

Comparing Chinese-English Bilingual and Heritage Speaker Stroop Effects: Evidence for the BIA+ Model and a Heritage Bilingual Advantage.

GRACE SHAN

UNIVERSITY OF CAMBRIDGE

ABSTRACT Studies on heritage speakers can provide insight on bilingual models of processing, native language processing after a drop in input, and a possible “bilingual advantage.” This is the first study to examine heritage speaker performance in a Stroop task, which measures selective attention through textual stimuli—fitting for the varying amounts of textual input heritage speakers receive. The selective attention of English-speaking heritage speakers of Chinese (the “heritage group”) is compared to native Chinese speakers with English as an L2 (the “native group”) through an online Stroop experiment following a language experience questionnaire. The Stroop task consists of within- and between-language conditions as well as a novel no-language “colour response” condition where participants visually identify but do not need to name their colour response. Language exposure and proficiency are found to inversely impact the amount of interference experienced in the Stroop task, compatible with the BIA+ model (Dijkstra & van Heuven 2002). Linear regressions also reveal a positive correlation between onset age of acquiring and reading English and interference for Chinese stimuli, suggesting better selective attention (Bialystok & Craik 2010) for earlier ages of L2 onset, which is especially pertinent to heritage speakers due to their early acquisition of both their L1 and L2.

1 INTRODUCTION

1.1 Heritage speakers

Heritage speakers are typically second-generation immigrants who grow up in a multilingual environment from an early age but in most cases never reach their parents’ level of proficiency, or native-like attainment in adulthood, in their L1. They are exposed to their L1, the heritage language, as well as the societally or institutionally dominant language of the country, their L2, simultaneously or sequentially in early childhood. Eventually, the L2 becomes the primary and more dominant language after the onset of schooling, leading to comparatively weaker L1 abilities, particularly in reading and writing. Their L1 may suffer from incomplete acquisition or attrition due to decreased input and use after the overwhelming start of L2 input at preschool (Montrul 2015). While reaching native-like attainment in this L2, heritage speakers may be deemed as non-native speakers by other speakers of

©2024 Shan

This is an open-access article distributed by Section of Theoretical & Applied Linguistics, Faculty of Modern and Medieval Languages and Linguistics, University of Cambridge under the terms of a Creative Commons Non-Commercial License (creativecommons.org/licenses/by-nc/3.0).

their L1 because they demonstrate difficulties with lexical retrieval, code-switching to fill lexical gaps, divergent pronunciation, morphological errors, avoidance of certain structures, and overuse of others due to transfer from the dominant language (Benmamoun, Montrul & Polinsky 2013) as well as non-native-like grammaticality judgements (Seliger 1996).

Research on heritage speakers can provide insight on many of the essential questions within the field of linguistics when compared to monolinguals and other types of bilinguals. How does frequency and quantity of input in each language affect processing in each language? Do the conditions of bilingual acquisition affect the strength of a possible cognitive control advantage? This study will compare heritage speakers to advanced L2ers to look at the impact of factors like speech and text input, age of acquisition, and language of education on language processing and is the first to examine the performance of heritage speakers in a Stroop task, which measures executive function. Filling this gap in the literature is pertinent because the Stroop task targets text processing, and one of distinctive features of heritage speakers is that their ratio of speech to text input is heavily skewed towards speech compared to typical L1 development since they are usually not schooled in their L1. Another distinctive feature of heritage speakers is the amount of variation in adulthood L1 proficiency, due to different amounts and types of input during childhood, contrasting with typical L1 development which consistently leads to native-adult-like proficiency. Looking at heritage language Stroop interference can help us identify the effects of varying proficiency and exposure to speech on performance.

1.2 Bilingual processing and a bilingual advantage

Heritage speakers can be bilinguals of many types, simultaneous or early sequential, balanced or unbalanced, etc. A prominent paradigm for describing bilingual memory and processing is the Bilingual Interactive Activation Plus (BIA+) Model (Dijkstra & van Heuven 2002). In this model, the bilingual lexicon is integrated across languages and is accessed in a non-language-selective bottom-up way orthographically, phonologically, and semantically. While membership tags for words exist, it occurs too late for lexical identification purposes. Along with linguistic context, a number of lexical candidates are activated in parallel depending on their similarity to the input string and on the resting level activation of the individual items, which depends on subjective frequency, recency of use, and L2 proficiency. L2 words may be activated less due to lower frequency of usage and proficiency. When the L1 and L2 share a writing system, words from both can be activated and language membership is not always immediately discriminated within the mental lexicon. However, with different writing systems, like the case with Chinese and English, the activated set of neighbors is much smaller and more restricted to one language due to written features, especially in a reading task and when input is language specific. Also, cross linguistic effects are larger from L1 to L2 because of temporal delay assumption, where L1 tends to be activated first before L2 due to higher resting activation. While performing a task (such as lexical decision or a

Stroop task), an early preconscious, automatic level of processing may be followed by an attention-sensitive level in which precepts are selected with reference to contextual factors and linked to particular responses relevant to the task at hand; only the latter can be controlled. Non-linguistic contextual information, such as the instructions of the Stroop task, leads to an adaptation of decision criteria but would not affect the activity in the identification system itself, meaning there is no initial top-down inhibition.

The Bilingual L1 Lexical Disadvantage Hypothesis proposes that the presence of two or more lexicons in the brain results in competition and interactions between lexical candidates that creates delays in lexical access in comparison to monolinguals (Coderre, van Heuven & Conklin 2013). Similarly, the Reduced Frequency Hypothesis (Pyers, Gollan & Emmorey 2009) also proposes a bilingual disadvantage because both the L1 and the L2 are used less frequently because both languages are spoken. While many studies have presented data disproving a bilingual disadvantage in comparison to monolinguals in their native or dominant language, a delay in lexical access to the L2 in bilingual speakers compared to their L1 was found (Coderre et al. 2013), supporting the Bilingual L2 Lexical Disadvantage Hypothesis. The temporal delay assumption of the BIA+ model (Dijkstra & van Heuven 2002) also agrees with the Bilingual L2 Lexical Disadvantage Hypothesis, stating that the activation speed of orthographic, phonological and semantic codes is delayed in the L2 or less dominant language of unbalanced bilinguals compared to their L1 or more dominant language due to reduced L2 frequency and proficiency.

According to the order of acquisition, the L1 of a heritage speaker would be their heritage language, but if measured by current proficiency, the heritage language would be the L2, with the societally dominant language as their L1. An explanation using the latter interpretation is more consistent with the BIA+ model or Bilingual L2 Lexical Disadvantage Hypothesis. While heritage speakers share the experience of exposure to L1 since birth with native speakers, they also have similarities with second language learners in their heritage language, such as reduced input, differing types of input, and lower frequency of use in comparison to monolingual or native speakers. For example, when comparing unaccusativity in Spanish L2 learners to heritage speakers, Montrul (2005) found that although heritage speakers have an advantage in proficiency compared to L2 learners due to their primary linguistic input as a child, heritage speaker L1 grammar at a given stabilised state (likely endstate) resembles the incomplete (either stabilised or developing) grammars of intermediate or advanced L2 learners. Therefore, the L1 of heritage speakers, in this case Chinese, may be subjugated to a sort of L2 position due to a neglect of input and use after the start of preschool, which would lead to an increasingly lower activation level in their heritage language and an increasingly higher one in their L2, English, along with higher L2 proficiency and input. Thus, L1 activation must compete with L2 due to decreased input in both, but especially in the L1 (Schmid & Köpke 2017).

Being bilingual may have both costs, like lower formal language proficiency, and benefits, like enhanced executive control in tasks requiring conflict resolution or inhibitory control such as the Stroop and Simon task (Bialystok & Craik 2010).

Bilinguals are assumed to have a mechanism for controlling attention to their two linguistic systems for use and comprehension, which would boost the development of executive control during childhood (Bialystok 2007). Many studies support a bilingual advantage, where bilinguals are better than monolinguals and faster at suppressing task-irrelevant information during cognitive control tasks like the Stroop task, leading to less semantic interference (Blumenfeld & Marian 2014). The bilingual advantage was found to be task-specific, finding that bilinguals were significantly faster and more accurate on lexical retrieval and executive control tasks such as the Stroop, and also age-specific, with the bilingual advantage being larger for participants older than age 50, suggesting that bilingualism protects the aging-related decay of such processing (Ware, Kirkovski & Lum 2020, Bialystok, Craik & Luk 2008). On the other hand, other studies claim to not find any bilingual effects compared to monolinguals in such tasks (Kousaie & Phillips 2012, Paap, Anders-Jefferson, Mason, Alvarado & Zimiga 2018).

1.3 *The Stroop task in a bilingual context*

Because the Stroop test is a canonical test of selective attention, processing speed, and executive function, when adopted to a bilingual context, it is an effective measure of a possible bilingual advantage. In the Stroop task, a colour word is printed in coloured ink that is either congruent or incongruent with the word meaning itself. When participants are told to name the colour of the ink, participants are slower and less accurate in the incongruent condition than the congruent and control condition, as they need to consciously inhibit the influence of the meaning of the stimuli and engage in selective attention to produce the correct response. Jensen & Rohwer (1966) explains theories proposed to account for this effect – that word processing is faster than colour processing (processing speed), that colour recognition requires more attention than reading a word (selective attention), or that the stronger and more frequently used cognitive pathway of reading interferes with the weaker pathway of colour naming (parallel distributed processing). The most widely accepted explanation for the Stroop effect is the automatic processing of words as a result of habitual reading that conflicts with the colour naming, which is not automatic, in incongruent conditions (Posner & Snyder 1975); automatic reading does not require controlled attention, but uses enough attentional resources to reduce the amount of attention accessible for colour processing. Regardless of which interpretation has the most merit, heritage speaker executive control can be evaluated through their Stroop performance. Managing the interference requires selective attention and inhibition, which may be higher in bilinguals than monolinguals because this mechanism is well-oiled from managing two or more grammars; for heritage speakers, we can also explore the impact of factors like age of onset and amount of text exposure and proficiency on Stroop interference management.

