Transparency of first language calendar terms and its impact on calendrical calculations in the first and the second language

Ziyi Zhuang

A thesis submitted in fulfillment of the requirements for the degree of Research Master of Arts in Applied Linguistics, the University of Auckland, 2022

Abstract

Calendrical calculations, defined here as calculating the target weekday/month after a stimuli weekday/month, are an important component of problem-solving and temporal cognition and can vary cross-linguistically. English speakers, whose calendar naming system is irregular and opaque, rely on verbal-list processing. Unlike English speakers, Chinese speakers, with knowledge of numerically and linguistically transparent calendar terms, prefer a more efficient numerical operation system. Previous studies have focused on the effects of differences in calendar lexicon on speakers' development of simple calendar calculations. However, scant research extends the potential effects to cognition of bilinguals, who know L1 transparent and L2 opaque calendar terms. To fill this research gap, this study aims to provide an experimental basis for the effects of different levels of transparency of the calendar naming system on calendrical calculations when tested in L1 and L2. A total of 30 Chinese-English bilinguals and English monolinguals were recruited from universities in English-speaking countries. This study used the Calendrical Calculation Task (week, month, hour, and year) and the Self-reported Strategies Task to test participants' cognition process in calendrical reasoning. Variables, including Distance (short/long), Direction (forward/backward), Boundary (within/across), and Input (linguistic/numerical), were manipulated. The English speakers were invited to do the two tasks in L1 English, labeled as the English group. The Chinese-English bilinguals were required to do the two tasks in L1 Chinese and L2 English, labeled as the Chinese group and the Bilingual group, respectively. Linear mixed-effects models in R were used to analyze reaction times and response accuracy, while the self-reported strategies were analyzed descriptively. The results showed that the English group relied on verbal list processing, while the Chinese and English groups relied on numerical reasoning. Also, the Chinese group was the fastest, while the Bilingual group was as fast as the English group in week and month calculation questions. Negative effects of long distance and backward directionality were found in all groups. Positive effects of crossing boundary and negative effects of numerical input were only found in the English group. Results have implications for research on temporal reasoning, linguistic relativity, and bilingual cognition.

Acknowledgments

Above all, I would like to express my greatest thanks to my supervisor Dr. Norbert Vanek, who supervised me for an entire year in the global situation of Covid-19. Thanks to his patient instructions and invaluable suggestions on my ethics proposal, experimental design, data collection, and results analysis, I can complete the thesis.

I would particularly like to show my gratitude to my participants, without whom the present study would fail to be completed. Many thanks for their interest, patience, and enthusiasm.

Also, I would like to thank all other great teachers in Cultures, Languages, and Linguistics who have taught me and helped me develop my academic knowledge and English writing skills.

Additional thanks are due to my family, especially my mother, who always loves me and supports me in pursuing knowledge in my academic career.

At last, a special appreciation of mine goes to my friends, for their kind support, companionship, and encouragement through this long journey.

Table of Contents

1. Introduction

Differences in languages can impact the ways in which speakers make sense of the world (Whorf, 1956) because the same concept can be interpreted and represented differently across various language communities. If concepts are linguistically modulated, what cognitive implications are there for bilinguals? One aspect of bilinguals' cognition that deserves further exploration is temporal reasoning and daily problem-solving such as calendrical calculations, an area that has an ecological validity advantage over strictly laboratory-based tasks. As an essential component of temporal reasoning, calendrical calculations, occur routinely for most people. These calculations involve identifying at what time of which day in which month a specific task should be completed. Remarkably, the conventional sets of symbols used to represent days of the week and months of the year are different in the level of transparency across various languages. For instance, English establishes an opaque calendar naming system, in which it is hard to guess the meaning of the terms for people who have no experience of English. On the contrary, Chinese use a numerical system to represent weekdays and months. Following the idea known as the linguistic relativity hypothesis, the cognitive consequence of the different levels of transparency in English and Chinese calendar terms may result in different performances in calendrical calculation tasks by English and Chinese speakers.

Previous studies (e.g., Friedman, 1983, 1984; Huang, 1993; Kelly, Miller, Fang, and Feng, 1999) find that the knowledge of numerical and linguistically transparent calendar terms endows the Chinese speakers with a preference for numerical operation strategy, while the English monolinguals prefer overtly or covertly verbal listing when solving calendrical calculations. This efficiently means that with different transparency levels of calendar terms and various problem-solving methods, it is highly likely that speakers from different language communities tend to perform differently in calendrical calculation tasks. The research described here compares the performance of speakers of English and Chinese in calendrical calculation tasks and their respective preferred calculation-based problem-solving methods to test whether different levels of transparency of calendar terms across the speakers' dominant languages affect the ability to solve the calendrical calculation questions. Prior research report that speakers of language with a transparent calendar naming system outperformed speakers of language with an opaque calendar lexicon when tested in their corresponding L1 (Yang and Zhang, 2011). Then, whether calendrical calculation task is easier to solve for Chinese speakers than English speakers, since Chinese native speakers have knowledge of a transparent calendar naming system. If so, whether such advantage remains when the Chinese-English bilinguals are required to solve calendrical calculation questions with second language (L2) English as testing language. In order to investigate how bilinguals who simultaneously know an L1 with transparent and an L2 with opaque calendar terms perform in reasoning calendar calculations when tested in a second language, the present study also involves comparisons between English group and Chinese-English bilingual group and between Chinese group and Chinese-English bilingual group. The findings of the study aim to contribute to theoretical and experimental research on temporal reasoning, linguistic relativity, and bilingual cognition.

2. Literature Review

2.1 Theoretical Background: Cognitive Effects of Linguistic Transparency of Calendar Terms

2.1.1 Linguistic Transparency of Calendar Terms

Linguistic transparency is a broad concept with different aspects, which has been widely used in the domain of semantic transparency and numerical transparency. A better understanding of semantic transparency and numerical transparency is also a prerequisite for explaining the concept of linguistic transparency of calendar terms, as the constituents of calendar terms are a combination of numerical systems and semantic systems, especially in English. Taking the week-day names as examples, the seven days within one week is represented by linguistic words Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday, but means the 1^{st}, 2^{nd}, 3^{rd}, 4^{th}, 5^{th}, 6^{th}, and 7^{th} day in the week. Similar phenomenon can also be found in the 12 month names. Overall, the calendar terms, which imply a numerical order of dates (from the 1^{st} day to 7^{th} day for weekdays and 1^{st} month to 12^{th} month for months), are normally represented by semantical words in daily life. Semantic transparency, numerical transparency, and linguistic transparency of calendar terms are discussed in the following sections.

2.1.2.1 Semantic Transparency

Semantic transparency is a phenomenon that occurs in compound words, a type of complex word that makes productivity across languages achievable (Momonian, Cham, Amini, Radman, and Weekes, 2021). As a productive morphological word formation process, compounding combines two or more than two words together and makes the new word function as one word, semantically and grammatically (Sherko, 2015). The semantic transparency and semantic opacity refer to a boundary condition of whether the meaning of a multimorphemic compound word can be derived from the meanings of its constituent morphemes or not (Libben, Gibson, Yoon, and Sandra, 2003). To be specific, if the meaning of a compound word could be easily guessed by people who had never learned the word before in terms of the constituents' meanings, then the word is a transparent word. For example, the word "blueberry" is a transparent compound word because the morpheme "blue" and "berry" are both transparent members. In contrast, the compound "deadline" is semantically opaque because it is hard for a new learner to comprehend the entire strings' meaning due to the opacity of the morpheme "dead" and "line". Moreover, it is important to mention that the boundary condition does not make semantic transparency and semantic opacity a dichotomous notion, but gradually varies along a continuum from semantically transparent to semantically opaque (Chen, Koda, and Wiener, 2020). According to Libben, Gibson, Toon, and Sandra (2003), there are four fundamental degrees of morphosemantic transparency to represent all possible combinations of the relationship between the meaning of one bimorphemic compound and the meaning of the compound's constituent morphemes, including transparent-transparent (TT) (e.g., blueberry), opaque-transparent (OT) (e.g., eyewitness), transparent-opaque (TO) (e.g., jailbird), and opaque-opaque (OO) (e.g., deadline) with the schematic representation of each type (Sherko, 2015) as shown in Figure 1:

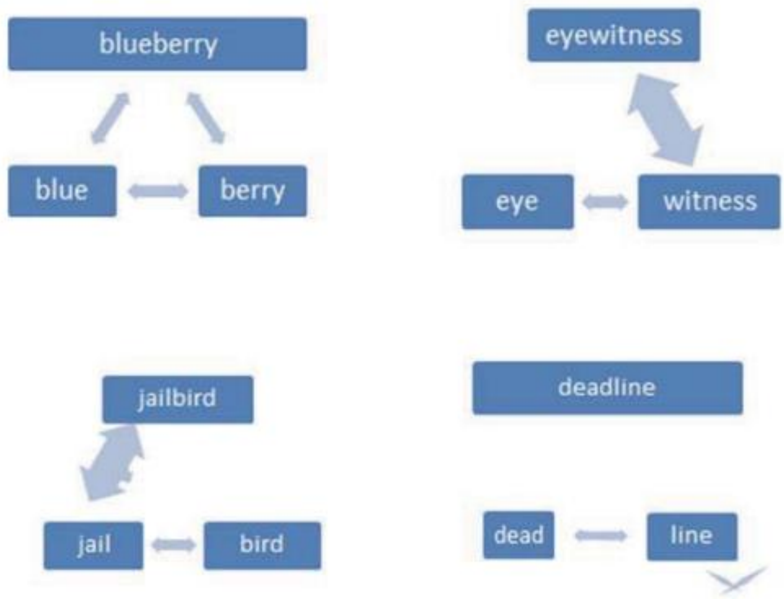

Figure 1: The schematic representation of TT, OT, TO, and OO compound words.

2.1.2.2 Numerical Transparency

Unlike semantic transparency, which pays attention to the transparency of compound words within one language, numerical transparency focuses on the different levels of numerical transparency across languages. Numerical transparency refers to a regular counting system with clear and consistent rules for combining primary numbers, which is common in many Asian number systems (Fuson and Kwon, 1991; Ng and Rao, 2010; Laski and Yu, 2013). For example, there is great regularity and transparency in Chinese number words from 11 to 20 (Miura, Okamoto, Kim, Steere, and Fayol, 1993) and from 10 to 99 (Ho and Fuson, 1998), in both written and spoken forms, which is also termed Chinese Number Advantage (Mark and Dowker, 2015). The Chinese numerical words between 11 and 19 are formed by compounding “ten” and “the unit word/ cardinality” (i.e., one to nine), numbers from 20 to 90 by compounding “the ten-digit code”, “ten” and “the unit word” (Miller, Smith, Zhu, and Zhang, 1995). Based on the organization rules, 11 and 12 are represented by “十一” (*shi yi*) and “十二” (*shi er*), literally “ten-one” and “ten-two” respectively, while 20 is spoken as “二十” (*er shi*), literally “two-ten”, 30 as “三十” (*san shi*), literally “three-ten”, and 45 as “四十五” (*si shi wu*), literally “four-ten-five”. With such a clear and transparent 10-based counting system, it tends to be easier for a new learner of Chinese to guess and speculate about the meaning and spelling of an unknown Chinese number term.

In contrast to the transparent counting systems, many western languages apply an opaque and irregular rule of numerical word naming, which cannot be mapped onto Arabic number systems directly (Dowker and Roberts, 2015). Taking the English number system as an example, English speakers have to memorize relatively arbitrary names like eleven and twelve, where there are no clear clues for their base system or the unit word. Although some may argue that “eleven” and “twelve” originate from the Old Saxon words “ellevan” and “twelif”, meaning “one-left” and “two-left” respectively, resulting from a regular subtraction of 10, it is extremely hard, if not impossible, for present English speakers and learners to find the unapparent information based on the spelling of "eleven" and "twelve". Moreover, various phonemic and morphological modifications of English numerical names further complicate the acquisition of the counting system (Mark and Dowker, 2015). To be specific, in teen numbers, ten is replaced by "-teen", three by "thir-", and five by "fif-", while ten becomes "-ty" for multiples of ten from 20 to 90. Thus, the number words from one to twelve and the transformations of teens and tens result in rote learning and English learners tend to have more difficulties than speakers of Chinese in acquiring the counting system (Rasmussen, Ho, Nicoladis, Leung, and Jeffrey, 2006; Dowker and Roberts, 2015) and mathematical development (Miller, Smith, Zhu, and Zhang, 1995; Dowker, Bala, and Lloyd, 2008; Siegler and Mu, 2008).

The greater numerical transparency is also believed to be beneficial to people's grasping of place value (Miura and Okamoto, 2003). Place value can be defined as an ability to figure out different digits' values based on their corresponding places in the multi-digit number (Mark and Dowker, 2015), so each "8" in "8888" represents different values and should be understood as "eight thousand", "eight hundred", "eight tens", and "eight units", respectively. While the Chinese numerical naming system shows a one-to-one correspondence with the Arabic number systems (e.g., 16 is "十六" as "ten-six" and 60 is "六十" as "six-ten"), English numerical units' values tend to be obscured by the three variants of ten (ten, -teen, and -ty) and the inconsistency of spelling and pronunciation orders with Arabic numbers (e.g., applying sixteen rather than teen-six / ten-six to indicate 16). As a result, the place values are masked by number names' irregularities and opacities, which may hinder the development of English speakers' and learners' mathematical calculation processing.

2.1.2.3 Linguistic transparency of Calendar Terms

The English and Chinese calendar terms include the week system and the month system. Although both Mandarin and English use the 7-day and 12-month solar calendar system, Mandarin and English apply different ways to form specific calendar terms. Mandarin applies transparent calendar terms, while English applies opaque weekday and month names. Start with the seven weekday names as shown in Figure 2 (Kelly, Miller, Fang, and Feng, 1999). English weekdays were originally named after seven planets in Hellenistic astrology, in the order of Sun, Moon, Mars, Mercury, Jupiter, Venus, and Saturn, which were also the Roman names of their gods. While names as Sun (Sunday), Moon (Monday), and Saturn (Saturday) remained, four Roman gods' names were replaced by Nordic gods with similarities, including Tyr for Mars (Tuesday), Odin for Mercury (Wednesday), Thor for Jupiter (Thursday), and Frigg for Venus (Friday) (Boorstin, 1985; Zerubavel, 1985). Though such historical relations between English weekday names and the ancient astronomy may be recognized by those well-versed in the forming history of names for days of the week, the relation is obscured, and the memorization of the gods'/planets' names is not an easy task. On the contrary, Chinese names for the seven days obey a transparent numerical combination rule, following a "星期 (*xing qi*, meaning week) + cardinal number corresponding to a particular day in week" format. For instance, the Mandarin term for Wednesday is *xing qi san*, literally week three. The one exception is the lexicalization of Sunday, which is termed as "*xingqi* + *ri/tian* (literally sun and sky, respectively)", instead of "*xing qi* + seven". It should be noted that *xing qi* can also be replaced by *zhou* and *li bai* (meaning week), but the formation structure, except for the non-numerical term Sunday, remains the same.

Weekday Names

Language	Structure	Items						
English	Planetary	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
		Moon	Mars (Tiw)	Mercury (Woden)	Jupiter (Thor)	Venus (Fria)	Saturn	Sun
Chinese	Numerical	星期一	星期二	星期三	星期四	星期五	星期六	星期日
		1 一	2 二	3 三	4 四	5 五	6 六	Sun 日

Figure 2: Names of the seven weekdays in English and Chinese.

Names for the 12 months of the year in English are also opaque, which results from a mélange of gods' names, Caesars' names, and Latin numerical words (Boorstin, 1985;

Grove, 1986), as shown in Figure 3 (Kelly et. al., 1999). Such a derivation of Gods' and Caesars' names makes English months from January to August opaque, similarly to English weekday names. Moreover, month names from September to December are derived from a Latin numerical system rather than English numbers, and the month names from September to December corresponds to Latin number 7 to 10 instead of 9 to 12. To be specific, Latin number 7 (septum), 8 (octo), 9 (novem), and 10 (decem) responds to September to December, respectively, because Roman used to apply a 10-month calendar (Kelly et al., 1999) where 7-10 are the last four months. Therefore, although there is a derivational structure of English 12 months names, the information is not apparent to many English learners and speakers. Chinese month names follow a regular format as those in weekday names, which is "the cardinal number of particular month + 月 (*yue*, meaning month)" as in Figure 2. For example, the Chinese word for February and August are 二月 (*er yue,* i.e., 2 + month) and 八月 (*qi yue*, i.e., 8 + month), respectively. Thus, it is reasonable to say that the calendar systems of English are opaque, while Chinese weekday and month names are transparent. To sum up, both Chinese weekday names and month names are TT (transparent-transparent) compounds, while English weekday names are OT (opaque-transparent) compounds and English month names are OO (opaque-opaque) compounds.

Month Names

Lan-guage	Structure	Items											
English	Gods, Caesars, Numbers	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
		Janus	Februa	Mars	Apru	Maia	Juno	Julius Casear	Augustus Casear	7	8	9	10
Chinese	Numerical	一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
		1 一	2 二	3 三	4 四	5 五	6 六	7 七	8 八	9 九	10 十	11 十一	12 十二

Figure 3: Names of the 12 months in English and Chinese.

2.1.2 Linguistic Relativity

Naturally, when speakers sharing typologically similar languages form various linguistic communities, whether patterns of thought vary among speakers from different linguistic communities draws researchers' attention and has been a long-standing intense debate in the field of philosophy, anthropology, linguistics, and psychology. The notion that speakers' perception and conception patterns of the external world are modulated and biased by the characteristics of various languages is referred to commonly as the linguistic relativity hypothesis. Given that the essential role of language in constructing the world was initially put forth in the work of Benjamin Whorf and his teacher Edward Sapir, the term "linguistic relativity" is also referred to interchangeably with the "Sapir-Whorf hypothesis" and "Whorfian hypothesis" in the literature. However, the application of the "Sapir-Whorf hypothesis" and "Whorfian hypothesis" appears to be falling out of favor in contemporary scholars because of the refinement of the relativity hypothesis in these years (Everett, 2013) based on the theoretical and methodological evolution in relevant research and new findings. Perhaps the most significant modification is that the original version underlining language's deterministic role on thought (Whorf, 1956) has been replaced by an influential role in present linguistic relativity (Sarantakis, 2014).

The deterministic role played by language on human minds is known as linguistic determinism, which argues that one's way of thinking is completely constrained (Wittgenstein, 1922; Sapir, 1949) and governed by the languages used in different language communities. As the strongest incarnation of the Sapir-Whorf hypothesis, linguistic determinism proposes that language determines the way reality is perceived, categorized, and acted on (Whorf, 1956). The strongest opposite view is that if language determines thought, then different languages give rise to different patterns of thought, which makes it impossible for speakers of different languages to fully understand each other. However, speakers from various language communities can communicate with each other through translation, indicating that crosslinguistic speakers' thoughts are not completely different and are determined by disparate linguistic characteristics in the world's languages. On the contrary, some researchers (e.g., Smith, Brown, Toman, and Goodman, 1947; Fodor, 1975; Jackendoff, 1983; Pinker, 1994) hold a 'universal' view that thought can operate independently of language and the conceptual models of human are invariant and universal across various languages and cultures.

Although there is no agreement on where the true or correct point lies between the two opposing views, only a few, if any, contemporary scholars would claim that language has no effect or plays a deterministic role on cognition (Athanasopoulos, 2009). The broad consensus is that both extremes are too absolute to be correct (Gleitman and Papafragou, 2013). Most scholars' statements are in the middle standpoints, where every linguistic community differs in their way of conceptualizing the world from every other to some extent in accordance with their native languages. The commonly received wisdom is the linguistic relativity hypothesis, which holds that human's conception, interpretation, and understanding of the world are influenced by various languages that differently carve up the world (Athanasopoulos, Bylund, and Casasanto, 2016). As a heated research topic in the disciplines of Anthropology, Psychology, and Linguistics (Lucy, 1997), the linguistic relativity hypothesis has been continually tested in the last couple of decades (Bylund and Dick, 2019). The effects of linguistic characteristics on speakers' perception and behavior are found flexible and dynamic (Athanasopoulos et al., 2016). To be specific, the subtle or not so subtle experimental conditions (e.g., verbal interventions and visual hemifield manipulations) tend to up- or down- regulate the linguistic effects on participants' thought. Consequently, several scholars have recently pointed out that the traditional all or nothing binary way (either language shapes human thoughts or it does not), in which researchers are used to answering the Whorfian question is counterproductive (Casasanto, 2008; Regier and Kay, 2009) and no longer tenable (Athanasopoulos and Casaonsa, 2020). Instead, recent theories have been looking to establish under which conditions and to what extent the linguistic or language effects show on cognitive processes, as well as proposing potential accounts of how language may influence human minds (Wolff and Holmes, 2011).

Nevertheless, while one may find that there have been numbers of areas studies, such as motions, color construal, space perception, and time (see Sarantakis, 2014; Everett, 2016, for a recent collection of theoretical and empirical papers), the very specific cognitive process of calendrical calculation is heavily underexplored and maybe super relevant because calendrical calculation cognition is a combination of speakers' numerical, temporal, and problem-solving perception and processing. According to Whorf (1994, p.210), calendars, accounting, and mathematics are "impresses of linguistic habit" and "receipts from culture and language". As a lens for investigating Whorfian relativity hypothesis, mathematical calculating can be regarded as a fundamental reasoning- and temporal-related cognitive product of specific language's notational features (Chrisomalis, 2021). Thus, researching the potential effects of different transparent degrees of numerical naming system (as described in section 2.1.2.2) and calendar naming structure (as mentioned in section 2.1.2.3) tends to be an interesting study.

2.2 Experimental Background: Cognitive Effect of Linguistic Transparency of Calendar Terms

Linguistic relativity hypotheses that speakers' perceptions are modulated or biased by languages (Athanasopoulos, 2006; Athanasopoulos, P., Bylund, E., Montero-Melis, G., Schartner, A., Kibbe, A., 2015; Lupyan, 2012; Samuel, Cole, and Eacott, 2019; Athanasopoulos and Bylund, 2021). A great deal of empirical research has investigated whether language acts as an attention and judgements directing mechanism. Evidence supporting this notion comes from various aspects. For example, in color matching or discrimination tasks, languages with different labels for the stimuli positively affect participants' performance, while languages subsuming similar colors under one term hinder participants' performance (Roberson, Pak and Hanley, 2008; Winawer, Witthoft, Frank, Wu, Wade, and Boroditsky, 2007). Linguistic differences across languages are also demonstrated to affect participants' perceptions of objects and substances, including objects themselves (Imai and Gentner, 1997) and object relations (Park and

Ziegler, 2014), as well as more abstract concepts such as time (Casasanto et al., 2004; Bylund and Athanasopoulos, 2017) and motion (Athanasopoulos and Bylund, 2013, 2015).

Recent research pays more attention to the effect of linguistic relativity on bilinguals' thought, that is, whether learning new concepts and categories in a second language can restructure the existing ones or whether linguistic characteristics in one's native language can result in the acquisition of new characteristics instantiated by the learned languages. As traditional bilingualism distinctions involve subordinate, coordinate, and compound bilinguals (Weinrich, 1953), the perceived outcomes of bilinguals could also be divided into three types, subordinate (native concepts used both in speaking L1 and L2), coordinate (two concepts, L1 concepts and L2 concepts used in speaking corresponding languages), and compound (an integrated concept of L1 and L2 concepts or a novel concepts which are built based on L1 and L2 concepts but more than them) (Bassetti et al., 2018). To be specific, with respect to numerical processing, there are three types of potential outcomes in the Chinese learners of English. For the subordinate type, the Chinese-English bilinguals may stick to the numerical reasoning strategy (Fuson and Kwon, 1992; Ma, 1999; Ho and Fuson, 1998; Lan et al., 2009; Chan and Ho, 2010) even tested in English, while the coordinate type tends to choose the numerical calculating process when tested in Chinese but counting (Cheng and Chan, 2005) more when tested in English. The compound type may combine the numerical process of Chinese and English and create a novel way to solve reasoning questions, such as counting for short trials but calculating for long trials.

However, studies on how knowledge of two languages affects bilinguals' concepts and thoughts mainly focus on participants' categorizations (Ervin, 1961; Caskey-Sirmons and Hickerson, 1977; Jameson and Alvarado, 2003; Athanasopoulos and Boutoneet, 2016; Kurinski and Sera, 2011; Vernich, 2017), attention (Vigil, Tyler, and Ross, 2006; Soveri, Laine, Hamalainen, and Hugdahl, 2011), memory (Boroditsky and Schmidt, 2000), and temporal sequencing (Tang, Vanek, and Roberts, 2021). Only limited researchers have investigated the effects of linguistic transparency on other cognition aspects, one of which is the calendrical calculation process, an everyday reasoning and problem-solving task. Nevertheless, some studies have been carried out on bilinguals' calendrical calculation processing. Two lines of previous research that may be relevant looked at the effects of bilingualism on temporal cognition and arithmetic calculations. In the next this section, research on temporal cognition, numerical cognition, and calendrical representation and processing will be discussed.

2.2.1 Research on Temporal Cognition in Monolinguals and Bilinguals

In the field of temporal cognition, researchers mostly focused on the effects of different reading-writing directionality and different time metaphors between languages on bilinguals' mental perception of the directionality of time. Firstly, the directionality of writing can be divided into two types: left to right reading-writing mode (e.g., English and French) and right to left reading-writing mode (e.g., Arabic and Hebrew). Investigation by Tversky, Kugelmass, and Winter (1991) shows that speakers' perception of time directionality corresponds to the language's reading-writing direction; that is, speakers of languages with left-to-right reading-writing mode are more likely to conceive of the time flow as from left to right, and vice versa. Moreover, Kugelmass and Winter (1991) further tested the bilinguals, finding that when exposed to two languages with opposite directions, participants accept both directionalities for the flow of time.

Secondly, spatial metaphors used to talk about time also relate to bilinguals' conception of time directionality. To be specific, English speakers more typically use front and back (horizontal) spatial metaphors to talk about time, while up and down (vertical) terms used to describe time are more common in Chinese (Boroditsky, 2001). For example, in English, meetings can be moved forward or pushed back, while there are "上个月" (*shang ge yue*, literally up month, meaning last or previous month) and "下个月" (*xia ge yue*, literally down month, meaning next or following month) in Chinese. Based on these distinctions, Boroditsky (2001) argued that English native speakers may prefer to embody time along the horizontal axis, while Chinese native speakers may be more likely to think about time vertically. However, Chen (2007) failed to replicate Boroditsky's research (2001), reporting that there were no significant differences between response times (RTs) for horizontal spatial priming and RTs for vertical spatial priming, neither in English nor Chinese native speakers. Moreover, also unlike Boroditsky's findings (2001), native speakers of English were observed to spend more time in processing temporal stimulus sentences following a horizontal mode than a vertical mode (January and Kako, 2006; Tse and Altattiba, 2014). This may be because besides vertical temporal metaphors, there are also horizontal terms (front and back) used by Chinese native speakers to talk about time, such as "前天" (*qian tian*, literally front day, meaning the day before yesterday) and "后天" (*hou tian*, literally back day, meaning the day after tomorrow) (Scott, 1989). Actually, horizontal spatial metaphors were found to be used even more frequently than vertical spatial metaphors by Chinese native speakers in a corpus analysis (Chen, 2007). Thus, evidence from behavioral studies on the effects of spatial and temporal metaphors on monolingual speakers' conceptualization of time directionality is inconsistent. Furthermore, Li, Casaponsa, Wu, and Thierry (2019) extends the investigation of temporal cognition to bilingualism, reporting that bilinguals' time conceptualization tends to be interfered by the temporal metaphors in their L1 when tested in L2 (Li, Casaponsa, Wu, and Thierry, 2019). More research investigating the temporal cognition of bilinguals could be used.

2.2.2 Research on Numerical Cognition in Monolinguals and Bilinguals

Cross-linguistic research on numerical cognition has mainly focused on two aspects, the first of which investigates the effects of the L1 numerical system's linguistic structure on participants' mental number line (MNL). The second aspect pays more attention to the effects of the counting system' numerical transparency on participants' numerical abilities, which is closer to the aim of present study.

Firstly, number representation specificities include inversion and calculation. Inversion means that in some languages, such as German and Dutch, tens and units are spoken and written in reversed order (Comrie, 2005). For instance, in Dutch, the number 21 is expressed as éénentwintig, literally "one-and-twenty", while 21 is represented by "twenty-one" in English. Moreover, in some languages (e.g., French), the expression of a number involves complex calculations. For example, in French, the number 91 is written quatre-vingt-onze, literally "four multiplies twenty and adds eleven – 4 * 20 + 11", while the number 91 is "九十一", literally "ninety-one" in Chinese. Following the idea of linguistic relativity, differences in the counting system's constitutive structure across languages are hypothesized to affect speakers' number line estimations on the mental number line. MNL refers to a mental scale, to which numbers are spatially aligned with a left-to-right orientation of increasing magnitude (Dehaene, Bossini, and Giraux, 1993). To be specific, zero is placed at the leftmost point (Dehaene, 1997; Fischer, 2003), whereas the bigger the number is, the further the number is represented to the right (Zorzi, Priftis, and Umiltà, 2002). In related research, speakers of languages with a right to left reading culture (e.g., Palestinians) were reported a reverse orientation of MNL (Shaki, Fischer, and Petrusic, 2009), where zero is positioned at the right endpoint.

Researchers (e.g., Booth and Siegler, 2006; Siegler and Opfer, 2003) further pointed out that the mental scale, along which numerical values are represented, is compressed and exists as a logarithmic line rather than a linear ruler in one's childhood. Over time, such spatial mapping of numbers would be less compressed and more accurate (Siegler and Opfer, 2003), which is believed to take place roughly from the age of 4 (Opfer and Furlong, 2011) to 7 (van Galen and Reitsma, 2008), at least for numbers ranging from 0 to 100 (Booth and Siegler, 2008). To assess speakers' spatial estimation abilities of number magnitude, the typical task requires participants to point out the spatial position of a number on a physical line flanked by two numbers that limit the range of the imaginary line (for range 0 – 100, see Opfer and Siegler, 2007; Muldoon, Simms, Towse, Burns, and Yue, 2011; for range 0 – 1,000, see Opfer and Siegler, 2007; for range 0 – 10,000, see Thompson and Opfer, 2010). Results show that children of languages with inversion and calculation counting systems performed worse than children of languages without inversion and calculation properties (e.g., Helmreich, Zuber, Pixner, Kaufmann, Nuerk, and Moeller, 2011). Additionally, the potential attribution of education is carefully obviated by recruiting pre-school children as participants (Siegler and Mu, 2008). However, Le and Noel (2020) found that there were no significant differences in performance of numerical tasks between Vietnamese and French-speaking preschool children, indicating that only limited advantages were provided by the transparent counting system, and Chinese preschool children's relatively high accuracy and fast reaction times could be a result of positive numeracy influence from family.

Laski and Yu (2014) further investigated bilinguals' number line estimation development, finding that Chinese-American children performed better than monolingual American children but worse than monolingual Chinese children. This was interpreted as an indication that the acquisition of a less regular number-naming system in English could impede bilinguals' knowledge of numerical scale mapping. Rinsveld, Schiltz, Landerl, Brunner, and Ugen's research (2016) supported Laski and Yu's study (2014), finding that number magnitude judgments in German-French bilinguals and German/French monolinguals could be qualitatively affected by the counting system's instruction rules (with or without inversion). Thus, the hypothesis that the linguistic structural properties of the counting system can influence children's mental conceptualization of number lines is verified. Nevertheless, compared to the amount of research on children, evidence on adults is very limited due to their relatively more sophisticated and mature numerical cognition. This limitation could be overcome by employing a calendar reasoning paradigm, since English calendar names are much less regular and transparent than English number terms.

Secondly, while the MNL task has been mainly used to assess children's ability in numerical spatial estimation, much interest has also been expressed in the field of numerical transparency on children's mathematical abilities in two typical tasks, a counting task (e.g., see Miller, Smith, Zhu, and Zhang, 1995; Dowker and Mark, 2015) and an arithmetic task (e.g., see Pica, Lemer, Izard, and Dehaene, 2004). The counting task can be further divided into abstract counting and object counting. In the abstract counting task, participants are required to count aloud from 0 or 1, while in the object counting task, researchers would show the participants a set of objects in small (3-6 |Leung, and Bisanz, 2006). The arithmetic task, on the other hand, sets up calculation questions, such as addition and subtraction, and collects children's accuracy of answers. Crosslinguistic comparison results of the tasks have demonstrated positive effects of numerical transparency on monolingual children's development of mathematical abilities, especially in comparisons of arithmetical performance between Chinese-speaking and English-speaking children over a long time span (Miller and Stigler, 1987; Miller, Smith, Zhu and Zhang, 1995; Geary, 1996; Miura, Okamoto, Vlahovic-Stetic, Kim, and Han, 1999; Miller, Major, Shu, and Zhang, 2000; Miller, Kelly, and Zhou, 2005; Ng and Rao, 2010; Chan, 2014). Dowker and Li (2019) further pointed out that only reaction times of Chinese children were significantly better than those of English children, while the two groups' accuracy scores showed no significant difference. However, despite the importance of numerical transparency, some researchers (e.g., Miller, Kelly, and Zhou, 2005; Ng and Rao, 2010) also emphasize the potential contribution of educational and cultural differences between children with various native languages. To control the potential influence of education and culture, Siegler and Mu (2008) designed a study, where participants' different achievements in arithmetic tasks could only be attributed to the various degrees of numerical transparency, with all else kept equal, demonstrating a link between languages' numerical transparency and the mathematical abilities of speakers of corresponding languages. Advantages of numerical transparency have also been observed in the performance of German (transparent numerical structure) and Italian (opaque numerical structure) children (Helmreich, Zuber, Pixner, Kaufmann, Nuerk, and Moeller, 2011), and Welsh children raised in English (irregular counting system) and Welsh (regular counting system) speaking environments (Dowker, Baa, and Lloyd, 2008; Dowker and Roberts, 2015).

