Back to the future? How Chinese-English bilinguals switch between front and back orientation for time

Abbreviated title: Cross-language interference on time conceptualisation

Yang Li[1], Aina Casaponsa[2], Yan Jing Wu[3], and Guillaume Thierry[1]

1. School of psychology, Bangor University, Bangor, Wales, UK
2. Department of Linguistics and English Language, Lancaster University, Lancaster, UK
3. Faculty of Foreign Languages, Ningbo University, Ningbo, China

Abstract

The ability to conceive time is a corner stone of human cognition. It is unknown, however, whether time conceptualisation differs depending on language of operation in bilinguals. Whilst both Chinese and English cultures associate the future with the front space, some temporal expressions of Chinese involve a configuration reversal due to historic reasons. For instance, Chinese refers to the day after tomorrow using the spatiotemporal metaphor *hou-tian* – 'back-day' and to the day before yesterday using *qian-tian* – 'front-day'. Here, we show that native metaphors interfere with time conceptualisation when bilinguals operate in the second language. We asked Chinese-English bilinguals to indicate whether an auditory stimulus depicted a day of the week either one or two days away from the present day, irrespective of whether it referred to the past or the future, and ignoring whether it was presented through loudspeakers situated in the back or the front space. Stimulus configurations incongruent with spatiotemporal metaphors of Chinese (e.g., "Friday" presented in the front of the participant during a session held on a Wednesday) were conceptually more challenging than congruent configurations (e.g., the same stimulus presented in their back), as indexed by N400 modulations of event-related brain potentials. The same pattern obtained for days or years as stimuli, but surprisingly, it was found only when participants operated in English, not in Chinese. We contend that the task was easier and less prone to induce cross-language activation when conducted in the native language. We thus show that, when they operate in the second language, bilinguals unconsciously retrieve irrelevant native language representations that shape time conceptualisation in real time.

Key words: Bilingualism, spatiotemporal metaphors, semantics, event-related brain potentials, unconscious processing

Introduction

Conceptualising the passing of time is a core aptitude of the human mind. One of the most common ways to represent time, an abstract concept, is to use space, a concrete concept. However, linguistic metaphors from different languages use spatial axes in different ways. For instance, spatiotemporal metaphors of Chinese frequently refer to the sagittal (front-back) and vertical (up-down) axes to represent time (e.g., Boroditsky, 2001; Boroditsky, Fuhrman, and McCormick, 2011; Lai and Boroditsky, 2013). Western languages, in contrast, tend to rely more exclusively on the sagittal axis.

Languages even differ in terms of orientation along the same axis. Whereas Aymara, like Moroccan, associates the past with the front space (*nayra*) and the future with the back space (*qhipa*), the majority of languages place the future in front and the past in the back (Núñez and Sweetser, 2006; see also, de la Fuente, Santiago, Román, Dumitrache, and Casasanto, 2014). Variations even exist within languages, as is the case in Chinese, which conforms to a future-in-front convention (e.g., *qian-tu* – 'future prospects' literally translates into "front-path") but features exceptions with a reverse orientation along the same axis (e.g., *hou-tian* – 'the day after tomorrow', which literally translates as "back-day", Table 1).

Table 1. Spatiotemporal metaphors of Mandarin Chinese conflicting with the future-in-front convention

Chinese	Pin Yin	English translation	Time	Literal translation
后 天	hou-tian	the day after tomorrow	future	'back day'
前 天	qian-tian	the day before yesterday	past	'front day'
后 年	hou-nian	the year after next	future	'back year'
前 年	qian-nian	the year before last	past	'front year'

One fundamental question, however, is whether such linguistic differences are mirrored by differences at a conceptual level, that is, the question at the centre of the linguistic relativity debate (Lupyan, 2012; Slobin, 1996; Thierry, 2016; Whorf, 1956). In the domain of time representation, **Boroditsky (2001) reported that native speakers of Chinese solved temporal problems (e.g., Is "March comes earlier than April" correct?) faster after viewing pictures of vertically arranged objects than horizontally arranged ones. In contrast, English native speakers verified temporal statements faster after presentation of horizontal layout than vertical ones. Boroditsky thus argued that** native speakers of