In bilingual Stroop tasks, there are: L1 stimuli and L1 responses (within-language), L2 stimuli and L2 responses (within-language), L1 stimuli and L2 responses (between-language), and L2 stimuli and L1 responses (between-language). Biederman & Tsao

(1979) reported a study in which they observed greater Stroop interference when the task was completely in Chinese by Chinese-English bilinguals, compared to an English Stroop task by monolingual English speakers. They argue for an orthographic variation hypothesis where logographic languages such as Chinese are processed differently in the brain than alphabetic languages such as English. On the other hand, Lee & Chan (2000) found that within-language Stroop effects of Chinese and English were similar despite differing orthographies. van Heuven, Conklin, Coderre, Guo & Dijkstra (2011) also found similar results for within-language conditions, but found that between-language Stroop interference correlates with cross-language similarity, with between-language interference for same-script languages like German, English, and Dutch being similar to within-language effects; but reduced between-language interference for differing scripts like Chinese, Uyghur, and English.

Furthermore, Chinese-English children of various ages showed greater within- than between-language interference when responding in their L1, Chinese; when responding in English, their L2, there was a developmental shift from more between- to more within-language interference, correlating with their building English abilities (Chen & Ho 1986). Fang, Tzeng & Alva (1981) found more within- than between-language interference for Chinese-English, Japanese-English, and Spanish-English bilinguals, as well as an inverse relationship between Stroop interference and degree of similarity between the orthographic structures of the two written languages, similar to van Heuven et al. (2011). These findings suggest that different processing mechanisms could be involved in reading logographic compared to phonetic scripts for bilingual processing.

This study focuses on Chinese-English bilingual Stroop effects with native and heritage speakers of Chinese. Both are bilingual groups but with very different linguistic experiences; the possible differences in performance can provide insight into how varying linguistic experience factors affect interference and cognitive control in a Stroop task, which relates to the idea of a bilingual advantage. This study also differs from previous bilingual Stroop experiments because there is a new control condition that looks at colour identification without requiring naming in a specific language. This condition may display the least interference because no access to language is necessary, only concepts. However, a similar amount of interference in this condition to naming in L1 or L2 would suggest an inextricable tie between the language system and concepts, where language is always accessed during conceptualisation, even without the task requiring it.

1.4 Aims and hypotheses

Firstly, we will test the robustness of previous bilingual Stroop literature findings through our own replication of Chinese-English between- and within-language Stroop. As there are contrasting claims on this subject, we will explore questions like which conditions have the highest interference and how valid is the orthographic variation hypothesis through our Stroop interference data. Further, the novel colour response conditions of this study will examine the degree of connection between

colour and linguistic processing since the task does not require naming the ink colour in any language, only visually identifying it.

Additionally, we will investigate how the amount of textual input during childhood impacts Stroop interference through the difference between bilingual native and heritage Chinese speakers. We would expect there to be lower interference for the language a participant is less proficient in and has had less textual input in, which is consistent with the BIA+ model of bilingual processing (Hypothesis 1). Therefore, heritage speakers would experience lower Stroop interference with Chinese stimuli compared to native speakers and higher interference with English stimuli due to their lower exposure to Chinese text across their lifetime.

We are also interested in how individual variation in the onset age of acquiring and reading L2 impacts Stroop interference. While this study will not look at the bilingual advantage (comparing bilingual to monolingual speakers), the question of if an advantage in selective attention is found in heritage speakers in comparison with late bilingual speakers (the native group) will be explored. We hypothesise that age of L2 acquisition and L2 reading is positively correlated with interference (Hypothesis 2). According to the BIA+ model, this would manifest particularly in the conditions where L1 is the Stroop stimuli because when L2 is learned later, the L1 usually has a higher activation level due to higher proficiency and use. Since the heritage group consists of simultaneous or early sequential bilingual speakers, they may perform better on the Stroop task compared to the native group, consisting of late sequential bilinguals, having longer practice managing two grammars and thus a more developed selective attention mechanism.

2 METHOD

2.1 Participants

This study targeted English-Chinese bilinguals that are heritage speakers of Chinese and Chinese-English bilinguals who are native speakers of Chinese and advanced L2 speakers of English. Participants must be older than 18 years, attend an English-speaking university, and have normal colour vision, no developmental disorders, and no reading disorders. Approval was obtained from the Ethics Research Committee of the Faculty of Modern and Medieval Languages and Linguistics at the University of Cambridge. Speakers of Chinese dialects were not excluded due to the use of the same writing system.

Data were collected from 49 eligible participants, with 21 participants from the native group and 28 from the heritage group (Table 1¹). The first group consists of university students who have grown up in China and attend an English-speaking university. They may have also attended an English speaking secondary school. All of these participants have passed the Test of English as a Foreign Language (TOEFL), demonstrating they have advanced English proficiency, but are Chinese-

¹ Note: The highest level of education completed by participants' parents was numericalized from 1 to 7: no schooling, elementary to middle school, high school, some college, trade/technical/ vocational training, Bachelor's degree, graduate degree.

dominant according to their questionnaire selection between the two, and will be referred to as the native speakers of Chinese. The second group consists of university students who have grown up in an English-speaking country who are first or second generation immigrants from a Chinese-speaking country. They all consider themselves to be bilingual in English and Chinese and able to read Chinese to varying degrees. The language they first acquired is Chinese, but they are English dominant, so will be referred to as heritage speakers of Chinese. Thus, Chinese is the first language both groups acquire, followed by English, but with large differences in age and circumstances. When asked which language the participant was overall better at (their dominant language), 20 out of 21 native speakers chose Chinese and 28 out of 28 heritage speakers chose English. There were no statistically significant differences in age or parental education between the two groups, but significant differences in values like years of living in Chinese-speaking and English-speaking environments (Table 1), among others which will be explored in the Results section.

	Mean Age in Years	Mean Years lived in a Chinese-speaking country	Mean Years lived in a English-speaking country	Mean/Median Parental Education Level
Native	21.381 SD = 4.914	18.333 SD = 3.527	3.143 SD = 2.689	Mean = 5.095 Median = 6 (bachelor's degree)
Heritage	20.714 SD = 2.141	2.321 SD = 4.128	18.429 SD = 4.316	Mean = 5.500 Median = 6 (bachelor's degree)
<i>p</i> -value	0.566	<0.001***	<0.001***	0.280

Table 1 Questionnaire Demographic Information of the Native and Heritage Participants.

2.2 Design

This study includes the classic monolingual (within-language) Stroop, the bilingual (between-language) Stroop, and the new colour response Stroop, leading to a total of six parts when accounting for both languages in each condition. While traditionally, participants respond verbally in Stroop tasks, we found reliable Stroop effects from an online format with keyboard responses, which has important potential for collecting large amounts of Stroop data quickly.

There are six parts (Figure 1), each with forty trials—twenty congruent and twenty incongruent. There are four English stimuli, yellow, green, red, or blue, and four equivalent Chinese stimuli, 黄, 绿, 红, and 蓝. The stimulus is written in the same colour it semantically codes for in the congruent condition, and in one of the other three colours in the incongruent condition. The four arrow keys on a computer keyboard are each associated with a colour — yellow, green, red, and blue — and these associations are changed after every part in order to prevent memorisation or habituation of the colour associations.

In the first two parts (Figure 1a and Figure 1b), the participant responds with arrow keys which represent the colour they are filled as, and the stimuli are in

English in the first part and Chinese in the second. In the next two parts (Figure 1c and Figure 1d), the arrow keys represent the four English colour words written in black ink, and the stimulus is in English, and in the following part, Chinese. In the last two parts (Figure 1e and Figure 1f), the arrow keys represent the four Chinese colour words written in black ink, and the stimulus is again in English, and then Chinese.



Figure 1 Example Screen for Each Part of the Task.

The response times for each trial are collected. The task is always to choose the colour of the text instead of its meaning. The interference is calculated as the difference in reaction time between the incongruent and congruent condition.

2.3 Procedure

The study, created with Gorilla Experiment Builder, was conducted online in order to receive responses from target participants in various locations. A link was created and shared on social media as a means of recruitment.

