A | SE | CP | AS | OL | BS | VL | 84 | 0.54 |
| 36 | 36.2 | Pattaratanakun | 2015 | A | SE | CP | AS | OL | WS | VL | 62 | 1.26 |
| 36 | 36.2 | Pattaratanakun | 2015 | A | SE | CP | AS | OL | WS | VL | 62 | 0.63 |
| 37 | 37.1 | Chiou & Mercado | 2016 | A | CH | UN | NA | OL | BS | VL | 98 | 0.40 |
| 37 | 37.1 | Chiou & Mercado | 2016 | A | CH | UN | NA | OL | BS | VL | 98 | 0.49 |
| 38 | 38.1 | Li | 2017 | A | CH | CP | AS | HK | BS | BH | 64 | 0.44 |
| 38 | 38.1 | Li | 2017 | A | CH | CP | AS | HK | BS | BH | 64 | 0.09 |
| 38 | 38.1 | Li | 2017 | A | CH | CP | AS | HK | BS | BH | 64 | 0.14 |
| 38 | 38.1 | Li | 2017 | A | CH | CP | AS | HK | BS | BH | 64 | 0.11 |
| 39 | 39.1 | Lee | 2018 | D | CH | MX | AS | OL | WS | BH | 37 | 0.36 |
| 39 | 39.1 | Lee | 2018 | D | CH | MX | AS | OL | WS | BH | 37 | 0.22 |
| 39 | 39.1 | Lee | 2018 | D | CH | MX | AS | OL | WS | BH | 37 | 0.08 |
| 39 | 39.1 | Lee | 2018 | D | CH | MX | AS | OL | WS | BH | 37 | 1.02 |
| 39 | 39.1 | Lee | 2018 | D | CH | MX | AS | OL | WS | BH | 37 | -0.84 |
| 39 | 39.1 | Lee | 2018 | D | CH | MX | AS | OL | WS | BH | 37 | -0.05 |
| 39 | 39.1 | Lee | 2018 | D | CH | MX | AS | OL | WS | BH | 37 | -0.25 |
| 40 | 40.1 | Smith et al. | 2018 | A | CH | UN | EU | OL | WS | CO | 44 | 0.26 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | 40.1 | Smith et al. | 2018 | A | CH | UN | EU | OL | WS | CO | 44 | 0.04 |
| 40 | 40.1 | Smith et al. | 2018 | A | CH | UN | EU | OL | WS | CO | 46 | 0.20 |
| 40 | 40.1 | Smith et al. | 2018 | A | CH | UN | EU | OL | WS | CO | 44 | -0.19 |
| 40 | 40.1 | Smith et al. | 2018 | A | CH | UN | EU | OL | WS | CO | 47 | -0.13 |
| 40 | 40.1 | Smith et al. | 2018 | A | CH | UN | EU | OL | WS | CO | 46 | 0.16 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | BH | 347 | 0.02 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | BH | 347 | 0.03 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | BH | 347 | 0.14 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | BH | 347 | 0.15 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | BH | 347 | -0.17 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | BH | 347 | -0.17 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 420 | -0.05 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 420 | -0.21 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 420 | -0.08 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 420 | -0.16 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 420 | 0.00 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 420 | -0.12 |
| 41 | 41.1 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 420 | -0.18 |
| 41 | 41.2 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 188 | 0.34 |
| 41 | 41.2 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 188 | 0.03 |
| 41 | 41.2 | Zhang | 2018 | D | CH | CR | AS | OL | BS | VL | 188 | 0.07 |
| 42 | 42.1 | Watkins & Gerong | 2019 | A | SE | CR | AS | OL | BS | SC | 166 | 0.42 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 42 | 42.1 | Watkins & Gerong | 2019 | A | SE | CR | AS | OL | BS | SC | 166 | -0.33 |
| 42 | 42.1 | Watkins & Gerong | 2019 | A | SE | CR | AS | OL | BS | SC | 166 | -0.69 |
| 43 | 43.1 | Cheng et al. | 2020 | A | CH | UN | AS | OL | WS | BH | 30 | 0.68 |
| 43 | 43.1 | Cheng et al. | 2020 | A | CH | UN | AS | OL | WS | BH | 30 | 0.31 |
| 44 | 44.1 | Liu et al. | 2021 | A | CH | CR | AS | OL | WS | BH | 59 | 0.26 |
| 44 | 44.1 | Liu et al. | 2021 | A | CH | CR | AS | OL | WS | BH | 59 | -0.35 |
| 44 | 44.1 | Liu et al. | 2021 | A | CH | CR | AS | OL | WS | BH | 59 | -0.40 |
| 44 | 44.1 | Liu et al. | 2021 | A | CH | CR | AS | OL | WS | BH | 59 | -0.10 |
| 45 | 45.1 | Zhou et al. | 2021 | A | CH | CR | EU | OL | BS | SC | 178 | 0.10 |
| 45 | 45.1 | Zhou et al. | 2021 | A | CH | CR | EU | OL | BS | SC | 178 | 0.09 |

