Language and Cognition (2025), 17, e26, 1–26
doi:10.1017/langcog.2024.75

ARTICLE

When the days are(n't) numbered: Calendar calculations in transparent and opaque systems

Ziyi Zhuang[1] and Norbert Vanek[1,2]

[1]School of Cultures, Languages and Linguistics, University of Auckland, Auckland, New Zealand and [2]Experimental Research on Central European Languages Lab, Charles University, Prague, Czechia
Corresponding author: Norbert Vanek; Email: norbert.vanek@auckland.ac.nz

(Received 07 January 2024; Revised 05 October 2024; Accepted 15 December 2024)

Abstract

Calendar calculations, the process of computing the target day or month, exhibit peculiar differences across languages. In systems like English, calendar labels are largely opaque (Tuesday, August), which invites calculations to rely more heavily on verbal listing. In transparent systems, like Chinese, habitual labeling of calendar terms numerically (Tuesday = Day 2, August = Month 8) facilitates fast numerical operations instead of verbal listing. This study examines the effects that different levels of transparency of the calendar naming system may have on calculations in the speakers' first and second language. Chinese–English bilinguals were tested alongside English and Chinese controls. Forced-choice calendar calculations (day, month, hour and year) and self-reported strategies were used as tasks to tap into participants' calculation speed, accuracy and temporal reasoning. In the calculation questions, we manipulated Distance (short/long), Direction (forward/backward), Input (linguistic/numerical) and Boundary (within/across). More complex Month calculations significantly differed across groups while easier day calculations did not. The English group reported reliance on verbal listing while the Chinese and the Bilingual groups preferred numerical reasoning. These findings bring new evidence for linguistic relativity in the form of modulations of calendar processing speed changing as a function of linguistic transparency, input type and task demand.

Keywords: bilingual cognition; calendar processing; linguistic relativity

1. Introduction

If numerical processing is linguistically modulated, what are the cognitive implications for bilinguals? A prominent aspect of bilingual cognition is temporal reasoning and daily problem-solving tasks such as calendar calculations, an area with an ecological validity advantage over strictly laboratory-based tasks. As an essential component of temporal reasoning, calendar calculations occur routinely in everyday contexts. These calculations involve identifying at what time of which day in which month a specific event was, is or will happen. 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. To illustrate, English has an opaque calendar naming system, as it is hard to guess the meaning of the terms for those with no experience with the language. On the contrary, Chinese uses numbers to represent days and months. The cognitive consequence of the different levels of transparency in English and Chinese calendar terms may bring about cross-linguistic differences in the processing routines that characterize calendar calculations.

Previous studies (e.g., Friedman, 1983, 1984; Huang, 1993; Kelly, et al., 1999) showed that the habitual use of numerical and linguistically transparent calendar terms endows Chinese speakers with a routinized numerical operation strategy, while English speakers more typically resort to verbal listing when solving calendar calculations. Cross-linguistic research showed that speakers used to transparent calendar naming outperformed speakers used to an opaque calendar lexicon (Yang & Zhang, 2011). Less is known whether such an advantage extends to bilingual populations, namely when Chinese–English bilinguals need to solve calendar calculation questions in their second language (L2) English.

2. Linguistic transparency of calendar terms and some key contrasts

Linguistic transparency is a broad concept with different aspects, including semantic transparency and numerical transparency. With immediate relevance to this study, semantic transparency is a linguistic phenomenon related to compounding (Momenian et al., 2021). As a productive morphological word formation process, compounding combines two or more words together and makes the new word function as one word, semantically and grammatically (Sherko, 2015). Semantic transparency denotes a condition when the meaning of a multimorphemic compound word can be derived from the meanings of its constituent morphemes (Libben et al., 2003). For example, the word ‘blueberry’ is a transparent compound due to ‘blue’ and ‘berry’, both functioning as transparent members. In contrast, ‘deadline’ is semantically opaque, because it is hard for a new learner to comprehend the entire string’s meaning due to the opacity of the morpheme combination of ‘dead’ and ‘line’. Transparency and opacity are not dichotomous but instead vary along a continuum from semantically transparent to semantically opaque. Libben et al. (2003) propose four degrees of semantic transparency in bimorphemic compounds: transparent–transparent (blueberry), opaque–transparent (eyewitness), transparent–opaque (jailbird) and opaque–opaque (deadline). Unlike semantic transparency, numerical transparency refers to a regular counting system with clear and consistent rules for combining primary numbers, which is common in many number systems across Asia (Laski & Yu, 2014; Ng & Rao, 2010). For example, regularity and transparency characterize Chinese number words from 11 to 20 and from 10 to 99, in both written and spoken forms. The Chinese numerical words between 11 and 19 are formed by compounding ‘ten’ and ‘the unit word/cardinality’ (i.e., one to nine) and numbers from 20 to 90 are formed by compounding ‘the ten-digit code’, ‘ten’ and ‘the unit word’ (Miller et al., 1995). Based on these organization rules, 11 and 12 are ‘十一’ (*shi yi*) and ‘十二’ (*shi er*), literally ‘ten-one’ and ‘ten-two’, respectively, while 20 is ‘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 degree of transparency in this

10-based counting system, the cognitive demand for a new learner of Chinese to guess the meaning of an unknown Chinese number term is relatively low.

In contrast to transparent counting systems, many Western languages apply opaque and irregular naming of numbers (Dowker & Roberts, 2015). Taking the English number system as an example, English speakers need to memorize relatively arbitrary names like eleven and twelve with no clear clues representing the ten-digit code or the unit word. Moreover, various phonemic and morphological modifications of English numerical names further complicate the acquisition of the counting system (Mark & Dowker, 2015). Specifically, 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 require rote learning, which is possibly why learners of English may have more difficulties than learners of Chinese when it comes to acquiring the counting system (Dowker & Roberts, 2015) and mathematical development (Dowker et al., 2008; Miller et al., 1995; Siegler & Mu, 2008). While Chinese may well represent a transparent counting system and English an opaque one, it is important to consider opacity–transparency as a spectrum rather than a binary distinction. To illustrate this point, Chinese *er shi* (two ten) is more transparent than *twenty*, but *twenty* is in turn more opaque than French *vingt*.

Returning to the Chinese–English distinctions, pronounced differences emerge in how English and Chinese label days and months. Although both Chinese and English use the 7-day and 12-month solar calendar system, Chinese uses transparent calendar terms, while English uses opaque day and month names. English days 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 names of gods. While names as Sun (Sunday), Moon (Monday) and Saturn (Saturday) remained, four Roman gods' names were replaced by Nordic gods, namely Tyr for Mars (Tuesday), Odin for Mercury (Wednesday), Thor for Jupiter (Thursday) and Frigg for Venus (Friday) (Boorstin, 1985; Zerubavel, 1985). Historical relations between English day names and ancient astronomy may be recognized by those well-versed in the history of English names for days of the week, but for many the etymology of English day terms is obscured. 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 Chinese 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 principle, except for the non-numerical term Sunday, remains the same.

The names of the 12 months in English are also opaque. They are traceable to a mélange of gods' and emperors' names (January to August), and Latin numerical labels (September to December) (Boorstin, 1985; Grove, 1986). For even less transparency, the English month names from September to December correspond to the Latin numbers of 7 to 10 instead of 9 to 12. This numbering stems from the Roman 10-month calendar system (Kelly et al., 1999) where 7–10 were the last four months. In sum, although the English names of the 12 months exhibit a derivational structure, the process is not transparent to many English learners and speakers. Chinese month names follow a regular compounding format like that in day names, which is 'the cardinal number of a particular month +月 (*yue*, meaning month)'. 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. To sum up, both Chinese day names and month names are TT (transparent–transparent) compounds, while English day names are OT (opaque–transparent) compounds and English month names are OO (opaque–opaque) compounds. The importance of relative transparency of calendar terms for cognitive processing is high, especially when one considers that mathematical calculations, one of the fundamental reasoning skills, are a cognitive product of linguistic notational features (Chrisomalis, 2021).