First, the participant is shown a screen that confirms the participant fulfills the requirements of the study. Then, there is a consent form with information on the study and its purpose, the data that will be collected and the researcher. Afterwards, the participants are shown a questionnaire in English asking demographic questions and language ability, exposure, usage, and attitude questions, taking inspiration from the LEAP-Q (Marian, Blumenfeld & Kaushanskaya 2007), the LHQ3 (Li, Zhang, Yu & Zhao 2019), and the Multilingual Language Use Questionnaire (Cohn, Bowden,

McKinnon, Ravindranath, Simanjuntak & Taylor 2013). Next, an introductory screen describes in Chinese and English the participants' task of pressing the arrow key that corresponds with the colour of ink that a series of colour words, shown one by one, are written in. In each trial, below the stimulus word is the arrow key correspondence diagram (Figure 1). Before each new type of response (colour, English, or Chinese), the participants complete a short practice session of six words, three in English and three in Chinese, to press the correct arrow key corresponding to the colour of the text. In each practice session, only two of the six stimulus words in the practice trials are colour words in order to reduce priming effects. Feedback on if the correct arrow key was pressed is given after each practice trial. The participants are informed with a screen before each of the six actual tasks begin, where there will no longer be correctness feedback.

After all the parts are completed, participants are shown a thank you screen and a link to share with friends who may be eligible for the study. In total, the study takes approximately 15 minutes.

2.4 Data analysis methods

The programming language R was used for statistical analysis, with the additional package 'Hmisc.' Statistical significance will be marked by asterisks, with * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

For the questionnaire, averages and heteroscedastic t-tests were used to compare the native and heritage group's numerical answers and examine statistical significance. For categorical questionnaire data, the assumptions of the chi-squared test were not met because the calculations resulted in almost all expected frequencies, or cell counts, being less than 5; instead, Fisher's exact test was used to compare the two participant groups and examine statistical significance.

For the task response time data, incorrect responses by participants were first removed, which was 5.442% of trials in the dataset. In order to perform paired t-tests comparing the congruent and incongruent conditions, reaction times were averaged for each participant per part, separated by congruent and incongruent conditions. Interferences were calculated by subtracting response time for incongruent stimuli from congruent stimuli. The average response times and interferences used were calculated from the averaged per participant data.

After using a correlation matrix to rule out multicollinearity, linear regression models were created for relationships between certain linguistic experience variables and interference per condition, using the original per participant questionnaire data and the average interferences per condition per participant.

3 RESULTS

3.1 Questionnaire

The tables highlight data from the questionnaire (other questionnaire results can be found in the [Appendix](#)).

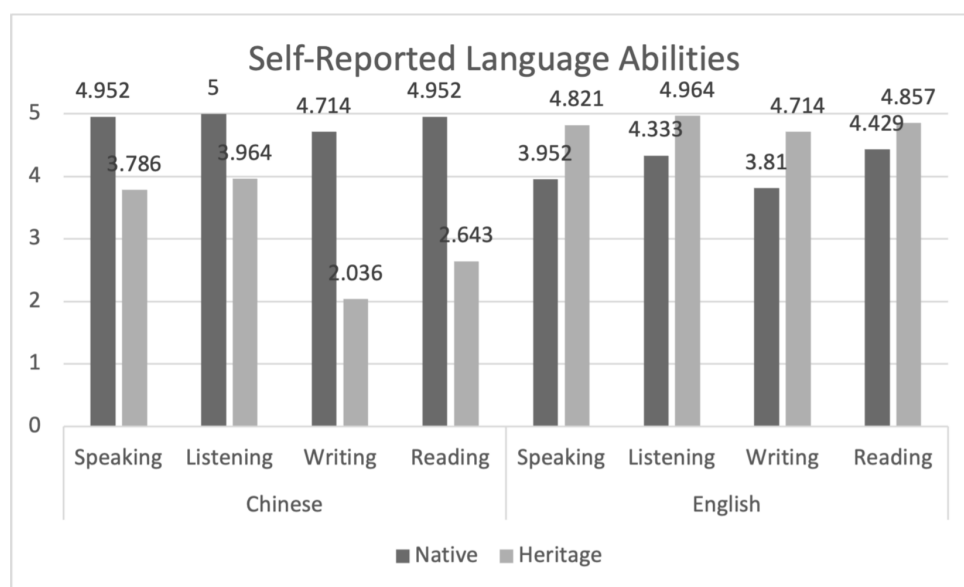


Figure 2 Average Self-reported Language Abilities in Chinese and English on a Scale of 1-5.

[Figure 2](#)² shows the self-reported Chinese and English language abilities from 1-5 in speaking, listening, writing, and reading, of the native and heritage group. The native group reported near perfect L1 Chinese abilities in all categories, and lower, but still high L2 English abilities due to their advanced English education. The heritage group reported near perfect English abilities in all categories, and lower abilities in Chinese, especially in reading and writing.

[Table 2](#) demonstrates that there is no statistically significant difference in when both groups began acquiring Chinese (at birth), so they both start as native speakers, though the heritage group does not reach complete native-like fluency in adulthood. The heritage speakers also learned to read in Chinese later than the native speakers, around mean age six instead of age three. After being primarily exposed to Chinese, the heritage group begins to acquire English at mean age two, and learns to read at a typical age English-monolingual children do; their English abilities soon overtake their Chinese due to overwhelming input and use outside of the home, a discrepancy that lasts into adulthood ([Figure 2](#)). The native group starts learning English at school age (speaking around age five, reading around age seven), but only during

² All differences between the native and heritage group for each category were statistically significant. The p -values for Chinese speaking, listening, reading, and writing, and English speaking, listening, and writing were less than 0.001***. The p -value for English reading was 0.007**.

designated English classes which last a small portion of the day. These findings validate the grouping methodology we used to separate the native and heritage group and conform with our expectations about each group.

(Means)	Chinese				English			
	Onset Age of Acquisition (yrs)	Onset age of Reading (yrs)	% of Speech Input	% of Text Input	Onset Age of Acquisition (yrs)	Onset Age of Reading (yrs)	% of Speech Input	% of Text Input
Native	0	3.238	77.667	72.571	5.381	7.381	22.333	27.429
Heritage	1.286	6.3571	40.857	20.464	2.25	4.4286	59.143	79.536
<i>p</i> -value	0.150	0.003**	<0.001***	<0.001***	<0.001***	0.001**	<0.001***	0.007**

Table 2 Participant Onset Age of Speaking and Reading in Each Language and Proportion of Input in Childhood.

Table 3³ also aligns with the linguistic and experiential profiles of the two participant groups. The null hypothesis using Fisher’s exact test is that being a member of the heritage or native group is independent from a given factor contributing to their acquisition of Chinese or English. The analyses suggest that there is a significant association between participant native/heritage group affiliation and the following factors’ influence on Chinese and English acquisition: family, language of schooling, language classes, friends, and overall immersion. For example, most participants from both groups learned Chinese from their families, but the heritage group also learned English from their families. Many from the native but not the heritage group learned Chinese from schooling, but both learned English from schooling. Environmental immersion contributed to the acquisition of Chinese and English in both groups, but more so for Chinese in the native group and English for the heritage group. Almost all from the native group learned English from language classes, and while some heritage speakers learned Chinese from supplemental language classes, they acquired English from school as it was the language of their education. More heritage speakers learned English from their friends than the native speakers.

Furthermore, the native group rated their English production (speaking and writing < 4) on average slightly lower than their comprehension (listening and reading > 4), a characteristic commonly associated with L2 learners. Table 2 shows that a fifth to a quarter of their speech and text input was in English during the whole of their childhood, likely concentrated in the later teenage years, but the rest of their input was in Chinese.

³ Each factor may contribute to one, both, or neither of a participant’s acquisition of Chinese and/or English; thus, the sums of the Chinese and English counts may add up to more than the number of participants in the group. *p* values were calculated with Fisher’s exact test.

Factor	Language	Native Speaker Count out of 21	Heritage Speaker Count out of 28	<i>p</i> -value
Family	Chinese	17	26	0.028*
	English	1	7	
Language of School Instruction	Chinese	12	2	<0.001***
	English	17	26	
Language Classes	Chinese	6	15	<0.001***
	English	20	4	
Friends	Chinese	10	6	<0.001***
	English	8	21	
Media (TV, films, books, social media)	Chinese	11	11	0.088
	English	11	22	
Overall environmental linguistic immersion	Chinese	19	10	<0.001***
	English	11	22	

Table 3 Factors Contributing to Participant Acquisition of Chinese and English.

Table 3 shows the native group acquired Chinese mainly from family, schooling, and linguistic immersion, and in about half of cases, friends and media; they acquired English mainly from schooling and English classes, and in about half of cases, friends, media, and linguistic immersion. This aligns with the story that they acquired Chinese in a Chinese-speaking environment with some English classes, and then later changed to an English-speaking environment and school. The heritage group rated their Chinese speaking and listening abilities higher (almost 4 on average) than their writing and reading skills (less than 3). Lower text compared to verbal proficiency is commonly found in heritage speakers (Benmamoun et al. 2013), and can be explained by Table 2 and Table 3. Table 2 demonstrates that while the majority of input heritage speakers received was in English, they received a high amount of speech input in Chinese (mostly from family), a mean of 41% of all speech input, but less Chinese text input, only around 20% of all text input. Most textual input occurs during school and thus would be in English for the heritage group. Table 3 shows that family is the main factor contributing to heritage speaker acquisition of Chinese, and in many cases, Chinese language classes, media, and/or linguistic immersion. The heritage group acquired English mainly from school, friends, media, and linguistic immersion. Therefore, as expected, the heritage group acquires Chinese mainly from parents, and English from everywhere else, including school. Along with the smaller amount of childhood Chinese text input, this explains the difference in proficiency between Chinese speaking/listening and reading/writing.

3.2 Stroop interference by condition and participant group

Table 4 and Table 5 display mean reaction times and interferences and their p -values for each condition, separated by the native and heritage group, and Figure 3 provides a visual representation for the interference data in these two tables.