*Note*. Columns indicate the following1: 1 = Author ID, 2 = Sample ID, 3 = authors, 4 = year, 5 = publication status, 6 = Language, 7 = Language acquisition style, 8 = Continent, 9 = City, 10 = Design, 11 = Outcome, 12 = Total N, 13 = *g*. Pubstat consists of Articles (A) and Dissertations (D). Languages consist of Chinese (CH), Japanese (JA), Korean (KO) and South East Asian Language (SE). Language acquisition style consists of Compound (CP), Coordinate (CR), Unclear (UN) and Mixed (MX). Continent consists of Asia (AS), Europe (EU) and North America (NA). City consists of Hong Kong (HK) and all locations outside of Hong Kong (OL). Design consists of Between-subjects (BS) and within-subjects (WS). Outcomes consist of Behaviour (BH), Cognition (CO), Self-concept (SC) and Values (VL). For the sake of comparability, an adjustment of .50 (as

an estimate of correlation) multiplied the d's of the between-subjects effects by 2*sqrt(1-r) was applied to all between—subjects effects in this table.

## Appendix B: Coding Guide

### Coding Guide

Coding guide

Here is a guide to the process of coding.

- I (Jordan) will go through each file first and do my own coding. In the process I will highlight the relevant information.
- I will then copy the PDF into the shared Google folder.
- Open the Excel file "Coding Book, Mallory" and the relevant PDF
- Look through the PDF for the relevant information. Most of it should be highlighted, but you may need to scan around a bit yourself.
- Include the following information on the spreadsheet
    - Authors, in order presented
        - Last name, first initial
            - E.g., Adams, J.
    - Year of publication
        - If there are multiple years/inconsistency include all years
    - Title
    - Publication/institution
        - Where was it published? If it is a thesis, what was the granting institution?
    - Format
        - **ARTICLE** (if it was published in a peer-reviewed journal)
        - **THESIS** (if it was a thesis or dissertation).

- Non-English language
    - If this study is included, it SHOULD involve speakers of English AND an East Asian language/EAL (Chinese, Korean etc.). English is assumed, but what is the East Asian language?
- Participant L1
    - What was the first language of the participants?
    - If this is not clear, just put a ?
- Participants L2
    - What was the second language of the participants?
    - If this is not clear, just put a ?
- Participants L…
    - Sometimes they will have more than two languages – include any pther languages in the order presented.
    - If this is not clear, just put a ?
- Language acquisition style
    - **COMPOUND** if participants learned both languages at the same time
    - **COORDINATE** if participants learned one language after being able to fluently speak another language
- Countries
    - Many of these studies take place in one country, and so just include the one country.
    - If there are multiple countries, include them.

        - Put the Western/English country first followed by the Eastern country
- Cities or states
    - This may not be readily available, but include it if you can. The main reason for this column is that MANY studies take place in Hong Kong, so this is where we would put that information. If there are multiple cities, put the Western/English city first.
    - If a city is not mentioned, but a state or province is, include that.
- Gender F/M
    - How many females/how many males.
        - If it is within, then you will just have one set of numbers
        - If it is between, report on the English condition first and then the EAL condition if possible
- Age range
    - If they just report the average age, you can include that.
    - If there are separate age ranges reported for the English/EAL samples, report English and then EAL
    - If they provide both, include both
- Design
    - **WITHIN** if the same participants did the task in English AND an EAL.
    - **BETWEEN** if participants did the task in English OR an EAL
- Outcome variable