Results of research on the effects of transparent and regular number systems on bilinguals' mathematical abilities have failed to reach a consensus. On the one hand, some researchers found an advantage of bilingualism in the development of numerical abilities. For example, Dowker, Baa, and Lloyd (2008) investigated the participants' performance in arithmetic tests and 2-digit number reading and comparing tasks (read stimulus numbers loudly before comparing their numerical magnitude). This study recruited Welsh-English bilinguals and Welsh/English monolinguals for the first experiment and Tamil-English bilinguals and Tamil/English monolinguals for the second experiment. The transparency of Tamil is in between of Welsh (transparent) and English (opaque). Dowker, et al. confirmed a positive effect of Welsh on bilinguals' accuracy in 2-digit numbers reading and comparing tasks, though there were no significant differences in arithmetic tests among the three groups, which was hypothesized to be relevant to participants' specific aspects of written abilities. Advantages in processing arithmetic tests were found in the performances of Tamil-speaking children, who were better than their English monolingual peers. On the other hand, the research of Rasmussen, Ho, Nicoladis, Leung, and Bisanz (2006) gained results pointing in the other direction. Through testing Chinese-English bilingual children's counting abilities and comparing the bilingual data with the English and Chinese monolingual data (gathered by Miller, Smith, Zhu, and Zhang, 1995), Rasmussen and his team reported no evidence of any effect of first language transfer or an advantage of the L1 transparent number-naming system on the bilinguals' number counting, neither in the abstract nor object counting task. When it comes to adult bilinguals, native Chinese speakers who were educated in Canadian schools with French as the medium of instruction outperformed French monolinguals in a simple arithmetic task (Campbell and Xue, 2001), though Chinese-French bilinguals' advantage in calculation questions could result from a cultural rather than numerical transparency factor, as the Chinese-French bilingual participants showed a stronger reliance on memorizing correct answers than French monolinguals. Similarly, German-English bilingual adults outperformed English monolinguals in multiplication tasks (Kraut and Pixner, 2022). However, such results could be affected by various factors, such as daily mathematical application and practice, language proficiency level (see Rasmussen et al., 2006 for evidence from bilingual children), and the instruction language of math class in early school years and later learning environment.

Furthermore, with the development of technologies and medical facilities, some investigators have recently turned to study bilinguals' brain activation patterns during solving mathematical questions, assisted by neuro-imaging techniques, such as fMRI (Functional Magnetic Resonance Imagination). The fMRI studies investigate speakers' numerical cognition by scanning the participants' brains while they are performing arithmetic computation and testing if crosslinguistic speakers activate different brain regions to solve mathematical questions. The fMRI studies reported that without high proficiency in a second language, bilinguals tended to retrieve arithmetic facts through activation of verbal codes from the first language (Wang, Lin, Kuhl, and Hirsch, 2007; Lin, Imada, Kuhl, 2011). Bilinguals might translate and process the questions in their first language even when tested in a second language. Behavioral studies on bilinguals' mathematical abilities also emphasized a higher accuracy and shorter RTs for solving problems in the participants' L1 than in L2 (Marsh and Maki, 1976; French-Mestre and Vaid, 1993) or at least in the instruction language used to teach arithmetic calculations (Bernardo, 2001; Van Rinsveld, Brunner, Landerl, Schiltz and Ugen, 2015). What these findings suggest is that the distinctions between the performance of monolinguals and bilinguals in arithmetical tasks tend to involve a translation process from the L2 to the L1 and back to the L2 by the bilinguals who are tested in L2. Nevertheless, the above studies paid more attention to late bilinguals, and research on bilinguals who acquired both languages much earlier in their life is needed. To fill the research gap, Van Rinsveld, Dricot, Guillaume, and Rossion (2017) recruited highly proficient and balanced German-French bilinguals with shared language learning history and scanned their neuroimages in simple and complex addition tasks with fMRI, reporting differential activation patterns in additive operations when tested in L1 and L2. However, it is hard to say whether L1 or L2 assimilates the bilinguals' performance, as no control group of German monolinguals and French monolinguals is involved. While there are numbers of research on numerical transparency, investigations on the effect of linguistic transparency on calendrical terms are very limited.

2.2.3 Research on Calendar Representation and Processing in Monolinguals and Bilinguals

According to Levin and Wilkening (1989), children spontaneously measure durations of an event by counting in rhythm before 5 years old, and the conventional time units in human society are provided to the children to achieve this goal. The acquisition of the calendar systems could be a rather difficult task and children could face many challenges when using the conventional temporal terms (Friedman, 1983, 1984, 1990). In Friedman's influential view of calendar reasoning, children initially learn lists of the calendar names in sequence, including terms of days within a week and months within a year. As a result, when doing a calendrical reasoning task, such as identifying the name of the day that comes two days before a given day, or determining which month comes five months after May, children need to recite the whole sequence of units and count them overtly or covertly to obtain a precise answer. This process is termed verbal-list processing. Four pieces of evidence of the existence of the verbal list system were reported by Friedman (1983). The first one was that interference strongly inhibited children's performance in simultaneous verbal-listing calendar naming tasks. The second was a distance effect, where it took participants longer response times and more effort to reach a further weekday or month away from the stimuli than a shorter time. The third evidence was a directionality effect, as it was easier to count backward than forward. The last evidence came from participants' self-reported overtly or covertly reciting, though participants could misunderstand or misreport their strategies. Furthermore, imagery (Friedman, 1986) was found to be used to help calendrical questions reasoning at the age of 10, when imaginary and verbal-listing co-exist. However, the findings and claims of Friedman were based exclusively on data from native English speakers. With the formation rules of calendar terms varying across different languages in the world, speakers of other languages may more or less suffer in the acquisition of conventional time terms and apply different strategies (e.g., counting, calculation, and memory) when reasoning the calendrical calculation questions. Due to the distinction of linguistic transparency between Chinese and English calendar systems (as shown in 2.1.2.1), a comparison between calendrical calculation abilities and calculational strategies could be informative.

Previous research has suggested that speakers of Chinese outperform speakers of English when reasoning about weekdays and months. A plausible explanation is that the Chinese calendar naming system is more transparent than that in English. Huang (1993) studied Chinese speakers' preferred reasoning strategies in the month calculation task and reported a different set of strategies from those reported in Friedman's research (1983). Instead of verbally listing the calendar terms, Chinese speakers used numerical processing, also known as arithmetic operations. For example, when Chinese speakers were asked to identify the month that comes three months after May (literally "month five"), they tended to add 3 to "month 5" and get the answer August (month 8), that is, 3 + 5 = 8. This may be because young leaners of Chinese firstly acquire the regular numerical counting system and then add the numbers to the root "星期" (pronounced as *xing qi*, meaning week) to express the target weekdays (e.g., 星期三, Wednesday is simply *"week-three"*) and root "月" (pronounced as *yue*, meaning month) to identify the target months (e.g., 四月, April is simply *"si-yue"*), rather than reciting the whole list of calendar terms by rote memorization (Cheng and Chan, 2005; Mark and Dowker, 2015). As a result, the regular and transparent conventional temporal representations lead to the application of numerical calculation in week and month calculation question tasks. Additionally, unlike English monolinguals, who more typically use a verbal listing strategy and need longer response time to solve reverse and longer distance questions, no distance or direction effects were found in Chinese monolinguals' month reasoning process, since addition and subtraction take similar lengths of time in the arithmetic calculation strategy. A boundary effect was still found in Chinese adults' calculations. Participants spent longer time when solving questions involving boundary crossing, which may result from an additional calculation process needed in performing boundary crossing trials. For instance, in order to obtain the month's name that is 10 months after March (month 3), one needs to add 10 to 3, get 13, then subtract 12 (as there are twelve months in a year), and get the resulting month 1 (January), that is, 10 + 3 – 12 = 1. The additional step of subtraction thus hinders Chinese monolinguals' speed of working out the answer. Similarly, Jiang and Fang (1997) also demonstrated that both Chinese school children and adults took advantage of numerical arithmetic operations and were influenced by the boundary effects in weekday calculation questions.

Furthermore, there is direct evidence that it is not cultural background but the transparency of Chinese calendar terms affects participants' performance in calendrical calculation processing. Although cultural is typically inseparable from language, Huang (1999) managed to overcome it by recruiting two groups of adult Chinese speakers from rural areas and asking them to do calculation tasks with solar months names and lunar months names, respectively. Lunar month names are a kind of traditional Chinese calendar representation, where calendar terms are opaque; for instance, January is represented by "正月" (*zheng yue*) instead of "一月" (*yi yue,* month 1) and December is "腊月" (*la yue*, *la* is something relevant to the sacrificial ceremony) rather than "十二月" (*shi er yue*, month 12). The lunar calendar system is mainly used in Chinese rural areas, where the solar month system is applied at the same time, and participants reported equal proficiency and frequency of the application of the two kinds of calendar naming systems. Based on different ways of labeling months within the same language Chinese, results reported different cognitive routines, as the solar group outperformed the lunar group in both accuracy and reaction times. Additionally, distance and direction effects were found in the lunar group, while only the boundary effect worked in the solar group, which was in line with the results of Friedman (1990) and Huang (1993). The self-reported strategy of the lunar group was verbal listing, while the strategy of the solar group was arithmetic operations. Therefore, it appears that linguistic relativity can be supported even within a language group where different subgroups are habituated to linguistic codes with different degrees of transparency.

The research mentioned above respectively investigated Chinese monolinguals and English monolinguals and Kelly, Miller, Fang, and Feng's (1999) study was the first direct comparison of the development of calendrical calculation abilities between Chinese speakers and English speakers, where both day-of-the-week and month-of-the-year calculation tasks were involved. Research reported that the Chinese speakers were overall faster than the English speakers when processing weekday and monthly calculations and showed a propensity for arithmetic calculation strategy. Moreover, the Chinese-speaking group was not affected by distance (short or long) or direction (forward or backward) but affected by boundary – the Chinese group spent longer time solving cross-boundary than within-boundary questions. Notably, the English group mostly reported the use of verbal listing strategy and was negatively affected by long distance and backward direction but not by the boundary effect. To sum up, it appears that the linguistic transparency of the calendar representation system leads to differences in calendrical calculation performance between languages with or without opaque calendar terms. Nevertheless, it is still unclear how the linguistic transparency of calendar lexicons affects the calendrical calculation process of bilinguals, who not only acquire the transparent calendar terms but also know the opaque calendar terms.

Perhaps the first research investigating bilinguals' calendrical calculation process is Yang and Zhang's (2011) study, at least to the best of the author's knowledge. Yang and Zhang tested whether the existence of a specific linguistic label for a temporal unit would positively affect the bilinguals' performance in the calendrical reasoning task. In the Cantonese language, there is a specific linguistic label used to represent the time unit "five minutes", called "一个字" (pronounced as *yi ge zi*, literally "one word"), which does not exist or have a substitutional expression in Modern Standard Chinese. Taking a quarter to five as an example, Cantonese speakers tend to say it as "四点九个字" (*si dian jiu ge zi*, literally "four o'clock and nine words") rather than "四点四十五" (*si dian si shi wu*, literally "four and forty five o'clock") as in Modern Standard Chinese. Modern Standard Chinese monolinguals and bilinguals who are fluent both in Cantonese and Chinese were asked to do calculations involving five-minute units in

Chinese. Results showed that Cantonese-Chinese bilinguals outperformed the Chinese monolinguals, indicating that the existence of a specific linguistic label for a temporal unit in one language tends to positively influence participants' performance in the "five-minute relevant calculation" task even when tested in another language. This may be because bilinguals spontaneously transfer from the instruction language to the language with more efficient representation lexicons. Nevertheless, the effects of the calendar system's linguistic transparency on bilinguals were still not explored.

Bassetti, Clarke, and Trenkic (2018) added the factor of bilingualism and compared day-of-week and month-of-year calculations in English monolinguals and Chinese-English bilinguals. The research found that reaction times of the bilingual group were shorter than those of the English group in the month calculation task but longer in the weekday calculation task, as the month calculation task (12-based calculations) is more demanding than the week calculation task (7-based calculations). Moreover, in the month calculation task, directionality effects were found in the English group but not in the Chinese group. Crossing boundaries negatively affected bilinguals' performance in the backward direction calculations but not English speakers either in the forward or backward direction. In the week calculation task, crossing calculations negatively affected the bilingual group both in the forward and backward direction but had no effects on the native English speakers. The results showed no directionality effects on either group. Additionally, Chinese-English bilinguals mostly self-reported a numerical arithmetical operation when processing the calculation task, while most native English speakers relied on verbal listing, which was in accordance with previous studies (see Huang, 1993; Jiang and Fang, 1997 for evidence of Chinese speakers' reasoning strategy; see Friedman, 1983 for evidence of English speakers' reasoning strategy; see Kelly, Miller, Fang, and Feng, 1999 for evidence of both groups).

2.3 Research Gap

Although Bassetti and her colleagues' (2018) paper is important in filling some research gaps in the calendrical reasoning field, there are some modifications that could be done in the tested group and research design.

Firstly, there are only two groups in Bassetti et al.'s (2018) research, including an English native speakers group tested in L1 English and a Chinese-English bilingual group tested in L2 English, lacking a Chinese native speakers group tested in L1 Chinese. A direct comparison across the English, Chinese, and Bilingual group can more establish potential L1 transfer effects on numerical processing more firmly.

Secondly, although the design of calendar calculation tasks is replicable, some calculation questions used may have been opaque not (only) for numerical but for semantic reasons. In the original study, participants were first informed that it takes four days or seven months for seeds to sprout or blossom. Then participants needed to answer when the seeds will sprout or blossom knowing when they had been panted, or when the seeds had been planted knowing when they sprout or blossom. It is possible that the Chinese speakers needed to make more effort for the semantic decoding of the question, which could have prolonged their response speed. This factor was not controlled. Moreover, even though the original calculation tasks apply an online task, trying to examine the students' automatic reactions, the participants were asked to give their answers orally and press the "next" button to move on while being recorded. A more time-sensitive on-line processing task may be needed (Blom and Unsworth, 2010, p. 139 - 142) to exclude potential effects linked to verbalization speed.

Thirdly, the original study tested the participants' mathematical abilities by asking them to solve 7-based and 12-based calculation questions, which are both common arithmetical exercises with a decimal addition and subtraction. However, the week and month calculation questions are both uncommon calculations, where decimal addition and subtraction are not applicable, but follow a 7-based and 12-based calculational rules. An Hour Calculation Task, which involves uncommon 24-based instead of regular 10-based calculation questions and relates to calendrical reasoning, could be an informative addition and a useful control condition for crosslinguistic comparisons.

Fourthly, two factors that have not been considered earlier are Distance (short or long) and Input (linguistic and numerical). Distance is an important factor, the effect of which differs in transparent and opaque calendar terms. Performance of speakers of a language with a transparent calendar naming system is known not to be affected by the factor of Distance, while the performance of speakers with knowledge of opaque calendar terms tends to be negatively influenced by long distance (Friedman, 1983; Huang, 1993; Kelly et al., 1999). The factor Input type is first proposed in the present study. Linguistic input refers to the calendar lexicons, while numerical input transfers the calendar lexicons into Arabic numbers. For instance, the numerical version of "Monday" is "1st day". Speakers with knowledge of opaque calendar terms could be positively affected by the numerical input when performing calendrical calculation tasks, since the numerical input directly provides them the numbers for calculating. In sum, this dissertation is a partial replication of Bassetti, Clarke, and Trenkic's (2018), with four innovative components (list them here).

2.4 Aims and scope of this dissertation

This dissertation investigated the effects of transparent vs. opaque calendar naming systems on respective L1 speakers' and bilinguals' (knowing both transparent and opaque calendar names) performance in calendrical calculation task. The strategies different language communities used to solve the calendrical calculation questions were also studied.

To fill the four research gap (Section 2.3), three groups were tested, the English group, the Chinese group, and the Bilingual group. In order to eliminate the potential differences in mathematical abilities of various participants, the bilingual speakers were invited to do the calendrical calculation task in English as the Bilingual group and in Chinese as the Chinese group. This is achievable and reasonable because only their native language is activated when tested in Chinese. The two tests were two weeks apart to ensure the participants would not remember the questions or answers. Another improvement proposed here relates to construct validity and internal validity. This dissertation project asked the participants to choose the right option by pressing a button rather than by saying the answer out loud, so multiple factors such as variation in articulation speed and second language anxiety (Mackey and Gass, 2005) could be controlled for. Moreover, with the addition of the Hour and Year Calculation Task, the present study investigated whether different groups' mathematical abilities in uncommon and common calculations respectively were at comparable level. At last, the factors Distance and Input were added in the present study to test how these two factors might influence different groups' performance in calendrical calculation tasks. Overall, the present study was built on two main research questions and tested their corresponding hypotheses:

RQ1: To what extent do Chinese-English bilinguals differ from monolingual English speakers when each group performs calendrical calculations with their corresponding L1 as the language of testing?

H1: There will be significant differences between the two groups in two contexts (Weekday and Month calculations). The reaction times (RTs) of bilinguals are predicted to be shorter than those of monolingual English speakers because of lexical and numerical transparency in marking weekdays and months in L1 Chinese, while the corresponding English terms are opaque.

RQ2: To what extent do Chinese-English bilinguals differ from English monolinguals when both groups perform calendrical calculations with English as the language of testing?

H2: There will be significant differences between the two groups in two contexts (Weekday and Month calculations). The RTs of bilinguals are predicted to be longer than those of monolingual English speakers because of the negative effects of a weaker language.

3. Methodology

3.1 Participants

Thirty English native speakers (15 females) and thirty Chinese-English bilinguals (16 females) took part in this experiment. Both the English participants (Mean age = 24.2, max. = 35, min. = 19) and the Chinese-English bilingual participants (Mean age = 25.2, max. = 35, min. = 20) were recruited from universities in English-speaking countries to make sure the participants were under same language environment. Following Athanasopoulos et al. (2011), Park and Ziegler (2014), and Vanek and Selinker (2017), all participants were asked to complete a questionnaire to collect their language background information. The background information questions are shown in Appendix 1. The English participants were all English monolingual speakers who at the time of testing had no experience or little exposure to a second language in daily life. The bilingual participants were Chinese-dominant learners of English who at the time of testing had gained a score equal to or greater than 6.5 in the International English Testing System (IELTS). They started to learn English in kindergarten and at the time of testing used English in certain situations (such as work and study) and Mandarin Chinese in other situations (such as contact with family). All participants were right-handed and reported normal or corrected-to-normal vision.

3.2 Ethical Considerations

All potential participants were invited through an email of this study's advertisement (Appendix 2). Those participants who were interested in the present study would contact the researcher for the Participant Information Sheet (PIS, Appendix 3). After reading the PIS and ensuring that there had no questions about the study, participants who were still willing to do the experiments would sign the Participant Consent Form (Appendix 4) with electronic signatures and send them back to the researcher.

The study received ethical approval from the Human Participants Ethics Committee at the University of Auckland (Appendix 5). Participation was voluntary and participants were free to withdraw from the study at any time during data collection. The anonymity and confidentiality of participants' information were ensured. The participants' identities were kept confidential by the researcher. Information about age, gender, educational background, the length of learning English, and Chinese and English proficiency were reported in aggregated form, which made it less likely that the participants would be identifiable. Additionally, the participants were de-identified by coding participants as ES (English students) 1,2,3 … and BS (bilingual students) 1, 2, 3 … rather than their real names.

3.3 Materials and Tasks

3.3.1 Experiment 1 – Calendrical Calculation Task

Task 1 was a Calendrical Calculation Task, which was a production task. The materials involved 96 calendar reasoning questions (as shown in Appendix 6), including Week, Month, Hour, and Year calculations. The Week and Month Calculation Tasks were designed to test the effect of calendar terms' linguistic transparency on English speakers and Chinese-English bilinguals' reaction times and accuracy. Given that the Week (7-based calculations) and Month (12-based calculations) Calculation Tasks were uncommon arithmetic calculation tasks, the Hour Calculation Task was used to test whether the two groups had comparable reaction times and accuracy in 24-based calculations, while the Year Calculation tested participants' reaction times and accuracy in more common 10-based calculations. Four variables were manipulated to design the calculation questions in the first experiment (the calendrical calculation task), including Distance (short or long), Direction (Forward or Backward), Boundary (within or across), and Input (linguistic or numerical).

3.3.1.1 Week Calculation Task

With four independent variables, the Week Calculation Task involved 32 questions. For the variable Distance, if the numerical gap between the question and answer was less than 4 days (half of 7 weekdays), then it counted as a short trial. If the numerical gap was more than 4, then it counted as a long trial. For the variable Direction, in the forward condition, the target was after the stimuli, while the target was before the stimuli in the backward condition. For the variable Boundary, within-boundary trials were within the boundary of a week, while cross-boundary trials crossed the Sunday-Monday boundary. The variable Input included input linguistic and input numerical. All questions were displayed both in numerical and linguistic description versions of numbers. For example, "Monday + 1 day = ?" was a short, forward, within, linguistic trial, while "4th day – 6 days = ?" is a long, backward, across, numerical trial. There were 4 short forward (2 within and 2 across), 4 long forward (2 within and 2 across), 4 short backward (2 within and 2 across), and 4 long backward (2 within and 2 across) questions, 16 questions in total, which were displayed in both linguistic and numerical input.

Distance	Short	Monday + 1 day = ?
	Long	Monday + 6 days = ?
Direction	Forward	Wednesday + 2 days = ?
	Backward	Saturday – 1 day = ?
Boundary	Within	Monday + 1 day = ?
	Across	Friday + 3 days = ?
Input	Linguistic	Monday + 1 day = ?
	Numerical	1st day + 1 day = ?

Table 1. Examples of week calculation questions in different conditions.

3.3.1.2 Month Calculation Task

With four independent variables, the Month Calculation Task also involved 32 questions. For variable Distance, short trials referred to trials with a numerical gap between question and answer of less than 6 months (half of 12 months), while long trials referred to trials with a numerical gap of more than 6 months. For Direction, forward trials referred to calculation months chronologically following the stimuli, while backward trials calculated months chronologically preceding the stimuli. For condition Boundary, within-boundary referred to trials within the boundary of a year, across-boundary trials crossing the December-January boundary. All questions were presented in linguistic and numerical input as in Week Calculation. One example of "short, forward, within, and linguistic" trial could be "January + 2 months = ?" and an example of "long, backward, across, and numerical" trial could be "6th month – 9 months = ?". There were 16 linguistic trials and 16 corresponding numerically input trials as in the Week Calculation Task.

Distance	Short	January + 2 months = ?
	Long	May + 7 months = ?
Direction	Forward	March + 6 months = ?
	Backward	December – 3 months = ?
Boundary	Within	March + 6 months = ?
	Across	August + 5 months = ?
Input	Linguistic	January + 2 months = ?
	Numerical	1st month + 2 months = ?

Table 2. Examples of month calculation questions in different conditions.

3.3.1.3 Hour and Year Calculation Task

Differences in transparency in the numerical system between English and Chinese are much fewer than that in calendar terms, so the English and Chinese adult speakers may perform similarly in hour and year calculation question, where there are numbers instead of calendar terms (e.g., two o'clock or 2003). The hour and year calculation tasks were applied to test if the two groups had comparable arithmetic skills in uncommon (24-based calculations) and common (10-based calculations) calculation questions under different conditions, Distance (short or long), Direction (forward or backward), and Boundary (within or across).

There were 16 hour and 16 year calculation questions, each with 4 short forward (2 within and 2 across), 4 long forward (2 within and 2 across), 4 short backward (2 within and 2 across), and 4 long backward (2 within and 2 across) trials. The condition Input was not involved. All hour calculation questions were described in the linguistic version, while all year calculation questions were described in the numerical version. The rationale for having two control conditions and each with different types of input (linguistic or numerical) is as follows. Firstly, the present study was sufficiently robust because it included a control condition with numerically input in year calculations and another control condition with linguistically input in hour calculation questions, where between-group differences were not expected. Secondly, the year and hour calculation tasks were essential to test different groups' mathematical abilities in uncommon 24-based calculations (hour) and common 10-based calculation questions (year). For example, "eight + 3 hours = ?" is a short, forward, within, and linguistic hour (uncommon) calculation question, while "1800 – 1400 = ?" is a long, backward, across, and numerical year (common) calculation question.

Distance	Short	Two o'clock + 7 hours = ?
	Long	Four o'clock + 15 hours = ?
Direction	Forward	Two o'clock + 7 hours = ?
	Backward	Fifteen o'clock – 11 hours = ?
Boundary	Within	Two o'clock + 7 hours = ?
	Across	Eighteen o'clock + 10 hours = ?

Table 3. Examples of hour calculation questions in different conditions.

Distance	Short	10 + 60 years = ?
	Long	1300 + 600 years = ?
Direction	Forward	10 + 60 years = ?
	Backward	90 – 55 years = ?
Boundary	Within	10 + 60 years = ?
	Across	50 + 310 years = ?

Table 4. Examples of year calculation questions in different conditions.

3.3.2 Experiment 2 – Self-reported Strategies Task

Task 2 was the Self-reported Strategies task, which focused on the strategies used by the participants in different conditions of different calendrical reasoning questions described in Experiment 1. In total, the combination of the conditions in Exp. 1 involved 16 week and 16 month calculation questions, namely Distance (short or long) * Direction (forward or backward) * Boundary (within or across) * Input (linguistic or numerical), in other words a two-by-two-by-two-by-two design, and 8 hour and 8 year calculation questions, namely Distance (short or long) * Direction (forward or backward) * Boundary (within or across), in other words a two-by-two-by-two design. There were seven strategies available, including Memory, Transform and Calculate, Count, Translate and Calculate, Translate and Count, Estimate, and Other (Luo, 2012). The full materials of Experiment 2 are shown in Appendix 6.

3.4 Procedure

There were three experimental steps. In the beginning, the researcher sent the information sheet and consent form to potential participants and then asked the qualifying participants to sign the consent form with an electronic signature if they agreed to continue. Next, a questionnaire was emailed to the participants to collect their background information. The consent forms and questionnaires were sent back to the researcher when completed. Then, all participants clicked the link sent to them earlier to open the Calendrical Calculation Task and Self-reported Strategies Task. When the participants initiated the Calendrical Calculation Task, they would see an instruction message shown on the computer screen followed by four practice calculations. The question appeared at the center of the screen and three options appeared below the question at the same time, one was correct and the other two were incorrect, as shown below.

E.g., Monday + 3 days = ?

Wednesday

Thursday Friday

The correct answer appeared randomly on the right, left, or up, with about one-third in each location. The instruction required the participants to choose the correct answers as fast and accurately as possible by pressing the direction arrows on the keyboard, "↑", "←", and "→". Participants were expected to answer 96 calendrical reasoning questions. Their reaction times (RTs – from when the target word(s) appeared on the screen until the participant's button press) and accuracy were recorded and sent to the researcher automatically. The software used was PsychoPy 3.0.

After the Calendrical Calculation Task, there was a short break. All participants were asked to take a break, but they could start the next task, the Self-reported Strategies Task, whenever they were ready. The question appeared in the upper part of the computer screen and seven strategies with detailed explanations and examples appeared below the question at the same time as below. Participants were asked to press number keys '1', '2', '3', '4', '5', '6', and '7' on the keyboard to choose one corresponding strategy that they thought best reflected their mental processes. Participants learned that they could spend as long as needed on the Self-reported Strategies Task, they knew that their reaction times would not be recorded and that there were no correct or incorrect answers. Strategies chosen by the participants were also recorded and sent to the researcher automatically by PsychoPy 3.0.

E.g.,

What strategy did you use to solve the following question? Monday + 3 days = ?
1. Memory (Automatic recall of the fact, you didn't need to work out the answer) 2. Transform and calculate (e.g., Transform 'Monday + 2 days' into '1 + 2 = 3' and get the answer) 3. Count (e.g., 'March + 2 months' – count 'March', 'April', 'May', so the answer is May) 4. Translate and calculate (e.g., translate the question into Chinese and then calculate in Chinese) 5. Translate and count (e.g., translate the question into Chinese then count in Chinese) 6. Estimate (Guess the answer) 7. Other (Another method that is not listed here)

Notably, in the Calendrical Calculation Task (Expt. 1) and Self-reported Strategies Task (Expt. 2), all English participants were tested in their L1 English. Half of the Chinese-English bilinguals were tested in L1 Chinese first and L2 English two weeks later to ensure that the participants would not remember the questions or answer, which might affect the RTs and accuracy in solving calendrical calculation questions. The left half bilinguals were tested in L2 English first and L1 Chinese later. The order of the tested languages in the bilinguals was Chinese-English versus English-Chinese 50% vs. 50% to avoid potential order effects.

All questions and forms in the two tasks in Chinese version were directly translated from questions and forms in the English version tasks. The Experiment 1 and Experiment 2 in Chinese version were the same as the two tasks in English version, except that Strategies "Translate and calculate" and "Translate and count" were deleted. The design may seem imbalanced because of the unequal number of strategies across the two languages, but this was not a mistake. The reason for excluding "translate-related strategies" was based on the rationale that the instruction language for learning mathematics for the Chinese-English bilinguals were their L1 Chinese, and there was no need for translating when tested in L1. Examples of calendrical calculation questions and strategy questions are shown below. The whole experimental process is shown in Figure 4.

E.g., 周一 ＋3 天 ＝ ？

周三

周四 周五

E.g.,

What strategy did you use to solve the following question? 周一 ＋3 天 ＝?
1. Memory (Automatic recall of the fact, you didn't need to work out the answer) 2. Transform and calculate (e.g., Transform '周一 ＋2 天' into '1 + 2 = 3' and get the answer) 3. Count (e.g., '三月 ＋2 个月' – count '三月', '四月', '五月', so the answer is 五月) 4. Estimate (Guess the answer) 5. Other (Another method that is not listed here)

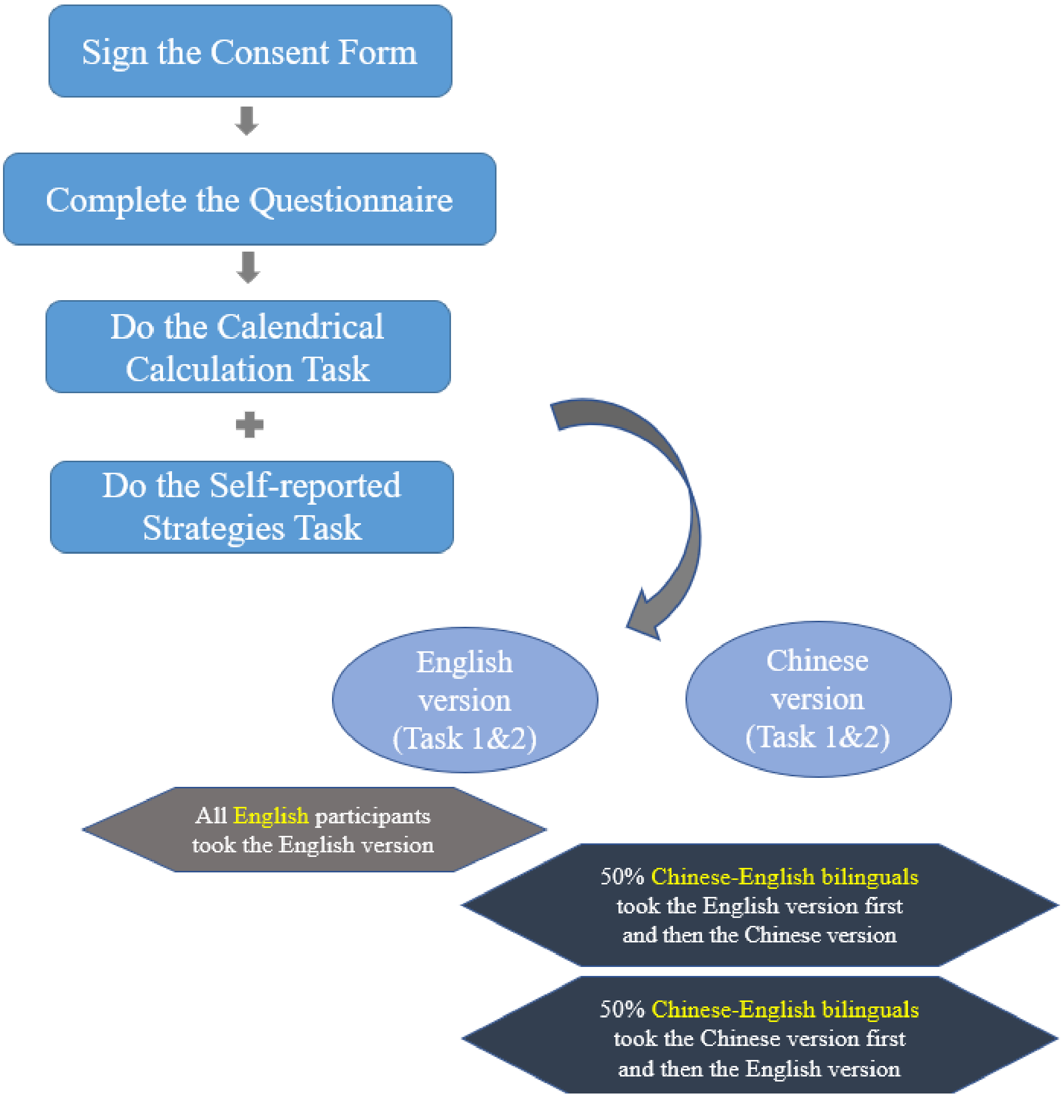

Figure 4. The experimental steps of present study.

3.5 Analysis Plan

The independent variables were reaction times (RTs) and accuracy. The main fixed factors were Calculation (Week, Month, Hour, Year) and Group (English, Chinese, Bilingual). Technically, there were only 2 groups (English monolinguals and Chinese-English bilinguals) in the present study, but the bilinguals were tested in two languages, so the Chapter Results of this study dealt with 3 groups. The English monolinguals tested in English were labeled as the English group. Data collected from bilinguals' performance in the Chinese version tasks was labeled as the Chinese group, while data from bilinguals' performance in the English version tasks was labeled as the Bilingual group. Thus, there is one analysis that is truly between-groups, referring to the Chinese-English bilinguals vs. English monolinguals. The second comparison is actually the same people – Chinese-English bilinguals, but tested in Chinese and English for two times. Moreover, the effects of Distance (short/long), Direction (forward/backward), Boundary (within/across), and Input (linguistic/numerical) is additionally analyzed.

4. Results

Participants were required to complete two experimental tasks, the Calendrical Calculation Task (Expt. 1) and the Self-reported Strategies Task (Expt. 2). The collected data from experiment 1 included Reaction Times (RTs) and Accuracy, while experiment 2 gathered strategies used by the participants in solving the calendrical calculation questions. Therefore, the Result chapter was divided into four parts, including the analysis of RTs, Accuracy, Self-reported Strategies, as well as a summary section of the main results.