Native	Congruent	Incongruent	Interference	p -value
English stimulus/colour response	916.905	1029.611	112.706	0.00019***
Chinese stimulus/colour response	813.791	1005.296	191.505	0.00070***
English stimulus/English response	1391.633	1607.041	215.409	0.00165**
Chinese stimulus/Chinese response	1099.506	1268.799	169.293	0.00958**
English stimulus/Chinese response	1251.147	1455.545	204.398	0.00146**
Chinese stimulus/English response	1424.232	1689.789	265.557	0.00107**

Table 4 Native Speaker Stroop Response Times and Interference per Condition in ms.

Heritage	Congruent	Incongruent	Interference	p -value
English stimulus/colour response	898.442	1013.604	115.162	0.00014***
Chinese stimulus/colour response	807.863	896.599	88.735	0.00194**
English stimulus/English response	1237.340	1430.993	193.652	0.00199**
Chinese stimulus/Chinese response	1168.731	1264.857	96.126	0.37292
English stimulus/Chinese response	1352.818	1509.838	157.020	0.03908*
Chinese stimulus/English response	1188.059	1386.469	198.410	<0.001***

Table 5 Heritage Speaker Stroop Response Times and Interference per Condition in ms.

The interferences, or the differences in reaction time between the incongruent and congruent conditions, were all significant, except for the Chinese stimulus/Chinese response condition for heritage speakers, which showed the second lowest mean interference of 96.126ms.

For the native speakers, interference ordered from lowest to highest is: (1) English stimulus/colour response, (2) Chinese stimulus/Chinese response, (3) Chinese stimulus/colour response, (4) English stimulus/Chinese response, (5) English stimulus/English response, (6) Chinese stimulus/English response (Table 4). For the heritage speakers, interference ordered from lowest to highest is: (1) Chinese stimulus/colour response, (2) Chinese stimulus/Chinese response, (3) English stimulus/colour response, (4) English stimulus/Chinese response, (5) English stimulus/English response, (6) Chinese stimulus/English response (Table 5). The differences between the interferences of each condition did not all reach significance (Table 6 and Table 7).

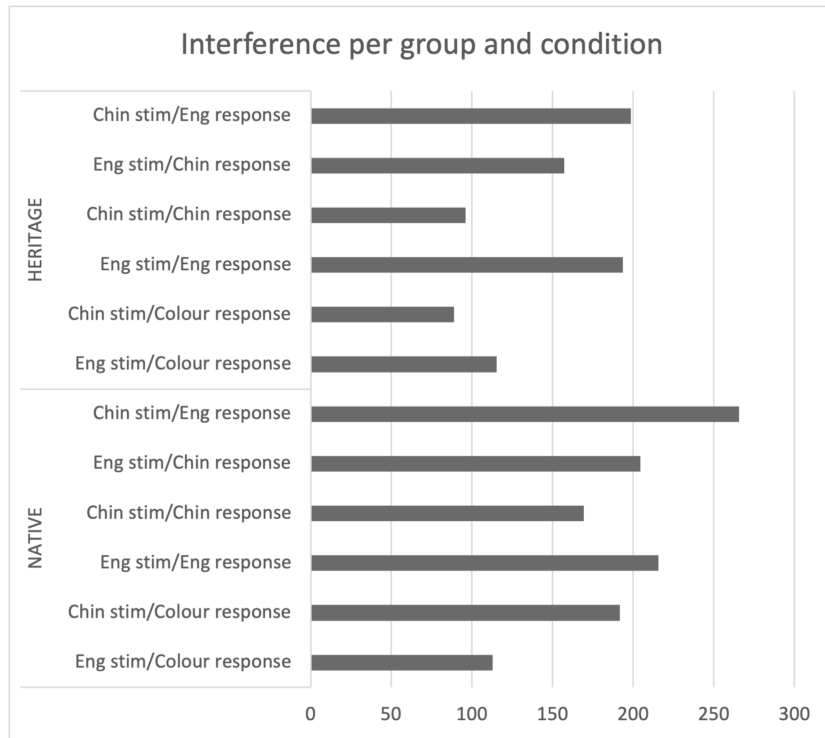


Figure 3 Stroop Interference in ms per Group and Condition.

	English stimulus/ colour response	Chinese stimulus/ colour response	English stimulus/ English response	Chinese stimulus/ Chinese response	English stimulus/ Chinese response	Chinese stimulus/ English response
English stimulus/ colour response	-	-	-	-	-	-
Chinese stimulus/ colour response	0.123	-	-	-	-	-
English stimulus/ English response	0.043*	0.744	-	-	-	-
Chinese stimulus/ Chinese response	0.263	0.775	0.545	-	-	-
English stimulus/ Chinese response	0.070	0.870	0.839	0.672	-	-
Chinese stimulus/ English response	0.043*	0.540	0.706	0.224	0.356	-

Table 6 Native Speaker Stroop Interference *p*-values between Conditions.

For the native group, two were statistically significant: the difference in interference between the English stimulus/colour response and English stimulus/English response conditions, and the difference in interference between the English stimulus/colour response and Chinese stimulus/English response conditions. In other words, there were significant differences between the lowest and highest interference conditions, and between the lowest and second highest inference conditions, for the native group. For the heritage group, the difference in interference between

the Chinese stimulus/colour response condition and Chinese stimulus/English response condition — which are the lowest and highest interference conditions for the heritage group — was significant.

	English stimulus/ colour response	Chinese stimulus/ colour response	English stimulus/ English response	Chinese stimulus/ Chinese response	English stimulus/ Chinese response	Chinese stimulus/ English response
English stimulus/ colour response	-	-	-	-	-	-
Chinese stimulus/ colour response	0.422	-	-	-	-	-
English stimulus/ English response	0.229	0.117	-	-	-	-
Chinese stimulus/ Chinese response	0.878	0.927	0.383	-	-	-
English stimulus/ Chinese response	0.615	0.420	0.681	0.393	-	-
Chinese stimulus/ English response	0.057	0.025*	0.948	0.194	0.615	-

Table 7 Heritage Speaker Stroop Interference p -values between Conditions.

Interestingly, the order of interferences for native and heritage speakers is the same, except that English stimulus/colour response and Chinese stimulus/colour response are interchanged. The mean interference for native speakers is lowest in the English stimulus/colour response condition, which corresponds to the later age at which they acquired and started reading English (Table 2), the lower exposure they had to English speech and text (Table 2), and their lower self-reported English abilities (Figure 2) compared to heritage speakers and compared to their own experience with Chinese. Similarly, for heritage speakers, their lowest mean interference was in the Chinese stimulus/colour response condition, and they reported lower exposure (Table 2) and language skills (Figure 2) in Chinese compared to native speakers compared to their own experience with English. Additionally, the Chinese stimulus/English response condition had the highest interference for both native and heritage speakers. While this aligns with Hypothesis 1 and the BIA+ model for the native speakers, who would experience high interference in their L1, Chinese, it is surprising that there was such high interference in the Chinese stimulus/English response condition for heritage speakers, since the stimuli in Chinese, their less-dominant language, would be expected to result in low interference — it did for the Chinese stimulus/colour response condition, but not for the Chinese stimulus/English response condition. There was also a statistically significant difference between the interferences of these two conditions, so the lowest and highest interference conditions for heritage speakers both had Chinese stimuli.

3.3 Individual variation of acquisition and reading onset ages

Before creating a possible model to represent relationships between questionnaire data and the interference times from the Stroop task, a correlation matrix was created out of the variables of interest in order to prevent multicollinearity (Table 8). Many of the variables were found to be highly correlated.

	Self-reported Chinese reading ability out of 5	Self-reported English reading ability out of 5	Onset Age of Chinese Acquisition in yrs	Onset Age of English Acquisition in yrs	Onset age of Reading in Chinese in yrs	Onset age of Reading in English in yrs	% of Childhood Exposure to Chinese (vs. English) Speech	% of Childhood Exposure to Chinese (vs. English) Text
Self-reported Chinese reading ability out of 5	1.00	-	-	-	-	-	-	-
Self-reported English reading ability out of 5	-0.33 <i>p</i> = 0.0193	1.00	-	-	-	-	-	-
Onset Age of Chinese Acquisition in yrs	-0.02 <i>p</i> = 0.8705	0.12 <i>p</i> = 0.3954	1.00	-	-	-	-	-
Onset Age of English Acquisition in yrs	0.59 <i>p</i> = 0.0000	-0.37 <i>p</i> = 0.0094	-0.25 <i>p</i> = 0.0887	1.00	-	-	-	-
Onset age of Reading in Chinese in yrs	-0.19 <i>p</i> = 0.1883	0.02 <i>p</i> = 0.8723	0.68 <i>p</i> = 0.0000	-0.25 <i>p</i> = 0.0829	1.00	-	-	-
Onset age of Reading in English in yrs	0.32 <i>p</i> = 0.0262	-0.15 <i>p</i> = 0.2992	-0.25 <i>p</i> = 0.0821	0.50 <i>p</i> = 0.0003	-0.27 <i>p</i> = 0.0655	1.00	-	-
% of Childhood Exposure to Chinese (vs. English) Speech	0.53 <i>p</i> = 0.0001	-0.33 <i>p</i> = 0.0226	-0.43 <i>p</i> = 0.0023	0.48 <i>p</i> = 0.0005	-0.62 <i>p</i> = 0.0000	0.37 <i>p</i> = 0.0095	1.00	-
% of Childhood Exposure to Chinese (vs. English) Text	0.70 <i>p</i> = 0.0000	-0.40 <i>p</i> = 0.0048	-0.26 <i>p</i> = 0.0697	0.53 <i>p</i> = 0.0000	-0.46 <i>p</i> = 0.0009	0.32 <i>p</i> = 0.0243	0.83 <i>p</i> = 0.0000	1.00

Table 8 Correlation and *p*-value Matrix for Questionnaire Variables of Interest.