- Outcomes categorized as *values* will require participants to make moral judgements and/or provide their opinion. This can be in response to a series of statements, or a situation. Outcomes demonstrating individualism and collectivism, or tradition and modernity would fall into this category.
- Responses categorized as *self-concept* will measure participants' sense of identity. This may be identity in a broad sense, such as affiliation to a group or nationality, or identity in a narrow sense such as personality traits. This will include outcomes demonstrating affiliation and belonging to a larger cultural group, as well as outcomes demonstrating self-reported traits and non-cultural identity. Outcomes demonstrating ethnic identity, national affiliation and acculturation would all fall into this category, along with measures of Big Five, or Five-Factor personality or trait intelligence.
- Outcomes categorized as *cognition* will demonstrate performance on mental tasks that require evaluation of a stimuli, or completion of a cognitive task. Outcomes demonstrating intelligence, mental ability, attention and social cognition will all be categorized as cognition. This category will also include examples of social cognition, including any assessment or evaluation of another person.

        - Outcomes categorized as *behaviour* will require some kind of physical action in response to a stimulus. Examples from the preliminary search include gesturing (So, 2010), standing position relative to a superior (Sechrest, 1968), and length of a spoken conversation (Bond & Lai, 1986).
    - Explanation of Effect
        - In a sentence or two, explain what the tasks involved. If they are established scales (E.g., the NEO-PI-R) make sure to include the labels.
    - English N/EAL N
        - How many in each condition? We will need to know this for the stats anyway, so we might as well write that down.
    - Stats used
        - What statistics are available for the calculation of the effect size?
            - If there are multiple options, use the most basic stats
                - E.g., if there are means and standard deviations, use those instead of an ANOVA.
    - *d, 95% C.I upper, 95% C.I lower, v*
        - Follow this link: Practical Meta-Analysis Effect Size Calculator David B. Wilson, Ph.D., George Mason University
        - Choose the appropriate calculator and include the information

        - The participants speaking English are to be the TREATMENT group, The participants speaking an EAL are to be in the CONTROL group.
        - Add in the numbers
    - Cultural accommodation or Ethnic Affirmation
        - Do they behave more Western when speaking English – ACCOMMODATION
        - Do they behave more Eastern when speaking English - AFFIRMATION
        - Unclear put ?
    - Explanation - why do you think it is accommodation or affirmation?

## Appendix C: Analyses Conducted with Results Presented by Design

## Results

### Overall Effect Size

To determine the overall estimated effect size in the full sample, I used the two-level null model, which I determined was the best fit for these data in the previous analysis. The intercept of this two-level null model revealed a mean estimated effect size of $g$ = .27, (95% CI: .16, .38), which was significantly different from zero ($p$ < .001). This small but significant positive effect suggests that participants, in general, do provide responses that are more consistent with established East Asian cultural practices when responding in their East Asian language than when responding in English.

I also examined the test of heterogeneity to determine if there was significant variation in effects between samples. This test was also significant, $Q(268) = 1751.03$, $p$ < .001 indicating significant heterogeneity between samples, which suggests that the overall effect is not a good representation of the sample, and systematic analyses of moderators are appropriate to account for this heterogeneity.

### Moderator Analysis

Each test of the moderators was conducted using the rma.mv command followed with the *robust* procedure from the *metafor* package (Viechtbauer, 2010) in R statistical software (R Core Team, 2019) version 4.4.0. Because the typical standard error fails to account for the possibility of dependence between samples in multilevel analyses, revised (robust) standard errors were necessary for a more accurate test of moderators in multilevel models (Rogers & Pustejovsky, 2020).

#### *Effect of Design*

The first moderator tested was Design to ensure that the adjustment I applied to effect sizes from within- and between-subject design resulted in comparable groups. If the groups were found to differ, moderator analyses would be conducted separately for the two types of designs. When design was entered as moderator, the test of moderators was not significant, $F(1, 60) = .89$, $p = .348$. The 95% confidence intervals for the estimated mean effects did not contain zero for within-subjects designs ($k = 106$, $g = .34$, 95% $CI = .16, .52$) or between-subjects designs ($k = 163$, $g = .24$, 95% $CI = .10, .37$).