4.1 Reaction Times

4.1.1 Preliminary Analysis

The Reaction Times (i.e., RTs between the onset of the stimulus and the point of pressing the chosen button) from Expt. 1 were analyzed in this part. In order to compare the cognitive demands of the answering systems in each group, RTs were only included from correct answers. RTs from incorrect responses were eliminated from the RT analysis of the week calculation task (6.8% trials, n = 196), month calculation task (6.8% trials, n = 197), hour calculation task (10.14% trials, n = 146), and year calculation task (3.3% trials, n = 48). Moreover, there were a few outliers in each group. Following Keating and Jegerski (2015) and Norris (2015), for the English participants, 70 data entries (28 week calculation questions, 22 month calculation questions, 10 hour calculation questions, and 10 year calculation questions) (2.43% of total RTs) were more than 2.5 standard deviations away from the group mean in each condition. These outlier RTs were eliminated from the RTs analysis. For the Chinese-English bilinguals who were tested in Chinese (L1), there were 70 outliers (34 week calculation questions, 19 month calculation questions, 11 hour calculation questions, and 6 year calculation questions) (2.43% of collected RTs), and they were not part of the analyzed dataset. For the bilinguals who were tested in English (L2), 58 outlier RTs (25 week calculations, 12 month calculations, 9 hour calculations, and 12 year calculations) (2% of total RTs) were eliminated from the RTs analysis.

There were two statistical steps to analyze participants' RTs, including the analysis of the effect of two main effect factors (Calculation and Group) and an additional analysis of the effect of different conditions (Distance, Direction, Boundary, and Input). Linear mixed effects models were used to do the analysis to take both fixed effect factors and random effect factors into consideration (Winer, 2014). The main fixed effect factors were Calculation (Week, Month, Hour, and Year) and Group. There were three groups, the English group, the Chinese group, and the Bilingual group. The English group involved English monolingual speakers who were tested in English (L1). The Chinese group involved Chinese-English bilinguals who were tested in Chinese (L1), so only Chinese (their mother tongue) were activated during the experiment, while the Bilingual group involved Chinese-English bilinguals who were tested in English (L2) and both Chinese (L1) and English (L2) could be activated. The effect of Distance (long or short), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical) were also taken into consideration and additionally analyzed as within-group factors. The random effect factors were Participant and Item.

4.1.2 RTs of Different Groups in the Calendrical Calculation Task

The analysis of reaction times of different groups in the Calendrical Calculation Task would involve four steps. Firstly, different groups' RTs in the Week, Month, Hour, and Year Calculation Task would be visualized. Secondly, the effect of factor Calculation and Group and their potential interaction would be explored. Thirdly, pairwise comparisons of RTs in Week and Month Calculation Tasks were between every two groups (English group vs. Chinese group, English group vs. Bilingual group, and Chinese group vs. Bilingual group). No pairwise comparisons would be processed based on data from Hour and Year Calculation Task, as there was no significant difference across groups' RTs in the two calculation tasks (as shown in 4.1.2.2). At last, the effects of Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical) on participants' RTs in the Week and Month Calculation Tasks would be additionally analyzed as within-group factors. The different conditions' effects would not be explored in the Hour and Year Calculation Tasks, since they were only used to test the general arithmetic abilities in common (10-based calculations) and uncommon (24-based calculations) mathematic questions.

4.1.2.1 Visualization of Different Groups' RTs in the Calendrical Calculation Task

This part examined the effect of the two main fixed factors (Calculation and Group) on participants' RTs in the calendrical calculation task (Expt. 1). RTs of each group in different calculation tasks (week, month, hour, and year) are illustrated in Figure 5. Overall, the Hour Calculation Task (M = 7715, SD = 5411) took the longest RTs, while the Year Calculation Task took the shortest RTs (M = 3265, SD = 1791). The Month Calculation Task (M = 5301, SD = 3466) took longer responses times than the Week Calculation Task (M = 4455, SD = 2807). Moreover, the Chinese group (M =3758, SD = 2452) was faster than the English group (M = 4829, SD = 3062) and the Bilingual group (M = 4806, SD = 2756) in the Week Calculation Task. In the Month Calculation Task, the Chinese group (M = 4529, SD = 3401) was still the fastest, while the Bilingual group (M = 5406, SD = 2994) was slightly faster than the English group (M = 6064, SD = 3829). The Hour (24-based calculations) Task took shortest RTs of the Chinese group (M = 6898, SD = 5427) to choose the correct answer, while the English group (M = 7980, SD = 5677) and the Bilingual group (M = 8293, SD = 5015) performed similarly. RTs of the three groups were at ceiling level in the Year Calculation Task (10-based calculations), English group (M = 3505, SD = 1869), Bilingual group (M = 3208, SD = 1729), and Chinese group (M = 3087, SD = 1747).

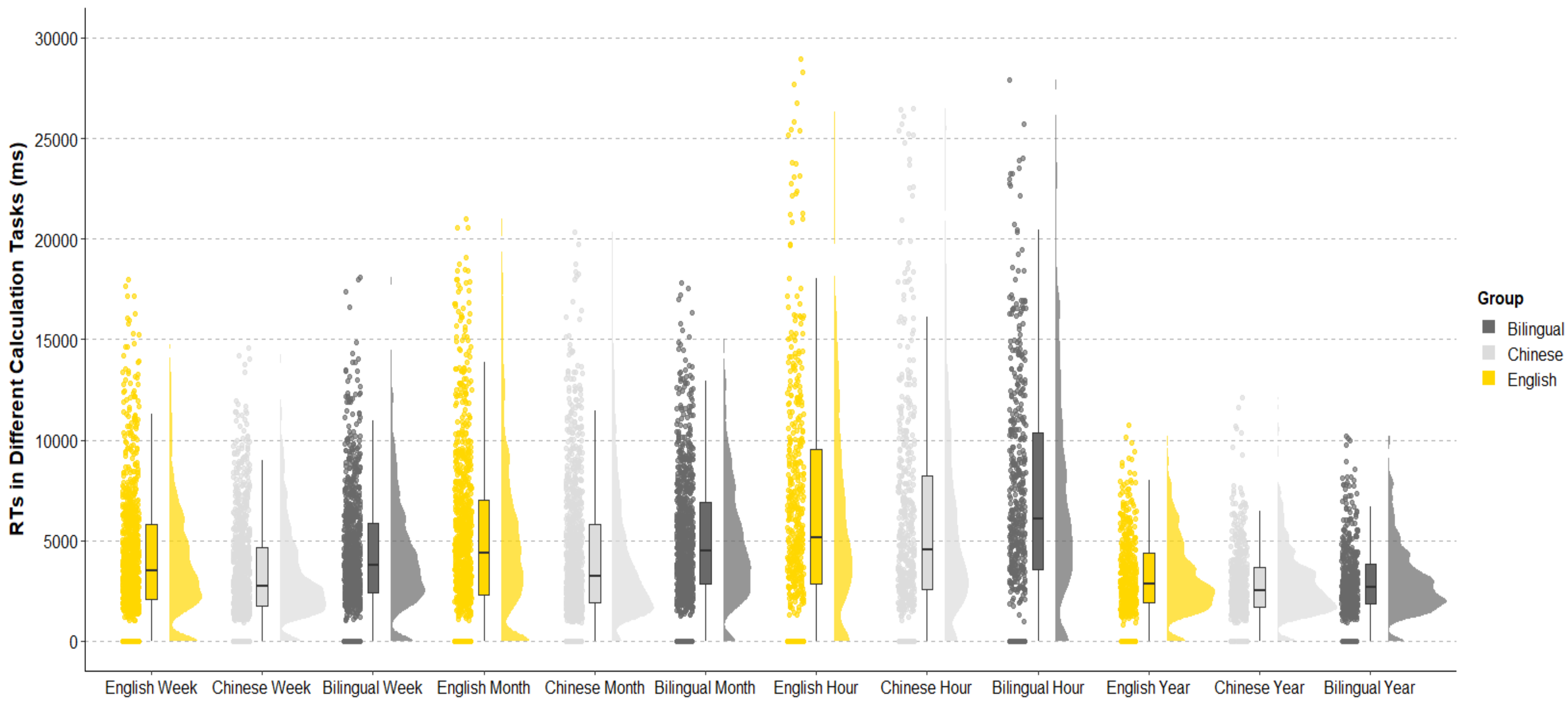

Figure 5: RTs of English speakers tested in English (English), Chinese-English bilinguals tested in Chinese (Chinese), and Chinese-English bilinguals tested in English (Bilingual) in each Calculation Task (Experimental 1) (Error Bars = 95% Confidence Interval).

4.1.2.2 Effects of Factor Calculation and Group and the Interaction between Calculation and Group

To test the effect of Group on reaction times of participants when solving calendrical calculation questions, this study built the mixed-effects regression models using the lme4 package (Baayen et al., 2008) in the R software (Version 4.1.3 R Development Core Team, 2021). As fixed factors, Calculation (week, month, hour, or year) and Group (English, Chinese, and Bilingual) were entered into the model. The dependent variable was Reaction times (RTs), and the random effect factors were Participant and Item. The model included all possible random effects (Barr et al., 2013), with random slopes over calculation by participant and random slopes over calculation, group, and their interaction by item as follows:

RTs ~ Calculation * Group +

(1 + Calculation | Participant) +

(1 + Calculation * Group | Item)

The multiple signs of the model were then replaced by the plus sign, and the new model was compared with the original one to test if there was a significant interaction between Calculation and Group, and to test whether this interaction significantly improved the model fit (Winter, 2014). The results reported that Calculation and Group were not inter-dependent on each other, $x^2(63) = 59.368$, $p = 0.6065$), so the below model was used as the full model. The results are shown in Table 5. The full dataset with RTs per participant, per calculation can be found in https://doi.org/10.17605/OSF.IO/HJZRG.

RTs ~ Calculation + Group +

(1 + Calculation | Participant) +

(1 + Calculation + Group | Item)

To further explore the effect of factor Calculation, a full model including Calculation and a reduced model excluding Calculation were compared to statistically test whether participants' reaction times significantly differed in processing week, month, hour, and year tasks. This comparison showed that the model fit was significantly improved with the presence of Calculation, $x^2(3) = 23.496$, $p < 0.001$, confirming that Calculation is a significant predictor of how RTs varied.

Then, the analysis zoomed in on each Calendrical Calculation Task to investigate the influence of factor Group through comparing a model including Group with a reduced model without Group in the data for week calculation questions only. This comparison confirmed a between-group difference in participants' RTs when answering week calculation questions, $x^2(2) = 13.372$, $p = 0.001248 < 0.01$. Such comparisons between full models with reduced models without Group repeatedly proceeded in data for the month calculation questions, hour calculation questions, and year calculation questions. The results also reported contribution of the factor Group to different groups' reaction times in the Month Calculation Task ($x^2(2) = 6.3759$, $p = 0.04126 < 0.05$), but no contribution to RTs in the Hour ($x^2(2) = 4.9412$, $p = 0.08454$) or Year Calculation Task ($x^2(2) = 3.5251$, $p = 0.1716$), though the average speed of Chinese speakers (M = 6898, SD = 5427) were faster than the English speakers (M = 7980, SD = 5677) when tested in their respective native language and the Bilingual group's RTs (M = 8293, M = 5015) were slightly slower than the English group when tested in English. No differences were found in RTs of year calculation questions across the Chinese (M = 3086, SD = 1747), Bilingual (M = 3208, SD = 1729), and English group (M = 3504, SD = 1869). It appeared that pairwise comparisons would be necessary for further investigating whether there were significant differences in participants' response times between each two groups, that is, between English and Chinese, between English and Bilingual, and between Chinese and Bilingual group.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	6677.2	491.8	13.074	$< 0.001^{***}$
Calculation (week)	-2663.1	226.8	-5.074	0.015^{*}
Calculation (month)	-1603.3	546.86	-3.006	0.057
Calculation (year)	-3196.8	526.39	-6.083	0.009^{**}
Group (Chinese)	-525.8	238.38	-2.428	0.14
Group (English)	104.1	225.90	0.152	0.893
Random effects	**Variance**		**SD**	
Participants (intercept)	2904162		1704	
Calculation (week)	1160616		1077	
Calculation (month)	438219		662	
Calculation (year)	1516691		1232	
Item (intercept)	2615964		1616	
Calculation (week)	817147		904	
Calculation (month)	2915964		1708	
Calculation (year)	1345131		1160	
Group (Chinese)	343449		586	
Group (English)	212529		461	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 5. Coefficients from a mixed-effects model fitted to the RTs of English speakers tested in English, Chinese-English bilinguals tested in Chinese, and Chinese-English bilinguals in tested English in the Calendrical Calculation task (Expt. 1)

4.1.2.3 Pairwise Comparisons of RTs in Week and Month Tasks between Groups

Tukey-adjusted pairwise comparisons were run in R to explore more closely how RTs of each group differed from each other in the Week and Month calculation Tasks, since the Factor Group showed no significant contribution to participants' RTs in the Hour and Year calculation Tasks. The results are displayed in Table 6. There was a significant difference between the English group and the Chinese group (different mean = 729.53, lower = 413.37, upper = 1045.69, p < 0.001) and between the Chinese group and the Bilingual group (different mean = -880.41, lower = -1196.57, upper = -564.25, p < 0.001) in the Week Calculation Task. Similarly, there was a significant difference between the English group and the Chinese group (different mean = 757.34, lower = 370.13, upper = 1144.55, p < 0.001) and between the Chinese group and the Bilingual group (different mean = -744.93, lower = -1132.13, upper = -357.71, p < 0.001) in the Month Calculation Task, but there was no difference between the English group and the Bilingual group (different mean = -150.88, lower = -467.04, upper = 165.28, p = 0.502 in week calculations, different mean = 12.41, lower = -374.8, upper = 399.62, p = 0.997 in month calculations).

Overall, these results showed that it took English speakers (week, M = 4829, SD = 3062; month, M = 6064, SD = 3829) longer than Chinese-English bilinguals (week, M = 3758, SD = 2452; month, M = 4529, SD = 3402) to solve the week and month questions when tested in their first language. When tested in the same language - English, the English speakers (week, M = 4829, SD = 3062; month, M = 6064, SD = 3829) and the bilinguals (week, M = 4806, SD = 2756; month, M =5406, SD = 2994) spent similar response times to give correct answers.

Calculation	Group	diff	lwr	upr	p value
Week	English vs. Chinese	729.53	413.37	1045.69	$< 0.001^{***}$
	English vs. Bilingual	-150.88	-467.04	165.28	0.502
	Chinese vs. Bilingual	-880.41	-1196.57	-564.25	$< 0.001^{***}$
Month	English vs. Chinese	757.34	370.13	1144.55	$< 0.001^{***}$
	English vs. Bilingual	12.41	-374.8	399.62	0.997
	Chinese vs. Bilingual	-744.93	-1132.14	-357.71	$< 0.001^{***}$

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 6. Tukey-adjusted pairwise comparisons of participants' RTs in the Week and Month Calculation Tasks between groups (Confidence Level = 95%).

4.1.3 RTs of Each Group in Different Conditions

Additional analysis of the effect of Distance, Direction, Boundary, and Input on each group's RTs in week and month calculation questions was further discussed in this section. The conditions effects would not be investigated in the hour and year calculation questions, since the two tasks were only provided to test whether the general arithmetic abilities of participants from the three groups are in equal level when solving common (10-based questions) and uncommon (24-based questions) calculations. The conditions were all two-level factors, including Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical).

4.1.3.1 Influence of Distance, Direction, Boundary, and Input on Each Group's RTs in Week Calculation Questions

4.1.3.1.1 Visualization of Each Group' RTs in Different Conditions

Figure 6 visualizes each group's reaction times in answering the week calculation questions under different conditions, including Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic, numerical). Based on Table 7, it took the participants more reaction times to process long trials (English, M = 5869, SD = 3411; Chinese, M = 4147, SD = 2657 ; Bilingual, M = 5450, SD = 2760) than short trials (English, M = 3907, SD = 2361; Chinese, M = 3355, SD = 2159; Bilingual, M = 4221, SD = 2152), backward trials (English, M = 5307, SD = 3329; Chinses, M = 4207, SD = 2782; Bilingual, M = 5441, SD = 2765) than forward trials (English, M = 4386, SD = 2720; Chinese, M = 3340, SD = 2009; Bilingual, M = 4218, SD = 2833), and across trials (English, M = 5422, SD = 3241; Chinese, M = 4707, SD = 2593; Bilingual, M = 5982, SD = 2758) than within trials (English, M = 4272, SD =2772; Chinese, M = 2897, SD = 1951; Bilingual, M = 3707, SD = 2750) in all three groups. Input type did not have a significant influence on the Chinese-English bilinguals either tested in Chinese or English, but numerical input negatively affected the English speakers' RTs (M = 5384, SD = 3395).

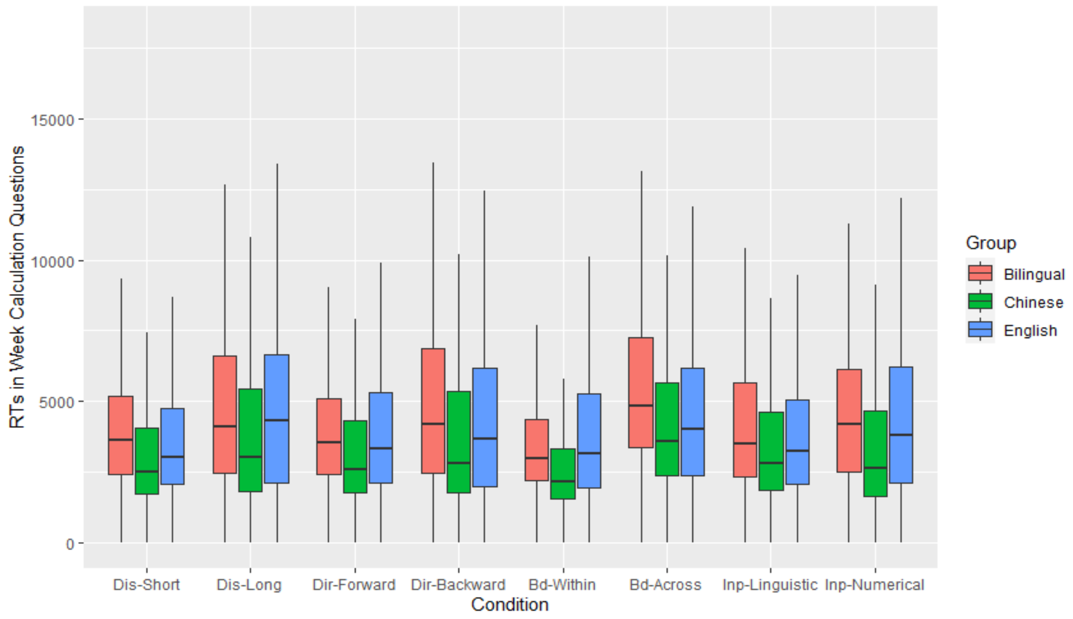

Figure 6: RTs in different conditions by the English, Chinese, and Bilingual Group in Week Calculation task (Experimental 1) (Error Bars = 95% Confidence Interval).

("Dis" for Distance, "Dir" for Direction, "Bd" for Boundary, and "Inp" for Input)

Group / Condition	English	Chinese	Bilingual
Distance - Short	**Mean = 3907** SD = 2361	**Mean = 3355** SD = 2159	**Mean = 4221** SD = 2152
Distance - Long	**Mean = 5869** SD = 3411	**Mean = 4174** SD = 2657	**Mean = 5450** SD = 2760
Direction - Forward	**Mean = 4386** SD = 2720	**Mean = 3340** SD = 2009	**Mean = 4218** SD = 2833
Direction - Backward	**Mean = 5307** SD = 3329	**Mean = 4207** SD = 2782	**Mean = 5441** SD = 2765
Boundary - Within	**Mean = 4272** SD = 2772	**Mean = 2897** SD = 1951	**Mean = 3707** SD = 2750
Boundary - Across	**Mean = 5422** SD = 3241	**Mean = 4707** SD = 2593	**Mean = 5982** SD = 2758
Input - Linguistic	**Mean = 4305** SD = 2606	**Mean = 3780** SD = 2446	**Mean = 4613** SD = 2760
Input - Numerical	**Mean = 5384** SD = 3395	**Mean = 3737** SD = 2458	**Mean = 5001** SD = 2766

Table 7: Summary of the Mean RTs and SDs by the English, Chinese, and Bilingual group in different conditions in the Week Calculation Task.

4.1.3.1.2 Influence of Distance, Direction, Boundary, and Input on the English Group's RTs

Next, the four fixed factors' effect on participants' response times in different conditions within each group were analyzed. Based on the data for the English group, Distance, Direction, Boundary, and Input were entered in the model as fixed effect factors, with Reaction Times as dependent variable and Participants and Items as random effect factors, RTs ~ Distance * Direction * Boundary * Input + (1 | Participant) + (1 | Item). The results are shown in Table 8. Distance, Direction, Boundary, and Input were significantly inter-dependent, $x^2(11) = 22.273$, $p = 0.02234 < 0.05$). In order to explore the effect of the fixed factors on English speakers' RTs, a series of reduced models excluding Distance, Direction, Boundary, and Input were built to be compared respectively with the full model. The results confirmed a significant contribution of Distance ($x^2(8) = 36.059$, $p < 0.001$), Boundary $x^2(8) = 28.281$, $p < 0.001$, and Input ($x^2(8) = 27.499$, $p < 0.001$) to participants' response times in week calculation questions, but no significant contribution of Direction to participants' different reaction times, $x^2(8) = 11.358$, $p = 0.1822$.

Moreover, there was a significant interaction between Distance (short) and Input (numerical) (Estimate = -1870, SE = 819, $p = 0.029 < 0.05$), indicating that the differences between RTs of short trials and long trials were larger in linguistic input than in numerical input.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	4671.57	409.50	11.408	< 0.001***
Distance (short)	-472.58	579.12	-0.816	0.415
Direction (forward)	-318.37	579.12	-0.550	0.583
Boundary (within)	-112.79	579.12	-0.195	0.846
Input (numerical)	1571.63	579.12	2.714	0.007**
Distance (short) × Direction (forward)	-405.04	819.00	-0.495	0.621
Distance (short) × Boundary (within)	-1323.74	819.00	-1.616	0.107
Direction (forward) × Boundary(within)	131.67	819.00	0.161	0.872
Distance (short) × Input (numerical)	-1870.00	819.00	-2.283	0.023*
Direction (forward) × Input (numerical)	-60.91	819.00	-0.074	0.941
Boundary (within) × Input (numerical)	-1161.63	819.00	-1.418	0.156
Distance (short) × Direction (forward) × Boundary (within)	362.49	1158.24	0.313	0.754
Distance (short) × Direction (forward) × Input (numerical)	773.87	1158.24	0.668	0.504
Distance (short) × Boundary (within) × Input (numerical)	2949.22	1158.24	2.546	0.011*
Direction (forward) × Boundary (within) × Input (numerical)	-1310.20	1158.24	-1.131	0.258
Distance (short) × Direction (forward) × Boundary (within) × Input (numerical)	477.47	1638.00	0.291	0.771
Random effects	**Variance**		**SD**	
Participants (intercept)	0.00		0.05	
Item (intercept)	8938128.00		2989.70	

Note. *p < 0.05; **p < 0.01; ***p < 0.001

Table 8. Coefficients from a mixed-effects model fitted to the RTs of the English group in different conditions within the week calculation task.

4.1.3.1.3 Influence of Distance, Direction, Boundary, and Input on the Chinese Group's RTs

Similar statistical steps were taken based on data for the Chinese group, reporting a not inter-dependent result, $x^2(11) = 16.483$, $p = 0.1241$, so "RTs ~ Distance + Direction + Boundary + Input + (1 |Participant) + (1 | Item)" was applied as the full model (Table 9). Reduced models excluding Distance, Direction, Boundary, and Input were compared with the full model respectively. The results confirmed a significant contribution of Distance ($x^2(1) = 11.764$, $p < 0.001$), Direction ($x^2(1) = 9.2952$, $p = 0.002298 < 0.01$), and Boundary ($x^2(1) = 27.114$, $p < 0.001$) to the Chinese group's reaction times in solving the week calculation questions, as well as no significant contribution of Input, $x^2(1) = 0.0427$, $p = 0.8362$.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	4771.87	175.77	27.148	$< 0.001^{***}$
Distance (short)	-656.76	157.22	-4.177	$< 0.001^{***}$
Direction (forward)	-558.81	157.22	-3.554	$< 0.001^{***}$
Boundary (within)	-1335.07	157.22	-8.492	$< 0.001^{***}$
Input (numerical)	-32.43	157.22	-0.206	0.837
Random effects	**Variance**		**SD**	
Participants (intercept)	380414.00		616.78	
Item (intercept)	6385.00		79.91	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 9. Coefficients from a mixed-effects model fitted to the reaction times of the Chinese group in different conditions based on data for the week calculation questions.

4.1.3.1.4 Influence of Distance, Direction, Boundary, and Input on the Bilingual Group's RTs

Analysis of participants' reaction times in different conditions based on data of the Bilingual group's reaction times in week calculation task was the last statistical step. Result of the interaction test showed the four fixed factors were not significantly inter-dependent, $x^2(11) = 10.046$, $p = 0.5262$, so "RTs ~ Distance + Direction + Boundary + Input + (1 |Participant) + (1 | Item)" was chosen as the full model. The results are shown in Table 10. A series of reduced models excluding Distance, Direction, Boundary, and Input were built and compared with the full model. The results confirmed that all fixed factors significantly increased the model fit, Distance ($x^2(1) = 13.725$, $p < 0.001$), Direction ($x^2(1) = 15.75$, $p < 0.001$), Boundary ($x^2(1) = 36.599$, $p < 0.001$) to the Bilingual group's response times in week calculation task, while numerical input had no significant influence on the Bilingual group's performance, $x^2(1) = 3.4581$, $p = 0.06249$.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	5825.0	201.4	28.916	$< 0.001^{***}$
Distance (short)	-701.7	180.2	-3.894	$< 0.001^{***}$
Direction (forward)	-776.8	180.2	-4.311	$< 0.001^{***}$
Boundary (within)	-1771.4	180.2	-9.831	$< 0.001^{***}$
Input (numerical)	321.2	180.2	1.783	0.075
Random effects	**Variance**		**SD**	
Participants (intercept)	631661		794.8	
Item (intercept)	0.00		0.00	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 10. Coefficients from a mixed-effects model fitted to the reaction times of the Bilingual group in different conditions based on data for the week calculation questions.

4.1.3.2 Influence of Distance, Direction, Boundary, and Input on Each Group's RTs in Month Calculation Questions

4.1.3.2.1 Visualization of Each Group in Different Conditions

The four fixed factors' effect on participants' performance in the Month Calculation Task was then explored. The response times of participants from the three groups (English, Chinese, and Bilingual) are visualized in Figure 7 and the mean RTs and SDs are shown in Table 11. The results showed that the RTs of Bilingual group was in-between of the RTs of the Chinese group and RTs of the English group. As for the four fixed factors, Distance (long), Direction (backward), and Boundary (across) appeared to have a negative influence on participants' reaction times in all three groups, while numerical input negatively affected the English group's RTs (M = 6220, SD = 3838 in numerical input trials; M = 5901, SD = 3805 in linguistic input trials) but showed no significant influence on the Chinese-English bilinguals, either tested in Chinese or English. These results were in line with the those of the week calculation questions.

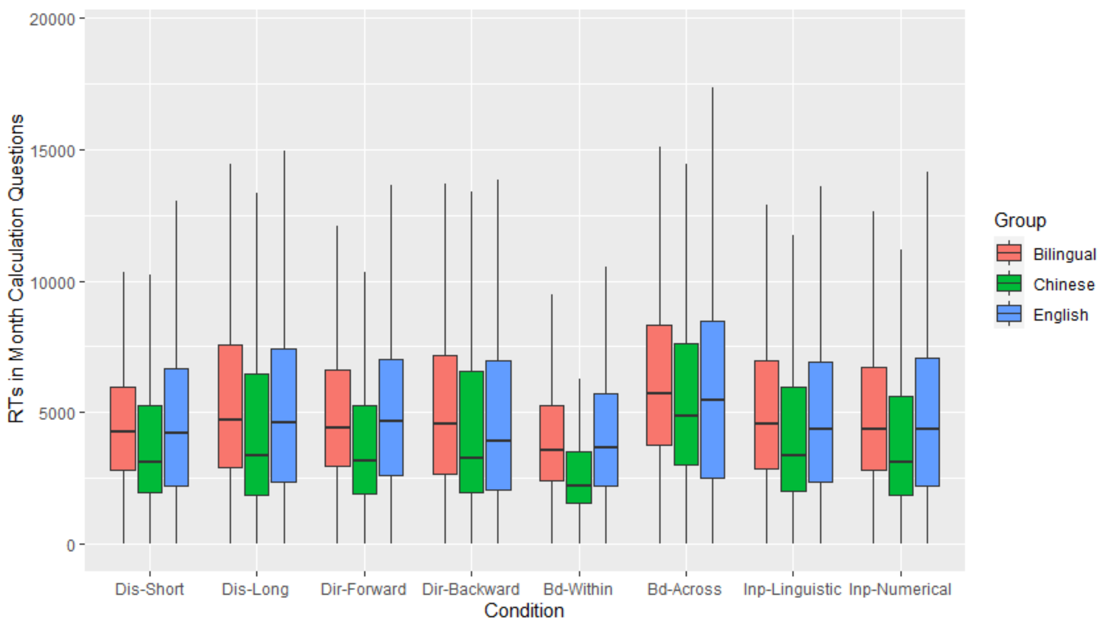

Figure 7: RTs in different conditions by English, Chinese, and Bilingual in Month Calculation task (Experimental 1) (Error Bars = 95% Confidence Interval).

("Dis" for Distance, "Dir" for Direction, "Bd" for Boundary, and "Inp" for Input)

Group / Condition	English Group	Chinese Group	Bilingual Group
Distance-Short	**Mean = 5779** SD = 3817	**Mean = 4166** SD = 3406	**Mean = 4940** SD = 2995
Distance-Long	**Mean = 6352** SD = 3830	**Mean = 4899** SD = 3407	**Mean = 5890** SD = 2995
Direction-Forward	**Mean = 5903** SD = 3819	**Mean = 4176** SD = 3411	**Mean = 5180** SD = 2998
Direction-Backward	**Mean = 6238** SD = 3834	**Mean = 4884** SD = 3384	**Mean = 5640** SD = 2993
Boundary-Within	**Mean = 4926** SD = 3814	**Mean = 3021** SD = 3405	**Mean = 4166** SD = 2996
Boundary-Across	**Mean = 7311** SD = 3828	**Mean = 6123** SD = 3404	**Mean = 6705** SD = 2996
Input-Linguistic	**Mean = 5911** SD = 3805	**Mean = 4572** SD = 3411	**Mean = 5415** SD = 3002
Input-Numerical	**Mean = 6220** SD = 3838	**Mean = 4484** SD = 3371	**Mean = 5396** SD = 2999

Table 11: Summary of the Mean RTs and SD by the English, Chinese, and Bilingual group in different conditions in the Month Calculation Task.

4.1.3.2.2 Influence of Distance, Direction, Boundary, and Input on English Group's RTs

To further explore the nature of the four fixed factors, mixed-fixed regression models in R were built and analyzed based on RTs of the month calculation task within each group, starting from the English group, followed by the Chinese group and the Bilingual group. The mixed-factors models took Distance (long or short), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical) as fixed factors, Participant and Item as the random factors, and RTs as the dependent variable. The analysis of each group's RTs included two statistical steps, including an interaction test and comparisons between the full model and reduced models. To start with the English group, the interaction test was run through comparing model "RTs ~ Distance

* Direction * Boundary * Input + (1 | Participant) + (1 | Item)” and model “Accuracy ~ Distance + Direction + Boundary + Input + (1 |Participant) + (1 | Item)”, resulting in not inter-dependent, $x^2(11) = 18.822$, $p = 0.06437$, so model “Accuracy ~ Distance + Direction + Boundary + Input + (1 |Participant) + (1 | Item)” was applied as the full model. The results are shown in Table 12. Then, the reduced models, excluding Distance, Direction, Boundary, and Input, respectively, were built and compared with the full model. The results confirmed that only Boundary significantly increased the model fit, $x^2(1) = 16.879$, $p < 0.001$, but there was no significant contribution of Distance ($x^2(1) = 2.1257$, $p = 0.1448$), Direction ($x^2(1) = 0.239$, $p = 0.625$), or Input ($x^2(1) = 0.1878$, $p = 0.6648$).

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	5945.20	294.9	20.159	< 0.001***
Distance (short)	-419.20	263.8	-1.589	0.112
Direction (forward)	136.30	263.8	0.517	0.606
Boundary (within)	-1533.60	263.8	-5.814	< 0.001***
Input (numerical)	120.70	263.8	0.457	0.647
Random effects	Variance		SD	
Participants (intercept)	1810577.00		1345.60	
Item (intercept)	62603.00		250.20	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 12. Coefficients from a mixed-effects model fitted to the RTs of the English group in different conditions based on data for the month calculation questions.

4.1.3.2.3 Influence of Distance, Direction, Boundary, and Input on the Chinese Group’s RTs

Similar statistical step was run in R based on reaction times in month calculation questions of the Chinese group. Given that the four fixed factors were not interreacted with each other ($x^2(11) = 15.315$, $p = 0.1685$), “RTs ~ Distance + Direction + Boundary

+ Input + (1 |Participant) + (1 | Item)" was entered in as the full model (Table 13). The results reported that Distance ($x^2(1) = 7.6915$, $p = 0.005548 < 0.01$), Direction ($x^2(1) = 8.3495$, $p = 0.003858 < 0.01$), and Boundary ($x^2(1) = 41.137$, $p < 0.001$) significantly influenced the Chinese group's performance, while there was no significant difference between response times of the Chinese group to answer month calculation questions with numerical input vs. with linguistic input, $x^2(1) = 0.8421$, $p = 0.3588$.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	6415.80	227.10	28.252	< 0.001***
Distance (short)	-607.40	203.10	-2.991	0.003**
Direction (forward)	-640.20	203.10	-3.152	0.002**
Boundary (within)	-2725.10	203.10	-13.416	< 0.001***
Input (numerical)	-179.00	203.10	-0.881	0.378
Random effects	Variance		SD	
Participants (intercept)	944495.00		971.90	
Item (intercept)	0.00		0.00	

Note. *p < 0.05; **p < 0.01; ***p < 0.001

Table 13. Coefficients from a mixed-effects model fitted to the RTs of the Chinese group in different conditions based on data for the month calculation questions.