Chinese predominantly conceptualise time along the vertical axis, whereas English natives predominantly embody time along the horizontal axis. **However, using the same paradigm as Boroditsky (2001), Chen (2007) failed to find significant reaction time differences between horizontal and vertical spatial priming in Chinese native speakers or English native speakers. In addition, in a corpus analysis, Chen (2007) observed that Chinese native speakers more frequently used horizontal spatial metaphors than vertical ones when expressing time (with the notable exception of temporal expressions containing "week"). This led Chen (2007) to argue that Chinese speakers, like English speakers, predominantly conceptualize time horizontally, despite the existence of vertically oriented spatiotemporal metaphors in Chinese. In addition, also against observations made by Boroditsky (2001), January and Kako (2007) and Tse and Altarriba (2008) showed that English native speakers take less time to respond to temporal sentences following a vertical than a horizontal prime. Therefore, data from behavioural studies have so far failed to reach a consensus on spatiotemporal interactions between language-specific metaphors and time conceptualisation.**

In order to assess how specific linguistic expressions such as spatiotemporal metaphors influence how speakers of different languages conceive time, we need an implicit, automatic, and unconscious index of conceptual processing that is resilient to strategic effects and does not rely on verbalisation (Thierry, 2016). A well-established such index is the N400 peak of event-related brain potentials (Kutas and Hillyard, 1980, 1984; Kutas, Lindamood, and Hillyard, 1984). Here, we set out to test whether spatiotemporal metaphors specific to Chinese that conflict with the future-in-front convention[1] selectively affect time conceptualisation in fluent Chinese-English bilinguals **operating in English or Chinese. It is well-established that lexical access in bilinguals is largely language non-selective and that the bilingual lexicon is highly integrated rather than fragmented by language (See the bilingual interactive activation model, van Heuven, Dijkstra, and Grainger, 1998; BIA+ model, Dijkstra and van Heuven, 2002). Previous research using the N400 as an**

[1] We chose the sagittal (front–back) axis for three reasons: (i) The sagittal axis is the most frequently used; (ii) It is common to Mandarin Chinese and English, which is critical because we tested Chinese-English bilinguals in the UK; (iii) Exceptional violations of the future-in-front convention only occur in Chinese.

index of cross-language activation established that there are automatic competition effects within and across languages at the lexical level, even when bilinguals operate in a monolingual language context (Thierry and Wu, 2004, 2007; Wu and Thierry, 2010, 2012; Hoshino and Thierry, 2012; Wen, Filik, and van Heuven, 2018; Meade et al., 2017; Lee, Meade, Midgley, Holcomb & Emmorey, 2019). Therefore, we predicted that Chinese-English bilinguals operating in English could suffer interference from spatiotemporal metaphors specific to Chinese.

We engineered a conflict between metaphor orientation and stimulus presentation along the front-back axis in the space around the participant. To our knowledge, no previous study has physically presented a stimulus in the back space surrounding participants, since all previous studies involved stimuli presented in the visual domain. In *Experiment 1*, we used days of the week as stimuli. For instance, when a participant was tested on a Wednesday, we presented the auditory stimulus ‘Friday’ through loudspeakers situated in the front of the participant, potentially clashing with the corresponding spatiotemporal metaphor of Chinese as compared to the same stimulus presented in their back, since the Chinese expression for ‘the day after tomorrow’ literally translates as “back-day” in English. We asked participants to make interval judgements (‘Is the date you hear one or two days away from today?’). Critically, sound origin in space was irrelevant as was the future or past reference afforded by the stimuli, and spatiotemporal metaphors were never presented or mentioned.

We expected that Chinese-English bilinguals would experience interference from conflicting metaphors of Chinese in the case of 2-day gaps, but not in the case of 1-day gaps since *ming-tian* – ‘tomorrow’ and *zuo-tian* – ‘yesterday’ are not spatiotemporal metaphors in Chinese (Table 2).