The finding that this moderator was not significant suggests that effects from within-subjects designs are not significantly larger than the effects from between-subjects designs. This, along with the application of the adjustment recommended by Borenstein and colleagues (2009), gave me confidence to proceed with my analyses using both within-subjects and between-subjects effects. Nonetheless, to address any remaining concerns about compatibility I conducted separate analyses for within-subjects effects and between-subjects effects and report them below. When only between-subjects effects were included in the model, the test of heterogeneity was significant $Q(162) = 823.00$, $p < .001$. The test of heterogeneity was also significant when only within-subjects effects were included in the model $Q(105) = 770.45$, $p < .001$, meaning that there was sufficient diversity between the effects from both designs that moderator analyses were necessary. Therefore, the moderator analyses for East Asian language, Task Type, Language Acquisition Style, Continent and City were conducted with the total sample, with only the between-subjects effects, and with only the within-subjects effects.

***East Asian Language***

East Asian Language was a significant moderator ($p$ = .03). The 95% confidence intervals did not contain zero for Chinese ($k$ = 213, $g$ = .32, 95% $CI$ = .20, .43), but did contain zero for Other (k = 56, g = .13, 95% CI = -.04, .30). Thus, the effect appears to be significant only for speakers of Chinese, and not speakers of the other East Asian languages used.

When analyses were conducted using only between-subjects effects, East Asian Language was not a significant moderator ($p$ = .12). East Asian language was also not a significant moderator when analyses were conducted using only within-subjects effects ($p$ = .07)[5]

***Task Type***

Task Type was a significant moderator ($p$ = .01). Mean estimates are included in Table 5. The confidence intervals of the mean estimates did not contain zero for any of the task types, suggesting that each task demonstrates a meaningful difference between language conditions. Tukey multiple comparisons indicated that the estimates for behaviour were significantly larger than the estimates for values ($p$ = .003), which suggests that the effect of language is stronger on behavioural tasks than on value-related tasks.

---

[5] When conducting analyses within subjects, and when outliers were included, East Asian Language was a significant moderator ($p$ = .02). The 95% confidence intervals did not contain zero for Chinese ($k$ = 90, $g$ = .47, 95% $CI$ = .24, .70), but they did contain zero for Other ($k$ = 19, $g$ = .26, 95%$CI$ = -.01. .54).

**Table 5**

*Results for Task Type as a Moderator, with outliers excluded.*

| Category | Effects | *g*(95% CI) |
|---|---|---|
| Self-concept | 72 | .20 (.02, .39) |
| Values | 57 | .26 (.05, .47) |
| Cognition | 99 | .32 (.13, .51). |
| Behaviour | 41 | .37 (.15, .59) |

*Note.* The table contains the number of effect sizes (Effects) and the mean weighted *g* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with $p < .05$ when the 95% CI for *g* does not contain zero.

Task Type was also a significant moderator when analyses were conducted using only the between-subjects effects ($p < .001$). Mean Estimates are included in Table 6. The confidence intervals of the Mean Estimates did not contain zero for behaviour or cognition, but they did contain zero for self-concept and values, suggesting that the self-concept and values tasks did not demonstrate a meaningful difference between language tasks. Tukey multiple comparisons indicated that estimates for behaviour were significantly larger than estimates for values ($p < .001$) suggesting that the effect of language is stronger in behavioural tasks than value-related tasks. Multiple comparisons also indicated that estimates for cognition were significantly larger than estimates for self-concept ($p < .001$), suggesting that the effect of language is stronger in cognitive tasks than self-concept related tasks.

**Table 6**

*Results for Task Type as a Moderator, in the Between-Subjects Design with outliers excluded.*

| Category | Effects | g(95%CI) |
|---|---|---|
| Self-concept | 57 | .19(-.003, .39) |
| Values | 49 | .16(-.05, .38) |
| Cognition | 44 | .40(.23, .56) |
| Behaviour | 13 | .25(.06, .44) |

*Note.* The table contains the number of effect sizes (Effects) and the mean weighted *g* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with $p < .05$ when the 95% CI for *g* does not contain zero.

Task type was not a significant moderator when analyses were conducted using only the within-subjects effects ($p = .07$).

***Non-significant Moderators***

Language acquisition style was not a significant moderator ($p = .13$), nor was Continent ($p = .83$), or City ($p = .81$).