4.1.3.2.4 Influence of Distance, Direction, Boundary, and Input on Bilingual Group's RTs

When it comes to analysis for the Bilingual group, there was a significant interaction among the four fixed factors, $x^2(11) = 26.667$, $p = 0.005157 < 0.01$, so "RTs ~ Distance * Direction * Boundary * Input + (1 | Participant) + (1 | Item)" was chosen as the full model, as shown in Table 14. The influence of Distance ($x^2(8) = 25.618$, $p = 0.00122 < 0.01$), Direction ($x^2(8) = 24.828$, $p = 0.001662 < 0.01$), and Boundary ($x^2(8) = 52.298$, $p < 0.001$) reached a statistical significance, while there was no significant contribution of Input to the Bilingual group's RTs in month calculation questions, $x^2(8) = 13.334$, p

= 0.1009.

Moreover, there was a significant interaction between Distance and Boundary (Estimate = 2410.3, SE = 720.6, p = 0.004 < 0.005), indicating that the RTs difference between long trials and short trials was higher in within boundary condition than in across boundary condition. To further understand a significant interaction, one efficient way is to condition on the level of one variable and look for the effect of the other. Therefore, RTs of within trials were pull out and the effect of Distance (short or long) were explored by simple effect test. The outcomes reported that the RTs of short trials were significant faster than RTs of long trials in boundary crossing trials (F (1) = 9.818, p = 0.002 < 0.001), but no significant differences between short and long trials in the condition of within boundary (F (1) = 3.628, p = 0.0574). Additionally, there was also a significant interaction between Direction (forward) and Boundary (within) (Estimate = 2850.6, SE = 720.6, p < 0.001), indicating that forward trials were answered faster than the backward trials compared in across boundary than in within boundary. To further explore the interaction, the variable Boundary (within or across) was conditioned and only data from within trials were selected to compare the effect of forward and backward through simple effect test, showing no significant effect of Direction in the within boundary condition, F (1) = 1.738, p = 0.188. Data from across boundary trials were then pull out and similar process was taken, the result of which reported that short trials were answered significantly faster than long trials, F (1) = 5.914, p = 0.015 < 0.05.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	7021.80	391.90	17.917	< 0.001***
Distance (short)	-1542.30	509.50	-3.027	0.004**
Direction (forward)	-916.40	509.50	-1.799	0.088
Boundary (within)	-3705.50	509.50	-7.272	< 0.001***
Input (numerical)	375.90	509.50	0.738	0.484
Distance (short) × Direction (forward)	926.90	720.60	1.286	0.222
Distance (short) × Boundary (within)	2410.30	720.60	3.345	0.002**
Direction (forward) × Boundary(within)	2850.60	720.60	3.956	<0.001***
Distance (short) × Input (numerical)	475.30	720.60	0.66	0.531
Direction (forward) × Input (numerical)	-326.60	720.60	-0.453	0.667
Boundary (within) × Input (numerical)	508.10	720.60	0.705	0.503
Distance (short) × Direction (forward) × Boundary (within)	-3187.50	1019.00	-3.218	0.003**
Distance (short) × Direction (forward) × Input (numerical)	-711.80	1019.00	-0.698	0.507
Distance (short) × Boundary (within) × Input (numerical)	-1740.70	1019.00	-1.708	0.105
Direction (forward) × Boundary (within) × Input (numerical)	-1768.80	1019.00	-1.736	0.1
Distance (short) × Direction (forward) × Boundary (within) × Input (numerical)	2753.80	1441.10	1.911	0.07
Random effects	**Variance**		**SD**	
Participants (intercept)	713486.00		844.70	
Item (intercept)	0.00		0.00	

Table 14. Coefficients from a mixed-effects model fitted to the RTs of the Bilingual group in different conditions based on data for the month calculation questions.

4.2 Accuracy

4.2.1 Preliminary Analyses

All participants achieved ≥ 70 % accuracy in the Calendrical Calculation Task (Expt. 1), so no result was excluded from the analysis. The correct answers were coded as 1, while the wrong answers were coded as 0. Participants' accuracy was represented by percentages, so all codes multiplied 100. Linear mixed effects models were used to do the analysis to take both fixed effect factors and random effect factors into consideration (Winter, 2014). The main fixed effect factors were Calculation (Week, Month, Hour, Year) and Group (English speakers tested in English, Chinese-English bilinguals tested in Chinese, and Chinese-English bilinguals tested in English). The effect of Distance (long or short), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical) were also taken into consideration and analyzed as within-group factors. The outcome variables were Answer (right or wrong), and the random effect factors were Participant and Item.

4.2.2 Accuracy of Different Groups in Calendrical Calculation Task

The analysis of different groups' accuracy in the Calendrical Calculation Task would involve four steps. Firstly, different groups' accuracy in the Week, Month, Hour, and Year Calculation Task would be visualized. Secondly, the effect of factor Calculation and Group and their potential interaction would be explored. Thirdly, pairwise comparisons of accuracy in Week and Month Calculation Tasks were taken between every two groups (English group vs. Chinese group, English group vs. Bilingual group, and Chinese group vs. Bilingual group). No pairwise comparisons would be processed based on data from Hour and Year Calculation Task, as there was no significant difference across groups' RTs in the two calculation tasks (as shown in 4.2.2.2). At last, the effects of Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical) on participants' achievement of accuracy would be additionally analyzed as within-group factors. The different conditions' effects would not be explored in the Hour and Year Calculation Tasks since they were only used to test the general arithmetic abilities in common (10-based calculations) and uncommon (24-based calculations) mathematic questions.

4.2.2.1 Visualization of Different Groups' Accuracy in Calendrical Calculation Task

This part examined the effect of the two main fixed factors on participants' accuracy in the Calendrical Calculation Task. Figure 8 shows the proportion of correct answers achieved by group English, Chinese, and Bilingual in different calculation tasks. In the Week calculation task, the English group (M = 90.1%, SD = 29.86%) was less accurate than the Bilingual group (M = 93.33%, SD = 24.94%), while the Chinese group (M = 96.15%, SD = 19.25%) achieved the highest accuracy. The differences in participants' accuracy across groups in the Month calculation task were similarly to that in the Week calculation, though the mean accuracy gap between the English group (M = 86.35%, SD = 34.33%) and the Chinese group (M = 97.81%, SD = 14.81%) was larger. Moreover, the English group achieved higher accuracy in the week calculation questions (M = 90.1%, SD = 29.86%) than in the month calculation questions (M = 86.35%, SD = 34.33%), while the Bilingual group and the Chinese group performed similarly during week (Bilingual, M = 93.33%, SD = 24.94; Chinese, M = 96.15%, SD = 19.25%) and month calculations (Bilingual, M = 95.31%, SD = 21.14%; Chinese, M = 97.81%, SD = 14.81%). In the Hour and Year calculation task, accuracy tended to be at ceiling level across English group (M = 87.71%, SD = 32.83% in hour questions, M = 95.42%, SD = 20.91% in year questions), Bilingual group (M = 89.58%, SD = 30.55% in hour questions, M = 98.13%, SD = 13.56% in year questions), and Chinese group (M = 92.29%, SD = 26.67% in hour questions, M = 96.46%, SD = 18.48% in year questions). It was also noted that all participants achieved higher accuracy in the Year task (10-based calculations) (English, M = 95.42%, SD = 20.91%; Bilingual, M = 98.13%, SD = 13.56%; Chinese, M = 96.46%, SD = 18.48%) than in the Hour task (24-

based calculations) (English, M = 87.71%, SD = 32.83%; Bilingual, M = 89.58%, SD = 30.55%; Chinese, M = 92.29%, SD = 26.67%).

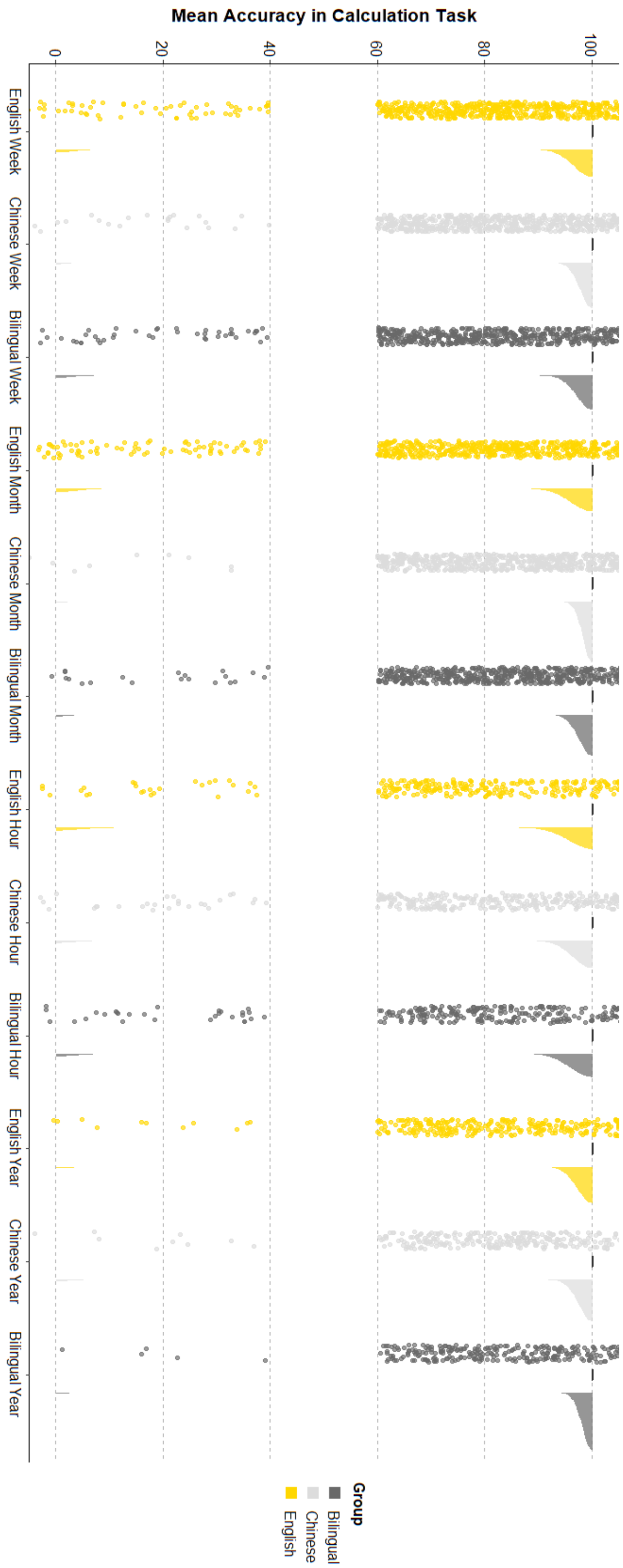

Figure 8: Mean Accuracy by English speakers in English (English), Chinese-English bilinguals in Chinese (Chinese), and Chinese-English bilinguals in English (Bilingual) in Calculation task (Experimental 1) (Error Bars = 95% Confidence Interval).

4.2.2.2 Effects of Factor Calculation and Group and the Interaction between Calculation and Group

To further test the effect of Factor Group on participants' accuracy of responses to calendrical calculation questions, the mixed-effects regression models in R was built. Calculation (week, month, hour, or year) and Group (English speakers tested in English, Chinese-English bilinguals tested in Chinese, or Chinese-bilinguals tested in English) were specified as fixed effect factors. The binary dependent variable was Answer (right or wrong), and the random effect factors were Participant and Item. The model included a maximal random effects structure, including random slopes over calculation by participant and random slopes over calculation, group, and their interaction by item as follows:

Accuracy ~ Calculation * Group +

1+ Calculation | Participant) +

(1 + Calculation * Group | Item)

The results confirmed significant contribution of factor Calculation and Group and confirmed that Calculation and Group were significantly inter-dependent on each other, as shown in Table 15. The material is shown in Appendix 6, and the full dataset with responses for each participant in each calculation task can be found in https://doi.org/10.17605/OSF.IO/HJZRG.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	95.31	0.80	118.664	< 0.001***
Type (week)	-1.98	1.14	-1.742	0.081
Type (hour)	-5.98	1.39	-4.118	< 0.001***
Type (year)	2.81	1.39	3.45	0.043*
Group (Chinese)	2.5	1.14	4.94	0.028

Group (Bilingual)	-8.96	1.14	4.38	< 0.001***
Type (week) × Group (Chinese)	0.31	1.61	-1.97	0.84
Type (hour) × Group (Chinese)	0.21	1.97	-2.09	0.92
Type (year) × Group (Chinese)	-4.16	1.97	-3.99	0.034*
Type (week) × Group (English)	5.73	1.61	-2.74	< 0.001***
Type (hour) × Group (English)	7.08	1.97	2.53	< 0.001***
Type (year) × Group (English)	6.25	1.97	-2.57	0.015**

Random effects	Variance	SD
Participants (intercept)	26.32	5.13
Type (week)	8.06	2.84
Type (hour)	10.76	3.28
Type (year)	12.97	3.60
Item (intercept)	3.09	1.76
Type (week)	49.10	7.01
Type (hour)	33.27	5.77
Type (year)	9.94	3.15
Group (Chinese)	2.13	1.77
Group (English)	20.07	4.48
Type (week) × Group (Chinese)	17.75	4.21
Type (hour) × Group (Chinese)	2.10	1.45
Type (year) × Group (Chinese)	3.27	1.81
Type (week) × Group (English)	35.38	5.95
Type (hour) × Group (English)	18.80	4.34
Type (year) × Group (English)	41.28	6.43

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 15. Coefficients from a mixed-effects model fitted to the accuracy of English speakers tested in English, Chinese-English bilinguals tested in Chinese, and Chinese-English bilinguals tested in English in the Calendrical Calculation task (Expt. 1)

The full model returned two significant two-way interactions, namely between Calculation and Group for English versus Chinese, and also between Calculation and Group for English versus Bilingual. To further explore the nature of the interactions, a reduced model excluding Calculation was built. A comparison of the reduced model with the full model showed that the presence of Calculation increased the model fit, $x^2(9) = 25.955$, $p = 0.002078 < 0.01$, confirming that participants answered week, month, hour, and year calculation questions differently. Then, the analysis proceeded with a forward variable selection and zoomed in on each type of calculation questions, comparing a model including Group with a reduced model without Group in the data for the week calculation questions only. This comparison indicated the contribution of Group to the variation in accuracy of week calculation questions, $x^2(2) = 6.7889$, $p = 0.03356 < 0.05$. Such comparisons between full models with reduced models without Group repeatedly proceeded in data for the month calculation questions, hour calculation questions, and year calculation questions. The results confirmed a significant contribution of the factor Group to the different degrees of accuracy achieved in month calculation questions, $x^2(2) = 13.641$, $p = 0.001091 < 0.01$. On the contrary, comparisons based on data from hour questions and year questions presented no significant difference across the English group, Chinese group, and Bilingual group, $x^2(2) = 2.3258$, $p = 0.3126$ for the hour calculation task, and $x^2(2) = 2.7271$, $p = 0.2558$ for the year calculation task.

4.2.2.3 Pairwise Comparisons of Accuracy in Week and Month Tasks between Groups

As the final statistical step, the Tukey-adjusted pairwise comparisons between groups were run (Table 16) to examine how Accuracy of week and month calculations differ between groups. In the Week calculation task, there was a significant difference between the English group and the Chinese group (mean difference = -6.042, lower = -8.726, upper = -3.358, $p < 0.001$) and between the English group and the Bilingual group (mean difference = -3.323, lower = -5.913, upper = -0.545, $p = 0.013 < 0.05$) as well as between the Chinese group and the Bilingual group (mean difference = 2.813, lower = 0.129, upper = 5.496, $p = 0.037 < 0.05$). In the Month calculation task, there was a significant difference between the English group and the Chinese group (mean difference = -11.458, lower = -14.11, upper = -6.307, $p < 0.001$) and between the English group and the Bilingual group (mean difference = -8.958, lower = -11.61, upper = -6.307, $p < 0.001$), but there was no difference between the Chinese group and the Bilingual group (mean difference = 2.5, lower = -0.151, upper = 5.151, $p = 0.069$). These results indicated that the Bilingual group's accuracy in week and month calculation tasks was in-between those of the English group and the Chinese group controls. However, such result was not in line with the RTs result, where the English group differed from the Chinese group, but spent similar RTs to the Bilingual group.

Calculation	Group	diff	lwr	upr	p
Week	English vs. Chinese	-6.042	-8.726	-3.358	$< 0.001^{***}$
	English vs. Bilingual	-3.323	-5.913	-0.545	0.013^{*}
	Chinese vs. Bilingual	2.813	0.129	5.496	0.037^{*}
Month	English vs. Chinese	-11.458	-14.11	-6.307	$< 0.001^{***}$
	English vs. Bilingual	-8.958	-11.61	-6.307	$< 0.001^{***}$
	Chinese vs. Bilingual	2.5	-0.151	5.151	0.069

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 16. Tukey-adjusted pairwise comparisons of accuracy in week and month calculation questions between groups (Expt. 1) (Confidence Level = 95%).

4.2.3 Accuracy of Each Group in Different Conditions

Additional analysis of different conditions' effects was further discussed based on data for the week, month, hour, and year calculation questions in this part. The conditions effects would not be investigated in the hour and year calculation questions, since the two tasks were only provided to test whether the general arithmetic abilities of participants from the three groups were in equal level when solving common (10-based questions) and uncommon (24-based questions) calculations. The conditions were all two-level factors, including Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical).

4.2.3.1 Influence of Distance, Direction, Boundary, and Input on Each Group's Accuracy in Week Calculation Questions

This section focused on each group's accuracy in different conditions based on data for week questions. After an analysis of the visualization of different groups' accuracy scores in different conditions, the mixed-factors regression models would be applied to examine the influence of Distance, Direction, Boundary, and Input on each group's performance in the Week Calculation Task.

4.2.3.1.1 Visualization of Each Group's Accuracy in Different Conditions

Figure 9 shows the proportion of correct answers to week calculation questions in different conditions answered by English speakers in English (the English group), Chinese-English bilinguals in Chinese (the Chinese group), and Chinese-English bilinguals in English (the Bilingual group). The English group was descriptively less accurate in all conditions than the Bilingual group, while the Chinese group achieved the highest accuracy, as shown in Table 17. Moreover, Table 17 also displayed that short trials (English, M = 94.38%, SD = 29.92%; Chinese, M = 96.25%, SD = 19.29%; Bilingual, M = 96.04%, SD = 24.99%) were more accurate than long trials (English, M = 85.83%, SD = 29.92%, Chinese, M = 96.04%, SD = 19.29%; Bilingual, M= 90.63%, SD = 24.99%), forward trials (English, M = 93.13%, SD = 29.97%; Chinese, M = 98.33%, SD = 19.33%; Bilingual, M = 95.83%, SD = 25.04%) more accurate than backward trials (English, M = 87.08 %, SD = 29.97%; Chinese, M = 93.96%, SD = 19.33%; Bilingual, M = 90.83%, SD = 25.04%), and within trials (English, M = 92.5%, SD = 29.89%; Chinese, M = 99.17%, SD = 19.27%; Bilingual, M = 95.42%, SD = 24.97%) more accurate than across trials (English, M = 87.71%, SD = 29.89%; Chinese, M = 93.13%, SD = 19.27%; Bilingual, M = 91.25%, SD = 24.97%) in all three groups. One exception occurred for the fixed factor “Input”. The English group was slightly more accurate with linguistic input (M = 91.46%, SD = 30.08%) than with numerical input (M = 88.75%, SD = 30.08%), while the Chinese-English bilinguals performed similarly with linguistic and numerical input, both in the Chinese version test (M = 96.25 % & 96.04%, SD = 19.41% & 19.15%, respectively) and in the English version test (M = 92.71% & 93.96%, SD = 25.14% & 24.77%, respectively), which was in line with the RTs results in week calculation questions.

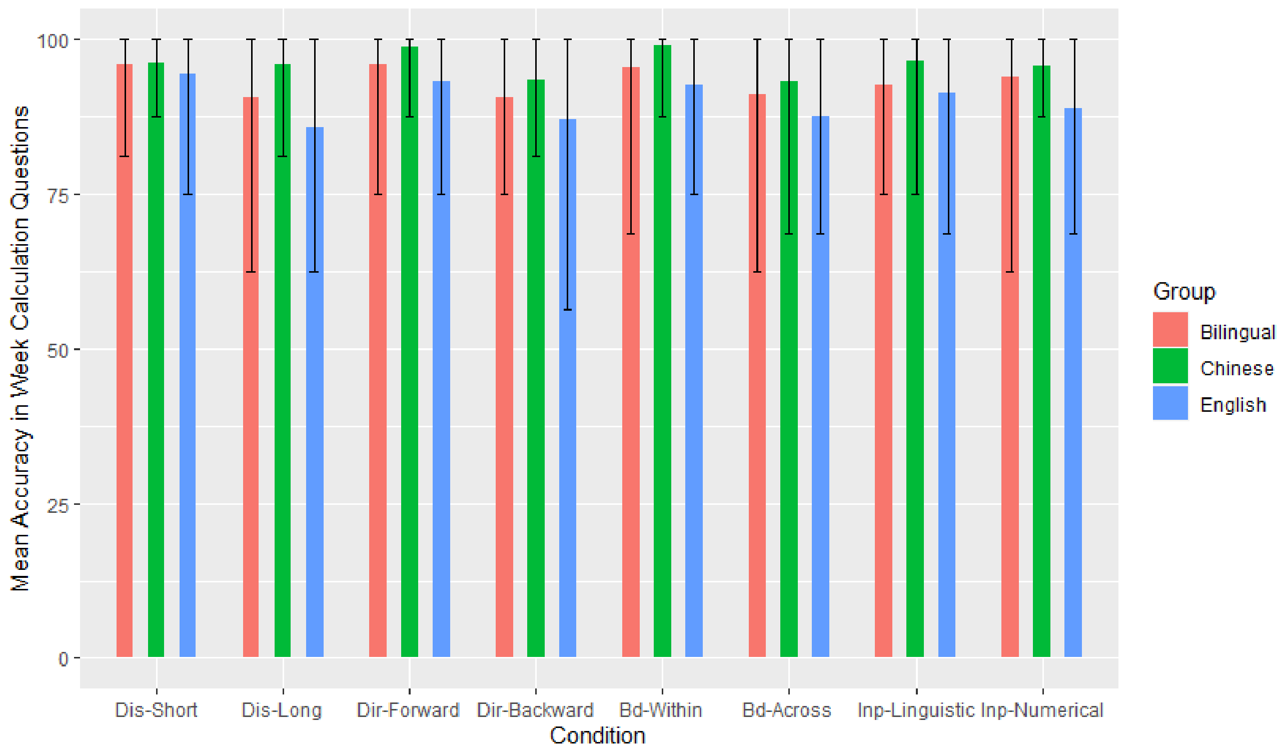

Figure 9: Mean Accuracy of responses to questions in different conditions by the English, Chinese, and Bilingual in the Week Calculation task (Experimental 1) (Error Bars = 95% Confidence Interval).

(“Dis” for Distance, “Dir” for Direction, “Bd” for Boundary, and “Inp” for Input)

Condition \ Group	English	Chinese	Bilingual
Distance-Short	**Mean = 94.38%** SD = 29.92%	**Mean = 96.25%** SD = 19.29%	**Mean = 96.04%** SD = 24.99%
Distance-Long	**Mean = 85.83%** SD = 29.92%	**Mean = 96.04%** SD = 19.29%	**Mean = 90.63%** SD = 24.99%
Direction-Forward	**Mean = 93.13%** SD = 29.97%	**Mean = 93.96%** SD = 19.33%	**Mean = 95.83%** SD = 25.04%
Direction-Backward	**Mean = 87.08%** SD = 29.97%	**Mean = 93.96%** SD = 19.33%	**Mean = 90.83%** SD = 25.04%
Boundary-Within	**Mean = 92.5%** SD = 29.89%	**Mean = 99.17%** SD = 19.27%	**Mean = 95.42%** SD = 24.97%
Boundary-Across	**Mean = 87.71%** SD = 29.89%	**Mean = 93.13%** SD = 19.27%	**Mean = 91.25%** SD = 24.97%
Input-Linguistic	**Mean = 91.46%** SD = 30.08%	**Mean = 96.25%** SD = 19.41%	**Mean = 92.71%** SD = 25.14%
Input-Numerical	**Mean = 88.75%** SD = 30.08%	**Mean = 96.04%** SD = 19.15%	**Mean = 93.96%** SD = 24.77%

Table 17: Summary of the Mean Accuracy and SD by the English, Chinese, and Bilingual group in different conditions in the Week Calculation Task.

4.2.3.1.2 Influence of Distance, Direction, Boundary, and Input on the English Group's Accuracy

The following statistical step focused on the influence of the four fixed factors on participants' accuracies in different conditions within each group. Starting with the English group, the mixed-effects regression models were built based on the data for the English group. Distance, Direction, Boundary, and Input were entered as fixed effect factors, while Answer (right or wrong) was entered as the binary dependent variable. As random effect factors, intercepts for Participant and Item have been taken into consideration (Barr et al., 2013) as follows:

Accuracy ~ Distance * Direction * Boundary * Input + (1 | Participant) + (1 | Item)

The model was then compared with the model below to test if the fixed factors interact or not, confirming that Distance, Direction, Boundary, and Input did significantly interact, $x^2(11) = 23.829$, $p = 0.01348 < 0.05$). The results are shown in Table 18.

Accuracy ~ Distance + Direction + Boundary + Input + (1 |Participant) + (1 | Item)

To further explore the nature of the interactions between fixed factors, a reduced model excluding Distance was built and compared with the full model. The result, $x^2(8) = 28.864$, $p < 0.001$, reported a significant contribution of Distance to the English group's accuracy in week calculation questions. Similar comparisons between the full model and reduced models, including reduced models without Direction, Boundary, and Input, were run, respectively. The results confirmed that fixed factors Direction ($x^2(8) = 24.519$, $p = 0.001874 < 0.001$) significantly increased the model fit. Boundary ($x^2(8) = 19.332$, $p = 0.01318 < 0.05$) and Input ($x^2(8) = 18.672$, $p = 0.01672 < 0.05$) also significantly contributed to the variation in English group's accuracy.

Furthermore, there was also a significant interaction between Distance and Direction (Estimate = -18.33, SE = 7.28, $p = 0.015 < 0.05$), indicating that accuracy of short trials was higher than long trials in the forward condition than in the backward condition. To further explore the nature of the interactions between factors Distance and Direction, the analysis controlled the variable Direction and zoomed in the forward condition only first (the backward condition later) and investigated the effect of factor Distance (short or long) on the accuracy of the English group. The results showed that there were significant differences between short trials and long trials in both forward condition ($F(1) = 5.54$, $p = 0.019 < 0.05$) and backward condition ($F(1) = 14.91$, $p < 0.001$).

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	75.00	3.82	19.641	$< 0.001^{***}$
Distance (short)	20.00	5.15	3.887	$< 0.001^{***}$
Direction (forward)	21.67	5.15	4.211	$< 0.001^{***}$
Boundary (within)	5.00	5.15	0.972	0.347
Input (numerical)	1.67	5.15	0.324	0.754
Distance (short) × Direction (forward)	-18.33	7.28	2.52	0.015^{*}
Distance (short) × Boundary (within)	-1.67	7.28	0.229	0.825
Direction (forward) × Boundary (within)	-11.67	7.28	1.603	0.121
Distance (short) × Input (numerical)	-11.67	7.28	1.603	0.121
Direction (forward) × Input (numerical)	-11.67	7.28	1.603	0.121
Boundary (within) × Input (numerical)	11.67	7.28	1.603	0.121
Distance (short) × Direction (forward) × Boundary (within)	8.33	10.29	0.81	0.434
Distance (short) × Direction (forward) × Input (numerical)	11.67	10.29	1.134	0.273
Distance (short) × Boundary (within) × Input (numerical)	-6.67	10.29	0.648	0.531
Direction (forward) × Boundary (within) × Input (numerical)	-3.33	10.29	0.324	0.754
Distance (short) × Direction (forward) × Boundary (within) × Input (numerical)	8.333	14.55	0.573	0.580
Random effects	**Variance**		**SD**	
Participants (intercept)	40.34		6.35	
Item (intercept)	0.00		0.00	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 18. Coefficients from a mixed-effects model fitted to the English group's accuracy in different conditions based on data for the week calculation questions.

4.2.3.1.3 Influence of Distance, Direction, Boundary, and Input on the Chinese Group's Accuracy

The analysis then zoomed in on the effect of the four fixed factors on the Chinese group's accuracy in week calculation questions. The full model had the same fixed factors, outcome variable, and random factors as those in the English group's model, and the four fixed factors were inter-dependent, $x^2(11) = 18.501$, $p = 0.07067$, so the full model for this step was "Accuracy ~ Distance + Direction + Boundary + Input + (1 | Participant) + (1 | Item)". The results are shown in Table 19. Furthermore, a series of reduced models one by one excluding Distance, Direction, Boundary, and Input were built and compared with the full model. The results showed that Distance did not affect the Chinese group's accuracy in week calculation questions, $x^2(1) = 0.0251$, $p = 0.8741$. The results also reported that fixed factors Input ($x^2(1) = 0.0251$, $p = 0.8741$) had no effects on the accuracy either, but confirmed the contribution of Direction ($x^2(1) = 8.4072$, $p = 0.003737 < 0.01$) and Boundary ($x^2(1) = 13.448$, $p < 0.001$) to Chinese group's accuracy in week calculation questions.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	90.94	1.52	59.82	$< 0.001^{***}$
Distance (short)	0.2	1.31	0.158	0.865
Direction (forward)	4.38	1.31	3.328	$< 0.001^{***}$
Boundary (within)	6.04	1.31	4.595	$< 0.001^{***}$
Input (numerical)	-0.21	1.31	-0.158	0.865
Random effects	**Variance**		**SD**	
Participants (intercept)	4.506		2.123	
Item (intercept)	1.063		1.031	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 19. Coefficients from a mixed-effects model fitted to the Chinese group's accuracy in different conditions based on data for the week calculation questions.

4.2.3.1.4 Influence of Distance, Direction, Boundary, and Input on the Bilingual Group's Accuracy

Next, the four fixed factors' influence on the accuracy of the Bilingual group in different conditions were analyzed based on data for week calculation questions. By typing Distance, Direction, Boundary, and Input as fixed effect factors, Answer (correct or wrong) as the binary dependent variable, and Participants and Items as random effect factors into the mixed-effects regression model, the interactions among the four fixed factors were verified ($x^2(11) = 24.463$, $p = 0.01092 < 0.05$). Therefore, the model "Accuracy ~ Distance * Direction * Boundary * Input + (1 | Participant) + (1 | Item)" was the full model for the Bilingual groups' data. The results are displayed in Table 20.

To further explore how the fixed factors affect the Bilingual group's accuracy in week calculation questions, reduced models excluding Distance, Direction, Boundary, and Input were built and compared with the full model, respectively. The results showed that Distance ($x^2(8) = 19.795$, $p = 0.01114 < 0.05$) increased the model fit, and there was also a significant contribution of Direction ($x^2(8) = 28.324$, $p < 0.001$) and Boundary ($x^2(8) = 26.402$, $p < 0.001$) to the variation in Bilingual group's accuracy, though Input ($x^2(8) = 4.116$, $p = 0.8465$) turned out to have no effects on the accuracy of Bilingual group.

Furthermore, it is also worth to note that there was an interaction effect between Direction and Boundary (Estimate = -10, SE = 6.14, $t = -3.257$, $p = 0.002 < 0.01$). With the simple effect tests, the nature of the interaction was further explored though pulling out the accuracy data on condition of within boundary, finding that the factor Direction (forward or backward) had no effect on participants' accuracy ($F(1) = 0.76$, $p = 0.384$). Similar test was proceed based on data from the condition of across boundary, where accuracy was found negatively influenced by backward trials when crossing boundary ($F(1) = 21.28$, $p < 0.001$). It appeared that directionality only affected the Bilingual group's RTs in the across condition but not in the within condition.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	80	3.19	25.103	< 0.001***
Distance (short)	8.33	4.34	1.919	0.062
Direction (forward)	18.33	4.34	4.222	< 0.001***
Boundary (within)	11.67	4.34	2.687	0.009**
Input (numerical)	0.00	4.34	0	1
Distance (short) × Direction (forward)	-10.00	6.14	-1.628	0.113
Distance (short) × Boundary (within)	0.00	6.14	0	1
Direction (forward) × Boundary (within)	-20.00	6.14	-3.257	0.002**
Distance (short) × Input (numerical)	5.00	6.14	0.814	0.428
Direction (forward) × Input (numerical)	-3.33	6.14	-0.543	0.598
Boundary (within) × Input (numerical)	3.33	6.14	0.543	0.598
Distance (short) × Direction (forward) × Boundary (within)	8.33	8.68	0.96	0.351
Distance (short) × Direction (forward) × Input (numerical)	0.00	8.68	0	1
Distance (short) × Boundary (within) × Input (numerical)	-10.00	8.68	-1.151	0.263
Direction (forward) × Boundary (within) × Input (numerical)	5.00	8.68	0.576	0.575
Distance (short) × Direction (forward) × Boundary (within) × Input (numerical)	0.00	12.28	0	1
Random effects	**Variance**		**SD**	
Participants (intercept)	21.86		4.68	
Item (intercept)	0.00		0.00	

Note. *p < 0.05; **p < 0.01; ***p < 0.001

Table 20. Coefficients from a mixed-effects model fitted the Bilingual group's accuracy in different conditions based on data for the week calculation questions.

4.2.3.2 Influence of Distance, Direction, Boundary, and Input on Each Group's Accuracy in Month Calculation Questions

This Section focused on each group's accuracy in different conditions based on data for month questions. After an analysis of the visualization of different groups' accuracy scores in different conditions, the mixed-factors regression models would be applied to examine the influence of Distance, Direction, Boundary, and Input on the accuracy scores of the English, Bilingual, and Chinese group in month calculation task.