Table 2. Temporal expressions of Mandarin Chinese neutral vis-à-vis the future-in-front convention

Chinese	Pin Yin	English translation	Relative time	Literal translation
明天	ming-tian	tomorrow	future	‘bright day’
昨天	zuo-tian	yesterday	past	‘yesterday’
明年	ming-nian	next year	future	‘bright year’
去年	qu-nian	last year	past	‘gone year’

In *Experiment 2*, conducted in late 2017 in the same session as Experiment 1, participants made interval judgements about years instead of days. Our predictions were the same as for Experiment 1 (Fig. 1).

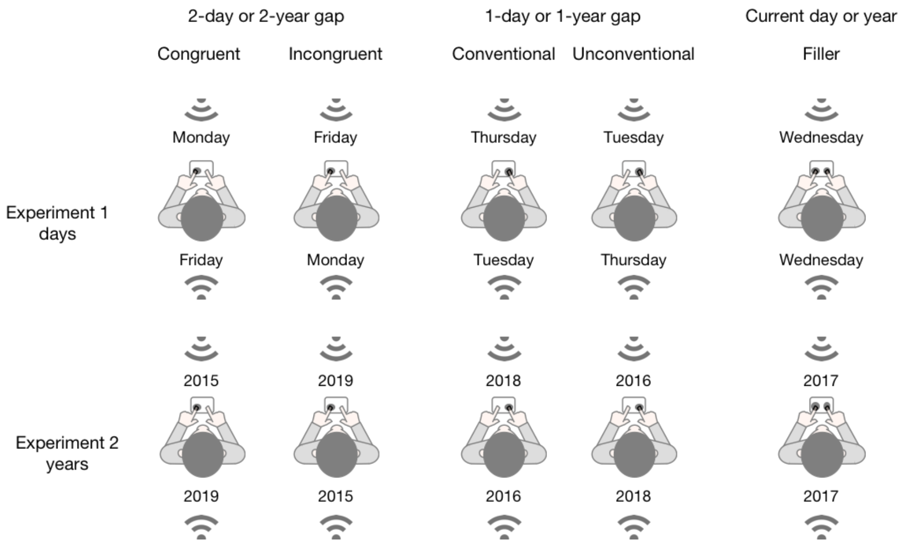

Figure 1. Experimental design. In experiment 1, participants heard days of the week presented through loudspeakers set in front of them and in their back. Stimuli depended on the day of testing (e.g., if the current day was Wednesday, stimuli were Monday, Tuesday, Wednesday, Thursday, and Friday in English and *xing-qi yi, xing-qi er, xing-qi san, xing-qi si,* and *xing-qi wu* in Chinese). Participants were instructed to press one button for stimuli one day away (in the future or the past) and the other button for stimuli two days away from the day of testing. For the current day, they had to press both buttons simultaneously (filler trial). In experiment 2, participants heard year labels: twenty-fifteen, twenty-sixteen, twenty-seventeen, twenty-eighteen, and twenty-nineteen (and *er-ling yi-wu, er-ling yi-liu, er-ling yi-qi, er-ling yi-ba, er-ling yi-jiu* in Chinese). Instructions were the same as in Experiment 1 but response was based on temporal distance in years, 2017 being the year of testing. Congruency is defined based on alignment between sound origin (front / back), temporal reference (future /past), and spatiotemporal metaphors of Mandarin Chinese.

Overall, we predicted that incongruent stimulus configurations involving 2-day or 2-year gaps presented from a location incompatible with the orientation embedded in native spatiotemporal metaphors of Chinese would differentially increase the amplitude of the N400 as compared to congruent configurations. In the case of 1-day or 1-year gaps, configurations violating the future-in-front convention were not expected to elicit semantic interference since no relevant spatial information was available, either in Chinese or in English.

Method

Participants

Twenty-four Chinese-English bilingual participants and 21 native speakers of English participated in this study. All participant took part in both Experiment 1 and Experiment 2. Data from 5 bilingual participants and 4 native speakers of English were discarded due to poor electrophysiological recording quality, excessively high impedances, excessive blinking, or insufficient number of trials per condition. All Chinese participants reported their International English Language Test System (IELTS) score (Mean = 6.3/ 9, SD = 0.4) and were resident in the UK at the time of testing. Bilingual participants self-reported their proficiency in both English and Mandarin Chinese (Fig. 2) and their language background is summarised in Table 3.