3. Effects of linguistic transparency on calculations

Few studies have investigated the effects of linguistic transparency on aspects of cognition, an area of which is calendar calculations, an everyday reasoning and problem-solving task. Two tasks characterize this domain: a counting task (e.g., Dowker & Roberts, 2015; Miller et al., 1995) and an arithmetic task (e.g., Pica et al., 2004). Cross-linguistic comparisons of the tasks have demonstrated positive effects of numerical transparency on monolingual children's development of mathematical abilities, especially in comparisons of arithmetic performance between Chinese-speaking and English-speaking children over a long time span (Geary, 1996; Miller & Stigler, 1987; Miller et al., 1995, 2000; Miura et al., 1999; Ng & Rao, 2010). 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, a link was found between numerical transparency of a language and the mathematical abilities of its speakers. Advantages of numerical transparency have also been observed in the performance of German (transparent numerical structure) and Italian (opaque numerical structure) children (Helmreich et al., 2011), and Welsh children raised in either English (irregular counting system) or Welsh (regular counting system) environments (Dowker & Roberts, 2015; Dowker et al., 2008). While numerical transparency can be linked to calendar calculations across many languages, it is important to acknowledge that calendric systems in some languages are not numerically based. For instance, Sinha et al. (2011) reported that the Amondawa language in Amazonia does not lexically encode weeks, months or years, and 'there is no word meaning time in Amondawa' (Sinha et al., 2011, p. 149) either. Results from calendar questionnaires and two calendar installation games showed that the ways in which Amondawa speakers conceptualize time intervals do not integrate the four-number Amondawa system. These findings inform relevant research by challenging the assumption of universal mappings between numerical and calendar constructions.

Recently, functional magnetic resonance imaging (fMRI) studies examined brain activation patterns when bilinguals were solving mathematical tasks. Numerical cognition examined via brain scans can provide a window into arithmetic computations with information on whether speakers of different linguistic systems potentially activate different brain regions to solve mathematical questions. One set of fMRI results showed that without high proficiency in a second language, bilinguals tended to retrieve arithmetic facts through the activation of verbal codes from the first language (Lin et al., 2011; Wang et al., 2007). These results suggest that low-proficiency bilinguals translate and process the questions in their first language when tested in a second language. Behavioral studies on bilinguals' mathematical abilities also emphasized a higher accuracy and shorter reaction times (RTs) when solving problems in the bilinguals' L1 compared to the L2 (Frenck-Mestre & Vaid, 1993; Marsh & Maki, 1976) or at least in the instruction language used to teach arithmetic calculations (Bernardo, 2001; Van Rinsveld et al., 2015). What these findings suggest is that the performance of bilinguals in arithmetic tasks tends to involve a translation process from the L2 to the L1, and back to the L2 by the bilinguals who are tested in their L2. Nevertheless, the surveyed studies paid more attention to late bilinguals, leaving the possibility open that results would differ for bilinguals who acquired both languages early in life. To fill that gap, Van Rinsveld et al. (2017) recruited highly proficient and balanced German–French bilinguals with shared language learning history and scanned their brains during simple and complex addition tasks. Differential activation patterns in additive operations were reported for tests in L1 and L2. However, firm L1/L2 effects are difficult to establish as no control groups (neither German nor French monolinguals) were involved in the testing. While research on numerical transparency abounds, investigations on the effect of linguistic transparency on calendar calculations are less well mapped.

In terms of research on calendar reasoning, Friedman's influential work (Friedman, 1983, 1984, 1990) showed that children initially learn lists of the calendar names in a sequence for both days and months. As a result, when faced with a calendar 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 recite the whole sequence of units and count them overtly or covertly to arrive at the answer. This process is called verbal listing. It should be noted that Friedman's claims were based on data exclusively from native English speakers. Considering the reviewed cross-linguistic variation in how calendar terms are formed, it is plausible that speakers of languages that promote calculation strategies other than verbal listing (e.g., counting, calculation and memory) would differ from English speakers in their calendar calculations. Support for this idea can be found, for instance, in the study by Huang (1993), who examined Chinese speakers' preferred reasoning strategies in a month calculation task and reported a different set of strategies from those reported in Friedman (1983). Instead of verbally listing the calendar terms, Chinese speakers employed 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'). Preference for arithmetic operations is supported already in young learners of Chinese, who first 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 days (e.g., 星期三, Wednesday is simply 'week-three'), and to the root '月' (pronounced as *yue*, meaning month) to identify the target months (e.g., 四月, April is simply '*si-yue*'), which is arguably easier than reciting lists of calendar terms by rote memorization (Mark & Dowker, 2015). In sum, regular and transparent conventional temporal representations of Chinese speakers were found to facilitate numerical calculation speed in day and month calculation tasks compared to English speakers. Additionally, English speakers needed more time to solve reverse and longer distance questions, but no distance or direction effects were found in Chinese speakers. However, a boundary effect (e.g., to name the month that is 10 months after March (month 3), was still found in Chinese adults' calculations. Both English and Chinese participants took longer to solve questions involving boundary crossing, which points to additional cognitive effort. Similarly, Jiang and Fang (1997) also demonstrated that both Chinese school children's and adults' calculations exhibited a boundary effect in day calculation questions.

To establish that it is not cultural background but the transparency of the Chinese calendar terms that affects participants' performance in calendar calculation processing, Huang (1999) recruited two groups of adult Chinese speakers from rural areas and asked them to do calculation tasks with solar months names and lunar months names, respectively. Lunar month names are a type of traditional Chinese calendar representation, where calendar terms are opaque. For instance, January is '正月' (*zheng yue*) instead of '一月' (*yi yue* 'month 1') and December is '腊月' (*la yue*, in which *la* refers to a sacrificial ceremony) rather than '十二月' (*shi er yue* 'month 12'). The lunar calendar system is mainly used in the rural areas of China alongside the solar month system. Participants reported equal proficiency and frequency of use of the two kinds of calendar naming systems. Results pointed to different cognitive routines across groups. The solar group outperformed the lunar group in both accuracy and RTs. Also, distance and direction effects were observed in the lunar group, while the solar group only exhibited the boundary effect. These results were in line with those of Friedman (1990) and Huang (1993). The dominant self-reported strategy of the lunar group was verbal listing, while the solar group preferred arithmetic operations. The stability of these results was later checked cross-linguistically with Chinese-dominant and English-dominant speakers in Kelly et al. (1999), involving both day-of-the-week and month-of-the-year calculation tasks. The Chinese speakers were found overall to be faster than the English speakers in day and month calculations, and they showed a propensity for the arithmetic calculation strategy. As in earlier related studies, the Chinese-speaking group was not affected by distance (short/long) or direction (forward/backward), but there was an effect of boundary (cross-boundary questions took longer than within-boundary questions). Regarding strategies, the English group was inclined to verbal listing. In sum, it appears that variations in linguistic transparency across calendar systems can give rise to differences in calendar calculation performance between groups with and without opaque calendar terms. One remaining point to address is how linguistic transparency of calendar lexicons affects calendar calculations of bilinguals who have both opaque and transparent calendar terms at their disposal.

Research investigating calendar calculation in bilinguals is scant but present. Yang and Zhang (2011) tested whether having a specific linguistic label for a temporal unit would positively affect bilinguals' performance in a calendar reasoning task. In Cantonese, 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 has no equivalent in Mandarin Chinese (henceforth Chinese). Taking quarter to five as an example, Cantonese speakers tend to say '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 Chinese. Chinese-speaking monolinguals and bilinguals fluent in both Cantonese and Chinese were asked to do calculations involving five-minute units in Chinese. Results showed that Cantonese–Mandarin bilinguals outperformed the Chinese-speaking monolinguals, indicating that the presence of a specific linguistic label for a relevant temporal unit in one language can positively influence participants' performance in the 'five-minute-relevant calculation' task even when tested in their other language. This finding hinted at the possibility that bilinguals may spontaneously resort to calculations in their language where the lexicon is more useful to solve the task at hand. Another intriguing question arising from Yang and Zhang (2011) is whether other specialized terms for temporal units, such as ‘fortnight’ or ‘quarter’ in English, can similarly facilitate term-relevant calculations. In a related study, Bassetti et al. (2018) examined bilinguals’ calendar calculations including day and month calculations of English-dominant speakers and Chinese–English bilinguals. RTs of the Bilingual group were found shorter than those of the English group in month calculations but longer in the day calculations. In more difficult month calculations (12-based), directionality effects emerged in the English group but not for the bilinguals. Boundary crossing negatively affected bilinguals’ but not English speakers’ performance in the backward direction calculations. In easier day calculations (seven-based), boundary crossing negatively affected bilinguals both in the forward and backward direction but had no effects on the native English speakers. As for calculation strategies, bilinguals mostly self-reported numerical arithmetic, while native English speakers relied more heavily on verbal listing, in line with previous studies (Friedman, 1983; Huang, 1993; Jiang & Fang, 1997; Kelly et al., 1999).