4.2.3.2.1 Visualization of Each Group's Accuracy in Different Conditions

Figure 10 shows the accuracy of the English, Chinese, and Bilingual group in different conditions based on data for the month calculation questions. The English group achieved the lowest accuracy, while the Chinese group achieved slightly higher accuracy than the Bilingual group in all conditions. Additionally, Table 21 reported that forward trials (English, M = 89.17%, SD = 34.45%; Chinese, M = 97.71%, SD =14.69%; Bilingual, M = 96.46%, SD = 21.22%) were more accurate than backward trials (English, M = 83.54%, SD = 34.35%; Chinese, M = 97.92%, SD = 14.64%; Bilingual, M = 94.17%, SD = 21.22%) and boundary within trials (English, M = 89.38%, SD = 34.36%; Chinese, M = 98.96%, SD = 14.64%; Bilingual, M =96.67%, SD = 21.16%) were more accurate than boundary crossing trials (English, M = 83.33%, SD = 34.36%; Chinese, M = 96.67%, SD = 14.64%; Bilingual, M = 93.96%, SD = 21.16%) in all three groups, who had similar accuracy in short trials (English, M = 86.46%, SD = 34.39%; Chinese, M = 98.13%, SD = 14.66%; Bilingual, M = 96.25%, SD = 21.18%) and long trials (English, M = 86.25%, SD = 34.39%; Chinese, M = 97.5%, SD = 14.66%; Bilingual, M = 94.38%, SD = 21.18%). Input tended to have no effect on the three groups as their achievements appeared to be at ceiling level (English, M = 85.63% & 87.08%, SD = 34.57% & 34.67%; Chinese, M = 98.13% & 97.5%, SD = 14.75% & 14.4%; Bilingual, M = 95.21% & 95.42%, SD = 21.31% & 21.31%).

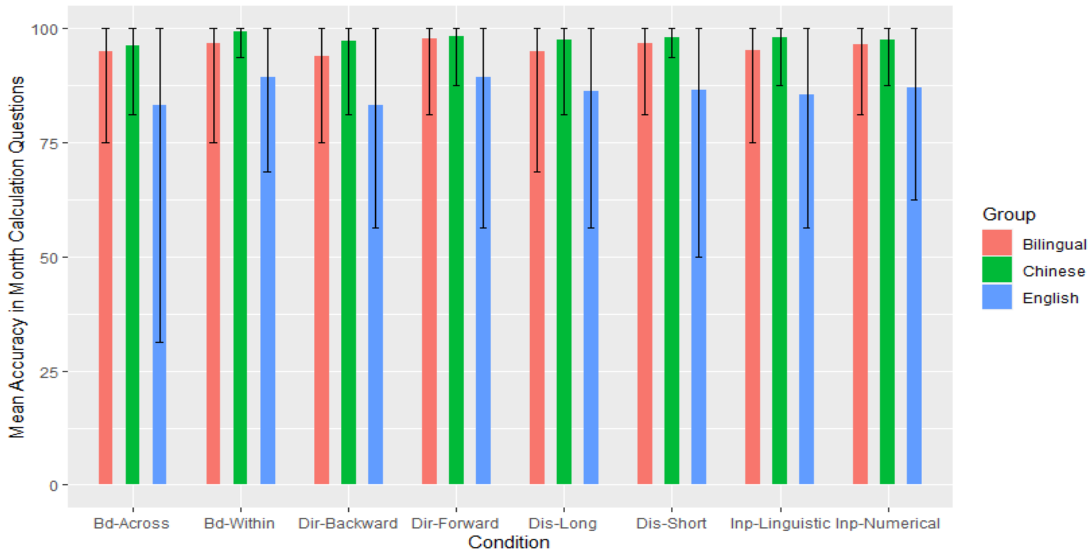

Figure 10: Mean Accuracy of responses to questions in different conditions by the English, Chinese, and Bilingual in Month Calculation task (Experimental 1) (Error Bars = 95% Confidence Interval).

("Dis" for Distance, "Dir" for Direction, "Bd" for Boundary, and "Inp" for Input)

Group / Condition	English Group	Chinese Group	Bilingual Group
Distance-Short	**Mean = 86.46%** SD = 34.39%	**Mean = 98.13%** SD = 14.66%	**Mean = 96.25%** SD = 21.18%
Distance-Long	**Mean = 86.25%** SD = 34.39%	**Mean = 97.5%** SD = 14.66%	**Mean = 94.38%** SD = 21.18%
Direction-Forward	**Mean = 89.17%** SD = 34.45%	**Mean = 97.71%** SD = 14.69%	**Mean = 96.46%** SD = 21.22%
Direction-Backward	**Mean = 83.54%** SD = 34.35%	**Mean = 97.92%** SD = 14.69%	**Mean = 94.17%** SD = 21.22%
Boundary-Within	**Mean = 89.38%** SD = 34.36%	**Mean = 98.96%** SD = 14.64%	**Mean = 96.67%** SD = 21.16%
Boundary-Across	**Mean = 83.33%** SD = 34.36%	**Mean = 96.67%** SD = 14.64%	**Mean = 93.96%** SD = 21.16%
Input-Linguistic	**Mean = 85.63%** SD = 34.57%	**Mean = 98.13%** SD = 14.75%	**Mean = 95.21%** SD = 21.31%
Input-Numerical	**Mean = 87.08%** SD = 34.67%	**Mean = 97.5%** SD = 14.4%	**Mean = 95.42%** SD = 21.31%

Table 21: Summary of the Mean Accuracy and SD by the English, Chinese, and Bilingual group in different conditions in the Month Calculation Task.

4.2.3.2.2 Influence of Distance, Direction, Boundary, and Input on the English Group's Accuracy

The subsequent statistical step explored how the fixed factors Distance, Direction, Boundary, and Input affected each group in different conditions. Starting from the English group, the mixed-effects regression models with fixed factors (Distance, Direction, Boundary, and Input), dependent variable (Answer), and random effect factors (Participant and Item) were built. The interactions among fixed factors were then tested through comparing the full model including interactions with a model without interactions, confirming that Distance, Direction, Boundary, and Input significantly interacted with each other, $x^2(11) = 23.975$, $p = 0.01284 < 0.05$). The results are shown in Table 22.

Accuracy ~ Distance * Direction * Boundary * Input +

(1 | Participant) + (1 | Item)

To further explore the nature of the interactions between the fixed factors, a reduced model excluding Distance was built and compared with the full model, $x^2(8) = 19.368$, $p = 0.01301 < 0.05$, reporting a significant contribution of Distance to the English group's accuracy in week calculation questions. Similar comparisons between the full and reduced models, including reduced models without Direction, Boundary, and Input, followed, respectively. The results confirmed that fixed factors Direction ($x^2(8) = 24.954$, $p = 0.001583 < 0.01$) and Boundary ($x^2(8) = 21.389$, $p = 0.006183 < 0.01$) significantly increased the model fit, but no contribution of Input ($x^2(8) = 18.672$, $p = 0.5948$) to English group's accuracy in month calculation questions.

Additionally, there was a significant interaction between factor Distance and Direction (Estimate = 25, SE = 8.43, t = 2.97, p = 0.004 < 0.01). Simple effect tests were processed respectively through pulling out data on condition of Direction forward and Direction backward, finding that there were significant differences between the accuracy of short trials and long trials in both forward condition (F(1) = 7.061, p = 0.00814 < 0.01) and backward condition (F(1) = 4.401, p = 0.0364 < 0.05), which were in line with the results of the accuracy achievement by the English group in the Week Calculation Task (as shown in 4.2.3.1.2).

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	91.67	4.47	20.53	< 0.001***
Distance (short)	-15.00	5.96	-2.52	0.016**
Direction (forward)	-12.67	5.96	-2.00	0.05*
Boundary (within)	-11.67	5.96	-1.96	0.06
Input (numerical)	-5.00	5.96	-0.84	0.42
Distance (short) × Direction (forward)	25.00	8.43	2.97	0.004**
Distance (short) × Boundary (within)	16.67	8.43	1.98	0.057
Direction (forward) × Boundary (within)	20.00	8.43	2.37	0.023*
Distance (short) × Input (numerical)	1.67	8.43	0.20	0.85
Direction (forward) × Input (numerical)	3.33	8.43	0.40	0.704
Boundary (within) × Input (numerical)	15.00	8.43	1.78	0.087
Distance (short) × Direction (forward) × Boundary (within)	-1833	11.92	-1.54	0.139
Distance (short) × Direction (forward) × Input (numerical)	0.00	11.92	0.00	1
Distance (short) × Boundary (within) × Input (numerical)	-5.00	11.92	-0.42	0.686
Direction (forward) × Boundary (within) × Input (numerical)	-6.67	1.192	-0.56	0.59
Distance (short) × Direction (forward) × Boundary (within) × Input (numerical)	-5.00	16.86	-0.30	0.775
Random effects	**Variance**		**SD**	
Participants (intercept)	65.33		8.08	
Item (intercept)	0.00		0.00	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 22. Coefficients from a mixed-effects model fitted the English group's accuracy in different conditions based on data for the month calculation questions.

4.2.3.2.3 Influence of Distance, Direction, Boundary, and Input on the Chinese Group's Accuracy

The analysis then moved to the fixed factors' effect on the accuracy of the Chinese group. The same fixed factors, random factors, and dependent variable were entered into R to build the mixed-effects regression model. Next, the interactions among fixed factors were examined in the same way as in the English group, which yielded no significant interactions between the fixed factors $x^2(11) = 16.369$, $p = 0.128$) leading to the choice of "Accuracy ~ Distance + Direction + Boundary + Input + (1 |Participant) + (1 | Item)" as the full model. The results are shown in Table 23. The full model was then compared with reduced models excluding Distance, Direction, Boundary, and Input, respectively. The results showed that only Boundary ($x^2(1) = 4.9847$, $p = 0.02557 < 0.05$) significantly increased the model fit, while other fixed factors had no effect on Chinese group's accuracy in month calculation questions, Distance ($x^2(1) = 0.4292$, $p = 0.5124$), Direction ($x^2(1) = 0.0483$, $p = 0.8261$), and Input ($x^2(1) = 0.4292$, $p = 0.5124$).

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	96.77	1.06	91.35	$< 0.001^{***}$
Distance (short)	0.63	0.95	0.66	0.508
Direction (forward)	-0.21	0.95	-0.22	0.825
Boundary (within)	2.29	0.95	2.42	0.015^{*}
Input (numerical)	-0.63	0.95	-0.66	0.508
Random effects	Variance		SD	
Participants (intercept)	0.00		0.00	
Item (intercept)	0.00		0.00	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 23. Coefficients from a mixed-effects model fitted the Chinese group's accuracy in different conditions based on data for the month calculation questions.

4.2.3.2.4 Influence of Distance, Direction, Boundary, and Input on the Bilingual Group's Accuracy

The mixed-effects regression model was built based on data of the Bilingual group's accuracy in the month calculation task, and the four fixed factors were verified to be not inter-dependent with each other, $x^2(11) = 12.944$, $p = 0.297$. The results are displayed in Table 24. The reduced models without Distance, Direction, Boundary, and Input were built and compared respectively with the full model to test the fixed factors' effects on the Bilingual group's accuracy. The results only confirmed the contribution of Boundary ($x^2(1) = 4.0264$, $p = 0.04479 < 0.05$) to the differences between Bilingual group' accuracy in short and long trials, while Distance ($x^2(1) = 1.9587$, $p = 0.1616$), Direction ($x^2(1) = 2.923$, $p = 0.08732$), and Input ($x^2(1) = 0.0242$, $p = 0.8764$) were found to have no significant effects on the accuracy of the Bilingual group.

Fixed effects:	Estimate	SE	t value	p value
(Intercept)	91.77	1.63	56.32	$< 0.001^{***}$
Distance (short)	1.88	1.34	1.40	0.169
Direction (forward)	2.29	1.34	1.71	0.093
Boundary (within)	2.71	1.34	2.02	0.047^{*}
Input (numerical)	2.08	1.34	0.16	0.878
Random effects	Variance		SD	
Participants (intercept)	12.43		3.53	
Item (intercept)	0.00		0.00	

Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 24. Coefficients from a mixed-effects model fitted the Bilingual group's accuracy in different conditions based on data for the week calculation questions.

4.3 Self-reported Strategies

The Self-reported Strategies (Expt. 2) were analyzed descriptively to see potential crosslinguistic differences in strategies applied to solve calendrical calculation questions. Participants' self-reported strategies were coded as "Memory", "Calculate", "Count", "Guess", and "Other". Self-reported strategies of the Chinese-English speakers who were tested in English (L2) also included "Translate and Calculate" and "Translate and Count".

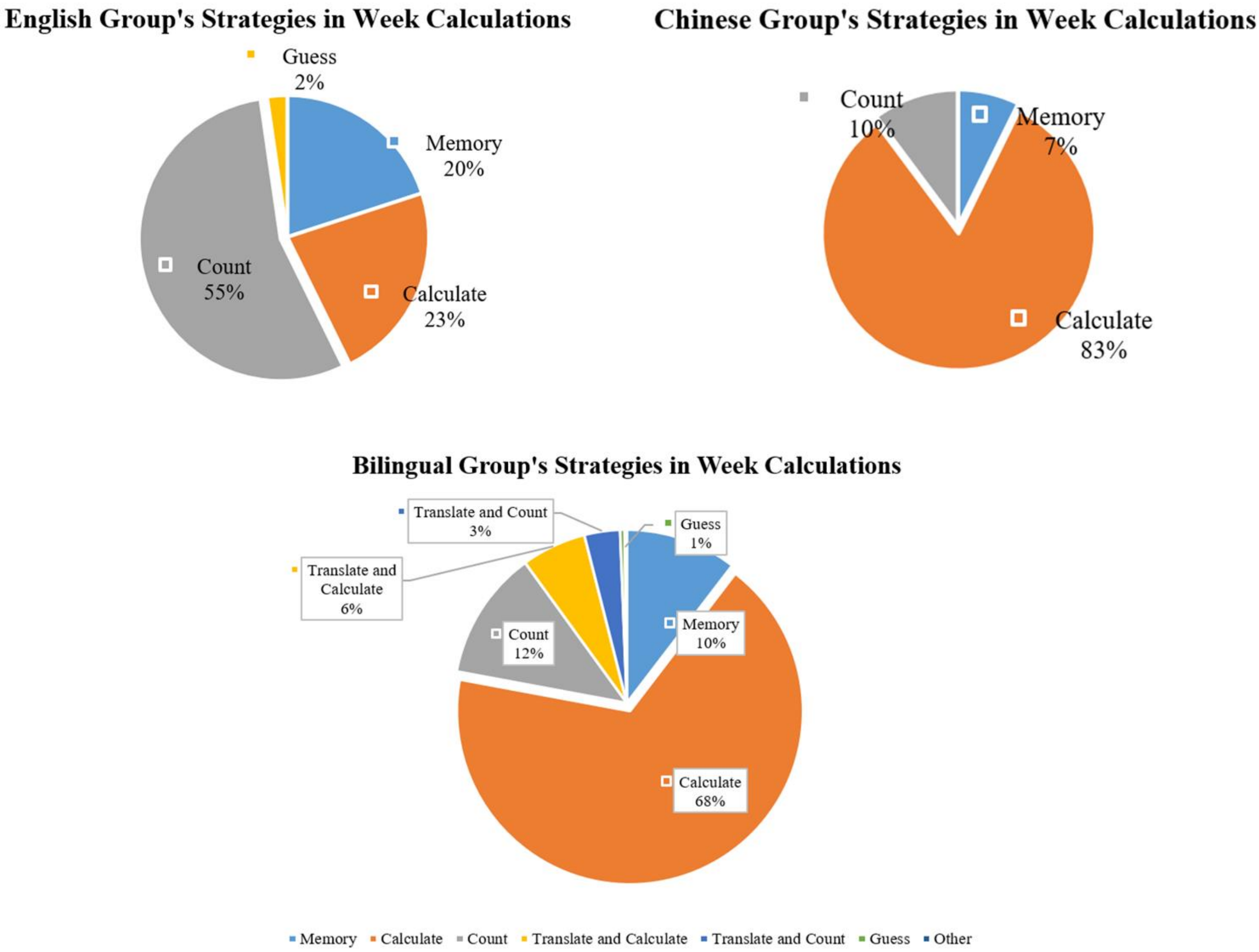

Figure 11. Self-reported strategies by the English, Chinese, and Bilingual group in the Week Calculation Task.

4.3.1 Self-reported Strategies in the Week Calculation Task

Overall, 55% of English speakers reported a counting strategy, while the remaining adopted a memory strategy (20%) and a calculating strategy (23%). When tested in their native language, Chinese, most Chinese-English bilingual respondents (83%)

preferred the calculating strategy, and only 10% reported a counting strategy, and 7% applied a Memory strategy. When tested in English, the bilinguals reported similar strategies as they chose in the Chinese version test, as shown in Figure 11.

The self-reported strategies were further analyzed for individual conditions, including Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical). In the short trials, English respondents reported more Memory strategy and less Calculate strategy than in the long trials (Figure 12). Moreover, the English speakers reported more Memory strategy and less Count strategy in within trials than that in across trials. Additionally, with long distance and numerical input, the English respondents applied more Calculate strategy and less Count strategy. There was no significant difference in self-reported strategies between the forward condition and backward condition. The Chinese-English bilinguals who were tested in Chinese (Figure 13) reported a preference for the Calculate strategy in all conditions. One notable difference occurred under the influence of Boundary, where the bilinguals reported more Count strategy and less Memory strategy in the across condition than in the within condition. The bilinguals who took the English version test provided similar strategic responses, as shown in Figure 14, except that some respondents translated the questions before they gave an answer.

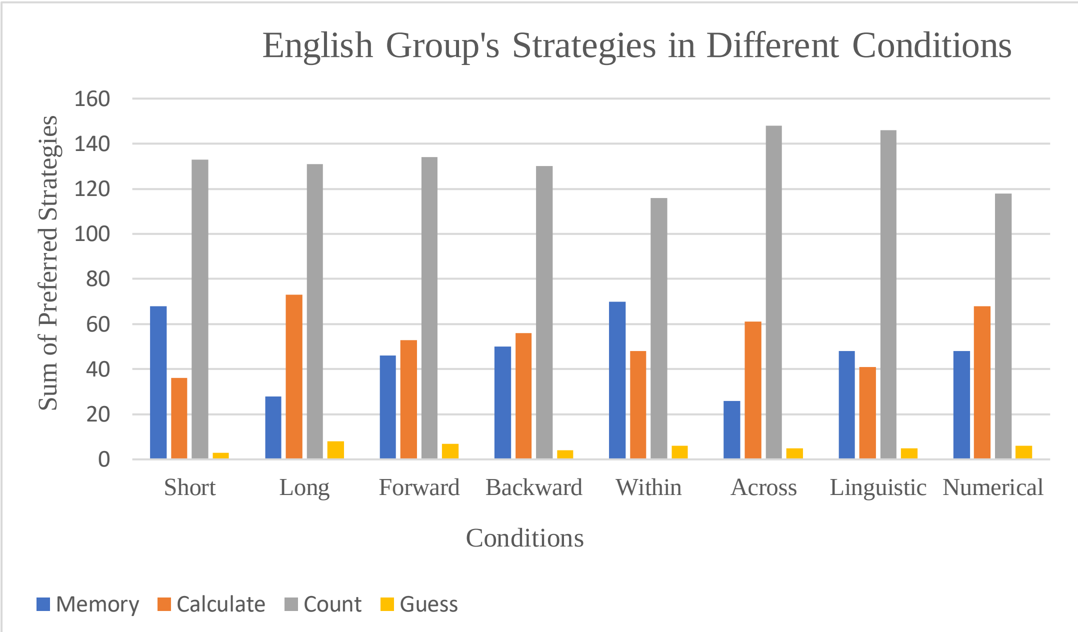

Figure 12. Self-reported strategies by the English group in different conditions, including Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical), based on data for week calculation questions.

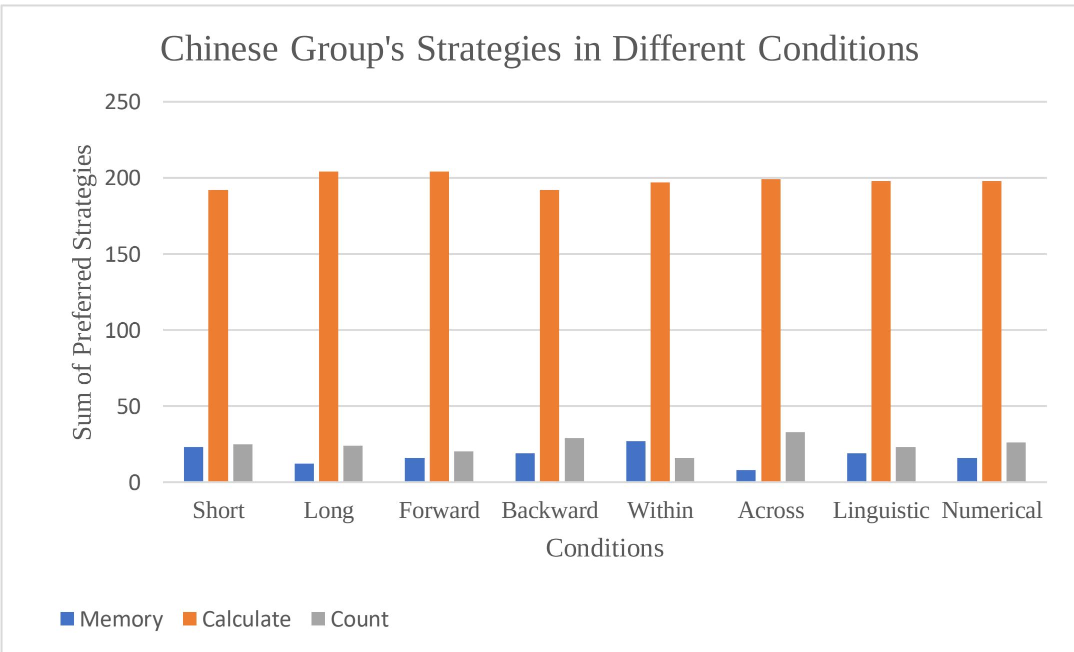

Figure 13. Self-reported strategies by the Chinese group in different conditions, including Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical), based on data for the Week calculations.

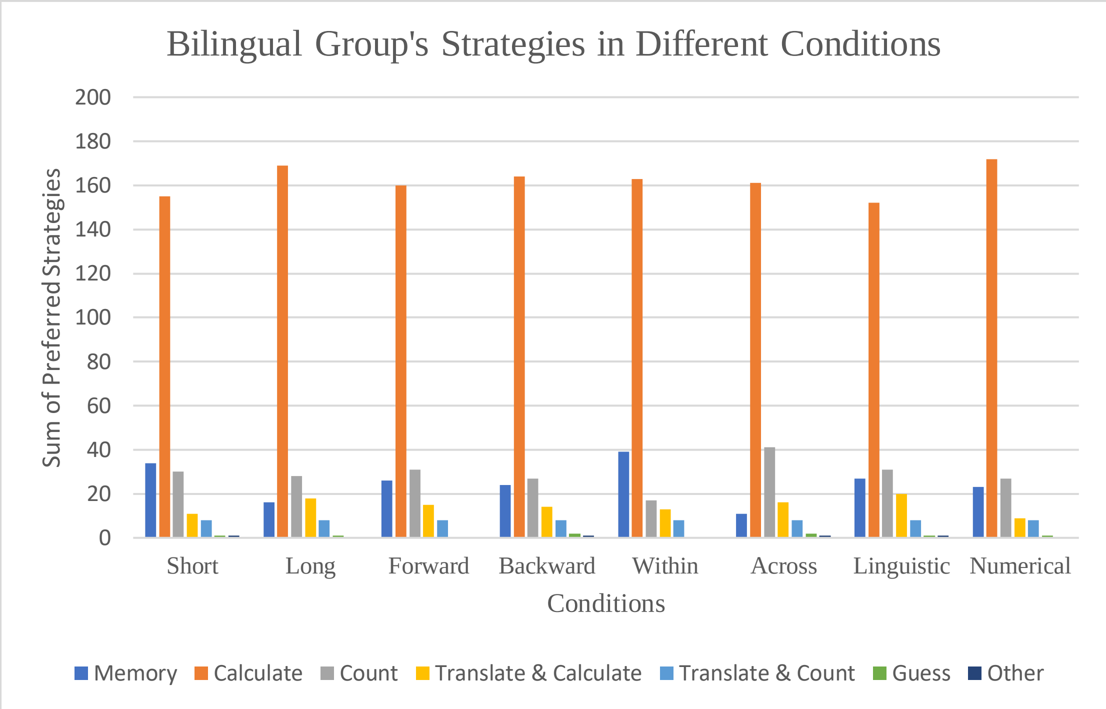

Figure 14. Self-reported strategies by the Bilingual group in different conditions, including Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical), based on data for week calculation questions.

4.3.2 Self-reported Strategies in the Month Calculation Task

Similar to the week calculation tasks, most bilinguals reported using the 'Calculate' strategy when tested in both Chinese ("81%") and English (69% calculate and 6% "translate and calculate"), as shown in Figure 15. Unlike the less demanding week calculation task, English respondents reported relying less on the Memory strategy but more on Calculate strategy, though there were still more than 50% of English speakers who reported a Count strategy (Figure 15).

The effects of Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical) on each group's self-reported strategies are displayed in Figure 16, Figure 17, and Figure 18. The results aligned with the self-reported strategies in the Week Calculation Task. In English group, strategy Calculate was more popular in long trials (88) than in short trials (65) and more popular in numerical input (73) than in linguistic input (80), while strategy Count was more popular in boundary crossing trials (130) than in boundary within trials (115).

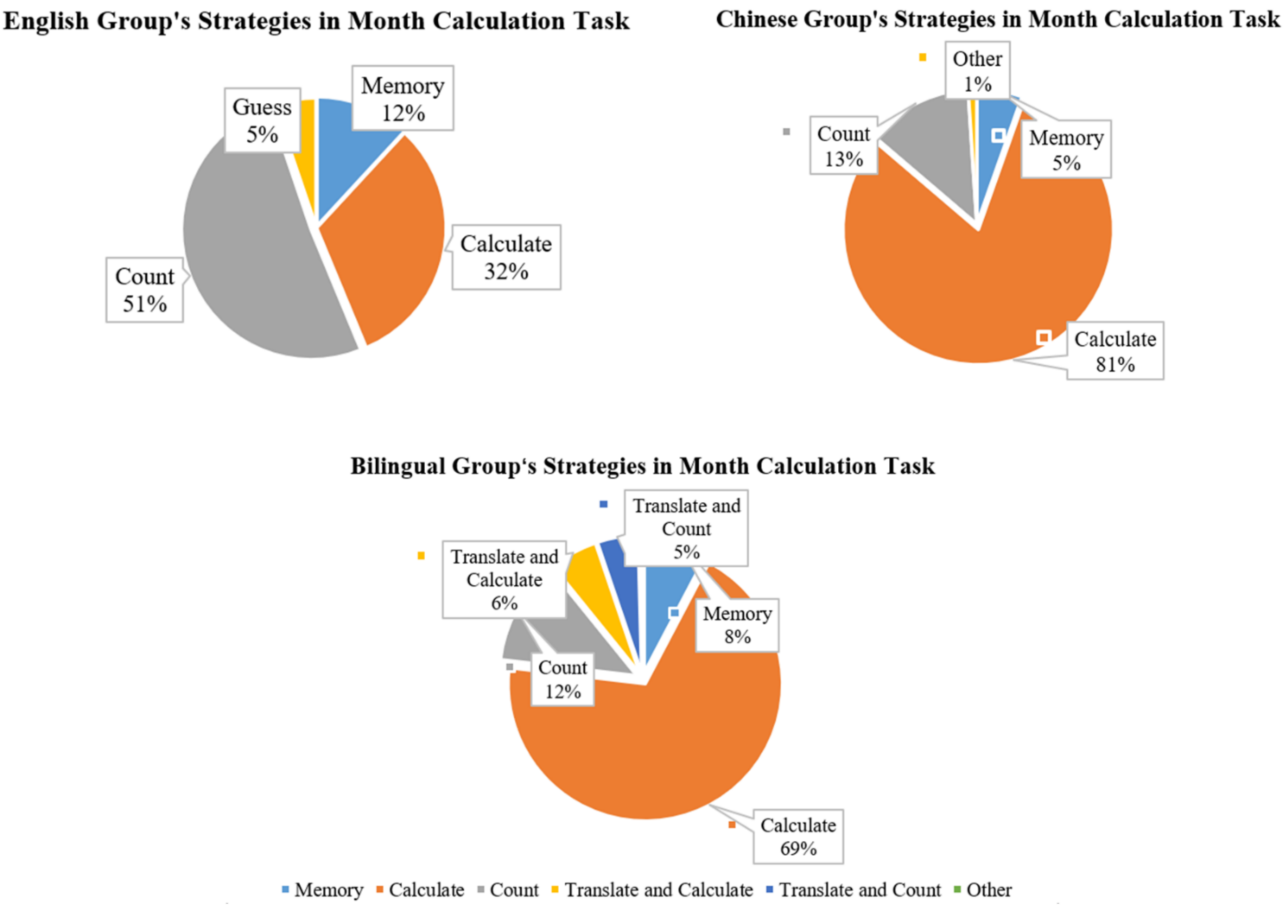

Figure 15. Self-reported strategies by different groups in the Month Calculation Task.

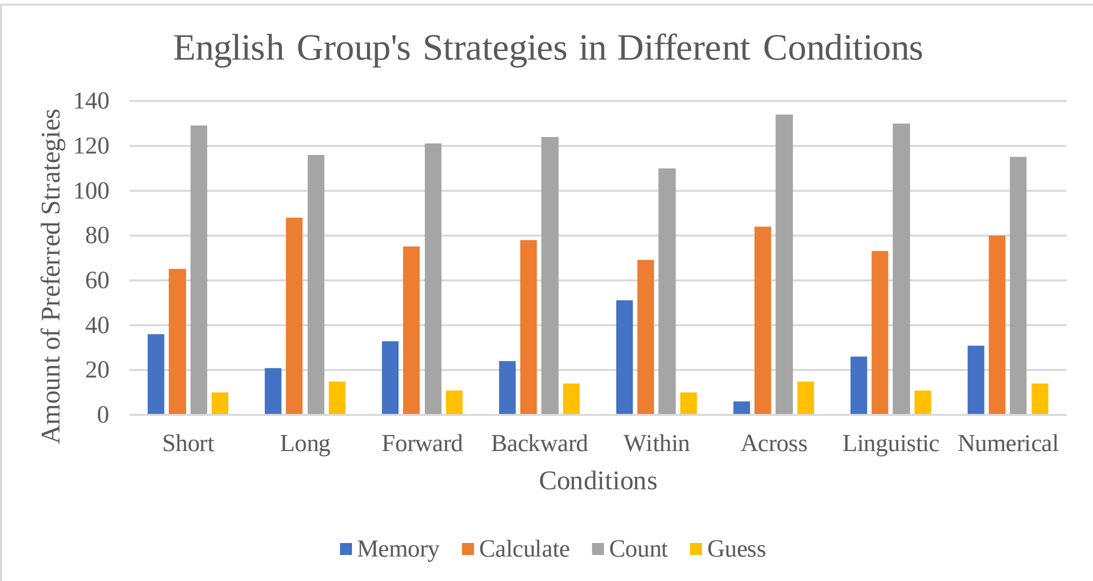

Figure 16. Self-reported strategies by the English group in different conditions based on data for the month calculation questions.

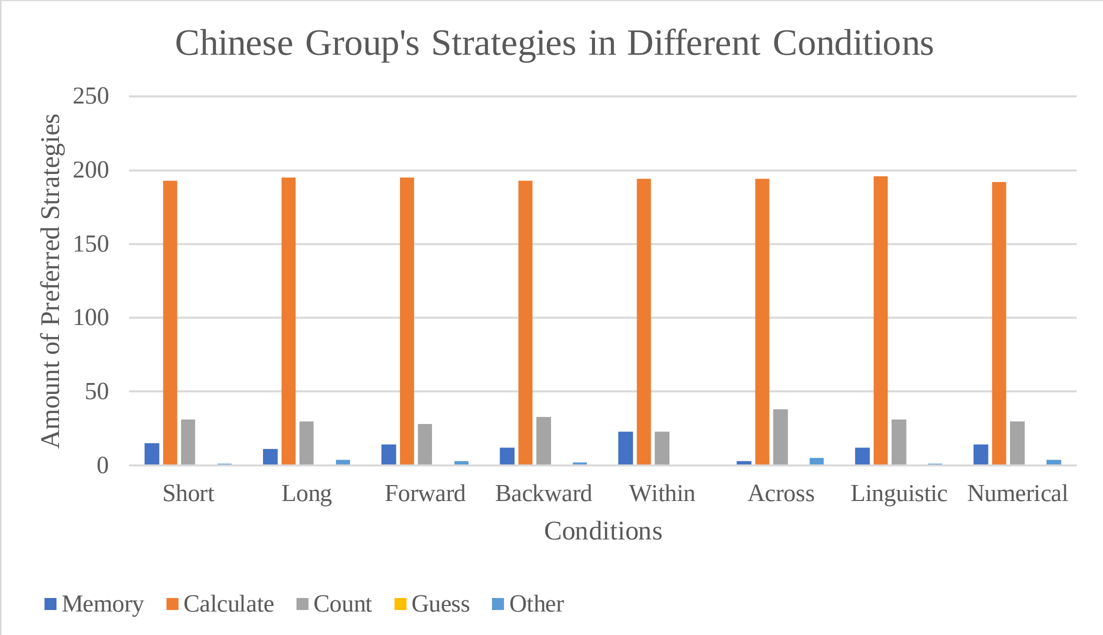

Figure 17. Self-reported strategies by the Chinese group in different conditions based on data for the month calculation questions.