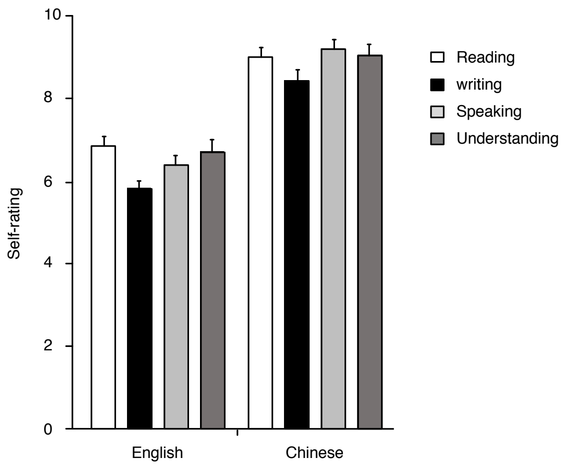

Figure 2. Chinese-English bilingual participants' self-estimation of their English and Chinese level (10 point-scale). Error bar represents stand error.

All participants had normal or corrected-to-normal vision and self-reported normal audition. Participants either received £15 or course credits for their participation in the study that was approved by the ethics committee of the School of Psychology at Bangor University. **We aimed at collecting more than 16 participants in each of the experimental groups in order to yield suitable statistical power for this experiment based on previous studies targeting similar effects in ERPs and spanning 9 years of research (e.g., Thierry and Wu, 2004, 2007).** We thus collected 21 participants in the native English group based on an average data attrition rate of ~10%, and 24 bilingual participants, since session duration was twice as long thus increasing data loss risks proportionally.

Table 3. Chinese-English bilingual participants' language background

Measure	Mean	SD
Age of L2 acquisition	9.42	2.36
Length of L2 learning (years)	14.26	3.57
Daily Chinese usage (%)	67	16.7
Daily English usage (%)	33	14.1

Materials

Stimuli consisted of digital audio files of days of the week and year numbers in Mandarin Chinese and English. All stimuli were recorded once in English by a native speaker of English and once in Chinese by a native speaker of Chinese. A cross-splicing procedure (using Adobe Audition™) was employed to ensure that participants could not guess the particular day or year stimulus presented in each trial on the basis of stimulus beginning alone[2]. Cross-splicing offered a good baseline and optimal accuracy in marking the onset of the critical information in the sound stream (Fig. 3).