4. The present study

This work aims to expand our understanding of temporal cognition in the domain of calendar reasoning by filling four research gaps left by relevant work (Bassetti et al., 2018). First, one limitation in the work of Bassetti et al. (2018) is the two-group design, only including a group of English-dominant speakers and Chinese–English bilinguals tested in L2 English, lacking a group of Chinese-dominant speakers. A direct three-group comparison (English L1, Chinese L1 and Chinese–English bilinguals) is important to establish potential L1 transfer effects more firmly. Second, a potential confound in the work of Bassetti et al. (2018) is that some calculation questions may have been opaque not (only) for numerical but (also) for semantic reasons. One example is when participants were first informed that it takes four days or seven months for seeds to sprout or blossom, and then they needed to calculate the sprouting or blossoming time based on information when the seeds had been planted. It is possible that Chinese–English bilinguals needed more effort just to decode the semantics of the question, potentially prolonging their response speed. This factor was not controlled. In addition, the participants were asked to give their answers orally and press the ‘next’ button to move on while being recorded. This is problematic due to participants’ potential variation in the level of verbal encoding. A more time-sensitive online processing experiment focusing purely on comprehension (Marinis, 2010) is beneficial to exclude confounds linked to differences in articulation speed. Third, Bassetti et al. (2018) tested the participants’ mathematical abilities by asking them to solve calendrical calculations using modular arithmetic (where a set of repeating cyclical numbers from 1 to N are used), such as seven-based Weekday calculations and 12-based Month calculation questions. The Weekday and Month calculation questions both represent modular arithmetic, where a set of repeating cyclical numbers from 1 to N (seven-based for Weekday calculation and 12-based for Month calculation) are used. Basetti et al. (2018) complemented these conditions with base arithmetic tasks (typical for year calculations, where there is no upper bound) to test between-group comparability in simple additions and subtractions. We find an *Hour Calculation*

Task with modular 24-based arithmetic instead of base arithmetic calculation questions an informative addition and a more directly comparable control condition for cross-linguistic analyses. Fourth, two factors that were not considered in Bassetti et al. (2018) and are known to play a role are Distance (short/long) and Input type (linguistic/numerical). Distance is an important factor, the effect of which differs in transparent and opaque calendar terms. The performance of speakers of a language with a transparent calendar naming system is known not to be affected robustly by distance, while the performance of speakers of languages with opaque calendar terms tends to be negatively influenced by long distance (Friedman, 1983; Huang, 1993; Kelly et al., 1999). Also, the factor of Input type deserves a closer look. Linguistic input refers to the calendar lexicons, while numerical input translates the calendar lexicons into Arabic numbers. For instance, the numerical version of 'Monday' is 'first day'. Speakers used to opaque calendar terms could be positively influenced by numerical input in their calendar calculations, as such an input directly provides the numbers needed to facilitate calculation speed. Conversely, speakers used to transparent calendar terms may be slowed down by linguistic input as it downregulates the availability of numbers to assist their calculations. Such influences on calendar calculations align with the Ad Hoc Cognition Framework (Casasanto & Lupyan, 2015), within which contextual changes can induce changes in cognitive processes that depend on them. In sum, this study conceptually replicates Bassetti et al. (2018) with the addition of four innovative components to make an informative contribution.

To fill the four research gaps, three groups were tested: the English group, the Chinese group and the Bilingual group. One improvement proposed here relates to construct validity and internal validity. This study asked participants to choose the correct answer by a button-press rather than through verbalization, eliminating potential noise induced by variation in articulation speed and second language anxiety (Mackey and Gass, 2005). Also, by adding an Hour and Year Calculation task, the present study added an important control to see whether different groups' mathematical abilities in base and modular arithmetic, respectively, were at comparable levels. Finally, Distance and Input type were added here to test how these two factors might influence different groups' performance in calendar calculations. Two main research questions and their corresponding hypotheses guided the study.

RQ1: To what extent do Chinese-dominant and English-dominant speakers differ when performing calendar calculations in their native languages?

H1: Significant differences were predicted between the two monolingual groups in two contexts (Day and Month calculations).

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

H2: Significant differences between the two groups were predicted in two contexts (Day and Month calculations). The reaction times (RTs) of bilinguals were predicted to be shorter than those of monolingual English speakers because of greater lexical and numerical transparency in marking days and months in L1 Chinese compared to more opaque English.

5. Methodology

5.1. Participants

Thirty English native speakers (15 females), 30 Chinese-dominant speakers (15 females) and 30 Chinese–English bilinguals (16 females) took part in this study. The English participants (M_{AGE} = 24.2, range 19–35) and the Chinese–English bilingual participants (M_{AGE} = 25.2, range 20–35) were recruited from universities in English-speaking countries, to control for language environment, while the participants of the Chinese group (M_{AGE} = 23.3, range = 19–32) were recruited from Chinese colleges. Following Athanasopoulos et al. (2011), Park and Ziegler (2014) and Vanek and Selinker (2017), all participants were asked to complete a questionnaire about their language background information (available at https://osf.io/3h97t/). The English participants were all English-dominant speakers who had no experience or little exposure to a second language in daily life at the time of testing. The Chinese participants were functionally monolingual, with some knowledge of basic English limited to a few phrases. The bilingual participants were Chinese-dominant learners of English who had gained a score equal to or greater than 6.5/9 in the International English Testing System at the time of testing. They started to learn English in kindergarten and, at the time of testing, regularly used English (for work and study) as well as Chinese in other situations (contact with family). All participants were right-handed and reported normal or corrected-to-normal vision. The study received ethics approval from the Human Participants Ethics Committee at the University of Auckland (UAHPEC #23368).

6. Materials and tasks

6.1. Calendar Calculation task

Task 1 was a Calendar Calculation task. The materials involved 96 calendar reasoning questions (available on the project page at https://osf.io/3h97t/), including Day, Month, Hour and Year calculations. The Day and Month Calculation Tasks were designed to test the effect of calendar terms' linguistic transparency on English speakers and Chinese–English bilinguals' RTs and accuracy. Given that the Day (7-based) and the Month (12-based) Calculation Task could have been modular arithmetic calculation tasks, an additional Hour Calculation Task was used as a control to test whether the two groups had comparable RTs and accuracy in 24-based calculations, and a further Year Calculation Task tested participants' RTs and accuracy in base calculations. Four variables were manipulated, including Distance (short or long), Direction (forward or backward), Boundary (within or across) and Input (linguistic or numerical).

6.2. Day calculations

With four independent variables, as shown in Table 1, the Day Calculation Task involved 32 questions. For Distance, if the numerical gap between the question and answer was less than 4 days (considering 7 days), then it counted as a short trial. If the numerical gap was more than 4 days, then it counted as a long trial. For Direction, in the forward condition, the target was after the stimuli, while the target was before the stimuli in the backward condition. For Boundary, within-boundary trials were within

Table 1. Examples of Day calculation questions in different conditions

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 2. Examples of Month calculation questions in different conditions

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 =?

the boundary of a week, while cross-boundary trials crossed the Sunday–Monday boundary. Input varied as either linguistic input or numerical input. 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 'fourth day – 6 days = ?' was a long, backward, across, numerical trial. There were four short-forward (two within and two across), four long-forward (two within and two across), four short-backward (two within and two across) and four long-backward (two within and two across) questions, 16 questions in total, which were presented using both linguistic and numerical input.

6.3. Month calculations

Month calculations also spread across 32 questions. As shown in Table 2, regarding 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 referred to calculation months chronologically preceding the stimuli. Regarding Boundary, within-boundary referred to trials within the boundary of a year and across-boundary referred to trials crossing the December–January boundary. All questions were presented in linguistic and numerical input as in the Day Calculation Task. One example of a 'short, forward, within, linguistic' trial is 'January + 2 months = ?' and an example of 'long, backward, across, numerical' trial is 'sixth month – 9 months = ?'. There were 16 linguistic trials and 16 corresponding numerical trials as in the Day calculation task.

6.4. Hour and year calculations

Differences in transparency in the numerical system between English and Chinese are far fewer than those in calendar terms, which guided the prediction that English and Chinese speakers may perform similarly in hour and year calculation questions, 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 modular (24-based calculations) and base (10-based calculations) questions. Just like for the critical calculations (Month and Day), these two control calculation tasks (Hour and Year) were manipulated for Distance (short/long), Direction (forward/backward) and Boundary (within/across).