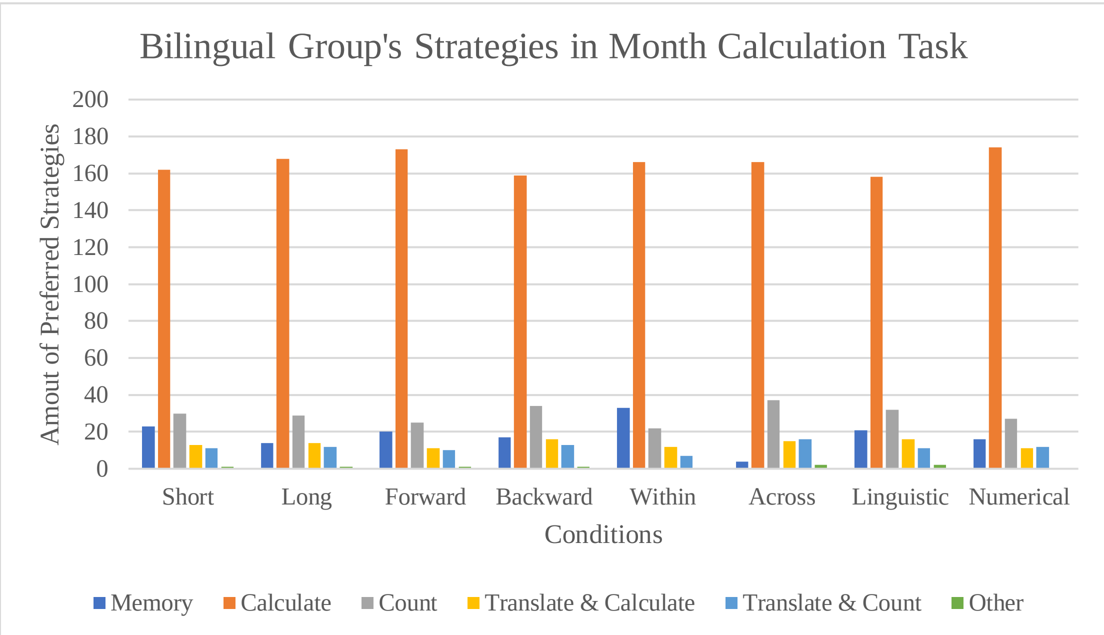

Figure 18. Self-reported strategies by the Bilingual group in different conditions based on data for the month calculation questions.

4.3.3 Self-reported Strategies in the Hour and Year Calculation Task

The arithmetic calculation task included the 24-based calculation questions from the hour task and the 10-based calculation questions from the year task. Firstly, all respondents reported a preference for the Calculate strategy in processing the hour calculation questions, though the proportion of the bilingual groups (88% and 83.3%) was higher than that of the English group (55.8%), some of whom also counted to solve the questions (27.9%). Secondly, most participants from the three groups relied on the Calculate strategy to answer the 10-based year calculation questions, English (69.6%), Chinese (82.5%), and Bilingual (69.6%).

4.4 Summary of the Main Results

Overall, there were two experiments in the present study. Experiment 1 was the Calendrical Calculation Task, where there were six factors, including two main factors, Calculation (Week, Month, Hour, and Year Calculation Tasks) and Group (the English group, the Chinese group, and the Bilingual group), and four within-group factors, including Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical). The dependent variables were reaction times (RTs) and accuracy. Experiment 2 was the Self-reported Strategies Task, where participants were required to choose the strategies that they used to solve calendrical calculation questions. RTs and accuracy from Experiment 1 were analyzed by mixed-effects regression models, while the proportion of self-reported strategies was descriptively analyzed, based on means and standard deviations (SDs).

The factor Calculation was a significant predictor of how participants' RTs varied. The processing speed of the year calculation questions was the fastest, followed by the speed of week and month calculation questions, and the processing speed of hour calculation questions was the slowest. The calculation also significantly influenced participants' accuracy in the Calendrical Calculation Task. Participants achieved the highest accuracy in the Year Calculation Task and lowest accuracy in the Hour Calculation Task, as well as similar accuracy in week and month calculation questions. Findings of the effects of Group and the four within-group factors would be summarized from the perspective of different levels of Calendrical Calculation Tasks (week, month, hour, and year tasks) as follows.

4.4.1 Week Calculation Task

In the Week Calculation Task, the factor Group had a significant influence on participants' performance. For data from RTs, the Chinese group took shorter RTs than the English group and the Bilingual group, though there were no significant differences between the English group and the Bilingual group. For data from accuracy, the week calculation questions were answered less accurately by the English group than the Chinese and Bilingual group, while the Chinese group achieved the highest accuracy and the Bilingual group was in-between. The effects of different conditions – Distance, Direction, Boundary, and Input, were found on participants' RTs and accuracy. Based on the data from RTs, the fixed factors Distance (short) and positively affected participants' reaction times when solving the week calculation questions among all three groups (the English, Chinese, and Bilingual group), while the fixed factors Direction (backward) and Boundary (across) only negatively affected the Chinese-English bilinguals but had no effect on the English speakers. The fixed factor Input (numerical) negatively affected the English group but showed no significant effect on the Chinese-English speakers, either tested in Chinese (L1) or English (L2). Moreover, the potential influence of interactions among the two-level fixed factors was also taken into consideration. There was a significant interaction between Distance and Input in the RTs of English speakers, who spent longer response times in longer trials compared to linguistic input than numerical input.

When it comes to the different conditions' effects on accuracy, long trials negatively affected the performance of the English group and the Bilingual group but not the Chinese group. Fixed factors Direction, and Boundary affected the accuracy scores in the week calculation questions among all three groups (the English, Chinese, and Bilingual group). On the contrary, the fixed factor Input only influenced the English group but had no significant effect on the Chinese-English speakers, either tested in Chinese (L1) or English (L2). Additionally, interactions between the conditional factors were also found in the English group and Bilingual group. The accuracy of the English group was significantly influenced by the interaction effect between Distance and Direction, and significant differences between the accuracy of short trials and long trials were found both in forward condition and backward condition. Interaction between Direction and Boundary was found in accuracy of the Bilingual group, where short trials positively influenced accuracy in conditions across the boundary (e.g., Friday + 3 days = ?) but not in conditions within the boundary. Table 25 summarizes if Distance, Direction, Boundary, and Input affected each group's RTs and accuracy in a statistically significant way in the Week Calculation Task.

Strategy Calculate was preferred by the Chinese and Bilingual group, while the English monolinguals counted a lot to get the answer. It is worth noticing that the proportion of Count strategy was highest chosen by the English group and the proportion of Bilingual group was in-between.

	Group	Distance	Direction	Boundary	Input
RTs	English	√	×	×	√
	Chinese	√	√	√	×
	Bilingual	√	√	√	×
Accuracy	English	√	√	√	√
	Chinese	×	√	√	×
	Bilingual	√	√	√	×

Table 25. An overview of whether the fixed factors significantly affect each group's RTs and accuracy in week calculation questions (√ means yes, × means no).

4.4.2 Month Calculation Task

The factor Group was an important predictor of participants' performance when doing the Month Calculation Task. On the one hand, RTs of the English group were similar to those of the Bilingual group, both significantly longer than the Chinese group, which was in line with RTs in the Week Calculation Task. On the other hand, the accuracy of the Chinese and Bilingual group was significantly higher than the English group.

The effects of Distance (short or long), Direction (forward or backward), Boundary (within or across), and Input (linguistic or numerical) would also be summarized from the perspective of RTs and accuracy. Starting with RTs, crossing boundary significantly inhibited the speed of working out the correct answers of all participants in the English, Chinese, and Bilingual groups. In contrast, the fixed factor Input showed no significant influence on participants' RTs in solving month calculation questions. Factors Distance and Direction showed no statistically significant effects on the English speakers, but only affected the Chinese-English bilinguals, both tested in Chinese (the Chinese group) and in English (the Bilingual group). Furthermore, interactions were found in the Bilingual group, where there was a significant interaction between Distance and Boundary, and between Direction and Boundary. Results of simple effect tests claimed that short trials were answered significantly faster than the long trials, and forward trials faster than the backward trials, in the crossing boundary condition but not in the within boundary condition.

For accuracy, Distance and Direction only significantly influenced the accuracy of the English group, while Boundary had significant effects on the Chinese group and the Bilingual group but not on the English group. Factor Input was not a significant predictor of accuracy of all three groups. Moreover, in the English group, Distance was significantly interacted with Direction, which was in line with the finding in the Week Calculation Task. Table 26 displayed an overview of whether the conditional factors significantly influenced the RTs and accuracy performed by participants from the English, Chinese, and Bilingual groups.

Results of self-reported strategies still showed Chinese speakers' reliance on calculating as in the Week Calculation Task, though the English speakers expressed a higher preference for calculating and a less preference for counting during solving month calculations, especially in the condition of Distance long and Direction backward.

	Group	Distance	Direction	Boundary	Input
RTs	English	×	×	√	×
	Chinese	√	√	√	×
	Bilingual	√	√	√	×
Accuracy	English	√	√	×	×
	Chinese	×	×	√	×
	Bilingual	×	×	√	×

Table 26. An overview of whether the fixed factors significantly affect each group's RTs and accuracy in month calculation questions (√ means yes, × means no).

4.4.3 Hour and Year Calculation Task

Results of the Hour and Year Calculation Task mainly involve three points, including comparisons between the Hour and Year Calculation Task, comparisons across the three groups within the Hour or Year Calculation Task, and the strategies reported to solve hour and year calculation questions. Firstly, participants' performance in the Hour Calculation Task was slower and less accurate than performance in the Year Calculation Task. Secondly, participants' RTs and accuracy were at ceiling level across groups both in the Hour and Year Calculation Tasks, though the RTs of Chinese speakers were slightly slower than the English speakers when tested in L2 English. Thirdly, strategy Calculate was primarily chosen by all groups' participants to solve the hour calculations in different conditions, though some also counted when processing the Hour Calculation Task, especially in solving across boundary trials. The proportion of strategy Calculate was used more frequently to solve the year calculation questions, and only a few participants chose strategy Count and Memory.

5. Discussion

This study investigated the effects of having a first language with linguistically and numerically transparent calendar systems on speakers' performance in calendrical calculation tasks when tested in a second language with opaque calendar names. Building on previous evidence about differences in reaction times and calculation strategies between Chinese and English monolinguals tested in their native languages, the present study tested how Chinese-English bilinguals who are native users of Chinese with a transparent calendar lexicon, and late learners of English, solve calendrical calculation questions when tested in L2 English. Results show that Chinese-English bilinguals' problem-solving process was positively influenced by their L1 Chinese in both Calendrical Calculation Task (RTs and accuracy) and Self-reported Strategies Task. Though tested in a weaker language, performance of the Bilingual group in week calculation questions equaled to the English group and was even better than the English group in the more demanding Month Calculation Tasks. Additionally, in the Self-reported Strategies Task, the Chinese group showed a reliance on numerical operations whereas most English monolinguals preferred a verbal encoding strategy. When tested in English, although the L2 English with opaque calendar terms was activated, the Chinese-English bilinguals still reported a preference on the numerical calculation method as the Chinese group rather than the verbal-listing strategy. Detailed similarities and differences among the Chinese, English, and Bilingual groups, and the potential explanations of the results are further discussed in this chapter.

In the chapter Discussion, results of the Hour and Year Calculation Task were discussed first, followed by discussion of results of the Week Calculation Task and Month Calculation Task. The logic for this sequencing is that hour and year calculation questions revealed the mathematical abilities in uncommon (24-based calculations) and common (10-based calculations) arithmetical questions of participants from the three groups. A comparable level of mathematical abilities was a precondition to discuss the potential different performances in 7-based week calculation questions and 12-based month calculation questions across the three groups. Discussion of results from the Week Calculation Task came before the discussion of month calculation questions results, as week calculation questions required relatively less effort than month calculation questions. At last, participants' self-reported calculation strategies applied to solve calendrical calculation questions in different conditions were discussed.

5.1 Hour and Year Calculation Tasks Findings and their Contextualization

The hour calculation questions were answered slower and less accurately than the year calculation questions, reflecting that the irregular 24-based hour calculations were more difficult to process than the regular 10-based year calculations for all three groups. Moreover, accuracy and RTs were at ceiling level across groups in both Hour and Year Calculation tasks, indicating that speakers of Chinese and English had very similar arithmetic calculation abilities in uncommon (10-based) and uncommon (24-based) calculation questions.

However, in the Hour Calculation Task, the average calculation speed of Chinese speakers was slightly slower than English speakers when tested in L2 English, while the Chinese group answered questions faster than the English group when tested in L1 Chinese. Firstly, when tested in a weaker language English, the Chinese-English bilinguals spent more response times to get the correct answer. This might be because the hour calculation questions were formulated linguistically, and the Chinese speakers might need to translate the English questions into Chinese or into Arabic numbers first, then calculate to get the answer, and translate the Chinese or Arabic numerical answer back into English. This is in line with previous evidence from Wang, Lin, Kuhl, and Hirsch (2007) and Lin, Imada, and Kuhl (2011). For example, to answer the hour calculation question "two o'clock + 7 hours = ?", results debriefing confirmed that Chinese participants tend to translate it into "两点 / 2 点 + 7 小时 = ?", get the answer "九/9 点", and translate it into "nine o'clock". Relating the findings to more recent models of the bilingual processing, the longer RTs in hour calculation questions might also be counted by RHM. RHM, the Revised Hierarchical Model, was first proposed by Kroll and Stewart (1994), claiming that high proficiency L2 learners tends to connect L1 to the concepts and L2 to the concepts directly rather than bridging L2 lexicons and the concepts through L1 words. Chinese-English bilinguals recruited by the present study were all advanced English learners, who were likely to bridge the English questions and the concepts directly. However, the longer response times of the Bilingual group than the English group reflected that connection between L1 words and concepts tended to be stronger than that between L2 words and concepts (Cheng, Wang, and Perfetti, 2011). Secondly, it took the Chinese-English bilinguals shorter RTs to solve the hour calculation questions with testing language Chinese (where only L1 Chinese was activated) than English (where both Chinese and English were activated). This is consistent with research on bilinguals' arithmetic calculation abilities, where bilinguals showed an advantage for solving questions in the L1 (Marsh and Maki, 1976; French-Mestre and Vaid, 1993) or at least in the language in which they learned mathematical skills (Bernardo, 2001; Van Rinsveld et al., 2015). Thirdly, the Chinese group was marginally faster than the English group even when both tested in their corresponding native languages, which might profit from the more transparent and regular number naming system in Chinese, including the awkward names (e.g., eleven and twelve) and a complex phonemic and morphological variant (e.g., ten transformed to "-teen" and "-ty"), as described in section 2.1.2.2. On the contrary, the means of RTs in the Year Calculation Task were similar in three groups, since the questions were displayed with numerical input and it was unnecessary to do any translation.

In sum, all three groups performed similarly in the Year Calculation Task. In the Hour Calculation Task, the English group slightly outperformed the Bilingual group, as the Bilingual group tended to involve an extra translation process to get the answer. The Chinese group was the fastest across the three groups in the irregular 24-based hour calculation questions, benefiting from the more transparent and regular Chinese number naming structure. Nevertheless, the differences across the three groups in hour calculations did not achieve a statistical significance. Therefore, it was reasonable to claim that the mathematical calculation abilities of the English, Chinese, and Bilingual group were at a comparable level in the uncommon 24-based Hour and 10-based Year Calculation Task. Moreover, despite the different calculation processes for solving hour and year calculation questions, speakers of Chinese, both tested in L1 Chinese or L2 English, as well as the English monolinguals, appeared to rely on numerical processing to calculate both hour and year calculation questions.

5.2 Week Calculation Task Findings and their Contextualization

When doing the Week Calculation Task, the English monolinguals showed a preference for overtly or covertly verbal listing, while speakers of Chinese tested in Chinse (the Chinese group) relied on numerical calculations. Under the influence of the transparent calendar terms of their first language, the Chinese-English bilinguals who were tested in English (the Bilingual group) also reported reliance on numerical representation and processing. The difference across the Chinese, Bilingual, and English groups in solving week calculation questions was demonstrated by differences in the three groups' RTs, distance effect, direction effect, boundary effect, input effect, and self-reported strategies. These differences are discussed below with reference to previous literature in the field.

5.2.1 Reaction times and Accuracy

Perhaps the most relevant previous study (Bassetti, Clarke, and Trenkic, 2018) found that Chinese speakers tested in English were just marginally slower than the English speakers tested in English, which led the authors to reason that verbal listing (of just seven items max.) is efficient enough for the English monolinguals to solve the relatively simple week calculation questions. However, the present study demonstrates that the Bilingual group achieved similar RTs to the English group, while the Chinese speakers who were tested in Chinese were significantly faster and more accurate than the English speakers. This might be because Bassetti et al.'s (2018) study asked participants description questions like "when the seeds will sprout" when telling the participants when the seeds had been planted and how many days it takes the speeds to sprout. Such description was easy to understand for English speakers but could cost extra time for late learners of English to process the contextual meaning. In the present study, questions have been replaced by "Monday + 1 day = ?", which is clear for both native speakers of English and bilinguals. In sum, in the Week Calculation Task, the verbal listing is believed to be efficient enough due to a low-complexity list and the Bilingual requires more time to translate the questions in numerical facts to get the answer, as they do in the Hour Calculation Task. Results still claimed that with knowledge of transparent calendar terms and efficient numerical operation strategy, the Bilingual group was as fast as and more accurate than the English group, even in the Week Calculation Task.

The advantage of knowing a linguistic and numerically transparent calendar system was apparent in all conditions, including short distance, long distance, forward direction, backward direction, within boundary, across boundary, linguistic input, and numerical input, but only when the Chinese speakers are tested in L1 Chinese. When tested in Chinese, the Chinese group was the fastest among the three groups in all conditions, especially in condition long (2 s faster than the English group), backward (1.1 s faster), within (1.4 s faster), and numerical (1.6 s faster). Such an advantage remained when Chinese speakers were tested in English, a language with an opaque calendar naming system, but was differently manifested. In short distance, both directions, across boundary, and linguistic input trials, there were no group differences in RTs or accuracy between the Bilingual group and the English group. This means that knowing transparent calendar terms allows L2 speakers to be as fast and accurate as the native speakers even when tested in their weaker language. Such RTs catch-up with target group has been observed in other areas of bilingual processing, such as negation processing (Zhang and Vanek). In long distance, within boundary, and numerical input trials, the Bilingual group performed better than the English group, achieving an average of 0.5s faster RTs and an average 5% higher accuracy.

5.2.2 Negative Effects of Long Distance in English and Chinese Speakers

Distance effects were found in the English group. Speakers of English were on average 2s faster and 8.6% more accurate in the short than in the long trials. Such distance effects are in line with the findings of Kelly, Miller, Fang, and Feng (1999). This is because, as Friedman (1983) argued, sequential activation of calendar names takes longer response times when the target is further away from the stimuli.

Distance did not affect the Chinese group's accuracy, which aligns with previous evidence that distance does not affect Chinese monolinguals when tested in L1 Chinese (Huang, 1993; Kelly et al., 1999). However, RT differences between short and long trials performed by the Chinese and the Bilingual group were statistically significant. This may be because not all day names within a week are numerical, and linguistically transparent. Sunday is an exception, which is "星期天 or 星期日", literally "week-day or week-sun" rather than "week-seven". Nevertheless, in spite of the fact that speakers of both English and Chinese were negatively influenced by long distance, the effects of distance on Chinese speakers (0.8 s difference gap in the Chinese group and 1.2 s difference gap in the Bilingual group) was markedly lower than that on the English speakers (2 s difference gap). This difference indicates that the numerical processing is more efficiently applied by the Chinese speakers even when tested in a language lacking a transparent calendar lexicon, namely English.

5.2.3 Negative Effects of Backward Directionality in English and Chinese Speakers

Directionality effects on the English group were found, as backward trials were answered less accurately (an average of 6.1%) and slower (an average of 1 s) than forward trials, though the RTs does not reach a statistically significant different. The negative effect of backward direction aligns with previous findings from research by Kelly et al. (1999), where calendar sequences were found to be more difficult to recite in backward order than in forward order.

There were also significant influences of directionality on RTs and accuracy of the Chinese and Bilingual group, which corroborates Bassetti et al.'s (2018) findings. Bassetti et al. proposed that the involvement of the nontransparent calendar term "Sunday" inhibits Chinese speakers' performance either tested in L1 Chinese or L2 English, as in the condition of Distance. Furthermore, there was an interaction effect between Direction and Boundary in the accuracy of the Bilingual group, which revealed that the negative effects of backward directionality only existed in the condition across boundary but not in within boundary. This is because although the Bilingual group reported a reliance on the numerical calculation strategy, there was an increase in Chinese speakers choosing the Count strategy when solving boundary trials instructed in English. The mixed choice of calculation strategy revealed that the Bilingual group who were familiar with both strategies, Count in English calendrical calculation and Calculate in Chinese calendrical calculation, tended to choose the most efficient way to process different calendrical calculation questions, rather than stick to the strategy typical of the language of testing.

5.2.4 Negative Effects of Crossing Boundary in Chinese Speakers

Negative effects of crossing boundary were found in the Chinese speakers but not in native speakers of English. The reason that the English group was not affected by the factor Boundary was that, as demonstrated by Kelly et al. (1999) and Bassetti et al. (2018), no extra demand was needed to solve boundary crossing trials compared with processing within boundary trials by English native speakers, as crossing boundary results in no additional costs for verbal or mental listing.

On the contrary, crossing boundary hindered Chinese speakers' RTs (an average of 1.8 s slower than solving within-boundary trials) and accuracy (an average of 6% less accurate than processing within-boundary trials) when tested in Chinese, which is in line with results of Huang's (1999) study. As the numerical or linguistically transparent calendar system leaded to a tendency of preferring numerical calculations, boundary crossing calculations were more demanding than within-boundary calculation.

Additionally, the week calculation questions were uncommon 7-based calculations, which greatly increased the difficulty of working out an answer, like the disadvantages of doing 24-based hour calculation questions discussed in section 5.1. Such negative effects of crossing a boundary extended to Chinese speakers who were tested in L2 English. There were significant differences in RTs and accuracy between cross-boundary and within- boundary trials in the Bilingual group. Cross-boundary trials were answered almost twice as slowly and 4% less accurately than within-boundary trials. This resulted from the reliance on numerical arithmetic strategy of the Chinese speakers when tested in L2 English like tested in L1 Chinese, which is consistent with previous evidence (Bassetti et al., 2018).

5.2.5 Negative Effects of Numerical Input in English speakers

Input effects were found in English speakers but not in Chinese speakers tested either in Chinese or in English. Surprisingly, the English monolinguals were on average 1 s slower and 2.7% less accurate in the numerical input trials than in the linguistic input trials. This is because the English calendar naming system is opaque and English native speakers may not be used to linking the weekday names with numbers. For instance, when solving a numerical input task like “1st day + 3 days = ?”, the English participants might have mentally translated “1st day” into “Monday” and then counted to get the answer “Thursday”. The additional mental operation could thus negatively influence the English group’s performance in numerical input trials.

The Chinese group spent similar RTs and achieved equal accuracy in numerical and linguistic input trials, since the weekday lexemes are numerical and linguistically transparent. For the Chinse group, processing the linguistic trial “周一 + 3 天 = ？” (literally “week-one + 3 days = ?” meaning “Monday + 3 days = ?”) is extracting the number “one” and “3” and adding the two numbers to get the answer “four”, leading to “week-four” (meaning Thursday), “周四”. Processing the numerical input trials “周 1 + 3 天 = ？” requires no additional effort, as “1” is the Arabic numerical form of the word “one”. Even when tested in a language without transparent calendar names, English, speakers of Chinese were still not affected by different input types and similarly performed in linguistic and numerical input trials. This is because when solving a linguistic trial, the Bilingual group may be translating English questions to Chinese before calculating them. When asked to process a numerical trial, the Bilingual group may tend to utilize the knowledge of the transparency of L1 Chinese calendar terms and extract the corresponding numbers to speed up simple additions and substructions. It was also notable that the Bilingual groups spent longer RTs (0.3 s) than the English group when solving the week calculations in linguistic input, which might be because bridge between the L2 mental lexicons and the concepts is weaker than that between the L1 and the concept (Brysbaert and Duyck, 2010).

5.2.6 Self-reported Strategies

Results of self-reported strategies support previous findings (Friedman, 1983 for English monolingual speakers; Huang, 1993, 1999 for Chinese monolingual speakers, Kelly et al., 1999 for both; and Bassetti et al., 2018 for Chinese-English bilinguals) that English speakers mostly relied on a verbal listing strategy (55%), while both the Chinese group (83%) and the Bilingual group (74%) mainly relied on a numerical operation strategy. Nevertheless, the English group applied alternative strategies in performing long distance, backward directionality, and numerical input trials. The Bilingual group also reported trying out more complex strategy combinations when different languages and conditions of the week calculation questions provided.

The English group reported a mixture of Count strategy (55%), Calculate strategy (23%), and Memory strategy (20%). The coexistence of strategy Memory is caused by the imagery of a calendar naming system emerging at English speakers' ten years old (Friedman, 1986). The virtual image refers to a picture with calendar terms in the English native speakers' minds, with which the English group can find out the answer in memory. Naturally, the shorter the list is, the easier the target answer is to be found. Therefore, English native speakers reported more Memory strategy and less Calculate and Count strategy in short trials than in long trials, and strategy Memory was particularly useful in the Week Calculation Task, as the weekday system is a 7-based list. Moreover, more Calculate strategy and less Count strategy were used by several English respondents in the numerical input trials compared with in the linguistic input trials. This is because numerical input directly provides transparent numbers that are convenient to do the calculations with, though most English speakers insist on transferring the numerical form of weekday names into a linguistic version (e.g., transfer "1st day" into "Monday") and then count the target day.

The Chinese group almost exclusively relied on the Count strategy in all conditions as the Chinese transparent calendrical terms would suggest, which is in line with Kelly et al.'s (1999) research results. In the Bilingual group, there was a marginal increasement of Memory (3% more than the Chinese group) and Count strategy (5% more than the Chinese group). Some participants from the Bilingual group found that using the Memory and Count strategy could prevent the difficulty of crossing boundaries in cross-boundary conditions and the difficulty of dealing with the opaque weekday term "Sunday". However, most bilingual respondents tested in English still sticked to the arithmetic calculation strategy, which aligns with Bassetti et al.'s (2018) results, indicating that participants find it unnecessary to try alternative strategies but preferred strategies they normally use when solving an easy calculation task.

5.3 Month Calculation Task Findings and their Contextualization

The present study also involved a Month Calculation Task, where participants were assumed to perform differently from the Week Calculation Task for two reasons. First, month calculation questions are more demanding to process than week calculation questions, as month calculation questions are established based on a 12-based list and week calculation questions based on a 7-based list. Second, unlike weekday names, where there is an opaque term "Sunday", month names are all numerically and linguistically transparent terms in Chinese. Participants performed in the Month Calculation Task differently in the Week Calculation Task in three aspects, the overall

Reaction Times and Accuracy, no effects of Distance, Direction, and Input on RTs of English speakers, negative effects of Boundary on RTs of English speakers, and different self-reported strategies. These findings are interpreted and contextualized separately.

5.3.1 Reaction Times and Accuracy

Given that the Month Calculation Task is more demanding than the Week Calculation Task, all three groups performed less accurately and spent longer RTs in the month calculation questions compared with week calculation questions. English speakers were on average 1s faster, while the Chinese and Bilingual group were also faster in a less pronounced way, with 0.8 s and 0.6 s on average, respectively. The relatively small RTs gap may be attributable to the Calculate strategy applied by the Chinese and Bilingual groups, with which processing 7-based week calculations and 12-based month calculations tends to take similar efforts. Moreover, differences in RTs between the Bilingual group and the English group and between the English group and the Chinese were bigger in the Month Calculation Task than that in the Week Calculation Task. This means that the Chinese speakers spent relatively little extra effort to solve month calculations compared with solving week calculations, while the English group found it much harder to process month than week calculations. Similar results were found in the accuracy scores across the three groups. The larger performance gap in month calculations of the English group than the Chinese and Bilingual groups was caused by the different strategies used by English speakers and Chinese speakers. As in the Week Calculation Task, more than half of the English speakers reported a Count strategy (51%), while Chinese speakers (81% for the Chinese group and 75% for the Bilingual group) showed heavier reliance on numerical arithmetic strategy, even when tested in English with an opaque calendar system. Therefore, it appears that the more demanding the calculation task was, the more efficient the Calculate strategy might be.

5.3.2 Effects of Distance, Direction, Boundary, and Input across Groups

No effects of Distance, Direction, and Input were found in English speakers' RTs in the Month Calculation Task. This may be because more English speakers reported the use of Calculate strategy when processing month calculation questions, especially in numerical input trials. English participants might find that imaging a picture with 12 month names and finding out the target month or counting in a range of 12 names is harder than doing so in the 7-based weekdays list. As a result, they tended to try an alternative efficient strategy Calculate, which also resulted in a significant negative effect of boundary crossing on RTs of the English group, similar to the Chinese speakers. Nevertheless, negative effects of Distance, Direction were still found in English speakers' accuracy, since there were still 51% of English participants who chose the Count strategy and 12% chose the Memory strategy.

In the Chinese group, RTs were reported to be influenced by the factors Distance and Direction, while accuracy was found not to be affected by distance and directionality. Moreover, in the Bilingual group, accuracy was also not influenced by distance and directionality, while RTs were negatively influenced by long distance trials and backward direction trials in boundary cross condition but not in boundary within condition. This is in line with the results of Bassetti et al.'s (2018) study. The interaction between distance and boundary crossing and between direction and boundary crossing helped to further clarify the reason of the negative cross-boundary effect; that is, long and backward boundary crossing trials require an additional calculation step, as the simple subtraction answer needs to be further processed to get the final answer. One example for long direction is when solving a long boundary crossing trial "September + 10 months = ?", Chinese speakers tend to add 10 to 9 and get the answer 19, then subtract 12 to get the answer 7. i.e., July. In short boundary crossing condition, e.g., "September + 4 months = ?", the calculation is "9 + 4 = 13 – 12 = 1", where the additional subtraction step is easier to process. The strategy Count was also popular and efficient among the Chinese speakers when solving such calculation questions. Additionally, for example, in backward direction, when solving trial "May + 9 months = ?", the calculation is "5 + 9 = 14", which needs to be further processed by subtracting

12 and getting the answer 2, i.e., February. When solving trial "May – 9 months = ?", the calculation is "5 – 9 = -4", plus an additional step "-4 + 12 = 8 (August)". The results thus suggest that it is more difficult to process boundary crossing calculations that result in negative numbers than positive numbers over 12, and more difficult to process boundary crossing resulting in numbers further away from 12 than numbers near 12.

5.3.3 Self-reported Strategies

The analysis of self-reported strategies confirms the English group's preference for the Count strategy, yet again. However, compared with the Week Calculation Task (23%), more English respondents (32%) reported the numerical operation strategy in the Month Calculation Task. This is either because the English language conventionally represents months in a numerical way in written documents, like cheques (Kelly et al., 1999), or because month calculation questions were more demanding and the English group preferred a more efficient calculation method. By collecting the English groups' self-reported strategies separately for different conditions, the present study reveals that the strategy Calculate was more popular in long trials than in short trials, and the strategy Count was more popular in boundary crossing than within boundary. This finding was in line with the results of the Week Calculation Task. The mixed use of various strategies suggests that factors other than the transparency or opacity of calendar terms, such as the efficiency of the problem-solving method, have potential influences on reasoners' strategies used in calendrical calculation questions (Bassetti et al., 2018).

The Chinese group tended to prefer the numerical operation strategy, which extended to the Bilingual group. This agrees with Kelly et al.'s (1999) results of testing monolingual Chinese speakers, but differs from Huang's (1999) findings, where Chinese speakers preferred a mental listing strategy when dealing with numerical and linguistic opaque lunar calendar naming systems. Bassetti and her colleagues (2018) argue that this is because solar and lunar calendar represent different calendar units, while Chinese and English month lexicons are simply different expressions for same month units. To be specific, English and Chinese solar system naming the 12 months in a number sequencing, while Chinese lunar calendar terms are created based on *Jieqi*. In the Chinese lunar year, there are 24 *Jieqi*, with about 15 days (half a solar month) between two *Jieqi* (Parise, 1982). Specific nonnumerical terms are used to represent different *Jieqi* (Huang, 1999), such as *Xiazhi* (the Summer Solstice) and *Daxue* (the Heavy Snow). This may encourage the Chinese-English bilinguals to use the same strategy, numerical operation calculations, to solve month calculation questions in English and Chinese versions of the task. Another explanation relates to language dominance, which can affect bilinguals' performance in mathematical tasks (Rasmussen, Ho, Nicoladis, Leung, and Bisanz, 2006). As late Chinese learners of English, bilingual participants in the present study may have stuck to the strategy from their native language to represent and process calendrical calculation questions, maybe as a cognitive relief strategy.

Furthermore, although the translation process of the Bilingual group is widely discussed in the present study and previous research (Bassetti et al., 2018), the percentages of strategy Translate and Calculate (6% in week and 6% in month calculation questions) and strategy Translate and Count (3% and 5% in month calculation task) were rather low. Some bilingual respondents said that they hesitated to decide whether there was a translation process when solving the calendrical calculation questions, as the process could be subconscious.

5.4 Summary of the discussion

In summary, the present study finds that the linguistic transparency of L1 calendar naming system can affect bilinguals' performance in a second language where calendar terms are opaque. Moreover, while a numerically and linguistically opaque calendar system leads to a preference for verbal listing strategy in English speakers, the transparent calendar lexemes/labels lead to a greater reliance on numerical processing strategy in Chinese speakers. Bilingual speakers knowing both transparent and opaque calendar systems reported a numerical operation strategy that is more readily afforded by their mother tongue, which is in line with a previous study by Basseti et al., (2018). The self-reported reliance on numerical calculation strategy by the Bilingual group agrees with the absence effect of Distance and Direction on bilinguals' means of RTs and accuracy and the phenomenon that the Bilingual group performance was significantly negatively influenced by crossing boundary trials in week and month calculation questions.

The advantage of knowing a linguistic and numerically transparent calendar naming system helped the Chinese-English bilinguals perform equally well as the English native speakers in the Week Calculation Task, even in the task was performed in a weaker language. Additionally, in the Month Calculation Task, where there is a longer list of calendrical labels, the Bilingual group outperformed the English group. This reveals that strategy Calculate is more efficient in a more demanding task. Though prior studies claimed that L1 performance is generally worse than L1 performance, especially in unbalanced late bilinguals (van Gelderen et al., 2004; Silva and Clahsen, 2008; Jiang, 2013; Trenkic and Warmington, 2019), this study demonstrates that when processing in the L2, bilinguals can be as fast and accurate as native speakers under the facilitative influence of a transparent calendar naming system.