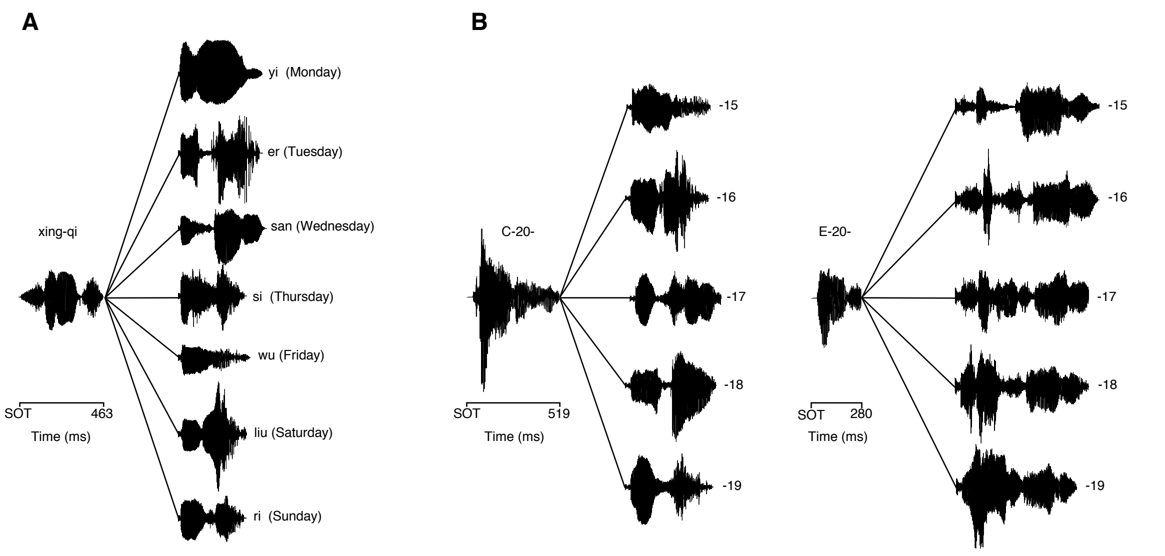

Figure 3. Cross-splicing procedure and stimuli presented. *(A)* Experiment 1 (days). *(B)* Experiment 2 (years).

[2] Note that for year stimuli in Chinese, we elected not to cross-splice between the decade digit (*yi* – 'one' in Chinese) and the final digit (5, 6, 7, 8, or 9) because of co-articulation in the case of *yi-wu* – 'fifteen', which would have created an artefact for that sound file. On average the duration of *yi* was 250 ms (range 230-272 ms), and thus RTs were artificially extended by the same duration in the corresponding condition.

In Experiment 1, stimuli consisted of the names for the 7 days of the week. For any participant, only 5 days of the week were presented in order to cover a time interval of two days before to two days after the day of testing. Average stimulus duration was 900 ±75 ms for days in Chinese and 845 ±66 ms for days in English day. Average auditory stimulus intensity was 48 dB (range 46–55 dB).

In experiment 2, stimuli were 4-digit numbers referring to 5 years surrounding the year of testing (2017), i.e., 2015, 2016, 2017, 2018, and 2019. Average stimulus duration was 1076 ±23 ms for years in Chinese and 1163 ±66 ms for years in English day. Average auditory stimulus intensity was 48 dB (range 47–52 dB).

Procedure

Participants first completed a language background and reading habits questionnaire whilst being fitted with the cap for electrophysiological recording. They were seated in the centre of a sound-attenuated testing booth, with two speakers located in the front and two speakers located behind them, set at a distance of between 1.4–1.6 meters from their ears. A 19-inch CRT monitor was placed 100 cm in front of their eyes and displayed a black fixation cross on a white background throughout the recording session. In experiment 1, participants were asked to judge whether each stimulus referring to a day of the week corresponded to a period of time situated one or two days away from the current day. In experiment 2, participants made the same judgements for stimuli referring to years. Responses were given by pressing designated left and right buttons on a response box. Response sides were counterbalanced between participants. Half of the stimuli were presented through the speakers located in front of the participant's chair, and the other half were presented in their back. When participants heard the current day or the current year, they were instructed to press both left and right buttons simultaneously. They heard 30 pseudo-randomly intermixed iterations of each individual stimulus condition. Apart from present day (one fifth of trials), half of the stimuli were one day away from the time of testing and the other half were two days away from the time of testing. Similarly, half of the stimuli referred to the future and half to the past, making a total of 300 trials per block in each experiment. Control native speakers of English performed the task in English only (600 trials in total) and Chinese-English participants performed the task once in English and once in Chinese (1200 trials in total) with order counterbalanced between languages **(all bilingual participants completed Experiment 1 or**

Experiment 2 first and then Experiment 2 or Experiment 1 accordingly. In addition, language order was counterbalanced between them). Every individual trial started with a pink fixation cross displayed in the centre of the screen for 300 ms. The fixation then turned to black after a pseudorandom inter-stimulus interval of 300–500 ms. The target auditory stimulus was then presented through loudspeakers either to the front or the back of the participant's chair whilst the black fixation stayed on the screen until participant's response with a maximum duration of three seconds from the onset of the sound stimulus. Participant's response immediately triggered a 200 ms inter-trial interval before the next pink fixation. Every 7 trials, the pink fixation lasted for four seconds, during which participants were encouraged to blink if they needed to, in order to minimise the occurrence of eye blink artefacts during the interval of time between auditory stimulation and response.