When analyses were conducted using only between-subjects effects, language acquisition style was not a significant moderator ($p = .13$), nor was Continent ($p = .83$), or City ($p = .88$).

When analyses were conducted using only within-subjects effects, language acquisition style was not a significant moderator ($p = .62$), nor was Continent ($p = .33$), or City ($p = .67$)

**Adjustment for Multiple Moderators**

Using a large number of separate analyses may increase the risk of type-one error. Therefore, for a more conservative perspective on the results, I applied the Holm-Bonferroni adjustment method to each of my moderators. This was done only for the significant moderators in the analysis. The pre-adjustment and post-adjustment *p*-values are presented in Table 7. After the adjustment, both of the significant moderators (East Asian language and task type) remained statistically significant ($p < .05$). This correction was not applied to the analyses using only the between-subjects effects, or the analyses using only the within-subjects effects, as they had one or fewer significant moderators.

**Table 7**

*Holm Test of Moderators for all effects*

| Moderator | *P* | Rank | Transformed |
|---|---|---|---|
| EAL | .03 | 1 | .03 |
| OUT | .01 | 2 | .02 |

Note: *p*-values (p) for each significant moderator (moderator) were ranked from largest to smallest (rank) and multiplied by their rank.

**Curriculum Vitae**

**Jordan Schriver**

Curriculum vitae

(Updated October 2024)

## Education

| | |
|---|---|
| 2013 to present | Ph.D Psychology (candidate)<br>University of New Brunswick, Department of Psychology<br>Supervisor: Dr. Elaine Perunovic<br>(Anticipated graduation Spring 2024) |
| 2008 to 2012 | BA Honours Psychology<br>University of New Brunswick, Department of Psychology<br>Supervisor: Dr. Elaine Perunovic<br>Overall GPA 4.3/4.3 Psychology: GPA 4.3/4.3 |

## Teaching Experience

| | |
|---|---|
| 2023 (Fall) | PS3103: Social Psychology, PS3803: Cultural Psychology, PS3873: Community Psychology (CU) |
| 2023 (Winter) | PS1023: Introduction to Psychology, PS4613: Advanced Research Methods (CU) |
| 2022 (Fall) | PS2233: Adult Development, PS3233: The Psychology of Close Relationships |
| 2022 (Winter) | PS1023: Introduction to Psychology, PS3230: Sport Psychology, PS4613: Psychology of Prejudice (CU) |

| | |
|---|---|
| 2021 (Fall) | PS2033: Personality Psychology, PS3073: Cultural Psychology, PS3103: Social Psychology (CU) |
| 2021 (Winter) | PS1023: Introduction to psychology, PS3873: Community psychology, PS4613: Advanced Research Methods. (CU) |
| 2020 (Fall) | PS2033: Personality Psychology, PS2233: Adult Development (CU) |
| 2020 (Winter) | PS1023: Introduction to Psychology, PS4203: Psychology of Prejudice, PS4613: Advanced Research Methods (CU) |
| 2019 (Fall) | PS2033: Personality Psychology, PS3103: Social Psychology (CU) |
| 2019 (Winter) | PSYC 2313A: Personality Psychology (STU) |
| 2018 (Fall) | PSYC 2103: Quantitative Research Methods (UNB) |
| 2018 (Winter) | PSYC 3453: Close Relationships (UNB) |
| 2017 (Fall) | PSYC 3053: Personality Psychology (UNB) |
| 2015 (Fall) | PSYC 2103: Quantitative Research Methods (UNB) |

**Awards**

| | |
|---|---|
| 2022 | Stephen & Ella Steeves Award for Excellence in Teaching |
| 2020 | Snodgrass Prize for Graduate Student Research in Psychology |
| 2020 | President’s Doctoral Tuition Award |
| 2018 | Third place: Oral presentation (UPEI Multidisciplinary Graduate Research Conference) |
| 2018 | Second place: Oral presentation, social sciences and humanities section (UNB Graduate Student Research Conference, 2018) |
| 2018 | (Nominated) Snodgrass Award for Service to the Psychology Department |