Comparing Chinese-English Bilingual and Heritage Speaker Stroop Effects

The correlation between onset age of Chinese acquisition and onset age of English acquisition was not significant, as well as the correlation between onset age of reading in Chinese and onset age of reading in English. Therefore, linear regressions were run on with these two pairs of variables (separately) in order to investigate a relationship between these participant variables and the interference per condition.

Linear regressions were run for onset age of acquisition, relating interference times with onset age of Chinese acquisition and onset age of English acquisition per condition; the coefficient of onset age of English acquisition was statistically significant for the Chinese stimulus/English response condition (Figure 4). Moreover, in linear regressions relating interference times with onset age of reading in Chinese and English, the coefficient of onset age of reading in English was statistically significant for the Chinese stimulus/English response condition (Figure 5).

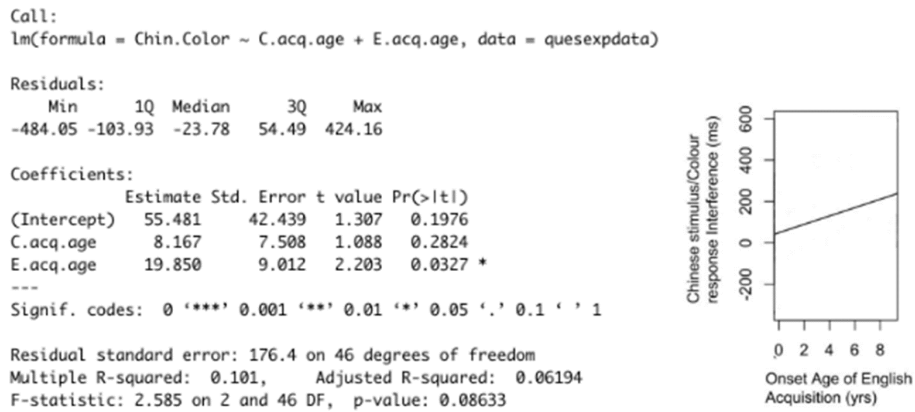


Figure 4 Chinese Stimulus/colour Response: Linear Regression with Onset Ages of Acquisition in Chinese and English.

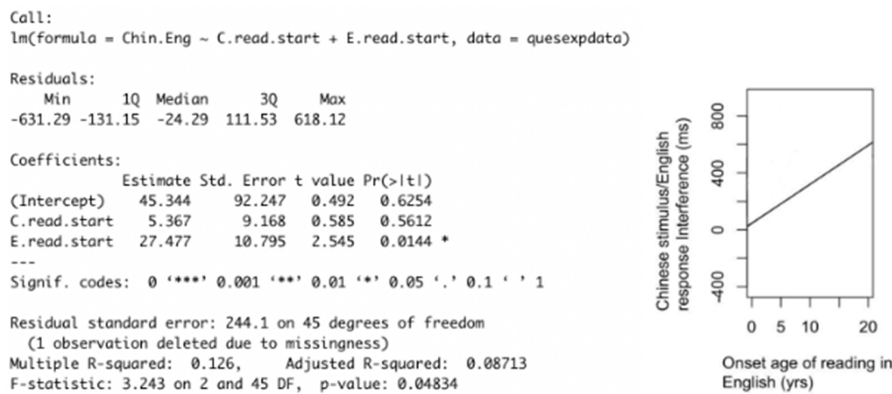


Figure 5 Chinese Stimulus/English Response: Linear Regression with Reading Onset Ages in Chinese and English.

The model in [Figure 4](#) demonstrates that for every yearly increase in onset age of English acquisition, the mean interference time in the Chinese stimulus/colour response condition increases by 19.850 ms, while holding all other variables, such as onset age of Chinese acquisition, constant ($p < 0.05$). Similarly, the model in [Figure 5](#) demonstrates that for every yearly increase in onset age of reading in English, the mean interference time in the Chinese stimulus/English response condition increases by 27.477 ms, while holding all other variables, such as onset age of reading in Chinese, constant ($p < 0.05$). This suggests that later ages of starting to acquire and read English is associated with more interference when the stimulus is in Chinese. In order to investigate if this trend was evidenced explicitly within the native or heritage group, separate linear regressions for each group's data were run. When the data were divided by participant group, in both the Chinese stimulus/colour response acquisition model in [Figure 4](#) and the Chinese stimulus/English response reading model in [Figure 5](#), the resulting coefficients were not significant. Therefore, the relationships found between acquisition or reading onset age and interference represent the differences between and throughout the native and heritage groups.

The data in [Table 2](#) also supports this proposition, showing a significant difference in the onset age of reading in English and the onset age of acquisition in English between the native and heritage group, with native speakers reporting later ages for both. In addition, native speakers have statistically significant higher exposure to Chinese text and speech during childhood ([Table 2](#)), and rate themselves as having higher Chinese abilities in all areas ([Figure 2](#)). The higher exposure to and language abilities in Chinese of native speakers compared to heritage speakers may lead to higher interference when the stimulus is in Chinese, as it would be harder to suppress the automaticity of reading Chinese. Earlier L2 acquisition and reading onset, which heritage speakers experience with English, is related to lower interference in these two models.

4 DISCUSSION

4.1 *Within and between-language Stroop findings*

We will first discuss the four conditions already present in the literature—within-and between-language Stroop, and how they compare to previous research findings. The identical ordering of mean interference for these four conditions in both groups, namely (from least to most interference): Chinese stimulus/Chinese response, English stimulus/Chinese response, English stimulus/English response, and Chinese stimulus/English response, is interesting to investigate, although not all of these differences reached statistical significance ([Table 4](#) through [Table 7](#)), possibly due to the averaging involved in the data analysis. [Marian, Blumenfeld, Mizrahi, Kania & Cordes \(2013\)](#) found that multilinguals were faster and more accurate in within-than between-language Stroop tasks, indicating that additional processing costs are required when stimulus and response languages differ. If Chinese is considered to be the L1 of both heritage and native speakers, defined by the first language

a child acquires, this finding would be applicable, with Chinese within-language Stroop having the least interference of these four conditions. However, if L1 is defined by the higher proficiency language, this finding would only apply to the native group, as English is the dominant language of heritage speakers. Fang et al. (1981) found greater within- than between-language interference in L2 English speakers who have Spanish, Japanese, or Chinese as their L1. Chen & Ho (1986) also found greater within- than between-language interference when L2 English participants responded in their L1, Chinese. These conclusions contrast with our findings, which demonstrate that the Chinese within-language Stroop condition resulted in the lowest interference for native and heritage groups, particularly if we consider Chinese the L1 for both. If we consider the L1 of heritage speakers to be English, the English-within language interference was still not the highest, and instead the interference of a between-language condition was. Therefore, previous research stipulating the difference between L1 and L2 within- and between-language interference is not completely consistent with the results of our study. Replicating the study with more participants would provide insight into these results, and if they resulted from the averaging of response times for each participant or are reliable patterns to be further investigated.

Another way to interpret the similar orderings of these four conditions between the two groups is to attribute the pattern to the nature of the tasks themselves with the orthographic variation hypothesis, where logographic languages and alphabetic languages are processed differently (Biederman & Tsao 1979). In our study, Chinese-English bilinguals, or the native group, as well as the heritage group, demonstrated the lowest interference in the Chinese within-language condition out of the four traditional conditions, challenging the claim that Chinese creates greater interference. When Lee & Chan (2000) asked bilinguals to complete a Chinese Stroop and an English Stroop and English monolinguals to complete an English Stroop, they found no significant difference when the task was conducted in English compared to in Chinese within or between the bilingual and monolingual group. In addition, Chen (2000) conducted a meta-analysis that contradicted the orthographic variation hypothesis, not finding varying within-language Stroop interference between different writing systems. Our findings do not support the orthographic variation hypothesis because the Chinese within-language Stroop does not lead to the greater interference than English within-language Stroop, which Lee & Chan (2000) and Chen (2000) both propose. However, the reverse might be possible — that English within-language Stroop creates greater interference than Chinese — indicating that there still may be a difference in processing the two language systems.

English has an alphabetic writing system where a standard set of letters represent speech sounds which can be manifested in different ways depending on the phonological environment. On the other hand, Chinese has a logographic writing system where each character represents a semantic unit such as a word or morpheme, instead of a phonetic element. According to Coltheart, Rastle, Perry, Langdon & Ziegler (2001)'s dual-route model of reading, meaning can be accessed from letters to phonemes to phonology to meaning, or directly from letters to orthography to

meaning. This dual model is most compatible with alphabetic languages like English, but for logographic languages like Chinese, there could be nuanced differences with the script to phonology/meaning route. There is not extensive research comparing the process of reading alphabetic compared to logographic languages. Reading may be equally automatic in either language, but different elements of the language are highlighted through their script, leading to differences in processing.