There were 16 Hour and 16 Year calculation questions, each with four short-forward (two within and two across), four long-forward (two within and two across), four short-backward (two within and two across) and four long-backward (two within and two across) trials. The condition Input (linguistic/numerical) was manipulated across rather than within these two control conditions to keep the experiment length manageable. All Hour calculation questions, as illustrated in Table 3, were described in the linguistic version, while all Year calculation questions, as shown in Table 4, were presented in the numerical version. There are two further reasons to rationalize having two control conditions, each with a different type of input (linguistic or numerical). First, sufficient robustness of control is deemed to come from the inclusion of two control conditions, where one is with numerical input in Year calculations and the other is linguistic input in Hour calculation questions. No between-group differences were expected in either of these two conditions. Second, involving both Year and Hour calculation tasks was important to test different groups' mathematical abilities in modular calculations (Hour) and base calculation questions (Year). For example, '8 + 3 hours = ?' is a short, forward, within and linguistic Hour (modular) calculation question, while '1800 – 1400 = ?' is a long, backward, across and numerical Year (base) calculation question.

6.5. Self-reported strategies

In addition to calculations, each participant completed a Self-reported Strategies task. This task monitored the strategies used by the participants in different conditions of different calendar reasoning questions. In total, the combination of the conditions

Table 3. Examples of Hour calculation questions in different conditions

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 4. Examples of Year calculation questions in different conditions

Distance	Short	10 + 60 years =?
	Long	1300 + 600 yeas =?
Direction	Forward	10 + 60 years =?
	Backward	90–55 years =?
Boundary	Within	10 + 60 years =?
	Across	50 + 30 years =?

resulted in (i) 16 Day and 16 Month calculation questions, namely Distance (short/long) * Direction (forward/backward) * Boundary (within/across) * Input (linguistic/numerical), in other words, a two-by-two-by-two-by-two design and (ii) 8 Hour and 8 Year calculation questions, namely Distance (short/long) * Direction (forward/backward) * Boundary (within/across), in other words, a two-by-two-by-two design. There were seven strategies available to select from, including Memory, Transform and Calculate, Count, Translate and Calculate, Translate and Count, B and Other (Luo, 2012). The full task used for the self-reports is available at https://osf.io/3h97t/.

6.6. Procedure

To start the experiment, participants were instructed to click on the link saying Calendar Calculation task. The click triggered an instruction message shown on the computer screen followed by four practice calculations. The question (e.g., ‘Monday + 3 days = ?’) appeared in the center of the screen and three options (e.g., ‘Wednesday’, ‘Thursday’, ‘Friday’) appeared below the question at the same time, out of which one was correct and the other two were incorrect. The correct answer appeared randomly on the right, left or up, with about one-third in each of the three possible locations. The instruction asked the participants to choose the correct answer as fast and accurately as possible by pressing one of the direction arrows on the keyboard, ‘↑’, ‘←’ and ‘→’. The task comprised 96 calendar reasoning questions in total. Their RTs (from when the target word(s) appeared on the screen until the button-press) and accuracy were recorded using PsychoPy 3.0.

Following the Calendar Calculation task, there was a short break, after which participants could start the Self-reported Strategies Task, whenever they felt 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 shown in Table 5. Participants were asked to press the number keys ‘1’, ‘2’, ‘3’, ‘4’, ‘5’, ‘6’ or ‘7’ on the keyboard to choose one strategy that they thought best reflected their mental processes. Participants could spend as long as needed for the Self-reported Strategies Task, it was made explicit that their RTs would not be recorded and that there were no (in)correct answers.

The Calendar Calculation task and Self-reported Strategies Task were presented in different languages across groups. The English group and the Bilingual group were tested in English, while the Chinese group was tested in Chinese. All questions and forms in the two tasks given in the Chinese version were direct translations of the materials used in the English version of the tasks. The tasks in Chinese were full equivalents of the tasks in English.

Table 5. Illustration of the Self-reported Strategies task

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. **B** (Guess the answer)
7. **Other** (Another method not listed here)

7. Results

Participants were required to complete two experimental tasks: the Calendar Calculation task and the Self-reported Strategies Task. The measured variables in Calendar Calculations were RTs and Accuracy, while Self-reported Strategies gathered information about how participants thought they solved the calendar calculation questions. The overall accuracy, out of the total of 2880 trials, was 89.34% in the English group (SD = 0.31), 95.31% in the Chinese group (SD = 0.21) and 94.34% in the Bilingual group (SD = 0.23).

We next analyzed RTs to compare the cognitive demands of the answering systems in each group. RTs were only included from correct answers. As for the outliers, we followed Keating and Jegerski (2015) and Norris (2015), and excluded RTs that were more than 2.5 SDs away from the group mean in each condition. In the English group, outliers represented 2.95% of all RTs (namely 31-day, 27-month, 14-hour and 14-year calculation questions). In the Chinese group, outliers represented 2.92% of all RTs (namely 29-day, 25-month, 15-hour and 15-year calculation questions). In the Bilingual group, outliers represented 2.47% of all RTs (namely 22-day, 24-month, 9-hour and 16-year calculations). For RT analyses, we used linear mixed-effects models following Winter (2013). The main fixed-effect factors were Calculation (Day and Month) and Group (English, Chinese, Bilingual). The effect of different conditions, that is, Distance (short/long), Direction (forward/backward), Boundary (within/across) and Input (linguistic/numerical), were treated as within-group factors. The random-effect factors were Participant and Item.

7.1. RTs in Day and Month calculations

The analysis of RTs of different groups in the Calendar Calculation task comprised four steps. We first visualize the RT distributions, explore the effects of main factors, run between-group pairwise comparisons and then analyze the effects of specific manipulations in different conditions. The RTs for the Day and Month Calculation Tasks per group are illustrated in Figure 1. Overall, Month calculations (M = 5401, SD = 3744) took longer than Day calculations (M = 4735, SD = 4340). The Chinese group (M = 4570, SD = 3603) was faster than the Bilingual group (M = 4799, SD = 2983) and the English group (M = 4843, SD = 3186) in the Day Calculation Task. In the Month Calculation Task, the Chinese group was the fastest (M = 4780, SD = 3903), while the Bilingual group (M = 5422, SD = 3112) was slightly faster than the English group (M = 6074, SD = 4070).

To test the effect of Group on RTs when solving calendar calculation questions, we built mixed-effects regression models using the lme4 package (Baayen et al., 2008) in R (Version 4.1.3 R Development Core Team, 2021). As fixed factors, we entered Calculation (Day and Month) and Group (English, Bilingual and Chinese). The dependent variable was 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 a 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 multiplication sign in the model was then replaced by the plus sign and the two models were compared to test if there was a significant interaction between Calculation and Group, and also to test whether this interaction significantly improved the model fit (following Winter, 2013)).

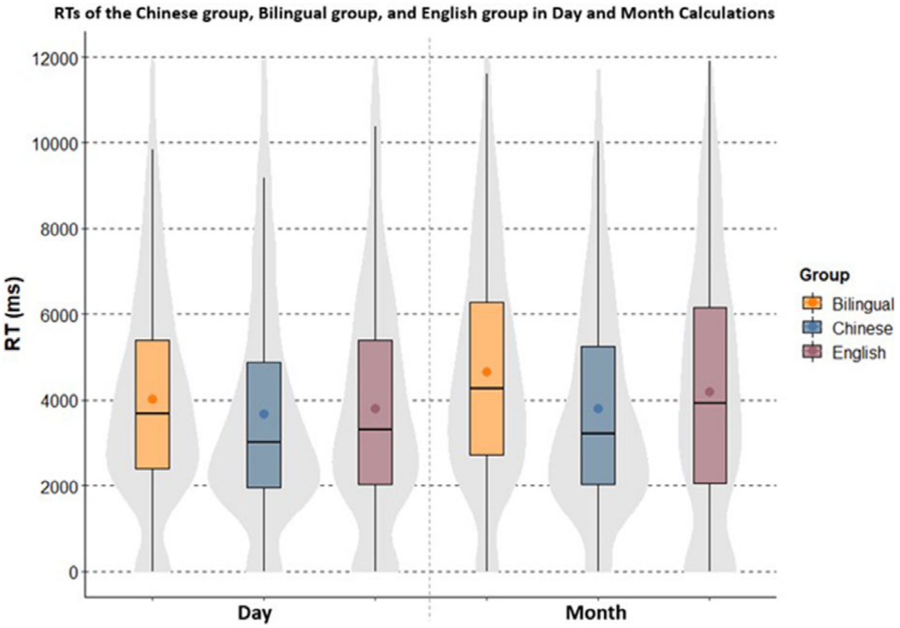

Figure 1. RTs of the Chinese group, Bilingual group, and English group in Day and Month calculations (Error Bars=95% CI)

As detailed in Table 6, the results showed no significant difference, $X^2(13) = 4.695$, $p = 0.9813$, so the reported final model excluded interactions. The dataset with RTs per participant is available at https://osf.io/3h97t/.