6. Conclusion

6.1 The design and main findings of the research project

This study involved two experiments, the Calendrical Calculation Task and the Self-reported Strategies. The first experiment was further divided into Week, Month, Hour, and Year Calculation Tasks. The week and month calculation questions investigated participants' performance in less and more demanding calculations, while the hour and year calculation questions tested participants' mathematical abilities in uncommon and common arithmetical calculations. The second experiment collected participants' calculation strategies used to solve the calendrical calculation questions, where the English speakers mainly reported a verbal listing strategy while most Chinese speakers relied on numerical operation strategy, no matter if tested in L1 Chinese or L2 English.

The findings demonstrate that knowing a linguistically and numerically transparent calendar naming system can positively affect performance in calendrical calculation tasks. Answer for the first research question is that there are significant differences between the English group and the Chinese group, when tested in their respective native languages. The RTs of the Chinese group were shorter than those of monolingual English speakers, arguably because of the transparent Chinese calendar terms and the preference for numerical operation strategy. The second research question targeted whether the performance in calendrical calculation task of Chinse-English bilinguals differs from that of English native speakers when tested in English. The results showed that there was no significant difference between RTs of the English group and the Bilingual group, indicating that the disadvantage of doing calculation tasks in a weaker language may be counteracted by the advantage of knowing transparent calendar lexemes and applying a numerical operation strategy.

6.2 Implications of this study

Implications of this work involve three areas, including implications for temporal reasoning research, for linguistic relativity research, and for bilingual cognition research.

6.2.1 Implications for temporal reasoning research

This study corroborates previous findings in that there are cross-linguistic differences in calendar calculations. By looking at the English monolingual speakers separately, the problem-solving pattern emerging from the present study was in line with the model proposed by Friedman (1990), where the verbal listing process and imaginary picture of the calendar terms list co-existed. However, the model preferred by English native speakers was not universal but presumably specific to speakers of languages with an opaque calendar naming system. The strategy used by the Chinese was quite different, where strategy Calculate was most popular and there was almost no movement from one strategy to multiple strategies (Kelly et al., 1999). The numerical operation strategy, which is established at an early age for calculations, tends to persist into adulthood and has become the Chinese group's dominant strategy.

Furthermore, the preference for verbal listing strategy was also not found in Chinese-English bilinguals even when tested in English. Having the knowledge of numerically and linguistically transparent calendar terms from Chinese and opaque calendar names from English, the Chinese-English bilinguals still showed an almost exclusive reliance on the numerical operation strategy, though strategy Count (verbal listing) and Memory (imagery) was sometimes used for higher efficiency, such as in short distance and cross-boundary trials.

Additionally, participants' different performance in calendrical calculation tasks appeared to depend on the level of transparency of English and Chinese calendar terms for two reasons. Firstly, unlike different performances found in mathematical calculation tasks (Cheng and Chan, 2005; Miller, Kelly, and Zhou, 2005; Rasmussen et al., 2006; Ng and Rao, 2010), the cross-linguistic differences in calendrical calculation tasks demonstrated by the present study were caused by linguistic instead of cultural or educational factor, because calendar calculations are not taught or tested in one's school days (Bassetti et al., 2018). Secondly, the English, Chinese, and Bilingual group performed similarly in the control conditions, namely Hour and Year Calculation Tasks, in RTs, accuracy, and self-reported strategies, so different groups' preferences for various strategies tended not to be linked to differences in arithmetic habits.

6.2.2 Implications for linguistic relativity research

The present study contributes to the theoretical development of linguistic relativity by revealing the effects of linguistic transparency of calendar terms on an important aspect of temporal reasoning and problem-solving, calendrical calculation, which has so far been limited to previous studies with bilinguals (e.g., Bassetti et al., 2018). Moreover, research on bilinguals is essential to the investigation of the relationship between language and thought, as speakers of a specific language and bilinguals can be tested in the same language. As a result, the potential influence of various testing languages can be eliminated, and different conceptual processes are more likely to be regarded as the reason for different performances of speakers with various language backgrounds.

The results of this study support the linguistic relativity hypothesis in that language can affect speakers' cognitive processes that are not necessarily linguistic. Firstly, under the influence of different levels of transparency of calendar terms in English and Chinese, native English speakers preferred the verbal listing process, while speakers of Chinese relied on the numerical operation strategy. Secondly, the application of the numerical calculation process due to Chinese transparent calendar terms was still being used by Chinese-English bilinguals, even when they performed calendrical calculation tasks in a second language with opaque calendar terms. It appears that linguistic differences across various languages can not only affect their native speakers' thoughts or cognitive routines in the corresponding native language environment, but also influence the bilinguals' thoughts or cognitive routines when they face a second language with opposite (opaque rather than transparent) linguistic characteristics.

6.2.3 Implications for bilingual cognition research

Perhaps the greatest contribution of the present study to bilingual cognition research is that this study overturns the traditional statements that native speakers tend to perform better (faster and more accurately) than the late or adult L2 learners. RTs and accuracy achieved by late bilinguals were closely aligned with, if not faster and higher than, those of the native speakers in week and month calculation tasks. Additionally, this study also finds that the English group found it much harder to solve month than week calculation questions, while the Chinese speakers did not need to make extra effort. Similar evidence comes from research of Bassetti et al. (2018), who argues that the disadvantages inherent in performing in a weaker language could be eclipsed when the native language provides a more efficient problem-solving method and the task is sufficiently complex.

The present study also provides an alternative explanation of cognitive development in bilingualism. The traditional bilingualism distinctions of subordinate, coordinate, and compound bilinguals (Weinrich, 1953) implies a cognitive consequence of bilinguals' failure to assimilate the L2 concepts, failure to acquire two separate concepts, and failure to create a novel concept on the basis of L1 and L2 concepts. In addition, the traditional expressions of "translate" and "transfer" indicate that bilinguals are unable to behave like native speakers of the target language. However, the results of this study proposed that instead of being unable to acquire the target language's concepts, the bilingual may intentionally adopt the strategy from their native language and reject the application of a strategy from the target language, simply because multi-competent individuals (Cook, 2012) can autonomously choose the more efficient strategy provided by their native language.

6.3 Limitations of the research

There are at least two limitations in the present study. Firstly, due to the Covid-19, participants were invited to perform the Calendrical Calculation Task and Self-reported

Strategies Task online and all by themselves. Therefore, the researcher could only provide written instructions online instead of face-to-face guidance and monitoring. Plus, failing to recruit participants in a controlled environment, the researcher could not guarantee that all participants took the two tasks in a quiet and separated room with no disturbance, which could inhibit the participants' best performance. Secondly, the Self-reported Strategies Task could be too subjective, and participants could be unsure or may have wanted to conceal the true strategies used to solve the calculations. For instance, several bilingual participants responded that they were struggling in deciding whether they translated the questions or not as the calculation time was too short for a competent introspection. It is also possible that the bilinguals thought the translation process implied they were less proficient language learners and the researcher might want more proficient learners of English. As a result, the situations that participants may try to please the researcher by giving the responses that they think are expected, known as halo effect, cannot be ruled out and may have negatively affected the internal validity of the present study.

6.4 Suggestions for future research

Evidence from the present study and prior study (Bassetti et al., 2018) shows that the disadvantages of performing in a weaker language can be neutralized when the task is complex enough and the native language provides a more efficient way of problem solving. Therefore, it could be a fruitful direction if future research on late second language learners replaced simple tasks with more complex tasks to investigate bilinguals' cognitive processes under greater demand for attentional resources. Additionally, while the present study investigated the effects of L1 Chinese transparent calendar terms on the Chinese-English bilinguals' performance in L2 English with opaque calendar lexicons, there is no research in the opposite direction, at least to the best of the researcher's knowledge, on how bilinguals who know an L1 with opaque calendar terms perform in an L2 with a transparent calendar naming system in calendrical calculation task.

7. Appendices

Appendix 1: Background Information Questionnaire

For English Speakers

School of Cultures, Languages, and Linguistics
Faculty of Arts | University of Auckland
1010 | Auckland | New Zealand

Arts 2 Building
18 Symonds Street
Auckland 1010, New Zealand
Telephone 64 9 373 7599
The University of Auckland
Private Bag 92019
Auckland 1142 New Zealand

Calendrical calculations in first and second language: the effects of transparency of first language calendar terms

Ziyi Zhuang
MA candidate, Applied Linguistics, University of Auckland

Dear participants:

I am an MA candidate in Applied Linguistics at the University of Auckland in New Zealand, and I am now collecting data for my Masters thesis. The research topic of my thesis is the calendrical calculations in first and second language: the effects of transparency of first language calendar terms. This questionnaire is confidential and only used for the data analysis of the study. It will treat personal privacy and personal views with confidence, and it will not be used for any commercial purposes. It only takes you a little time to fill in the questionnaire. Please fill it in carefully. Your participation is very important to my research. Thank you very much for your cooperation.

Ziyi Zhuang
November 2021

Basic information
● Gender: ( ) A. Male B. Female C. Another gender (please specify): ____________
● Age： Nationality: Majors：
● Education background: ( ) A. Bachelor B. Master C. Ph.D.
● Which day do you consider the start of the week? ( ) A. Sunday B. Monday

English level
● Is English your first language?
A. Yes B. No

- When did you formally start learning English? (   )

  A. kindergarten   B. 1st-2nd grade in elementary school

  C. 3rd-5th grade in elementary school   D. middle school   E. High school   F. university

- What is the current frequency of your English use? (   )

  A. I currently use English in daily life

  B. I currently use English in certain situations (such as work and study) and other language(s) in other situations

  C. I currently only use language(s) other than English in daily life

(If you choose B or C, please answer questions in next section.)

Second Language level:

- What is your second language? /What are your foreign languages?

  ____

- When did you formally start learning your second language: (   )

  A．Kindergarten   B. 1st-2nd grade in elementary school

  C. 3rd-5th grade in elementary school   D. Middle school   E. High school   F. University

- Have you passed any test of proficiency in your second language?

  ____

- Can you count from 1 to 100 in your second language without dictionary assistance? (   )

  A. Yes   B. No

- Can you count from January to December in your second language without dictionary assistance? (   )

  A. Yes   B. No

- Can you name the days from Monday to Sunday in your second language without dictionary assistance? (   )

  A. Yes   B. No

- How many hours (self-estimated) are you in contact with your second language in one week (e.g., class preparation at work, study, reading, web browsing on mobile phone, communication with others, etc.)? (   )

  A．over 80   B.80-40   C.40-10   D. below 10

Approved by the University of Auckland Human Participants Ethics Committee on 5 November 2021 for three years. Reference Number UAHPEC23368.

For Bilingual Speakers

School of Cultures, Languages, and Linguistics
Faculty of Arts | University of Auckland
1010 | Auckland | New Zealand

Arts 2 Building
18 Symonds Street
Auckland 1010, New Zealand
Telephone 64 9 373 7599
The University of Auckland
Private Bag 92019
Auckland 1142 New Zealand

Calendrical calculations in first and second language: the effects of transparency of first language calendar terms

Ziyi Zhuang
MA candidate, Applied Linguistics, University of Auckland

Dear participants:

I am an MA candidate in Applied Linguistics at the University of Auckland in New Zealand, and I am now collecting data for my Master's thesis. The research topic of my thesis is the calendrical calculations in first and second language: the effects of transparency of first language calendar terms. This questionnaire is confidential and only used for the data analysis of the study. It will treat personal privacy and personal views with confidence, and it will not be used for any commercial purposes. It only takes you a little time to fill in the questionnaire. Please fill it in carefully. Your participation is very important to my research. Thank you very much for your cooperation.

Ziyi Zhuang

November 2021

Basic information

- Gender: ( ) A. Male B. Female C. Another gender (please specify): _____________
- Age： Nationality: Majors：
- Education background: ( ) A. Bachelor B. Master C. Ph.D.
- Which day do you consider the start of the week? ( ) A. Sunday B. Monday

Chinese level

- Is Chinese your first language?

  A. Yes B. No
- When did you formally start learning Mandarin Chinese? ( )

  A. Kindergarten B. 1st-2nd grade in elementary school

  C. 3rd -5th grade in elementary school D. Middle school E. High school F. University
- What is the current frequency of your Mandarin Chinese use? ( )

  A．I currently use Mandarin Chinese in daily life

  B. I currently use Mandarin Chinese in certain situations (such as work and study) and English in other situations

  C. I currently use English only or mostly in daily life

- How many hours (self-estimated) are you in contact with Mandarin Chinese in one week (e.g., class preparation at work, study, reading, web browsing on mobile phone, communication with others, etc.)? ( )

  A．over 80 B.80-40 C.40-10 D. below 10

English level：

- When did you formally start learning English? ( )

  A．Kindergarten B. 1st -2nd grade in elementary school

  C. 3rd -5th grade in elementary school D. Middle school E. High school F. College

- English test：( )

(If you have attended courses of English Pathway for Postgraduate Studies (EPPS) and got a score over B-, please choose C.)

(If you have attended English Academic Program (EAP) and got a score over B+, please choose C.)

(If you have not taken IELTS within the last 3 years, please answer the next question.)

  A. IELTS 5.5/ TOEFL 46-59 B. IELTS 6/ TOEFL 60-78 C. IELTS 6.5/ TOEFL 79-93 D. IELTS 7/ TOEFL 94-101

  E. ILETS over 7/ TOEFL over 102

- When did you formally start learning in a school using English as a medium of instruction: ( )

  A．Kindergarten B. 1st -2nd grade in elementary school

  C. 3rd -5th grade in elementary school D. Middle school E. High school F. University

- How long have you been living in an English-speaking country?

  A. less than 1 year B. 1 year C. 2 years D. 3 years D. 4 years E. 5 years F. more than 5 years

- What is the current frequency of your English use? ( )

  A. I currently use English in daily life

  B. I currently use English in certain situations (such as work and study) and Mandarin Chinese in other situations

  C. I currently use Mandarin Chinese only or mostly in daily life

- How many hours (self-estimated) are you in contact with English in one week (e.g., class preparation at work, study, reading, web browsing on mobile phone, communication with others, etc.)? ( )

  A．over 80 B.80-40 C.40-10 D. below 10

Approved by the University of Auckland Human Participants Ethics Committee on 5 November 2021 for three years. Reference Number UAHPEC23368.

Appendix 2: Advertisement Email Script

School of Cultures, Languages, and Linguistics
Faculty of Arts | University of Auckland
1010 | Auckland | New Zealand

Arts 2 Building
18 Symonds Street
Auckland 1010, New Zealand
Telephone 64 9 373 7599
The University of Auckland
Private Bag 92019
Auckland 1142 New Zealand

Email script-students

Project title: Calendrical calculations in first and second language: the effects of transparency of first language calendar terms
Name of Supervisor: Dr Norbert Vanek
Name of Researcher: Ziyi Zhuang

Dear potential participant,
I am Ziyi Zhuang, a graduate student of MA in Applied Linguistics at the University of Auckland. I am the researcher for this project on the effects of transparency of first language (L1) calendar terms on participants' calendrical calculation performance in L1 and second language (L2). This study aims to investigate the effects of transparency of first language (L1) calendar terms on calendar calculations in a second language (L2). It will contribute to the research of linguistic relativity, numerical and temporal reasoning process, and bilingual cognition. If you are Chinese-dominant learners of English, it will benefit you in your understanding of how the transparency of their L1 calendar terms affect the calendrical calculations in L2. If you are a monolingual English speaker, it will help you get a deep understanding of the mechanism of linguistic relativity from the perspective of temporal reasoning.

You will complete questionnaires collecting their background information, including age, gender, teaching subjects, nationality, L1 and L2 proficiency. It will take you approximately 3 minutes to complete it. You will also complete two tasks (i.e., Calendar Calculation and Self-reported Strategies Task). The two tasks will take you approximately 35 minutes to complete. I would like to invite you to participate in this research project. If you are a Chinese-English bilingual, you need to do the two tasks again in Chinese one week later, it will take you approximately 35 minutes to complete.

If you have any questions about this research, please contact me at zzhu737@aucklanduni.ac.nz. If you are interested in this research or require more information, please read the Participant Information Statement with further information and assurances about the project as attachment. If you agree to participate, please sign on the Consent Form as attached and send the electronic Consent Form back to me.

Thank you very much for considering this request,
Ziyi Zhuang

Graduate student of MA in Applied Linguistics
School of Languages, Cultures, and Linguistics
Faculty of Arts, the University of Auckland
Private Bag 92019
Auckland 1142
Email: zzhu737@aucklanduni.ac.nz

Approved by the University of Auckland Human Participants Ethics Committee on 5 November 2021 for three years. Reference Number UAHPEC23368.

Appendix 3: Participant Information Sheet

School of Cultures, Languages, and Linguistics

Faculty of Arts | University of Auckland

1010 | Auckland | New Zealand

Arts 2 Building
18 Symonds Street
Auckland 1010, New Zealand
Telephone 64 9 373 7599
The University of Auckland
Private Bag 92019
Auckland 1142 New Zealand

Project Information Sheet - Student

Project title: Calendrical calculations in first and second language: the effects of transparency of first language calendar terms
Name of Supervisor: Dr Norbert Vanek
Name of Researcher: Ziyi Zhuang

Dear Participant,

I am Ziyi Zhuang, a graduate student of MA in Applied Linguistics at the University of Auckland. I am the researcher for this project on the effect of transparency of first language (L1) calendar terms on participants' calendrical calculation performances in L1 and second language (L2).

Project description and invitation
Linguistic relativity hypothesizes that differences in languages impact the way the speakers make sense of the world (Whorf, 1940) because the same concept can be interpreted and represented differently across various language communities. One aspect of linguist relativity that varies cross-linguistically but has been seldomly investigated tends is numerical (Cheng and Chan, 2005) and temporal (Bassetti, Clarke, and Trenkic, 2018) reasoning, which can be tested through calendrical calculation task and self-reported strategies task. The study aims to investigate the

effects of transparency of L1 calendar terms on calendar calculations in a L2 (with opaque calendar items). This study will contribute to the research of linguistic relativity, temporal cognition, and bilingual cognition. It will benefit Chinese-dominant learners of English in their understanding of how the transparency of their L1 calendar terms affects the calendrical calculations in L2 English. It will also help native English speakers get an understanding of the mechanisms of crosslinguistic influence in the area of calendrical reasoning.

This study aims to recruit 20 Chinese-dominant learners of English and 20 native English speakers. If you are an advanced Chinese-dominant leaner of English (6.5 scores in IELTS), or if you are a native English speaker who has no experience or little exposure to a second language in your daily life, I would like to invite you to take part in this study.

Project procedures

Your participation will involve a questionnaire and two tasks. The tasks will be web-based so you can choose the optimal time at your convenience. If you are a native English speaker, the research will take you approximately 40 minutes in total to complete. If you are a Chinese-English bilingual, the research will take you approximately 80 minutes in total to complete, as you will be asked to do the tasks in Chinese one week after you have done them in English.

I would like you to complete a questionnaire, only collecting your basic background information, including age, gender, education background, nationality, and L1 and L2 proficiency. Codes will be used on the questionnaires so that participants are de-identified. All the information will be confidential and only accessible to the researcher. It will take every participant approximately 3 minutes to complete.

I would also like you to complete Task 1. Task 1 is Calendar Calculation. It involves 96 calendar reasoning exercises, including Weekday, Month, Hour, and Year calculations, each with 4 questions in four condition (short Forward, long Forward, short Backward, and long Backward) and two description versions (numerical and linguistic version). Forward condition refers to calculating dates/time chronologically, while Backward condition refers to calculations with reversed chronology. The participants need to answer questions shown on the computer screen by pressing one of the three buttons (left, right and upward arrow) as fast and accurately as possible. A specialized software will be used for this purpose. A weblink to the experiment will be emailed to the participants by the researcher once they have expressed consent to participate. The computer will record their reaction times (RTs) and responses (button press choices). This task will take every participant approximately 20 minutes to complete.

I would also like you to complete Task 2. Task 2 is Self-reported Strategies task. It involves 48 questions about the strategies used by the participants in different types of calendar reasoning. Participants need to decide which strategy shown on the computer screen they used. The participants can take their time since reaction times for this task will not be recorded. This task will take every participant approximately 5 minutes to complete.

If you are an advanced Chinese-dominant learner of English, I would like you to

complete Task 1 and 2 again in Chinese one week after you have completed these tasks in English.
If you express interest, I will email you a PIS and a consent form as attachments and respond to their questions should there be any. Interested participants will be asked to submit the consent form to the researcher via email, based on which they will be sent a Qualtrics weblink to the questionnaire and a PsychoPy weblink to Tasks 1 and 2.

Data storage/ retention / destruction/ future use
Once you have completed the Calendar Calculation task, you will have the right to know your accuracy and RTs. The data will be stored as electronic data. The consent forms and the results of questionnaires, Calendar Calculation task, and Self-reported Strategies task will be transferred to a University of Auckland (UoA) administered Google Drive Account, in accordance with University policy, and accessed through a password-protected computer used only by the researcher. All the data will be confidential and only accessible to the researcher and her supervisor. The de-identified data may be used in presentations, in research reports, in publications, online or similar. RTs records and self-reported strategies will not be published online in public. The data will be kept for six years, after which point it will be destroyed.

Voluntary participation
Participation in this study is optional. If you decide to take part, you will be given a copy of this information sheet for your records and will be asked to send me an electronic consent form. If you do not wish to participate, please let me know now and do not sign the consent form. Your participation will contribute to the study of the effect of transparency of L1 calendar terms on temporal reasoning and numerical cognition. It is important to know that if you decide to participate or not to participate, the Dean of the Faculty of Arts has provided an assurance that participation or non-participation will have no consequences on your learning experience, grades or the relationship between you and the school. Likewise, the researcher's own students' participation or nonparticipation will not affect their grades and their relationship with their teacher and the university alike.

Right to withdraw from participation
You can withdraw your participation in the study or the data from the study at any time before the end of data collection without giving a reason.

Anonymity and confidentiality
Your identity will be kept confidential throughout the study. The researcher will de-identify and analyse the data. In the reports of this study, only group averages will be given, no individual responses or reaction times. Participants will be referred to as a group, not individually. Information about age, gender, the length of learning English, Chinese and English proficiency, and foreign language proficiency will be aggregated, again, lowering the likelihood of identifiability. Any information used in

the MA thesis, in any conference presentation, or publication from this research will not identify you or your institution as its source.
The risk of you being identifiable is low as there is no intention to name the Faculty in research outputs. Instead, a generic description, i.e., 'Chinese-English bilingual students and English-speaking students from a university in New Zealand,' will be used, and the participants will be coded as BS (bilingual students)1,2,3….. And MES (monolingual English students)1,2,3…… rather than their real names. The data gathered from the study (i.e., the completed questionnaire, accuracies and RTs, and self-reported strategies) will be stored under a code number. Any information that identifies students will be stored separately from the data and de-identified after completing data collection.

Contact Details

If you require more information about the study, please contact the researcher. Māori students are welcome to use cultural support services available at the Faculty by emailing tuakanaarts@uoa.auckland.ac.nz or artsengagement@auckland.ac.nz.

Researcher	**Supervisor**	**Head of Schools**
Ziyi Zhuang Graduate student of MA in Applied Linguistics School of Languages, Cultures, and Linguistics The University of Auckland Private Bag 92019 Auckland 1142 Email: zzhu737@aucklanduni.ac.nz	Dr. Norbert Vanek Applied Linguistics and Language Teaching School of Languages, Cultures, and Linguistics The University of Auckland Private Bag 92019 Auckland 1142 Email: norbert.vanek@auckland.ac.nz	Prof. Martin East School of Cultures, Languages and Linguistics The University of Auckland Private Bag 92019 Auckland 1142 Email: m.east@auckland.ac.nz

UAHPEC Chair contact details

For any queries regarding ethical concerns, you may contact the Chair, The University of Auckland Human Participants Ethics Committee, Office of Research Strategy and Integrity, The University of Auckland, Private Bag 92019, Auckland 1142. Telephone 09 373-7599 ext. 83711. Email: humanethics@auckland.ac.nz.

Approved by the University of Auckland Human Participants Ethics Committee on 5 November 2021 for three years. Reference Number UAHPEC23368.

Appendix 4: Consent Form

School of Cultures, Languages, and Linguistics
Faculty of Arts | University of Auckland
1010 | Auckland | New Zealand

Arts 2 Building
18 Symonds Street
Auckland 1010, New Zealand
Telephone 64 9 373 7599
The University of Auckland
Private Bag 92019
Auckland 1142 New Zealand

Consent Form – Students

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS

Project title: Calendrical calculations in first and second language: the effects of transparency of first language calendar terms
Name of Supervisor: Norbert Vanek
Name of Student Researcher: Ziyi Zhuang

I have read the Participant Information Sheet and have understood the nature of the research project. I have had the opportunity to ask questions about the research and have had them answered to my satisfaction.

- I agree to take part in this research.
- I understand that my participation is voluntary and that the principal of my school has given an assurance that the participation or non-participation will have no consequences on my learning experience or grades.
- I agree to allow the researcher to collect my background information only for research purposes. I understand that it will take approximately 3 minutes for me to complete.
- I agree to complete a Calendar Calculation Task. I understand that it will take approximately 25 minutes for me to complete. I understand that my reaction times in answering the calculation questions shown on the screen will be recorded by yhe computer.
- I agree to complete a Self-reported Strategies task. I understand that it will take 10 minutes for me to complete.

- (For Chinese-English bilinguals) I agree to complete the Calendar Calcualtion and Self-reported Strategies tasks again in Chinese. I understand that it will take extra 35 minutes for me to complete.
- I understand that the data collected for this research will be kept for six years and then destroyed. I also understand that my Consent Form will be kept separately from any data.
- I understand that I am free to withdraw my participation at any time and to withdraw any data traceable to me up to 1st December.
- I understand that no identifying data on me and my institution will be reported in the MA thesis or in any publication or presentation resulting from this research.
- I wish/ do not wish to receive a summary of research findings.

Name: ___________________________

Signature: ___________________________ Date: _______________

Approved by the University of Auckland Human Participants Ethics Committee on …… for three years. Reference Number 23368.

Appendix 5: Ethics Approval

The University of Auckland
Private Bag 92019
Auckland, New Zealand

Level 3, 49 Symonds Street
Auckland, New Zealand
Telephone 86356
Facsimile +64 9 373 7432

UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE (UAHPEC)

05/11/2021

Dr Norbert Vanek

Re: Application for Ethics Approval (Our Ref. UAHPEC23368): Approved

The Committee considered your application for ethics approval for the study entitled "**Calendrical calculations in first and second language: the effects of transparency of first language calendar terms**".

We are pleased to inform you that ethics approval has been granted for a period of three years.

The expiry date for this approval is **05/11/2024**.

Completion of the project: In order that up-to-date records are maintained, you must notify the Committee once your project is completed.

Amendments to the approved project: Should you need to make any changes to the approved project, please follow the steps below:

- Send a request to the UAHPEC Administrators to unlock the application form (using the Notification tab in the Ethics RM form).
- Make all changes to the relevant sections of the application form and attach revised documents (as appropriate).
- Change the Application Type to "Amendment request" in Section 13 ("Submissions and Sign off").
- Add a summary of the changes requested in the text box.
- Submit the amendment request (PI/Supervisors only to submit the form).

If the project changes significantly, you are required to submit a new application.

Funded projects: If you received funding for this project, please provide this approval letter to your local Faculty Research Project Coordinator (RPC) or Research Project Manager (RPM) so that the approval can be notified via a Service Request to the Research Operations Centre (ROC) for activation of the grant.

The Chair and the members of UAHPEC would be happy to discuss general matters relating to ethics approvals. If you wish to do so, please contact the UAHPEC Ethics Administrators at humanethics@auckland.ac.nz in the first instance.

Additional information:

- Do not forget to fill in the 'approval wording' on the PISs, CFs and/or advertisements, using the date of this approval and the reference number, before you use the documents or send them out to your participants.

All communications with the UAHPEC regarding this application should indicate this reference number: **UAHPEC23368**.

UAHPEC Administrators

University of Auckland Human Participants Ethics Committee

c.c. , Ziyi Zhuang

Page 1 of 1

Appendix 6: Stimuli

The English Version

The Practice Trials

exercise	answer
Monday + 3 days = ? Tuesday Wednesday Thursday	right
August - 12 months = ? August September October	up
ten o'clock + 18 hours = ? three o'clock four o'clock five o'clock	left
2012 + 7 years = ? 2017 2018 2019	right

The Test Trials

word	corrAns
Monday + 1 day = ? Tuesday Wednesday Thursday	up
Wednesday + 2 days = ? Thursday Friday Saturday	left
Friday + 3 days = ? Sunday Monday Tuesday	left
Sunday + 2 days = ? Tuesday Wednesday Thursday	up
Monday + 6 days = ? Saturday Sunday Monday	left
Wednesday + 4 days = ? Sunday Monday Tuesday	up

word	corrAns
1st month + 2 months = ? February March April	left
3th month + 6 months = ? July August September	right
8th month + 5 months = ? January February March	up
11th month + 4 months = ? February March April	left
2nd month + 9 months = ? October November December	left
5th month + 7 months = ? October November December	right

Friday + 5 days = ? Tuesday Wednesday Thursday	left
Saturday + 6 days = ? Wednesday Thursday Friday	right
Saturday - 1 day = ? Wednesday Thursday Friday	right
Friday - 2 days = ? Tuesday Wednesday Thursday	left
Monday - 2 days = ? Friday Saturday Sunday	left
Monday - 3 days = ? Wednesday Thursday Friday	right
Saturday - 4 days = ? Tuesday Wednesday Thursday	up
Friday - 4 days = ? Monday Tuesday Wednesday	up
Thursday - 6 days = ? Wednesday Thursday Friday	right
Tuesday - 5 days = ? Wednesday Thursday Friday	left
1st day + 1 day = ? Tuesday Wednesday Thursday	up
3rd day + 2 days = ? Thursday Friday Saturday	left
5th day + 3 days = ? Sunday Monday Tuesday	left
7th day + 2 days = ? Tuesday Wednesday Thursday	up

9th month + 10 months = ? June July August	left
10th month + 8 months = ? April May June	right
7th month - 2 months = ? April May June	left
12th month - 3 months = ? September October November	up
2nd month - 6 months = ? July August September	left
4th month - 5 months = ? September October November	right
12th month - 10 months = ? January February March	left
8th month - 7 months = ? January February March	up
6th month - 9 months = ? July August September	right
3rd month - 11 months = ? March April May	left
two o'clock + 7 hours = ? eight o'clock nine o'clock ten o'clock	left
eight o'clock + 3 hours = ? eleven o'clock twelve o'clock thirteen o'clock	up
eighteen o'clock + 10 hours = ? two o'clock three o'clock four o'clock	right
twenty o'clock + 11 hours = ? five o'clock six o'clock seven o'clock	right

1st day + 6 days = ? Saturday Sunday Monday	left
3rd day + 4 days = ? Sunday Monday Tuesday	up
5th day + 5 days = ? Tuesday Wednesday Thursday	left
6th day + 6 days = ? Wednesday Thursday Friday	right
6th day - 1 day = ? Wednesday Thursday Friday	right
5th day - 2 days = ? Tuesday Wednesday Thursday	left
1st day - 2 days = ? Friday Saturday Sunday	left
1st day - 3 days = ? Wednesday Thursday Friday	right
6th day - 4 days = ? Tuesday Wednesday Thursday	up
5th day - 4 days = ? Monday Tuesday Wednesday	up
4th day - 6 days = ? Wednesday Thursday Friday	right
2nd day - 5 days = ? Wednesday Thursday Friday	left
January + 2 months = ? February March April	left
March + 6 months = ? July August September	right

four o'clock + 15 hours = ? eighteen o'clock nineteen o'clock twenty o'clock	left
one o'clock + 17 hours = ? seventeen o'clock eighteen o'clock nineteen o'clock	left
thirteen o'clock + 21 hours = ? ten o'clock eleven o'clock twelve o'clock	up
nineteen o'clock + 23 hours = ? sixteen o'clock seventeen o'clock eighteen o'clock	right
fifteen o'clock - 11 hours = ? three o'clock four o'clock five o'clock	left
twenty-three o'clock - 3 hours = ? twenty o'clock twenty-one o'clock twenty-two o'clock	up
six o'clock - 10 hours = ? nineteen o'clock twenty o'clock twenty-one o'clock	left
five o'clock - 7 hours = ? twenty o'clock twenty-one o'clock twenty-two o'clock	right
twenty-three o'clock - 12 hours = ? ten o'clock eleven o'clock twelve o'clock	left
seventeen o'clock - 15 hours = ? two o'clock three o'clock four o'clock	up
twelve o'clock - 20 hours = ? fourteen o'clock fifteen o'clock sixteen o'clock	right
six o'clock - 23 hours = ? five o'clock six o'clock seven o'clock	left
10 + 60 years = ? 70 80 90	up
45 + 6 years = ? 49 50 51	right

August + 5 months = ? January February March	up
November + 4 months = ? February March April	left
February + 9 months = ? October November December	left
May + 7 months = ? October November December	right
September + 10 months = ? June July August	left
October + 8 months = ? April May June	right
July - 2 months = ? April May June	left
December - 3 months = ? September October November	up
February - 6 months = ? July August September	left
April - 5 months = ? September October November	right
December - 10 months = ? January February March	left
August - 7 months = ? January February March	up
June - 9 months = ? July August September	right
March - 11 months = ? March April May	left

90 - 55 years = ? 25 35 45	left
60 - 18 years = ? 42 43 44	up
120 + 160 years = ? 260 270 280	right
340 + 500 years = ? 740 840 640	left
860 - 210 years = ? 640 650 660	left
900 - 700 years = ? 100 200 300	left
1985 + 5 years = ? 1980 1990 2000	left
1300 + 600 years = ? 1700 1800 1900	right
2050 - 1000 years = ? 150 1000 1050	right
2015 - 11 years = ? 2004 2005 2006	up
310 + 50 years = ? 340 350 360	right
2010 + 440 years = ? 2450 2460 2470	up
1800 - 1750 years = ? 30 40 50	right
1800 - 1400 years = ? 300 400 500	left

The Self-reported Strategies:

Question
Which strategy did you use to solve the following question? Monday + 1 day = ?
Which strategy did you use to solve the following question? Friday + 3 days = ?
Which strategy did you use to solve the following question? Monday + 6 days = ?
Which strategy did you use to solve the following question? Friday + 5 days = ?
Which strategy did you use to solve the following question? Saturday - 1 day = ?
Which strategy did you use to solve the following question? Monday - 2 days = ?
Which strategy did you use to solve the following question? Saturday - 4 days = ?
Which strategy did you use to solve the following question? Thursday - 6 days = ?
Which strategy did you use to solve the following question? January + 2 months = ?
Which strategy did you use to solve the following question? August + 5 months = ?
Which strategy did you use to solve the following question? February + 9 months = ?
Which strategy did you use to solve the following question? September + 10 months = ?
Which strategy did you use to solve the following question? July - 2 months = ?
Which strategy did you use to solve the following question? February - 6 months = ?
Which strategy did you use to solve the following question? December - 10 months = ?
Which strategy did you use to solve the following question? June - 9 months = ?
Which strategy did you use to solve the following question? two o'clock + 7 hours = ?
Which strategy did you use to solve the following question? eighteen o'clock + 10 hours = ?
Which strategy did you use to solve the following question? four o'clock + 15 hours = ?
Which strategy did you use to solve the following question? thirteen o'clock + 21 hours = ?