Since there is an inextricable connection between phonology and meaning, there may also be phonological in addition to semantic interference in the incongruent Stroop condition. The lack of encoded phonemes in logographic languages may suggest that the phonological element of words may be less prevalent when reading compared to an alphabetic language. [Mingjin, Hasko, Schulte-Körne & Bruder \(2012\)](#) found that ‘automatic association of characters and lexical tones in experienced Chinese adult readers requires more processing time than for alphabetic languages.’ If there is a component of phonological, in addition to semantic, interference in Stroop, the lack of explicit phonology in Chinese script might decrease interference in Stroop tasks where the stimulus text is in Chinese compared to English; despite both leading to semantic interference, there could be less phonological interference in the Chinese stimulus conditions for both participant groups due to the nature of the script.

However, it is difficult to make claims about the automaticity of reading one language compared to another from our results because the order of interference alternates by stimulus language, and the differences in interferences were mostly not significant between conditions in this experiment.

4.2 The new colour response condition

The two colour response conditions resulted in the lowest and third lowest interferences in both the native and heritage group, with Chinese stimulus/colour response being the lowest for the heritage group and English stimulus/colour response being the lowest for the native group. This new response condition seemed to result in a distinct interference effect as intended because there were significant differences between the colour response conditions and conditions with the same stimulus type but a traditional linguistic response; for example, there was a statistically significant difference between the English stimulus/colour response and the English stimulus/English response interferences for native speakers ([Table 6](#)), and between the Chinese stimulus/colour response and the Chinese stimulus/English response interferences for heritage speakers ([Table 7](#)). This suggests that language may not be automatically accessed during colour identification; this one less linguistic task may decrease reaction time and interference. They require visual decision-making based on colour but not explicit language retrieval to name the colour of the text in any particular language, isolating the interference effect to the stimulus itself when there is contradictory ink colour and meaning, possibly without further linguistic interference from the response options.

4.3 *The BIA+ Model: The Impact of Linguistic Experience on Interference*

[Hypothesis 1](#) predicted that there would be lower interference for the language a participant is less proficient in and has had less textual input in, consistent with the BIA+ model of bilingual processing and the Bilingual L2 Lexical Disadvantage Hypothesis. This hypothesis is strongly supported by evidence from the new colour response condition. The within- and between-language conditions also reveal some supporting evidence when compared to the new colour response conditions, but taken alone are not completely consistent with the hypothesis. The linear regressions' statistically significant coefficients relating onset ages and interference also provide support for the BIA+ model.

Our results relating to the novel colour response conditions support [Hypothesis 1](#). The native group experienced the lowest interference in the English stimulus/colour response condition, and the heritage group experienced the lowest interference in the Chinese stimulus/colour response condition ([Figure 3](#)). Thus, participants experience less interference in their less dominant language, as [Table 2](#) shows that the native group had a later onset age in English, lower exposure to English text and speech, and lower self-reported proficiencies in English, and the heritage group displays lower exposure and proficiency for Chinese.

According to the BIA+ model, each lexical item has a resting activation, which impacts its prevalence in one's lexicon and therefore processing speed. Their resting activation, which depends on frequency and recency of use and language proficiency, in tandem with similarity to input string, becomes activated. The temporal delay assumption assumes that L1 tends to be activated before L2. In heritage speakers, Chinese would be used less frequently and recently, mostly in family contexts; they otherwise grew up, were educated, and now attend university in an English-speaking environment, leading to low frequency of use and proficiency in Chinese, relegating it as an L2 of sorts below English in terms of resting activation level. Native speakers, whose L2 is English, would also experience temporal delay, due to later onset age and less exposure ([Table 2](#)) as well as lower self-reported proficiency ([Figure 2](#)) in English below Chinese. The magnitude of each group's temporal delay is uncertain – the native group's may be mitigated because they are proficient enough in English to attend an English-speaking university and are exposed to English very frequently in their daily life; the heritage group's may be mitigated by earlier native-like onset age of Chinese.

The Stroop task calls for fast response to the colour of the stimulus of the text and does not require semantic processing of the text, which is only a distraction that leads to interference. Lower resting level activation in Chinese for heritage speakers and English for native speakers means that participants of each group would experience less interference in their respective lower proficiency or L2 languages because they are not activated as quickly as their dominant language; this temporal delay when there is L2 stimuli lessens Stroop interference. This is consistent with native speakers demonstrating the lowest interference for the English stimulus/colour response condition, and heritage speakers demonstrating the lowest interference for the Chinese stimulus/colour response condition. The text is processed slower

and therefore to a lesser degree in reaction-time-related tasks like the Stroop, and instead they are able to quickly complete the colour identification in incongruent conditions with less semantic interference from the stimulus, in combination with the lower interference from responding without colour names in colour response condition.

The significant difference in interference between English stimulus/colour response and Chinese stimulus/English response conditions for the native group (Table 6) also supports the Bilingual L2 Lexical Disadvantage Hypothesis; they experienced significantly less interference when the Stroop stimuli is in their L2, English, compared to when it is in their L1, Chinese, which supports Hypothesis 1. The concept of temporal delay also predicts this outcome since cross-linguistic interference is higher from their L1, Chinese, onto their L2, English, and the Chinese stimulus/English response condition creates interference from Chinese onto English.

Moreover, heritage speakers displayed relatively lower interference compared to the native group in the Chinese stimulus/colour response and Chinese stimulus/Chinese response. According to the BIA+ model, this is explained by heritage speakers' lower usage and exposure to and thus lower activation levels for Chinese compared to English lexical items. However, from the perspective of the orthographic variation, heritage and L2 speakers of Chinese have difficulty with learning to associate the correct phonemes with characters both when character meaning is and is not accessed. As the Stroop task does not require and in fact is slowed by understanding the stimulus text, this may have been helpful to them when the stimulus is in Chinese. The native group does not experience as little interference when their less dominant language, English, is the stimulus. In the BIA+ model, this could be explained by the frequent and recent usage of English by the native group in everyday life, but taking into account orthographic variation, this could be because phonemes are always encoded in English script through letter combinations, and this letter to sound association is pertinent to both L1 and L2 speakers at high proficiency levels, leading to both phonological and semantic interference.

When we look just at the within- and between-language conditions, without reference to the colour response conditions, the four identical orderings of interference for both groups are inconsistent by stimulus type and thus do not completely conform to the assumptions of Hypothesis 1, that participants would experience the greatest interference when stimuli are in their dominant language, Chinese for native speakers and English for heritage speakers. Because many differences in interference values between the Stroop conditions did not reach significance, the experiment would benefit from replication at larger scales. However, Hypothesis 1 is supported by the statistically insignificant interference in the Chinese stimulus/Chinese response condition for heritage speakers (Table 5). The BIA+ model would assert that the heritage language of speakers would have lower resting activation and therefore be less prevalent in the lexicon due to incomplete or attrited development, lower proficiency, and lower frequency of use. If we interpret the Chinese processing of heritage speakers like that of an L2, the statistically insignificant interference of the Chinese stimulus/Chinese response condition for heritage speakers supports the Bilingual L2 Lexical Disadvantage Hypothesis. The heritage

group would experience temporal delay in Chinese compared to English as unbalanced bilinguals, so in both incongruent and congruent conditions, the Chinese stimulus meaning is not accessed in time for it to interfere or affect the Stroop task decision-making in colour identification, leading to no statistically significant difference in response time between the two conditions.

Lastly, the linear regression models corroborate the BIA+ model. They revealed two significant correlations: one between age of English acquisition and interference for the Chinese stimulus/colour response condition, and one between age of starting to read in English and interference for the Chinese stimulus/English response condition. This trend was found to not be within the native or heritage group, but throughout all participants, indicating a difference between groups. According to the questionnaire, the native group has later onset acquisition and reading ages of English, their L2, compared to the heritage group. Thus, higher resting activation of their more dominant language and L1, Chinese, due to earlier acquisition, higher proficiency, and higher frequency of use and exposure, leads to faster processing for L1 input. This is a distractor in incongruent trials of the Stroop task, which leads to higher interference for Chinese stimuli by the native group (who acquired English later) compared to the heritage group.

4.4 A heritage advantage: relating onset ages and interference

[Hypothesis 2](#) predicted that onset age of L2 acquisition and L2 reading is positively correlated with interference, especially with L1 Stroop stimuli; in other words, the earlier the L2, the better the performance. This is supported by the linear regression models.

As previously mentioned, across participants from both groups, there was a significant correlation between age of English acquisition and interference for the Chinese stimulus/colour response condition, and between onset age of reading in English and interference for the Chinese stimulus/English response condition. Therefore, later acquisition of English speech and reading (L2) is connected with higher interference, particularly for Chinese (L1) stimuli, consistent with [Hypothesis 2](#). This seems to suggest the existence of a selective attention (or a sort of ‘bilingual advantage’) for heritage speakers in comparison to later bilinguals. In the Chinese stimulus/colour response condition, the model predicts that the later in life one acquires English, the L2 for both native and heritage speakers, the more interference from L1 (the Chinese stimulus) there is; in the Chinese stimulus/English response condition, the later one learns to read English, the more interference as well. The heritage group consists of simultaneous or early sequential bilinguals, who learn their L2 much earlier than the native group, which consists of later sequential bilinguals. Therefore, in these tasks, the heritage speakers who acquired and started to read their L2 earlier might have a more developed selective attention advantage – from juggling two grammars – where they are able to inhibit irrelevant information, like the meaning of the stimuli, to complete the Stroop decision more efficiently, leading to less interference. In the Chinese stimulus/English response condition, which is a bilingual task testing interference from Chinese onto English, there is

also a disadvantage when English is acquired later because the English grammar and lexicon would be weaker and Chinese would be more established and therefore exert more interference due to this imbalance. Consequently, in both a monolingual and bilingual Stroop task, there may be a cognitive control advantage for acquiring and reading an L2 earlier, supporting [Hypothesis 2](#). These data suggest that a possible advantage in selective attention is on a gradient, where there are more benefits for cognitive control when the onset of bilingualism is earlier, which would give heritage speakers and other simultaneous/early bilinguals a greater benefit.