To further explore the effect of Calculation, a full model including Calculation and a reduced model excluding Calculation were compared to statistically test whether participants' RTs significantly differed in processing Day and Month calculations. This comparison showed that the model fit was significantly improved with the presence of Calculation, $X^2(7) = 19.631$, $p = 0.006$, confirming that Calculation was a significant predictor of how RTs varied.

Table 6. Coefficients from a mixed effects model fitted to the RTs of the Bilingual, Chinese, and English group in the Day and Month calculations

Fixed effects	*b*	SE	df	*t*	*p*
(Intercept)	4336.35	354.02	29.81	12.25	< 0.001***
Calculation (month)	917.95	433.79	32.16	2.12	0.042*
Group (Chinese)	−321.76	266.15	92.76	−1.21	0.230
Group (English)	−53	274.73	93.14	−0.20	0.846
Random effects	**Variance**	**SD**			
Participants (Intercept)	564,970	751.6			
Calculation (month)	116,360	341.1			
Item (intercept)	1,698,351	1303.2			
Calculation (month)	4,723,735	2173.4			
Group (Chinese)	357,652	598			
Group (English)	493,481	702.5			

Significance levels: ***$p < 0.001$, *$p < 0.05$.

Table 7. Tukey-adjusted pairwise comparisons of participants' RTs in the Day and Month Calculation Tasks between groups (CI = 95%)

Calculation	Group	Difference	Lower	Upper	*p*
Day	English versus Chinese	27.73	−337.94	393.40	0.983
	English versus Bilingual	−161.77	−527.44	203.89	0.553
	Chinese versus Bilingual	−189.50	−555.17	176.17	0.444
Month	English versus Chinese	593.31	176.53	1010.10	0.003**
	English versus Bilingual	14.09	−402.69	430.88	0.997
	Chinese versus Bilingual	−579.22	−996.01	−162.43	0.003**

Significance level: $^{**}p < 0.01$.

Then, using forward variable selection, the analysis zoomed in on Day and Month calculations to investigate the influence of Group by comparing a model including Group with a reduced model without Group in the data for Day calculation questions only. This comparison confirmed a between-group difference in participants' RTs when answering Day calculation questions, $X^2(7) = 84.008$, $p < 0.001$. The next comparison between the full vs reduced models without Group with data from Month calculation questions also showed that Group was a significant contributor to different groups' RTs, $X^2(7) = 32.952$, $p < 0.001$.

Tukey-adjusted pairwise comparisons were run to explore more closely how the RTs of each group differed from one another in the Day and Month Calculation Tasks. The results are shown in Table 7. No significant between-group difference emerged for Day calculations (Chinese group, $M = 4570$, SD = 3603; Bilingual group, $M = 4799$, SD = 2983; English group, $M = 4843$, SD = 3186). However, for Month calculations, there was a significant difference between the English group and the Chinese group (Mean difference = 593.31, $p = 0.003$) and between the Chinese group and the Bilingual group (Mean difference = −579.22, $p = 0.003$). No significant difference emerged between the English group and the Bilingual group (Mean difference = 14.09, $p = 0.997$). Chinese speakers ($M = 4780$, SD = 3903) were faster than the English speakers ($M = 6074$, SD = 4070), but Chinese–English bilinguals ($M = 5422$, SD = 3122) did not differ from English monolinguals. Overall, Day calculation questions took similar time while Month calculations differed across groups.

7.2. Day calculations by factor

We next examined the influence of four factors (Distance, Direction, Boundary, Input) on English group's RTs for Day calculations (as visualized in Figure 2). For the subset of data from the English group, the four factors were used in the model as fixed-effect factors, and Participant and Item as random-effect factors, RTs ~ Distance*Direction*Boundary*Input + (1|Participant) + (1|Item). Distance, Direction, Boundary and Input returned a significant interaction, $X^2(11) = 25.218$, $p = 0.008471 < 0.01$. Numerical input showed a negative influence on English monolinguals' RTs ($\beta = 2589.88$, SE = 556.05, $t = 4.66$, $p < 0.001$). The other factors did not significantly predict RT variation in the English group. Another significant interaction emerged between Distance and Input ($\beta = -2466.75$, SE = 786.37, $t = -3.137$, $p = 0.002$), indicating that the differences between RTs in short trials and long trials were larger with linguistic input than with numerical input.

The same analyses were conducted with the Chinese group's RTs. Within this subset, there was a significant interaction between the four fixed factors $X^2(11) = 129.95$, $p < 0.001$. To unpack this interaction, the Chinese group performed

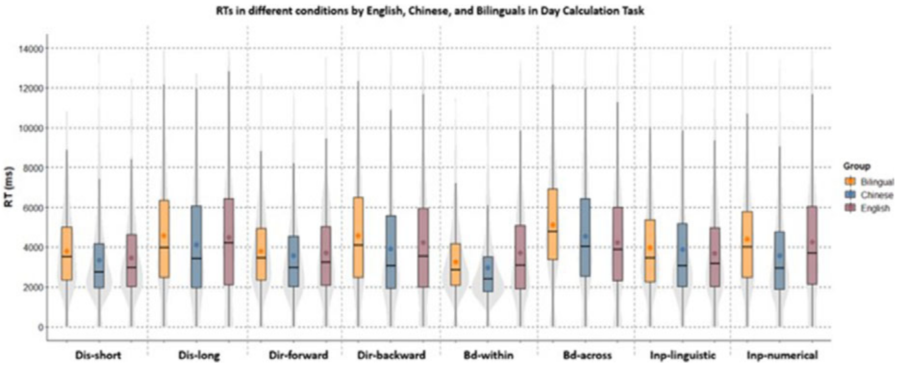

Figure 2. RTs in different conditions by the English, Chinese, and Bilingual group in Day calculations (Error Bars=95% CI). (“Dis”=Distance, “Dir”=Direction, “Bd”=Boundary, and “Inp”=Input)

significantly faster in short than long trials ($\beta = -2993.20$, SE = 590.30, $t = -5.071$, $p < 0.001$), forward than backward trials ($\beta = -3299.90$, SE = 590.30, $t = -5.590$, $p < 0.001$), within-boundary than across-boundary trials ($\beta = -5162.50$, SE = 590.30, $t = -8.746$, $p < 0.001$). Unlike in the English group, in the Chinese group, numerical trials were solved significantly faster than linguistic trials ($\beta = -2845.50$, SE = 590.30, $t = -4.820$, $p < 0.001$). As for interactions, a significant one was found between Distance and Boundary ($\beta = 2352.70$, SE = 834.80, $t = 2.818$, $p = 0.005$), indicating that the RT gap between short trials and long trials was larger in within-boundary trials than in across-boundary trials. Moreover, a significant interaction between Direction and Boundary ($\beta = 3511.00$, SE = 834.80, $t = 4.206$, $p < 0.001$) showed a larger RT gap between forward and backward trials in within-boundary trials than across-boundary trials. A significant interaction between Distance and Input was also found ($\beta = 2789.20$, SE = 834.80, $t = 4.206$, $p < 0.001$), showing that RT differences in short and long trials were larger with numerical input than with linguistic input. Furthermore, Direction and Input significantly interacted too ($\beta = 2539.20$, SE = 834.80, $t = 3.042$, $p = 0.002$), revealing that the RT gap between forward and backward trials was larger with numerical than linguistic input. Boundary and Input significantly interacted too ($\beta = 2315.40$, SE = 834.80, $t = 2.774$, $p = 0.006$), which indicated a bigger RT gap for within-boundary and across-boundary trials with numerical than linguistics input.