| | |
|---|---|
| | (University of New Brunswick) |
| 2017-2018 | School of Graduate Studies Merit Award (University of New Brunswick) |
| 2017 | First place: Oral presentation, social sciences and humanities section (UNB Graduate Student Research Conference, 2017) |
| 2016-2018 | President's Doctoral Tuition Award (University of New Brunswick) |
| 2015-2017 | Magee Third Century Postgraduate Merit Award (University of New Brunswick) |
| 2014-2017 | Doctoral award (New Brunswick Innovation Foundation) |
| 2014-2015 | Magee Fellowship in the Humanities and Social Sciences (University of New Brunswick) |
| 2014 | Snodgrass Student Travel Award (University of New Brunswick) |
| 2014 | Snodgrass Award for Service to the Psychology Department (University of New Brunswick) |
| 2013-2018 | Doctoral Research Award (Social Sciences and Humanities Research Council) 2013 Joseph-Armand Bombardier Canada Graduate Scholarship (Social Sciences and Humanities Research Council) |
| 2013 | University of New Brunswick Board of Governors Merit Award (University of New Brunswick). |
| 2012 | Lieutenant Governor's Silver Medal, Arts, Encaenia 2012 (University of New Brunswick) |

**Publications**

Schriver, J., Perunovic, W.Q.E., Brymer, K., & Hachey, T. (2019). Do

relatives with greater productive potential get help first?: A test of the inclusive fitness explanation of kind altruism. *Evolutionary Psychology, 17,* 1-12.

**Ongoing Research**

Schriver, J., Perunovic, W.Q.E. (TBD). Online dating profiles in Japan and the United States: Describing oneself in the past, present and future. *In progress.*

Sharp, R., Schriver, J. (TBD) The social problem: Bullying during adolescence and its relationship to explanatory style flexibility and peer attachment in Adulthood. *In progress.*

Schriver, J., Perunovoc, W.Q.E., Ronis, S., Malcolm, K. (TBD). Invisible minorities: Gender and responsiveness to non-native acculturation experiences. *In progress.*

Schriver, J., Perunovic, W.Q.E., Voyer, D., (TBD). The effects of task language on responses in east-Asian bilingual individuals: a meta-analysis. *In progress.*

**Conference Activity**

***Papers Presented***

2024 Schriver, J., Perunovic, W.Q.E., Voyer, D. (2024, June). *The effects of language use on responses in English/East-Asian bilingual individuals: A meta-analysis*. Oral presentation at the Annual Meeting of the Canadian Psychological Association. Ottawa, ON, Canada.

2019 Schriver, J., Perunovoc, W.Q.E., Ronis, S., Malcom, K. (2019, March).

*Invisible minorities: Gender and responsiveness to non-native acculturation experiences*. Oral presentation at the Annual University of New Brunswick Graduate Research Conference, Fredericton, NB, Canada.

2018 Schriver, J., Perunovic, W.Q.E. (2018, March). *Online dating profiles in Japan and the United States: Describing oneself in the past, present and future*. Oral presentation at the University of Prince Edward Island Multidisciplinary Graduate Research Conference, Charlottetown, PEI, Canada.

2018 Schriver, J., Perunovic, W.Q.E. (2018, March). *Online dating profiles in Japan and the United States: Describing oneself in the past, present and future*. Oral presentation at the Annual University of New Brunswick Graduate Research Conference, Fredericton, NB, Canada.

2017 Schriver, J., Perunovic, W.Q.E., Brymer, K., & Hachey, T. (2017, March). *Manipulating reproductive potential of a family member leads to greater helping behaviour: Support for inclusive fitness theory*. Oral presentation at the Annual University of New Brunswick Graduate Research Conference, Fredericton, NB, Canada.

2015 Schriver, J., Perunovic, W.Q.E., & Perunovic, M. (2015, April). *Trying and Failing or Failing to Try: The role of conscientiousness in judging people with perceived controllable stigmas.* Oral

presentation at the Annual University of New Brunswick Graduate Research Conference, Fredericton, NB, Canada.

2015 Schriver, J., Perunovic, W.Q.E., & Perunovic, M. (2015, March). *The role of self-control in weight bias.* Oral presentation at the Annual Conference of the Eastern Psychological Association, Philadelphia, PA, USA.