ERP recording and processing

Electrophysiological data were recorded at a rate of 1 kHz from 64 Ag/AgCl electrodes according to the extended 10-20 convention and referenced to electrode Cz. Impedances were kept below 5 kΩ. The electroencephalogram (EEG) was filtered using an online bandpass filter (0.05–200 Hz), and offline using a low-pass, zero phase-shift digital filter (0.1 Hz, 24 dB/oct–20 Hz, 28 dB/oct). Eye-blink artefacts were first manually removed through visual inspection of the data and the remaining artefacts were then mathematically corrected using the procedure advocated by Gratton, Coles and Donchin (1983). Epochs ranging from -200 to 1200 ms after stimulus onset were extracted from continuous EEG recordings. Epochs with activity exceeding ± 100 µV at any electrode site, except the vertical electroocculogram channels, were discarded. Baseline correction was performed in reference to pre-stimulus activity, and individual averages were digitally re-referenced to the global average reference.

Behavioural data analysis

Stimulus onsets were corrected to the onset of the critical information in the sound stream (Fig. 3). Reaction times (RTs) below 200 ms were removed from the analysis (0.05%). Trials with RTs that deviated 2.5 interquartile range below the 1^{st} and above the 3^{rd} quartile of each participant in each intra-subject variable were considered outliers and discarded from data analyses (1.49%). Accuracy data and RTs of correct answers were then analysed with logit and linear mixed-effect models respectively [*lme4* (Bates, Maechler, and Dai, 2008) package in R (R core Team, 2012)]. Collinearity was not an issue in the models: variance inflation factor (VIF) ranged from 1 to 1.5. All models included random intercepts for subjects and items and maximal random slopes for each within-subjects and within-items predictor respectively. Following Barr et al. (2013) and Barr (2013) when models with maximal random structure failed to converge, maximal within-items and within-subject interactions for random slopes were used. All fixed effects were contrast coded before analyses using sum coding so that each model's intercept represented the mean value of each predictor. Significance *P*-values and Type III *F*-statistics for main effects and interactions for continuous variables (RTs) were calculated using Satterthwaite approximations to denominator degrees of freedom as implemented in the *LmerTest* (Kuznetsova, Brockhoff, and Christensen, 2017) package, and planned comparisons and β estimates were calculated using *difflemeans* and *lsmeans* as implemented in the *lmerTest* package. Binary outcomes (accuracy data) were analysed using logit mixed-effects models (Jaeger, 2008). Type III Wald χ^2-statistics, *P*-values, planned comparisons and β estimates for main effects and interactions were calculated using *car* (Fox and Weisberg, 2014) and incorporated *lsmeans* packages (Lenth, 2016).

EEG data analysis

ERP amplitudes were measured at 6 centroparietal electrodes (C1, Cz, C2, CP1, CPz, and CP2) where the N400 is usually maximal (Kutas and Hillyard, 1980, 1984; Kutas et al., 1984). In experiment 1, for the English day block, mean N400 amplitude were computed between 350–500 ms, determined predictively based on previous literature (Kutas and Hillyard, 1980; 1984; Kutas et al., 1984). For the Chinese Day block, the N400 window was 813–963 ms (since *xing-qi* lasted 463 ms, Fig. 3). In experiment 2, for the English year block, the predicted time-window of the N400 was 630–780 ms after stimulus onset, given that the 'twenty-' portion of the auditory stream lasted 280 ms (Fig. 3). In the Chinese year block, the N400 time window was 869–1019 ms (since *er-ling* lasted for 519 ms).

Results

To analyse our results, we proceeded in four steps. First, we analysed behavioural measures and ERP results from Experiment 1 (days), starting with 2-day gaps, where spatiotemporal metaphor effects were anticipated. We then analysed data for the 1-day gaps where only effects of conventionality could be expected. Third, we analysed data collected in Experiment 2 (years), to establish whether the pattern of results obtained for days would also obtain for years (replication). Starting with 2-year gaps, we tested for spatiotemporal metaphor congruency and then for conventionality effects in the case of 1-year gaps. Reaction times, accuracy data, and ERP's time-windows were corrected to the onset of the critical information in the sound stream.