Analyses of RTs within the Bilingual Group’s Day calculations showed that the four fixed factors significantly interacted, $X^2(11) = 26.60$, $p = 0.005$. Long-distance ($\beta = 2422.20$, SE = 490.00, $t = 4.934$, $p < 0.001$), backward direction ($\beta = 1692.30$, SE = 490.00, $t = 3.448$, $p < 0.001$) and across-boundary trials ($\beta = 3374.50$, SE = 490.00, $t = 6.874$, $p < 0.001$) exerted a significant negative influence on Chinese–English bilinguals’ Day calculation speed. Moreover, significant interactions were between Distance and Boundary ($\beta = 1961.30$, SE = 694.20, $t = 2.825$, $p = 0.005$) and between Direction and Boundary ($\beta = 1682.20$, SE = 694.20, $t = 2.423$, $p = 0.016$), closely aligning with the interactions found for the Chinese group. Linguistic versus numerical input made no pronounced difference for the Bilingual group ($\beta = 627$, SE = 490.30, $t = 1.278$, $p = 0.202$).

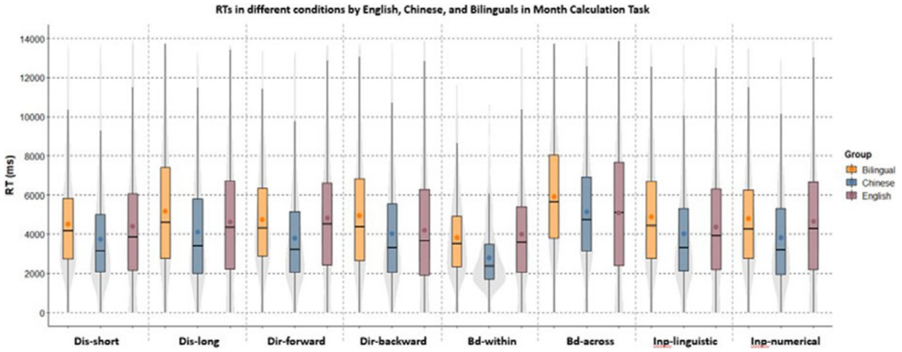

Figure 3. RTs in different conditions by the English, Chinese, and Bilingual group in Month calculations (Error Bars=95% CI). ("Dis"=Distance, "Dir"=Direction, "Bd"=Boundary, and "Inp"=Input)

7.3. Month calculations by factor

The month calculation RTs of participants from the three groups are visualized in Figure 3. Long distance, backward direction and across-boundary negatively affected all groups' RTs. As for between-group differences, numerical input negatively affected the English group's RTs ($M = 6479$, SD = 4608 for numerical; $M = 5668$, SD = 3397 for linguistic input) but showed no pronounced influence on the Chinese group ($M = 4724$, SD = 3931 for numerical; $M = 4836$, SD = 3887 for linguistic input) or the Bilingual group ($M = 5491$, SD = 3432 for numerical; $M = 5352$, SD = 2747 for linguistic input). These results were in line with those in Day calculations.

We next examined the influence of Distance, Direction, Boundary and Input, first on English group's RTs for month calculations. The results showed that only the cross-boundary trials showed a significant negative influence on the RTs of English monolinguals in month calculation questions ($\beta = -3329.41$, SE = 722.00, $t = -4.611$, $p < 0.001$). Moreover, there was a significant interaction between Direction and Boundary ($\beta = 2346.70$, SE = 1021.07, $t = 2.298$, $p = 0.022$), indicating that the RT differences between forward and backward trials were larger in within-boundary than in across-boundary trials.

The same analyses were run using the subset of the month calculation RT data from the Chinese group. The four fixed factors significantly interacted with each other $X^2(11) = 30.593$, $p = 0.001277$. Looking more closely, Distance ($\beta = -2647.90$, SE = 607.50, $t = -4.358$, $p < 0.001$), Direction ($\beta = -1896.90$, SE = 607.50, $t = -3.122$, $p = 0.002$) and Boundary ($\beta = -5053.40$, SE = 607.50, $t = -8.318$, $p < 0.001$) significantly influenced the Chinese group's performance. There was no significant difference between RTs in answering numerical and linguistic input trials ($\beta = 672.40$, SE = 607.50, $t = 1.107$, $p = 0.269$). Distance and Boundary significantly interacted ($\beta = 2714.50$, SE = 859.20, $t = 3.159$, $p = 0.002$) and so did Direction and Boundary ($\beta = 2158.50$, SE = 859.20, $t = 2.512$, $p = 012$), in line with Chinese group's results in the Day calculations. A further significant interaction was found between Distance and Direction ($\beta = 1699.20$, SE = 859.20, $t = 1.978$, $p = 0.049$), indicating that the RT gap between short and long trials was larger in forward than in backward trials.

For bilinguals' RTs in Month calculations, a significant four-way interaction emerged too, $X^2(11) = 34.42$, $p = 0.000306$, so 'RT ~ Distance * Direction * Boundary * Input + (1 | Participant) + (1 | Item)' was chosen as the reported model. The facilitatory influence of short distance ($\beta = -1542.20$, SE = 516.10, $t = -2.988$,

$p < 0.001$) and within-boundary ($\beta = -4045.40$, SE = 516.10, $t = -7.838$, $p < 0.001$) reached statistical significance. Furthermore, there were significant interactions between Distance and Boundary ($\beta = 2750.10$, SE = 729.90, $t = 3.768$, $p < 0.001$) and Direction and Boundary ($\beta = 2979.10$, SE = 729.90, $t = 4.081$, $p < 0.001$).

7.4. RTs in Hour and Year calculations

The Hour (Figure 4) and Year calculation (Figure 5) RTs of participants from the three groups (English, Chinese and Bilingual) were also visualized.

The same analytical procedures were applied to compare the arithmetic skills in modular Hour calculations and base Year calculation questions. The results are shown in Table 8.

To explore the effect of Calculation (Hour/Year), a reduced model excluding Calculation compared with the full model and the results confirmed a significant contribution of Calculation. That is, it took participants longer to solve the Hour questions than the Year questions. The analysis then zoomed in on Hour and Year Calculation Tasks, respectively. The hour calculation questions were answered slower and less accurately than the year calculation questions, reflecting that the modular hour calculations were more difficult to process than the base year calculations. The results showed no between-group difference in participants' RTs either in answering

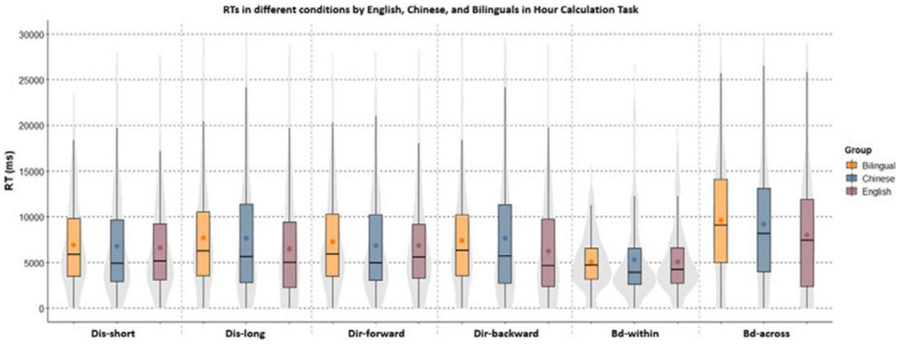

Figure 4. RTs in different conditions by the English, Chinese, and Bilingual group in Hour calculations (Error Bars=95% CI). ("Dis"=Distance, "Dir"=Direction, and "Bd"=Boundary)

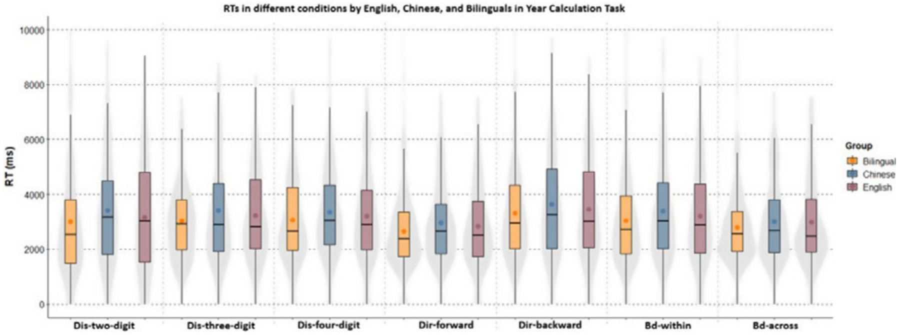

Figure 5. RTs in different conditions by the English, Chinese, and Bilingual group in Year calculations (Error Bars=95% CI). ("Dis"=Distance, "Dir"=Direction, and "Bd"=Boundary)

Table 8. Coefficients from a mixed effects model fitted to the RTs of the Bilingual, Chinese, and English group in the Hour and Year calculation

Fixed effects	*b*	SE	Df	*t*	*p*
(Intercept)	6987.93	782.18	13	8.93	< 0.001***
Calculation (year)	−3691.89	769.02	13.33	−4.80	< 0.001***
Group (Chinese)	503.26	322.02	51.05	1.56	0.124
Group (English)	−47.80	354.45	35.24	−0.14	0.893
Random effects	**Variance**	**SD**			
Participants (intercept)	4,216,711	2053.5			
Calculation (year) 2977917	2,977,917	1725.7			
Item (intercept)	4,857,938	2204.1			
Calculation (year)	3,018,465	1737.4			
Group (Chinese)	195,282	441.9			
Group (English)	467,092	683.4			

Significance level: ***$p < 0.001$.