Which strategy did you use to solve the following question? fifteen o'clock - 11 hours = ?
Which strategy did you use to solve the following question? six o'clock - 10 hours= ?
Which strategy did you use to solve the following question? twenty-three o'clock - 12 hours = ?
Which strategy did you use to solve the following question? twelve o'clock - 20 hours = ?
Which strategy did you use to solve the following question? 1st day + 1 day = ?
Which strategy did you use to solve the following question? 5th day + 3 days = ?
Which strategy did you use to solve the following question? 1st day + 6 days = ?
Which strategy did you use to solve the following question? 5th day + 5 days = ?
Which strategy did you use to solve the following question? 6th day - 1 day = ?
Which strategy did you use to solve the following question? 1st day - 2 days = ?
Which strategy did you use to solve the following question? 6th day - 4 days = ?
Which strategy did you use to solve the following question? 4th day - 6 days = ?
Which strategy did you use to solve the following question? 1st month + 2 months = ?
Which strategy did you use to solve the following question? 8th month + 5 months = ?
Which strategy did you use to solve the following question? 2nd month + 9 months = ?
Which strategy did you use to solve the following question? 9th month + 10 months = ?
Which strategy did you use to solve the following question? 7th month - 2 months = ?
Which strategy did you use to solve the following question? 2nd month - 6 months = ?
Which strategy did you use to solve the following question? 12th month - 10 months = ?
Which strategy did you use to solve the following question? 6th month - 9 months = ?
Which strategy did you use to solve the following question? 10 + 60 years = ?

Which strategy did you use to solve the following question? 60 - 18 years = ?
Which strategy did you use to solve the following question? 120 + 160 years = ?
Which strategy did you use to solve the following question? 900 - 700 years = ?
Which strategy did you use to solve the following question? 1985 + 5 years = ?
Which strategy did you use to solve the following question? 2050 - 1000 years = ?
Which strategy did you use to solve the following question? 310 + 50 years = ?
Which strategy did you use to solve the following question? 1800 - 20 years = ?

1. **Memory** (Automatic recall of the fact, you didn't need to work out the answer)
2. **Transform and calculate** (e.g., Transform 'Monday + 2 days' into '1+2 = 3' and get the answer)
3. **Count** (e.g., 'March + 2 months' is 'March', 'April', 'May', so the answer is May)
4. **Translate and calculate** (e.g., translate the question into Chinese then calculate in Chinese)
5. **Translate and count** (e.g., translate the question into Chinese then count in Chinese)
6. **Estimate** (Guess the answer)
7. **Other** (Another method that is not listed here)

The Chinese Version

The Practice Trials:

exercise		answer
周一 + 3 天 = ? 周二 周三	 周四	right
八月 - 12 个月= ? 八月 九月	 十月	up
十点 + 18 小时= ? 三点 四点	 五点	left
2012 + 7 年= ? 2017 2018	 2019	right

The Test Trials:

word	corrAns
周一 + 1 天 = ? 周二 周三 周四	up
周三 + 2 天 = ? 周四 周五 周 六	left
周五 + 3 天 = ? 周日 周一 周二	left
周日 + 2 天 = ? 周二 周三 周四	up
周一 + 6 天 = ? 周六 周日 周 一	left
周三 + 4 天 = ? 周日 周一 周 二	up
周五 + 5 天 = ? 周二 周三 周四	left
周六 + 6 天 = ? 周三 周四 周五	right
周六 - 1 天 = ? 周三 周四 周五	right

word	corrAns
周 3 + 2 天 = ? 周四 周五 周六	left
周 5 + 3 天 = ? 周日 周一 周二	left
周 7 + 2 天 = ? 周二 周三 周四	up
周 1 + 6 天 = ? 周六 周日 周一	left
周 3 + 4 天 = ? 周日 周一 周二	up
周 5 + 5 天 = ? 周二 周三 周四	left
周 6 + 6 天 = ? 周三 周四 周五	right
周 6 - 1 天 = ? 周三 周四 周五	right
周 5 - 2 天 = ? 周二 周三 周四	left

周五 - 2 天 = ? 周二 周三 周四	left
周一 - 2 天 = ? 周五 周六 周日	left
周一 - 3 天 = ? 周三 周四 周五	right
周六 - 4 天 = ? 周二 周三 周四	up
周五 - 4 天 = ? 周一 周二 周三	up
周四 - 6 天 = ? 周三 周四 周五	right
周二 - 5 天 = ? 周三 周四 周五	left
一月 + 2 个月 = ? 二月 三月 四月	left
三月 + 6 个月 = ? 七月 八月 九 月	right

周 1 - 2 天 = ? 周五 周六 周日	left
周 1 - 3 天 = ? 周三 周四 周五	right
周 6 - 4 天 = ? 周二 周三 周四	up
周 5 - 4 天 = ? 周一 周二 周三	up
周 4 - 6 天 = ? 周三 周四 周五	right
周 2 - 5 天 = ? 周三 周四 周五	left
1 月 + 2 个月 = ? 二月 三月 四 月	left
3 月 + 6 个月 = ? 七月 八月 九月	right
8 月 + 5 个月 = ? 一月 二月 三月	up

八月 + 5 个月 = ? 一月 二月 三月	up
十一月 + 4 个月 = ? 二月 三月 四 月	left
二月 + 9 个月 = ? 十月 十一月 十 二月	left
五月 + 7 个月 = ? 十月 十一月 十 二月	right
九月 + 10 个月 = ? 六月 七月 八月	left
十月 + 8 个月 = ? 四月 五月 六月	right
七月 - 2 个月 = ? 四月 五月 六月	left
十二月 - 3 个月 = ? 九月 十月 十一 月	up
二月 - 6 个月 = ? 七月 八月 九 月	left

11 月 + 4 个月 = ? 二月 三月 四月	left
2 月 + 9 个月 = ? 十月 十一月 十二 月	left
5 月 + 7 个月 = ? 十月 十一月 十二 月	right
9 月+ 10 个月 = ? 六月 七月 八 月	left
10 月 + 8 个月 = ? 四月 五月 六 月	right
7 月 - 2 个月 = ? 四月 五月 六 月	left
12 月 - 3 个月 = ? 九月 十月 十一月	up
2 月 - 6 个月 = ? 七月 八月 九月	left
4 月 - 5 个月 = ? 九月 十月 十一月	right

四月 - 5 个月 = ? 九月 十月 十一 月	right
十二月 - 10 个月 = ? 一月 二月 三月	left
八月 - 7 个月 = ? 一月 二月 三月	up
六月 - 9 个月 = ? 七月 八月 九 月	right
三月 - 11 个月 = ? 三月 四月 五 月	left
两点 + 7 小时 = ? 八点 九点 十点	left
八点 + 3 小时 = ? 十一点 十二点 十三点	up
十八点 + 10 小时 = ? 两点 三点 四点	right
二十点 + 11 小时 = ? 五点 六点 七点	right

12 月 - 10 个月 = ? 一月 二月 三月	left
8 月 - 7 个月 = ? 一月 二月 三月	up
6 月 - 9 个月 = ? 七月 八月 九月	right
3 月 - 11 个月 = ? 三月 四月 五月	left
六点 - 10 小时 = ? 十九点 二十点 二 十一点	left
10 + 60 年 = ? 70 80 90	up
45 + 6 年 = ? 49 50 51	right
90 - 55 年 = ? 25 35 45	left
60 - 18 年 = ? 42 43 44	up

四点 + 15 小时 = ? 十八点 十九点 二十点	left
一点 + 17 小时 = ? 十七点 十八点 十九点	left
十三点 + 21 小时 = ? 十点 十一点 十二点	up
十九点 + 23 小时 = ? 十六点 十七点 十八点	right
十五点 - 11 小时 = ? 三点 四点 五点	left
二十三点 - 3 小时 = ? 二十点 二十一点 二十二点	up
五点 - 7 小时 = ? 二十点 二十一点 二十二点	right
二十三点 - 12 小时 = ? 十点 十一点 十二点	left
十七点 - 15 小时 = ? 两点 三点 四点	up

120 + 160 年 = ? 260 270 280	right
340 + 500 年 = ? 740 840 640	left
860 - 210 年 = ? 640 650 660	left
900 - 700 年 = ? 100 200 300	left
1985 + 5 年 = ? 1980 1990 2000	left
1300 + 600 年 = ? 1700 1800 1900	right
2050 - 1000 年 = ? 150 1000 1050	right
2015 - 11 年 = ? 2004 2005 2006	up
310 + 50 年 = ? 340 350 360	right

十二点 - 20 小时 = ? 十四点 十五点 十六点	right
六点 - 23 小时 = ? 五小时 六小时 七小时	right
周 1 + 1 天 = ? 周二 周三 周四	up

2010 + 440 年 = ? 2450 2460 2470	up
1800 - 1750 年 = ? 30 40 50	right
1800 - 1400 年 = ? 300 400 500	left

The Self-reported Strategies:

Question
Which strategy did you use to solve the following question? 周一 + 1 天 = ?
Which strategy did you use to solve the following question? 周五 + 3 天 = ?
Which strategy did you use to solve the following question? 周一 + 6 天 = ?
Which strategy did you use to solve the following question? 周五 + 5 天 = ?
Which strategy did you use to solve the following question? 周六 - 1 天 = ?
Which strategy did you use to solve the following question? 周一 - 2 天 = ?
Which strategy did you use to solve the following question? 周六 - 4 天 = ?
Which strategy did you use to solve the following question? 周四 - 6 天 = ?
Which strategy did you use to solve the following question? 一月 + 2 个月 = ?
Which strategy did you use to solve the following question? 八月 + 5 个月 = ?
Which strategy did you use to solve the following question? 二月 + 9 个月 = ?

Which strategy did you use to solve the following question? 九月 + 10 个月 = ?
Which strategy did you use to solve the following question? 七月 - 2 个月 = ?
Which strategy did you use to solve the following question? 二月 - 6 个月 = ?
Which strategy did you use to solve the following question? 十二月 - 10 个月 = ?
Which strategy did you use to solve the following question? 六月 - 9 个月 = ?
Which strategy did you use to solve the following question? 两点 + 7 小时 = ?
Which strategy did you use to solve the following question? 十八点 + 10 小时 = ?
Which strategy did you use to solve the following question? 四点 + 15 小时 = ?
Which strategy did you use to solve the following question? 十三点 + 21 小时 = ?
Which strategy did you use to solve the following question? 十五点 - 11 小时 = ?
Which strategy did you use to solve the following question? 六点 - 10 小时 = ?
Which strategy did you use to solve the following question? 二十三点 - 12 小时 = ?
Which strategy did you use to solve the following question? 十二点 - 20 小时 =?
Which strategy did you use to solve the following question? 周 1 + 1 天 = ?
Which strategy did you use to solve the following question? 周 5 + 3 天 = ?
Which strategy did you use to solve the following question? 周 1 + 6 天 = ?
Which strategy did you use to solve the following question? 周 5 + 5 天 = ?
Which strategy did you use to solve the following question? 周 6 - 1 天 = ?
Which strategy did you use to solve the following question? 周 1 - 2 天 = ?
Which strategy did you use to solve the following question? 周 6 - 4 天 = ?
Which strategy did you use to solve the following question? 周 4 - 6 天 = ?

Which strategy did you use to solve the following question? 1 月 + 2 个月 = ?
Which strategy did you use to solve the following question? 8 月 + 5 个月 = ?
Which strategy did you use to solve the following question? 2 月 + 9 个月 = ?
Which strategy did you use to solve the following question? 9 月 + 10 个月 = ?
Which strategy did you use to solve the following question? 7 月 - 2 个月 = ?
Which strategy did you use to solve the following question? 2 月 - 6 个月 = ?
Which strategy did you use to solve the following question? 12 月 - 10 个月 = ?
Which strategy did you use to solve the following question? 6 月 - 9 个月 = ?
Which strategy did you use to solve the following question? 10 + 60 年 = ?
Which strategy did you use to solve the following question? 60 - 18 年 = ?
Which strategy did you use to solve the following question? 120 + 160 年 = ?
Which strategy did you use to solve the following question? 900 - 700 年 = ?
Which strategy did you use to solve the following question? 1985 + 5 年 = ?
Which strategy did you use to solve the following question? 2050 - 1000 年 = ?
Which strategy did you use to solve the following question? 310 + 50 年 = ?
Which strategy did you use to solve the following question? 1800 - 20 年 = ?

1. **Memory** (Automatic recall of the fact, you didn't need to work out the answer)
2. **Transform and calculate** (e.g., Transform '周一 + 2天' into '1 + 2 =3' and get the answer)
3. **Count** (e.g., '三月 + 2个月' is '三月', '四月', '五月', so the answer is 五月)
4. **Estimate** (Guess the answer)
5. **Other** (Another method that is not listed in this list)

8. List of References

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition, 9*, 89 –96.

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: the case of Greek blues. *Bilingualism (Cambridge, England), 12*(1), 83–95.

Athanasopoulos, P. and Boutonnet, B. (2016). Learning grammatical gender in a second language changes categorization of inanimate objects: Replications and new evidence from English learners of L2 French. In R. Alonso (Ed.), *Cross-linguistic influence in second language acquisition* (pp. 173–192). Bristol, UK: Multilingual Matters.

Athanasopoulos, P., Bylund, E., and Casasanto, D. (2016). Introduction to the Special Issue: New and Interdisciplinary Approaches to Linguistic Relativity. *Language Learning, 66*(3), 482-486.

Athanasopoulos, P. and Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology, 37*(5-6), 393-412.

Athanasopoulos, P., Damjanovic, L., Krajciova, A., and Sasaki, M. (2011). Representation of colour concepts in bilingual cognition: The case of Japanese blues. *Bilingualism: Language and Cognition, 14*(1), 9–17.

Athanasopoulos, P. and Bylund, E., (2013). Does grammatical aspect affect motion event cognition? a cross-linguistic comparison of English and Swedish speakers. *Cognition Science, 37*, 286–309

Athanasopoulos, P. and Bylund, E. (2021). Whorf in the wild: Naturalistic evidence from human interaction. *Applied linguistics, 41*(6), 947-970.

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., et al., (2015). Two languages, two minds: flexible cognitive processing driven by language of operation. *Psychol. Sci. 26* (4), 518–526.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bassetti, B., Clarke, A., and Trenkic, D. (2018). The linguistic transparency of first language calendar terms affects calendar calculations in a second language. *Acta Psychologica, 186*, 81-89.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255.

Bernardo, A.B.I., (2001). Asymmetric activation of number codes in bilinguals: Further evidence for the encoding complex model of number processing. *Memory Cognition, 29* (7), 968–976.

Blom, E., and Unsworth, S. (Eds.). (2010). *Using on-line processing methods in language acquisition research. In Experimental Methods in Language Acquisition Research.* John Benjamins Publishing Company.

Boroditsky, L., (2001). Does language shape thought? English and Mandarin speakers' conceptions of time? *Cognition, 118*(1), 123-129.

Boroditsky, L. and Schmidt, L. A. (2000). Sex, syntax, and semantics. In L. R. Gleitman and A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 42– 47). Mahwah, NJ: Erlbaum

Boorstin, D. J. (1985). *The discoverers.* New York: Random House.

Booth, J. L. and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 41,* 189-201.

Booth, J. L. and Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*, 1016-1031.

Brysbaert, M. and Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition, 13*(3), 359-371.

Bylund, E. and Athanasopoulos, P. (2015). Introduction: Cognition, Motion Events, and SLA. *The Modern language journal (Boulder, Colo.), 99*(S1), 1-3.

Bylund, E. and Athanasopoulos, P. (2017). The Whorfian time warp: Representing duration through the language hourglass. *Journal of experimental psychology, 146*(7), 911-916.

Campbell, J. I. D., and Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology. General, 130*(2), 299–315.

Casasanto, D. (2008). Who's afraid of the Big Bad Whorf? Cross-linguistic differences in temporal language and thought. *Language Learning, 58*, 63–79.

Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S. et al. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. In K. Forbus, D. Gentner, and T. Regier (Eds.)*, Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 186–191). Mahwah, NJ: Erlbaum.

Caskey-Sirmons, L. A. and Hickerson, N. P. (1977). Semantic shift and bilingualism: Variation in the color terms of five languages. *Anthropological Linguistics, 19,* 358 –367

Chan, W. W. L. (2014). Understanding and processing numbers among Chinese children. *Psychology & Neuroscience, 7*(4), 583–591.

Chan, B. M. Y. and Ho, C. S. H. (2010). The cognitive profile of Chinese children with mathematics difficulties. *Journal of Experimental Child Psychology, 107*, 260–279.

Chen, J.-Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition, 104,* 427-436.

Chen T., Koda K., and Wiener S., (2020). Word-meaning inference in L2 Chinese: an interactive effect of leaners' linguistic knowledge and words' semantic transparency. *Reading and Writing, 33*, 2639 – 2660.

Cheng, Z. and Chan, L. K. S. (2005) Chinese number-naming advantages? Analyses of Chinese pre-schoolers' computational strategies and errors. *International Journal of Early Years Education, 13*(2), 179-192.

Cheng, C., Wang, M., and Perfetti, C. (2011). Acquisition of compound words in Chinese-English bilingual children: Decomposition and cross-language activation. *Applied psycholinguistics, 32*(3), 583-600.

Chrisomalis, S. (2021). The scope of linguistic relativity in graphic and lexical numeration. *Language and Communication, 76*, 1-12.

Comrie, B. (2005). Endangered numeral systems. In J. Wohlgemuth & T. Dirksmeyer (Eds.), *Bedrohte vielfalt: Aspekte des sprach(en)tods [Endangered diversity: Aspects of language death]* (pp. 203-230). Berlin: Weißensee Verlag.

Cook, V. J. (2012). Multi-competence. In C. Chapelle (Ed.). *The encyclopedia of applied linguistics* (pp. 3768–3774). Oxford, UK: Wiley-Blackwell.

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and numerical magnitude. *Journal of Experimental Psychology: General, 122*, 371-

396.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics.* New York: Oxford University Press.

Dowker, A., Baa, S., and Lloyd, D. (2008). Linguistics influence of mathematical development: How important is the transparency of the counting system? *Philosophical Psychology, 21*(4), 523-538.

Dowker, A. and Li, A. M. (2019). English and Chinese Children's Performance on Numerical Tasks. *Frontiers in psychology, 9*, 2731-2

Dowker, A. and Roberts, M. (2015). Does the transparency of the counting system affect children's numerical abilities? *Frontiers in Psychology, 6,* 945.

Ervin, S. E. (1961). The connotations of gender. *Word, 18*, 249 –261.

Everett, C. (2013) *Linguistic Relativity: Evidence Across Languages and Cognitive Domains*, edited by Gitte Kristiansen, and de Mendoza Ibáñez, Francisco J. Ruiz, De Gruyter, Inc.

Everett, C. (2016). Mapping explorations of linguistic influences on nonlinguistic thought. *Languages (Paris), 201*, 65-76.

Fischer, M. (2003). Spatial representations in number processing: Evidence from a pointing task. *Visual Cognition, 10*, 493-508.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

Friedman, W. J. (1983). Image and verbal processes in reasoning about the months of the year. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 9*(4), 650.

Friedman, W. J. (1986). The development of children's knowledge of temporal structure. *Child Development, 57*, 1386–1400.

Friedman, W. J. (1984). Analog and semantic models of judgments about the months of the year. *Memory and Cognition, 12*(3), 306–313.

Friedman, W. J. (1990). *About time: Inventing the fourth dimension.* Cambridge, MA: MIT Press.

Frenck-Mestre, C. and Vaid, J. (1993). Activation of number facts in bilinguals. *Mem. Cogn. 21*(6), 809–818.

Fuson, K. C. (1992). Issues in place-value and multidigit addition and subtraction learning and teaching. *Journal for Research in Mathematics Education, 21*, 273–280.

Fuson, K. C. and Kwon, Y. (1991). Chinese-based regular and European irregular systems of number words: the disadvantages for English-speaking children. In K. Durkin and B. Shire (Eds.), *Language and Mathematical Education* (pp. 211–226). Milton Keynes: Open University Press.

Geary, D. C. (1996). Biology, culture, and cross-national differences in mathematical ability. In R. J. Sternberg, and T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 145-171). Mahwah, NJ: Lawrence Erlbaum Associates.

Gleitman, L. and Papafragou, A. (2013). Relations between language and thought. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 504–523). New York, NY: Oxford University Press.

Grove, B. G. (Ed.). (1986). *Webster's third new international dictionary (unabridged)*. Springfield, MA: Merriam–Webster.

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H-C., and Moeller, K. (2011).

Language effects on Children's Nonverbal Number Line Estimations. *Journal of Cross-Cultural Psychology, 42*(2), 59-613.

Ho, C. S. H. and Fuson, K. C. (1998) Children's knowledge of teen quantities as tens and ones: comparisons of Chinese, British, and American kindergartners. *Journal of Educational Psychology, 90*, 536–544.

Huang, W. (1993). 中国成人推理月份时间的加工模型 [Processing models of month calculations in Chinese adults]. *Xin Li Ke Xue, 2*, 84–89.

Huang, W. (1999). Reasoning about conventional time as a function of conventional time systems. *Memory and Cognition, 27*(6), 1080–1086.

Imai, M. and Gentner, D. (1997). A crosslinguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition, 62*, 169 –200.

Jackendoff, R. (1983). *Semantics and Cognition* (Second printing 1985). Cambridge, MA: The MIT Press.

Jameson, K. A. and Alvarado, N. (2003). Differences in color naming and color salience in Vietnamese and English. *Color Research and Application, 28*, 113–138.

January, D. and Kako, E. (2007). Re-evaluating evidence for linguistic relativity: reply to Boroditsky. *Cognition 104*(2), 417–426.

Jiang, T., and Fang, G. (1997). 小学儿童对习俗时间的周期性特点的认知 [Primary school pupils' cognition of cyclic aspects of conventional time]. *Xinli Kexue, 5*, 431–435.

Keating, G. D. and Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, *37*, 1–32.

Kelly, M. K., Miller, K., Fang, G., and Feng, G. (1999). When days are numbered:

Calendar structure and the development of calendar processing in English and Chinese. *Journal of Experimental Child Psychology, 73*, 289–314.

Kraut, C. and Pixner, S. (2022). Bilingual adults practicing multiplication tables – looking into bilingual arithmetic learning. *International Journal of Bilingual Education and Bilingualism, 25*(5), 1825-1837.

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

Kurinski, E. and Sera, M. D. (2011). Does learning Spanish grammatical gender change English-speaking adults' categorization of inanimate objects? *Bilingualism: Language and Cognition, 14*, 203–220.

Kutas, M. and Hillyard, S.A. (1980). Reading senseless sentences - brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205.

Kutas, M., Lindamood, T., and Hillyard, S. (1984). Word expectancy and event-related brain potentials during sentence processing. In: Kornblum, S., Requin, J. (Eds.), *Preparatory States and Processes* (pp. 217–237). Hillsdale Lawrence Erlbaum.

Laski, E-V. and Yu, Q., (2014). Number line estimation and mental addition: Examining the potential roles of language and education. *Journal of Experimental Child Psychology, 117*, 29-44.

Lan, X., Ponitz, C. C., Miller, K. F., Li, S., Cortina, K., Perry, M., et al (2009). Keeping their attention: Classroom practices associated with behavioral engagement in first grade mathematics classes in China and the United States. *Early Childhood Research Quarterly, 24*, 198–211.

Le, M-L. T. and Noel, M-P, (2020). Transparent number-naming system gives only limited advantage for preschooler's numerical development: Comparisons of

Vietnamese and French-speaking children. *PloS One, 15*(12), 1-24.

Levin, I., & Wilkening, F. (1989). Measuring time via counting: The development of children's conceptions of time as a quantifiable dimension. In I. Levin and D. Zakay (Eds.), *Advances in psychology: Vol. 59. Time and human cognition: A life-span perspective* (pp. 185–218). Amsterdam: Elsevier

Libben, G., Gibson, M, Toon, Y, and Sandra, D., (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language, 84*, 50 – 64.

Lin, J-F. L., Imada, T., and Kuhl, P.K. (2011). Mental addition in bilinguals: an fMRI study of Task-related and performance -related activation. *Cerebral Cortex 22*(8), 1851-1861.

Li Y., Casaponsa A., Wu Y-J., and Thierry G., (2019). Back to the future? How Chinese-English bilinguals switch between front and back orientation for time. *NeuroImage, 203*, 1-10.

Lucy, J. A. (1997). Linguistic relativity. *Annual Review of Anthropology, 26*, 291–312.

Luo, T. (2012), Mental Arithmetic across Three Language Groups. *Master's Theses.* 4200.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology, 3*, 54.

Mackey, A. and Gass, S. M. (2005). *Second language research: methodology and design.* Mahwah, NI: Lawrence Erlbaum.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States.* Mahwah, NJ: Lawrence Erlbaum.

Marsh, L.G. and Maki, R.H. (1976). Efficiency of arithmetic operations in bilinguals as a function of language. *Mem. Cogn. 4*(4), 459–464.

Mark, W. and Dowker, A. (2015). Linguistic influence on mathematical development is specific rather than pervasive: revisiting the Chinese Number Advantage in Chinese and English children. *Frontiers in Psychology, 6*, 1-9.

Miller, K F., Kelly, M., and Zhou, X. (2005). Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of preschool mathematical development. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 163-178). New York: Psychology Press.

Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science, 6*(1), 56-60.

Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (2000). Ordinal knowledge: Number names and number concepts in Chinese and English. *Canadian Journal of Experimental Psychology, 54*(2), 129-140.

Miller, K. F., and Stigler, J. W. (1987). Counting in Chinese: Cultural variations in basic cognitive skill. *Cognitive Development, 2*(3), 279-305.

Miura, I. T., and Okamoto, Y. (2003). Language supports for mathematics understanding and performance. In Baroody, A. J. and Dowker, A. (Eds.), *The Development of Arithmetic Concepts and Skills: Constructing adaptive expertise. Studies in mathematics thinking and learning* (pp. 229–242). Mahwah, NJ: Erlbaum.

Miura, I. T., Okamoto, Y., Kim, C. C., Steere, M., and Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: crossnational comparisons – France, Japan, Korea, Sweden, and the United States.

J. Educ. Psychol. 85, 24–30.

Miura, I. T., Okamoto, Y., Vlahovic-Stetic, V., Kim, C. C., and Han, J. H. (1999). Language Supports for Children's Understanding of Numerical Fractions: Cross-National Comparisons. *Journal of experimental child psychology, 74*(4), 356-365.

Momenian, M., Cham, S. K., Amini, J. M., Radman, N., and Weekes, B. (2021) Capturing the effects of semantic transparency in word recognition: a cross-linguistic study on Cantonese and Persian. *Language, Cognition and Neuroscience, 36*(5), 612-624.

Muldoon, K., Simms, V., Towse, J., Burns, V., and Yue, G. (2011). Cross-Cultural Comparisons of 5-Year-Olds' Estimating and Mathematical Ability. *Journal of Cross-Cultural Psychology, 42*(4), 669-681.

Muira, I, T., Okamoto, Y., Vlahovic-Stetic, V., Kim, C. C., and Hann, J. H. (1999). Language Supports for children's understanding of numerical fractions: Cross-national comparison. *Journal of Experimental Child Psychology, 74*(4). 356-365.

Ng, S. S. N. and Mu, Y. (2010). Chinese number words, culture, and mathematics learning. *Review of Educational Research, 80*(2), 180-206.

Norris, J. M. (2015). Statistical significance testing in second language research: Basic problems and suggestions for reform. *Language Learning, 65*, 97–126.

Opfer, J. E. and Furlong, E. E. (2011). How numbers bias preschoolers spatial search. *Journal of Cross-Cultural Psychology, 42*(4), 682-695.

Opfer, J. E. and Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*, 169-195.

PARISE, F. (Ed.) (1982). *The book of calendars.* New York: Facts on File.

Park, H. I. and Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean–English bilinguals. *Bilingualism: Language and Cognition, 17*, 410–430.

Pica, P., Lemer, C., Izard, W., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*, 499–503.

Pinker, Steven. (1994). *The language instinct*. New York: Harper Collins.

Rasmussen, C., Ho, E., Nicoladis, E., Leung, J., and Bisanz, J. (2006). Is the Chinses Number-Naming System Transparent? Evidence from Chinese-English Bilingual Children. *Canadian Journal of Experimental Psychology, 60*(1), 60-67.

Regier, T. and Kay, P. (2009). Language, thought and color: Whorf was half right. *Trends in Cognitive Sciences, 13*, 439–446.

Rinsveld, A. V., Schiltz, C, Landerl, K., Brunner, M., and Ugen, S. (2016). Speaking two languages with different number naming systems: What implications for magnitude judgments in bilinguals at different stages of language acquisition? *Cognition Process, 17*, 225-241.

Roberson, D., Pak, H., and Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition, 107*(2), 756-762.

Silva, R. and Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition, 11*(2), 245–260.

Samuel, S., Cole, G., and Eacott, M. (2019). Grammatical gender and linguistic relativity: A systematic review. *Psychonomic bulletin and review, 26*(6), 1767-1786.

Sapir, E. (1949). *The selected writings of Edward Sapir in language, culture, and* *personality (Edited by David G. Mandelbaum).* Berkeley: University of California Press.

Sarantakis, N. P. (2014). The influence of out native language on cognitive representations of colour, spatial relations and time. *Journal of European Psychology Students, 5*(3), 74-77.

Scott, A. (1989). The vertical dimension and time in Mandarin. *Australian Journal of Linguistics, 9*, 295-314.

Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin and Reviews, 16*, 328-331.

Sherko, E. (2015). Linguistic Transparency and Opacity in Compounding. Academic *Journal of Interdisciplinary Studies, 4*(3), 590- 593.

Siegler, R. S. and Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237

Siegler, R-S. and Mu, Y. (2008). Chinese Children Excel on Novel Mathematics Problems Even before Elementary School. *Psychological science, 19*(8), 759-763.

Smith, S., Brown, H., Toman, J., and Goodman, L. (1947). The lack of cerebral effects of d-tubocurarine. *Anesthesiology (Philadelphia), 8*(1), 1-14.

Soveri, A., Laine, M., HamaLainen, H., and Hugdahl, K. (2011). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism, 14*(3), 371-378.

Tang, M., Vanek, N., and Roberts, L. (2021). Crosslinguistic influence on English and Chinese L2 speakers' conceptualization of event series. *The international journal* *of bilingualism: cross-disciplinary, cross-linguistic studies of language behavior, 25*(1), 205-223.

Thierry, G. (2016). Neurolinguistic relativity: how language flexes human perception and cognition. *Lang. Learn. 66*(3), 690–713.

Thompson, C. A. and Opfer, J. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81*, 1768-1786.

Trenkic, D. and Warmington, M. (2019). Language and literacy skills of home and international university students: How different are they, and does it matter? *Bilingualism: Language and Cognition, 22*(2), 1–17.

Tse, C.-S. and Altarriba, J. (2014). Evidence against linguistic relativity in Chinese and English: a case study of spatial and temporal metaphors. *J. Cogn. Cult. 8*(3–4), 335–357.

Tversky, B., Kugelmass, S., and Winter, A. (1991). Cross-cultural and development trends in graphic productions. *Cognitive Psychology, 23*(4), 515-557.

Vanek, N. and Selinker, L. (2017). Covariation between temporal interlanguage features and nonverbal event categorisation. *International Review of Applied Linguistics in Language Teaching, 55*, 223–243.

van Galen, M. S. and Reitsma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology, 101*, 99-113.

Van Rinsveld, A., Brunner, M., Landerl, K., Schiltz, C., and Ugen, S. (2015)). The relation between language and arithmetic in bilinguals: insights from different stages of language acquisition. *Front. Psychol. 6*, 265-265.

Van Rinsveld. A., Dricot, L., Guillaume, Rossion, B., and Schiltz C. (2017). Mental arithmetic in the bilingual brain: Language matters. *Neuropsychologia, 101*, 17-29.

Vernich, L. (2017). Does learning a foreign language affect object categorization in native speakers of a language with grammatical gender? The case of Lithuanian speakers learning three languages with different types of gender systems (Italian, Russian and German). *International Journal of Bilingualism, 23*, 417–436.

Vigil, D. C., Tyler, A. A., and Ross, S. (2006). Cultural differences in learning novel words in an attention-following versus attention-directing style. *Journal of Multilingual Communication Disorders, 4*(1), 59-70

Wang, Y., Lin, L., Kuhl, P., and Hirsch, J. (2007). Mathematical and linguistic processing differs between native and second languages: an fMRI study. *Brain Imaging Behavior 1 (3–4*), 68–82.

Weinreich, U. (1953). *Languages in contact: Findings and problems.* New York: Linguistic Circle of New York

Weinrich U., and Martinet, A., (1968). *Languages in Contact: Findings and Problem*. The Hague; Paris: Mouton & Co.

Whorf, B. L. (1956). *Language, thought and reality: Selected writing of Benjamín Lee Whorf*. Cambridge, MA: MIT Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104*, 7780–7785.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus (Translated by Charles Kay Ogden).* New York: Harcourt, Brace & Company.

Wolff, P. and Holmes, K. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(3), 253–265.

Yang, C. and Zhang, J. (2011). 粤语-普通话双言者和普通话单言者周期性时间推理比较 [A comparison of cyclical temporal reasoning between Chinese-Mandarin diglossic speakers and Mandarin speakers]. *Xin Li Ke Xue, 34*(4), 782–787.

Zerubavel, E. (1985). *The seven day circle: The history and meaning of the week.* Chicago: Univ. of Chicago Press

Zorzi, M., Priftis, K., and Umiltà, C. (2002). Brain damage: Neglect disrupts the mental number line. *Nature, 417*, 138-139.