5 IMPLICATIONS

The presented findings can inform immigrant families and educational systems on the best way to support immigrant children, despite it not being the main aim of the study. Many parents exclusively speak the societally dominant language with their children in an attempt to quicken assimilation in the host country, leading to a decline or complete loss of the heritage language in future generations. However, literature supporting a bilingual advantage showing bilingualism has benefits for cognitive control and selective attention ([Bialystok & Craik 2010](#)) means that immigrant parents should aim to preserve the bilingualism and heritage language in their children. The results of this study also provide evidence for an executive control advantage that is scaled on how early the onset of bilingualism is (the earlier, the better), which bodes well for heritage speakers, who acquire their L2 early, at around preschool age. However, it is equally important to maintain L1 development, which is at risk due to decreased input, practice, and education, after this early L2 onset. On the whole, this may encourage all parents to have their children learn a second language starting at a young age.

In addition to the importance of heritage language input at home, education also has strong impacts on language proficiency. The heritage group reported lower reading and writing abilities in Chinese compared to their Chinese speaking and listening abilities because it requires explicit instruction, and heritage speakers acquire the language from family settings without support in school. On the other hand, the self-rated English reading and writing abilities of the native group are higher than the Chinese reading and writing abilities of the heritage group, while the respective speaking and listening abilities were rated similarly. Though the native group started acquiring English later than the heritage group did with Chinese, they were educated at high levels in English and later lived in an English-speaking country. This demonstrates the potential of input from schooling along with immersion in fostering language development and literacy, especially for the case of heritage languages, where the speaker already has a native-like L1 acquisition onset and grammatical background knowledge. In addition, [Pires & Rothman \(2009\)](#) looked at the use of inflected infinitives by heritage speakers, L2 learners, and native speakers of Brazilian Portuguese, all at highly proficient levels. The grammatical properties of inflected infinitives are no longer present in colloquial dialect, which heritage speakers are solely exposed to at home, and thus, the lack of exposure to the standard dialect and formal education led them to be outperformed by L2 learners

and native speakers on the distribution of inflected infinitives. Education provides input needed to acquire such structures and consolidate others, like grammatical gender or aspectual properties, in addition to promoting literacy.

Pure auditory linguistic input that tapers greatly at preschool age, like what heritage speakers experience, is usually insufficient for learners to sustain their heritage language and develop more advanced proficiency, and therefore nurture their bilingual advantage. In addition, the extent of L1 attrition is inversely related to the age of onset of bilingualism and is affected by frequency of input and use (Bylund 2009), so heritage languages are especially vulnerable to attrition after the onset of the L2. It is urgent for schools, communities, and families to be educated on the subject so that they can work together to encourage heritage language literacy and continued development, especially before attrition occurs; then, immigrant children will be able to maintain a stronger connection to their heritage as well as receive the possible cognitive benefits of bilingualism, especially with their early onset ages of bilingualism.

6 CONCLUSION

Our findings suggest that heritage speakers seem to benefit from an advantage related to cognitive control compared to other advanced bilinguals, due to early L2 onset. The linear regressions conducted demonstrate that an early age of L2 (English) acquisition and reading was connected with lower L1 (Chinese) interference, so the heritage speakers, who are simultaneous or early sequential bilinguals, had an advantage over the native speakers, who are sequential but advanced Chinese-English bilinguals. Also, the BIA+ temporal delay model and the Bilingual L2 Lexical Disadvantage Hypothesis are supported by the linear regression models as well as the lower interference demonstrated when the Stroop stimulus is in the participants' less dominant language – especially in the novel colour response conditions – due to lower input, frequency of use, and proficiency. The novel colour response condition also seems to highlight a separation between visual and linguistic colour processing, leading to less decision-making conflict, and therefore interference in these conditions.

The ordering of the four within- and between-language Stroop conditions, which when excluding the novel colour response conditions were identical for both the native and heritage group, is not completely compatible with any existing research or model regarding bilingual Stroop. Future research on differences between alphabetic and logographic script processing in monolinguals and bilinguals may shed light on these findings. Importantly, both the new online Stroop methodology of collecting responses through the computer arrow keys and the novel colour-response condition proved successful in reliably demonstrating the Stroop interference effect. These implementations have huge potential for collecting Stroop data faster, as well as provide new avenues for Stroop research related to bilingualism, cognitive control, or the tie between visual and linguistic processing and identification.

The questionnaire results re-affirms many findings and beliefs about heritage speakers regarding their language abilities and acquisition experiences, but also pro-

vides interesting information for future research, such as situational code switching relating to adaptive control, mixed languages like Chinglish, or the impact of parent, sibling, and friend language abilities on speaker L1 and L2 proficiency (Appendix).

This study provides further insight into the heritage speaker linguistic experience, a recent but expanding field of study. Crucially, it can be applied to how children can be nurtured linguistically, especially heritage speakers. Due to the potential of the bilingual advantage, this dissertation argues for both providing early input and education in an L2 along with nurturing the L1 through continued input from home and school.

REFERENCES

- Benmamoun, E., S. Montrul & M. Polinsky. 2013. Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39(3-4). doi:10.1515/tl-2013-0009.
- Bialystok, E. 2007. Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *International Journal of Bilingual Education and Bilingualism* 10(3). 210–223. doi:10.2167/beb441.0.
- Bialystok, E., F. Craik & G. Luk. 2008. Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34(4). 859–873. doi:10.1037/0278-7393.34.4.859.
- Bialystok, E. & F.I. Craik. 2010. Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science* 19(1). 19–23. doi:10.1177/0963721409358571.
- Biederman, I. & Y.-C. Tsao. 1979. On processing Chinese ideographs and English words: Some implications from Stroop-test results. *Cognitive Psychology* 11(2). 125–132. doi:10.1016/0010-0285(79)90007-0.
- Blumenfeld, H.K. & V. Marian. 2014. Cognitive control in bilinguals: Advantages in stimulus–stimulus inhibition. *Bilingualism: Language and Cognition* 17(3). 610–629. doi:10.1017/s1366728913000564.
- Bylund, E. 2009. Maturation constraints and first language attrition. *Language Learning* 59(3). 687–715. doi:10.1111/j14679922200900521.
- Chen, H.C. & C. Ho. 1986. Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12(3). 397–401. doi:10.1037/0278-7393.12.3.397.
- Chen, J.Y. 2000. The orthographic variation hypothesis revisited: Evidence from a meta-analysis of the intra-lingual stroop interference in Chinese and English. *Acta Psychologica Sinica* 32. 96–101.
- Coderre, E.L., W.J. van Heuven & K. Conklin. 2013. The timing and magnitude of Stroop interference and facilitation in monolinguals and bilinguals. *Bilingualism: Language and Cognition* 16(2). 420–441. doi:10.1017/s1366728912000405.
- Cohn, A.J., T. Bowden, M. McKinnon, R. Ravindranath, B. Simanjuntak & Y. Taylor. 2013. Multilingual language use questionnaire. https://conf.ling.cornell.edu/pdfs/LangUseQuesF_copy.pdf.