Hour questions ($X^2(7) = 13.457$, $p = 0.06173$) or in answering Year questions ($X^2(7) = 5.1302$, $p = 0.6441$). This confirmed the prediction that groups would not significantly differ in either of these control conditions.

To sum up the main patterns, the processing speed of the year calculations was the fastest, followed by the speed of day and month calculations, and the speed of hour calculation questions was the slowest. Group was a significant predictor of RTs in month calculations but not in day, hour or year calculations. In month calculations, the English group performed significantly differently from the Chinese group but not from the Bilingual group. The effects of different conditions — Distance, Direction, Boundary and Input – were found on participants' RTs in the day and month calculations. Table 9 summarizes if Distance, Direction, Boundary and Input affected each group's RTs in statistically significant ways in day and month calculations. In day calculations, the fixed factors Distance (short), Direction (forward) and Boundary (within) positively affected Chinese speakers' RTs but showed no significant influence in the English group. Input (numerical) negatively affected the English group but positively influenced the Chinese group, while no significant effect was found in the Bilingual group. Results in month calculations aligned with those in day calculations, except that the English group performed significantly better in within-trials than across-trials, and that Input showed no significant influence in any group.

7.5. *Self-reported strategies*

The participants' self-reported strategies were collected to shed more light on the cross-linguistic differences in calendar calculation questions. Participants' strategy

Table 9. An overview of whether the fixed factors significantly affected RTs in Day/Month calculations

	Group	Distance	Direction	Boundary	Input
Day	English	✖	✖	✖	✓
Calculation	Chinese	✓	✓	✓	✓
Task	Bilingual	✓	✓	✓	✖
Month	English	✖	✖	✓	✖
Calculation	Chinese	✓	✓	✓	✖
Task	Bilingual	✓	✖	✓	✖

options were coded as ‘Memory’, ‘Calculate’, ‘Count’, ‘Guess’ and ‘Other’. For day calculations, 55% of English speakers reported the Count strategy, while the rest reported using the Calculate strategy (23%) and the Memory strategy (20%). Most respondents (68%) from the Chinese group preferred Calculate and only 5% reported Count. The Bilingual group reported similar strategies as the Chinese group (74% Calculate, 15% Count and 10% Memory).

The self-reported strategies were further analyzed within conditions, namely Distance (short/long), Direction (forward/backward), Boundary (within/across) and Input (linguistic/numerical). Regarding Distance, for the short trials English respondents reported more Memory strategy (29%) and less Calculate strategy (15%) than in the long trials. Moreover, the English speakers reported more Memory strategy (30%) and less Count strategy (49%) in within-trials than that in across-trials (Memory strategy: 11.25% and Count strategy: 62%). Additionally, with numerical input, the English respondents applied more Calculate strategy (29%) and less Count strategy (50%) compared to the linguistic trials (Calculate strategy: 18% and Count strategy: 61%). The Chinese group reported a preference for the Calculate strategy in all conditions. One notable difference occurred under the influence of Boundary, where more Count strategy was reported in the across condition (7.5%) than in the within condition (0.83%). The bilinguals’ strategy reports were similar, except that some respondents translated the questions before they answered.

For Month calculations, like day calculations, most Chinese-dominant speakers (71%) and Chinese–English bilinguals (74% for Calculate) reported using the ‘Calculate’ strategy. Unlike the less demanding day calculation task, English respondents reported relying less on the Memory strategy (12%) but more on the Calculate strategy (32%), although there were still more than 50% of English speakers who reported a Count strategy (51%). The results aligned with the self-reported strategies in the day calculations. In the English group, the Calculate strategy was more popular in long trials (37%) than in short trials (27%) and more popular in numerical input (33%) than in linguistic input (30%), while the strategy Count was more popular in boundary crossing trials (56%) than in boundary within trials (46%).

For the Hour calculations, the pattern in self-reported strategies was somewhat different. In total, all respondents reported a preference for the Calculate strategy in processing, though the proportion in the Chinese group (74%) and Bilingual group (83%) was higher than that in the English group (52%), some of whom also counted to solve these questions (32%).

For year calculations, most participants from the three groups relied on the Calculate strategy (70% for the English group, 70% for the Chinese group and 71% for the Bilingual group).

In sum, in day calculations, the Calculate strategy was preferred by the Chinese and Bilingual group, while the English speakers preferred the Count strategy. When solving month calculations, results of self-reported strategies still showed Chinese speakers’ reliance on Calculate as in the day calculations, although the English speakers expressed a higher preference for calculating and less for counting in month calculations, especially in long distance and backward direction trials.

8. Discussion

The main finding is that a linguistically and numerically transparent calendar naming system can positively affect performance in calendar calculation tasks with increased cognitive demand. In response to the first research question, significant differences between the English group and the Chinese group emerged for Month calculations but not for Day calculations. One reason for the latter could be the Count strategy, used to a comparable extent in both groups and efficient enough to solve relatively simple seven-based calculations. In Month calculations, the Chinese group was faster than the English group, arguably because of the transparent Chinese calendar terms and the preference for numerical operation strategy. The second research question targeted the idea that Chinese–English bilinguals' performance in calendar calculations in an English context would differ from that of English monolinguals. The results showed no significant differences between calculation speed in the English group and the Bilingual group in the critical Day and Month calculations. One possible explanation is that slowdowns induced by doing calculations in a weaker language may swallow processing advantages that come with knowing transparent calendar terms, even when using a numerical strategy to calculate. The advantage of knowing a linguistically transparent calendar naming system helped the Chinese–English bilinguals perform as well as the English native speakers in Day calculations, despite doing the task in their weaker language. In Month calculations, which involve a longer list of calendar terms, the Bilingual group numerically outperformed the English group. However, a potential advantage of a Calculate strategy needs to be viewed with caution as the difference did not reach statistical significance. Even though much of prior research suggests that performance in L2 is generally worse than in L1, especially in late unbalanced bilinguals (van Gelderen et al., 2004; Silva & Clahsen, 2008; Trenkic & Warmington, 2019), this study demonstrates that a transparent calendar naming system of Chinese as L1 can provide some processing boost in performing calendar calculations in L2 English, at least to an extent that it gets on par with English native speakers' accuracy and response speed. We contextualize the main findings within research on temporal reasoning and the Ad Hoc Cognition framework (Casasanto & Lupyan, 2015).

8.1. Implications for temporal reasoning and bilingual cognition research

This study corroborates previous findings that documented non-trivial cross-linguistic differences in calendar calculations. Looking at the English-speaking group separately, the task-solving pattern emerging from the present study was in line with the model proposed by Friedman (1990), where the verbal listing process draws on the mental picture of the list of calendar terms. However, the strategy preferred by the Chinese group was different. Calculating was the most popular strategy and there was almost no oscillation away from it to alternative strategies (Kelly et al., 1999). Between-group differences in calendar calculation tasks can be attributed to the level of transparency in English and Chinese calendar terms for two reasons. First, unlike the performance differences found in mathematical calculations of Chinese versus English speakers (Ng & Rao, 2010), the cross-linguistic differences in calendar calculations shown in the present study can be linked to mainly language-based rather than cultural or educational factors, especially when one considers that calendar calculations are not taught or tested in one's school days (Bassetti et al., 2018). Second, the English, the Chinese and the Bilingual groups all performed similarly in the control conditions, namely in Hour and Year calculations, both in response speed and accuracy, as well as in self-reported strategies. These findings ascertain that different groups' preferences for various strategies were not linked to differences in arithmetic abilities.