***Poster Presentations***

2019 Schriver, J., Perunovoc, W.Q.E., Ronis, S., Malcom, K. (2019, June). *Invisible minorities: Gender and responsiveness to non-native acculturation experiences*. Poster presented at the Annual Meeting of the Canadian Psychological Association, Halifax, NS, Canada.

2015 Schriver, J., Perunovic, W.Q.E., & Perunovic, M. (2015, February). *Trying and failing or failing to try: The role of conscientiousness in judging people with perceived controllable stigmas*. Poster presented at the Annual Conference of the Society for Personality and Social Psychology, Long Beach, CA, USA.

2013 Schriver, J., Perunovic, W.Q.E., & Perunovic, M. (2013, June). *The impact of ego depletion on discrimination against various stigmatized groups.* Poster presented at the Annual Meeting of the Canadian Psychological Association, Quebec City, QC, Canada.

2013 Schriver, J., Perunovic, W.Q.E., & Perunovic, M. (2013, January). *The*

*effect of self-control on discrimination against various groups*. Poster presented at the Annual Conference of the Society for Personality and Social Psychology, New Orleans, LA, USA.

**Invited Talks**

2024 Schriver, J., Perunovic, W.Q.E., Voyer, D. (2024, February). *The effects of language use on responses in English/East-Asian bilingual individuals: A meta-analysis*. Oral presentation at the Crandall University Faculty Colloquium. Moncton, NB, Canada.

2023 Schriver, J. (2023, October). *Faith seeking understanding symposium: Christianity on the Mind*. Invited speaker for symposium at Crandall University. Moncton, NB, Canada.

2023 Schriver, J. (2023, February). *Challenges of Witnessing online.* Oral presentation at the Ratio Christi group at Crandall University. Moncton, NB, Canada.

2021 Schriver, J. (2021, November). *Witnessing online.* Oral presentation at the Ratio Christi group at Crandall University. Moncton, NB, Canada.

2019 Schriver, J. (2019, February). *Witnessing online.* Oral presentation at the Ratio Christi group at Crandall University. Moncton, NB, Canada.

2017 Schriver, J., Perunovic, W.Q.E., Brymer, K., & Hachey, T. (2017, September). *Manipulating reproductive potential of a family member leads to greater helping behaviour: Support for inclusive*

*fitness theory*. Oral presentation at Psychology department graduate student research showcase. Fredericton, NB, Canada.

**Departmental Service**

| | |
|---|---|
| 2023 (Summer) | Academic Standards and Appeals Committee (CU) |
| 2021 - present | Ethics Committee (CU) |
| 2021 - present | Social Committee (CU) |
| 2022 - 2023 | Crandall University Psychology Department Coordinator (CU) |
| 2022 - 2023 | Crandall University Psychology Society Faculty Representative (CU) |
| 2020 - 2021 | Academic Standards and Appeals Committee (CU) |
| 2014 - 2019 | Departmental Colloquium Committee, Student representative (UNB) |
| 2014 - 2015 | Experimental Student Representative (UNB) |

**Honours Students**

| | |
|---|---|
| 2023 - 2024 | Ashley Shaw (Honours Student) |
| 2022 - 2023 | Indira Trowsdale (Honours Student) |
| 2022 - 2023 | Samuel Chute (Honours Student) |
| 2019 - 2020 | Rachel Sharp (Honours Student) |
| 2019 - 2020 | Madison Homan (Honours Student) |

**Student Research Co-Supervision**

| | |
|---|---|
| 2018 | Brianna Lister |
| 2018 | Robert Joy |
| 2017 – 2018 | Marika Kuroki |

| | |
|---|---|
| 2017 | Allison Meng |
| 2016 – 2017 | Maurgan Leblanc |
| 2015 | Craig Shaw |
| 2014 | Krista Keetch |
| 2013 | Gabrielle MacMillian |

**References**

**Elissa Rodkey**

Associate Professor, Crandall University

333 Gorge Road

Moncton New Brunswick

Canada

E1G 3H9

506-858-8970 Ext: 123

elissa.rodkey@crandallu.ca

**Elaine Perunovic (Undergraduate/Graduate Supervisor)**

Associate Professor, University of New Brunswick

38 Dineen Drive

Room 110, Keirstead Hall

Fredericton, NB

E3B 5A3

(506) 458-7689

eperunov@unb.ca