- Coltheart, M., K. Rastle, C. Perry, R. Langdon & J. Ziegler. 2001. Drc: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* 108(1). 204–256. doi:[10.1037/0033-295x.108.1.204](https://doi.org/10.1037/0033-295x.108.1.204).
- Dijkstra, T. & W.J.B. van Heuven. 2002. The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition* 5(3). 175–197. doi:[10.1017/s1366728902003012](https://doi.org/10.1017/s1366728902003012).
- Fang, S.P., O.J. Tzeng & L. Alva. 1981. Intralanguage vs. interlanguage Stroop effects in two types of writing systems. *Memory & Cognition* 9(6). 609–617. doi:[10.3758/bf03202355](https://doi.org/10.3758/bf03202355).
- van Heuven, W.J., K. Conklin, E.L. Coderre, T. Guo & T. Dijkstra. 2011. The influence of cross-language similarity on within- and between-language Stroop effects in trilinguals. *Frontiers in Psychology* 2. doi:[10.3389/fpsyg.2011.00374](https://doi.org/10.3389/fpsyg.2011.00374).
- Jensen, A.R. & W.D. Rohwer. 1966. The Stroop color-word test: A review. *Acta Psychologica* 25. 36–93. doi:[10.1016/0001-6918\(66\)90004-7](https://doi.org/10.1016/0001-6918(66)90004-7).
- Kousaie, S. & N.A. Phillips. 2012. Ageing and bilingualism: Absence of a ‘bilingual advantage’ in Stroop interference in a nonimmigrant sample. *Quarterly Journal of Experimental Psychology* 65(2). 356–369. doi:[10.1080/17470218.2011.604788](https://doi.org/10.1080/17470218.2011.604788).
- Lee, T.M. & C.C. Chan. 2000. Stroop interference in Chinese and English. *Journal of Clinical and Experimental Neuropsychology* 22(4). 465–471. doi:[10.1076/1380-3395\(200008\)22:4;1-0;ft465](https://doi.org/10.1076/1380-3395(200008)22:4;1-0;ft465).
- Li, P., F. Zhang, A. Yu & X. Zhao. 2019. Language history questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition* 23(5). 938–944. doi:[10.1017/s1366728918001153](https://doi.org/10.1017/s1366728918001153).
- Marian, V., H.K. Blumenfeld & M. Kaushanskaya. 2007. The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research* 50(4). 940–967. doi:[10.1044/1092-4388\(2007/067\)](https://doi.org/10.1044/1092-4388(2007/067)).
- Marian, V., H.K. Blumenfeld, E. Mizrahi, U. Kania & A.-K. Cordes. 2013. Multilingual Stroop performance: Effects of trilingualism and proficiency on inhibitory control. *International Journal of Multilingualism* 10(1). 82–104. doi:[10.1080/14790718.2012.708037](https://doi.org/10.1080/14790718.2012.708037).
- Mingjin, H., S. Hasko, G. Schulte-Körne & J. Bruder. 2012. Automatic integration of auditory and visual information is not simultaneous in Chinese. *Neuroscience Letters* 527(1). 22–27. doi:[10.1016/j.neulet.2012.08.031](https://doi.org/10.1016/j.neulet.2012.08.031).
- Montrul, S. 2005. Second language acquisition and first language loss in adult early bilinguals: Exploring some differences and similarities. *Second Language Research* 21(3). 199–249. doi:[10.1191/0267658305sr247oa](https://doi.org/10.1191/0267658305sr247oa).
- Montrul, S. 2015. *The acquisition of heritage languages*. Cambridge University Press.
- Paap, K.R., R. Anders-Jefferson, L. Mason, K. Alvarado & B. Zimiga. 2018. Bilingual advantages in inhibition or selective attention: More challenges. *Frontiers in Psychology* 9. doi:[10.3389/fpsyg.2018.01409](https://doi.org/10.3389/fpsyg.2018.01409).
- Pires, A. & J. Rothman. 2009. Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. *International Journal of Bilingualism* 13(2). 211–238. doi:[10.1177/1367006909339806](https://doi.org/10.1177/1367006909339806).

- Posner, M.I. & C.R.R. Snyder. 1975. Attention and cognitive control. In *Information Processing and Cognition: The Loyola Symposium*, 55–85. Erlbaum, Lawrence, Associates.
- Pyers, J.E., T.H. Gollan & K. Emmorey. 2009. Bimodal bilinguals reveal the source of tip-of-the-tongue states. *Cognition* 112(2). 323–329. doi:[10.1016/j.cognition.2009.04.007](https://doi.org/10.1016/j.cognition.2009.04.007).
- Schmid, M.S. & B. Köpke. 2017. The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism* 7(6). 637–667. doi:[10.1075/lab.17058.sch](https://doi.org/10.1075/lab.17058.sch).
- Seliger, H.W. 1996. Primary language attrition in the context of bilingualism. In Tej K. Ritchie, William C. & Bhatia (ed.), *Handbook of Second Language Acquisition*, 605–626. Academic Press. doi:[10.1016/b978-012589042-7/50020-7](https://doi.org/10.1016/b978-012589042-7/50020-7).
- Ware, A.T., M. Kirkovski & J.A. Lum. 2020. Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Frontiers in Psychology* 11. doi:[10.3389/fpsyg.2020.01458](https://doi.org/10.3389/fpsyg.2020.01458).

Comparing Chinese-English Bilingual and Heritage Speaker Stroop Effects

APPENDIX

The Appendix consists of other highlights from the questionnaire part of our experiment, which provides further language experience information on the participants and may be interesting starting points for future research.

Appendix 1

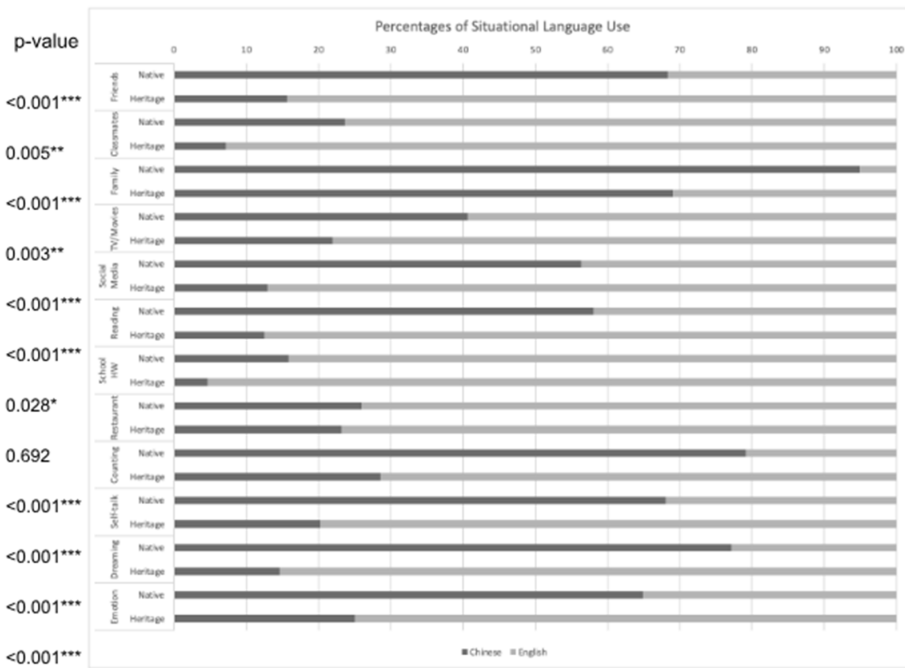


Figure A1 Percentage of Language Use per Situation by the Native and Heritage Group: Chinese vs. English.

Note: All differences between groups in situational language use were statistically significant, except for ordering at a restaurant.

Appendix 2

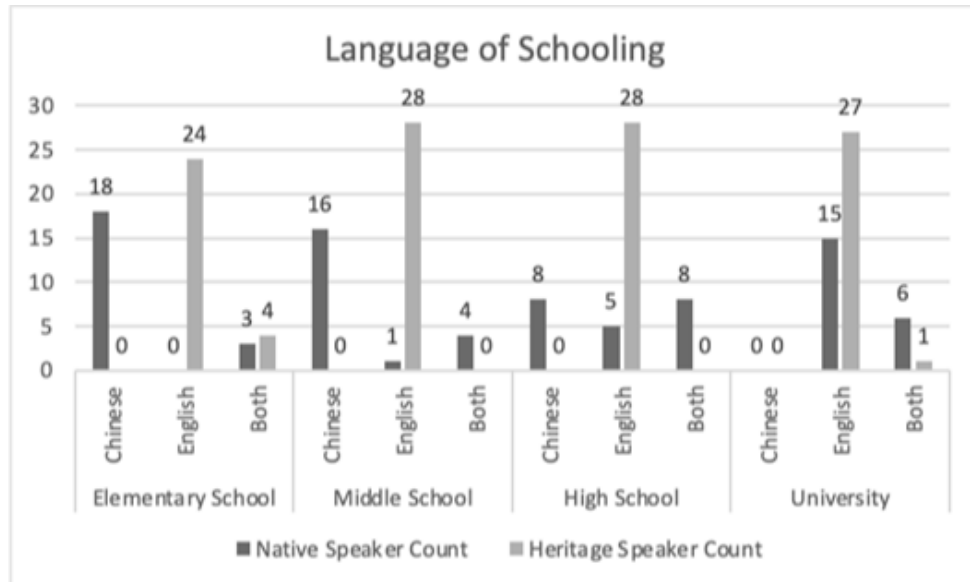


Figure A2 Languages of Schooling per Participant Group by Count.

Note: ‘Both’ includes bilingual instruction as well as total English instruction and total Chinese instruction within the same period of schooling due to school changes and immigration.

Appendix 3

	Chinese	English
Parents, Native	5	2.167
Parents, Heritage	4.518	3.554
Siblings, Native	4.833	3.167
Siblings, Heritage	2.958	4.75

Table A1 Reported Family Language Abilities from 1-5.

Note: All differences between the native and heritage group were statistically significant ($p < 0.05^*$ for parental Chinese abilities, $p < 0.001^{***}$ for parental English abilities, and $p < 0.001^{***}$ for sibling English and Chinese abilities).

Comparing Chinese-English Bilingual and Heritage Speaker Stroop Effects

Appendix 4

	Native Group	Heritage Group
% of Friends that can speak Chinese but not English	44.286	5.714
% of Friends that can speak English but not Chinese	27.620	74.286
% of Friends that can speak both Chinese and English	60.952	26.429
Frequency of Chinglish Used with Family	1.667	3.607
Frequency of Chinglish Used with Friends	3.905	2.143
Frequency of Chinglish Used with Others in Daily Life	3.048	1.893

Table A2 Participants' Friend Language Abilities and Frequency of Chinglish (Mixed Language) Use.

Note: Frequency was quantised from a scale of 1 (Never) to 5 (Always). All values were found to be statistically significant with $p < 0.001^{***}$.

Grace Shan
University of Cambridge
qgs20@cam.ac.uk