Another aspect these findings bring to light is that one's native language system, in this case of calendar terms, can affect cognitive processes that can go beyond just the verbal domain. First, under the influence of different levels of transparency of calendar terms in English and Chinese, native English speakers preferred verbal listing, while speakers of Chinese relied on the numerical calculation strategy. Second, numerical calculations facilitated by Chinese transparent calendar terms were still being applied by Chinese–English bilinguals, even when they performed calendar calculation tasks in a second language with opaque calendar terms. It appears that differences across languages can not only affect their native speakers' cognitive routines in the corresponding native language environment but also assist bilinguals when they operate in a second language in which calendar labels happen to be opaque. The findings point to greater difficulty for the English group to solve Month compared to Day calculations in English, while the Chinese group and the Bilingual group exhibited no such discrepancy. This set of patterns replicates Bassetti et al. (2018), jointly pointing out that the added cognitive effort inherent in performing in a weaker language could be eclipsed when the native language provides a more efficient problem-solving method, which is particularly useful when the task gets more complex. The strategy reports helped to elucidate that the process driving the calculations is conscious rather than automatic.

Day calculations exhibited most fragility to contextual manipulations of input. Changes from linguistic to numerical input asymmetrically affected response times, slowing down English speakers but speeding up Chinese speakers. Number-induced slowdowns in the English group can be explained as an effect of temporary/ad hoc (Casasanto & Lupyan, 2015) destabilization of the relevant experiential priors (verbal listing). By the same token, number-induced response speed increases in the Chinese group can be accounted for as an effect of upregulation of the relevance of experiential priors (arithmetic computations). Contextual changes of input did not affect bilinguals, arguably as their representations of day labels/numbers, and the related calculation processes, can adjust more flexibly than those of monolinguals. The absence of input-driven effects in Month calculations in both Chinese and English speakers could stem from the experience that months, unlike days, are commonly expressed both numerically and linguistically in written Chinese as well as in written English. To specify, in English, a month such as 'January' is commonly expressed in writing using a numerical label '01' in a date like '30/01/2024' or as a linguistic label 'January 30th, 2024'. Likewise, the same month in Chinese can be either '2024 年1月30号' (numerical label) or '二零二四年一月三十号' (linguistic label), where both 1月and 一月 read as 'yi yue'.

One might wonder whether calculations using calendars are scaffolded not only by numerical expressions but also by the artifact-based construction of calendars. Previous research shows that mental representations of calendars could potentially be stored and employed during calculations, just as mental abacus computations (i.e., a system for performing fast and accurate arithmetic by visualizing and manipulating images of a physical calculation device) work for East Asian users of the soroban abacus (Frank & Barner, 2012; Hatano & Osawa, 1983; Stigler, 1984). It is possible that visualizing calendars in the mind could assist calculations the same way that analog clock users use the mental representation of the clock in doing time calculations (especially those shorter than or around an hour). Visuospatial or abacus-like facilitation of calendrical calculations remains an exciting possibility, the epistemological implications of which are yet to be identified. One way to test how robust a role mental simulation of calendars plays in calculations could be via an adaptation of the blank screen paradigm (Vanek et al., 2024a, 2024b) to measure anticipatory eye movements in the presence versus absence of pictured abacus primes.

9. Limitations and suggestions for future research

First, participants performed all calendar calculations and strategy reports unsupervised. Only written instructions rather than face-to-face guidance and monitoring were involved. Possible effects of the testing environment left uncontrolled cannot be ruled out. Uniform testing conditions across groups could benefit future research. Second, the potential effects of participants' variation in expertise in calendar calculations cannot be neglected either. It is possible that the tested sample of young and educated adults differs in their ability to perform calculations, for instance, from older individuals. It could also be the case that certain professions train knowing months by number and writing them customarily as such (e.g., October 16 = 10/16) more than others. Considering this limitation in the current study, future studies on calendar calculations could benefit from a task-independent check of month conversion ability (e.g., by measuring the response speed of how quickly participants recognize that 10 in 10/16 stands for October). Average response times could then be incorporated as a factor in statistical analyses to control for potential individual differences in month-to-number/number-to-month conversion ability. Third, self-reported strategies may be too subjective, participants might have been unsure or may have wanted to conceal the actual strategies they used. It is also possible that the processes involved in calendar calculations are not readily available through introspection. During debriefs, several bilingual participants responded that they were struggling to decide whether they translated the questions or not as the calculation time was too short for a thorough self-analysis. It is also possible that the bilinguals might have thought choosing translation as a strategy might send the unwanted message that they lack proficiency in L2 English. Collecting confidence levels for (a selection of) strategy questions and building these into the analyses could aid internal validity in future studies. Informative further extensions could involve more complex calculation tasks to investigate bilinguals' cognitive processes under greater demand for attentional resources, or a bidirectional design also including a group of bilinguals with an opaque calendar lexicon in L1 whose L2 has transparent calendar terms, or possibly a training design manipulating the degrees of opacity in calendar labels ad hoc.

References

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

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

Bassetti, B., Clarke, A., & 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., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.

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.

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

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543–566). MIT Press.

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

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

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, 141(1), 134–149.

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

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), Article 650.

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*. MIT Press.

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

Grove, B. G. (1986). *Webster's third new international dictionary*. Merriam–Webster.

Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition*, 15(1–3), 95–110.

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2011). Language effects on children's nonverbal number line estimations. *Journal of Cross-Cultural Psychology*, 42(2), 598–613.

Huang, W. (1993). Zhongguo chengren tuili yuefen shijian de jiagong moxing [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.

Jiang, T., & Fang, G. (1997). Xiaoxue ertong dui xisu shijian de zhouqixing tedian de renzhi [Primary school pupils' cognition of cyclic aspects of conventional time]. *Xinli Kexue*, 5, 431–435.

Keating, G. D., & 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., & 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.

Laski, E.-V., & 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.

Libben, G., Gibson, M., Toon, Y., & 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., & Kuhl, P. K. (2011). Mental addition in bilinguals: An fMRI study of Task-related and performance-related activation. *Cerebral Cortex*, 22(8), 1851–1861.

Luo, T. (2012). Mental arithmetic across three language groups. [Unpublished doctoral dissertation]. San Jose State University.

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 139–162). John Benjamins.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum.

Mark, W., & 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.

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

Miller, K. F., Smith, C. M., Zhu, J., & 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., & 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., & Stigler, J. W. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Cognitive Development*, 2(3), 279–305.

Miura, I. T., Okamoto, Y., Vlahovic-Stetic, V., Kim, C. C., & 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., & 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.

Ng, S. S. N., & Rao, N. (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.

Park, H. I., & 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., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306, 499–503.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Siegler, R.-S., & Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychological Science*, 19(8), 759–763.

Silva, R., & 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.

Sinha, C., Sinha, V. D. S., Zinken, J., & Sampaio, W. (2011). When time is not space: The social and linguistic construction of time intervals and temporal event relations in an Amazonian culture. *Language and Cognition*, 3(1), 137–169.

Sherko, E. (2015). Linguistic transparency and opacity in compounding. *Academic Journal of Interdisciplinary Studies*, 4(3), 590–593.

Stigler, J. W. (1984). Mental abacus: The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology*, 16(2), 145–176.

Trenkic, D., & 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.

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

Vanek, N., Matić Škorić, A., Košutar, S., Matějka, Š., & Stone, K. (2024a). Mental simulation of the factual and the illusory in negation processing: Evidence from anticipatory eye movements on a blank screen. *Scientific Reports*, 14(1), Article 2844.

Vanek, N., Škorić, A. M., Košutar, S., Matějka, Š., & Stone, K. (2024b). Looks at what isn't there: Eye movements on a blank screen when processing negation in a first and a second language. *Frontiers in Human Neuroscience*, 18, Article 1457038.

Van Gelderen, A., Schoonen, R., Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology*, 96, 19–30.

Van Rinsveld, A., Brunner, M., Landerl, K., Schiltz, C., & Ugen, S. (2015). The relation between language and arithmetic in bilinguals: Insights from different stages of language acquisition. *Frontiers in Psychology*, 6, 265–265.

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

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

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

Yang, C., & Zhang, J. (2011). Yueyu - putonghua shuangyanzhe he putonghua danyanzhe zhouqixing shijian tuili bijiao [A comparison of cyclical temporal reasoning between Chinese - Chinese diglossia speakers and Chinese speakers]. *Xin Li Ke Xue*, 34(4), 782–787.

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

Cite this article: Zhuang, Z., & Vanek, N. (2025). When the days are(n't) numbered: Calendar calculations in transparent and opaque systems, *Language and Cognition*, 17, e26, 1–26. https://doi.org/10.1017/langcog.2024.75