Chinese spatiotemporal metaphors for days affect time conceptualisation

In experiment 1, we tested whether a change of language would affect congruency between spatiotemporal metaphors of Chinese and spatiotemporal configuration of the stimuli in Chinese-English bilinguals in the case of two-day intervals. Accuracy was at ceiling in the interval calculation task whether bilinguals heard day stimuli in Chinese or in English (Fig. 4A). We found no significant main effect of language (English, Chinese**; χ^2_1 = 2.06, *P* = 0.15)** or congruency (congruent, incongruen**t; χ^2_1 = 0.58, *P* = 0.45)** on accuracy and no interaction **(χ^2_1 = 0.1, *P* = 0.76).** As for Reaction Times (RTs), we found a main effect of language (*F* (1,19.53) = 24.66, *P* < 0.001) so that bilingual participants were slower responding to English (β = 1057, *SE* = 54) than Chinese stimuli (β = 861, *SE* = 37). There was no **significant main effect of congruency (*F* (1, 21.01) = 1.38, *P* = 0.25) and no interaction (*F* (1, 21.6) = 0.49, *P* = 0.49).**

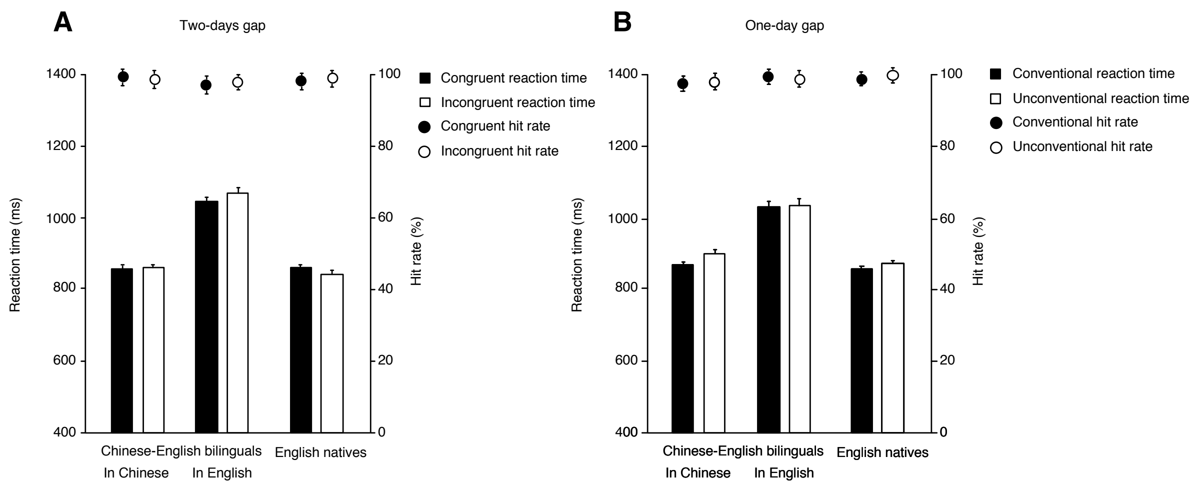

Figure 4. Behavioural results. (A) Two-days gap. (B) One-day gap. Bars represent reaction times and bullets represent accuracy. Error bars depict s.e.m.

We then analysed mean N400 amplitudes in the same Chinese-English bilinguals to determine whether spatiotemporal metaphors interfered with time conceptualisation during the task. A repeated measure ANOVA with language (Chinese, English) and congruency (congruent, incongruent) as within-subject factors revealed a significant effect of congruency (F (1,18) = 21.83, $P < 0.001$, $\eta^2_p = 0.55$). The effect of language was marginally significant (F (1,18) = 4.14, $P = 0.06$, $\eta^2_p = 0.2$) and the interaction between congruency and language was also significant (F (1,18) = 7.06, $P = 0.02$, $\eta^2_p = 0.28$). Planned comparisons showed that incongruent stimulus configurations elicited significantly more negative N400 amplitudes than congruent ones when bilingual participants operated in English (t (18) = 4.66, $P < 0.001$; Fig. 5). No such effect was found when participants responded to Chinese stimuli (t (18) = -0.53, $P = 0.3$).

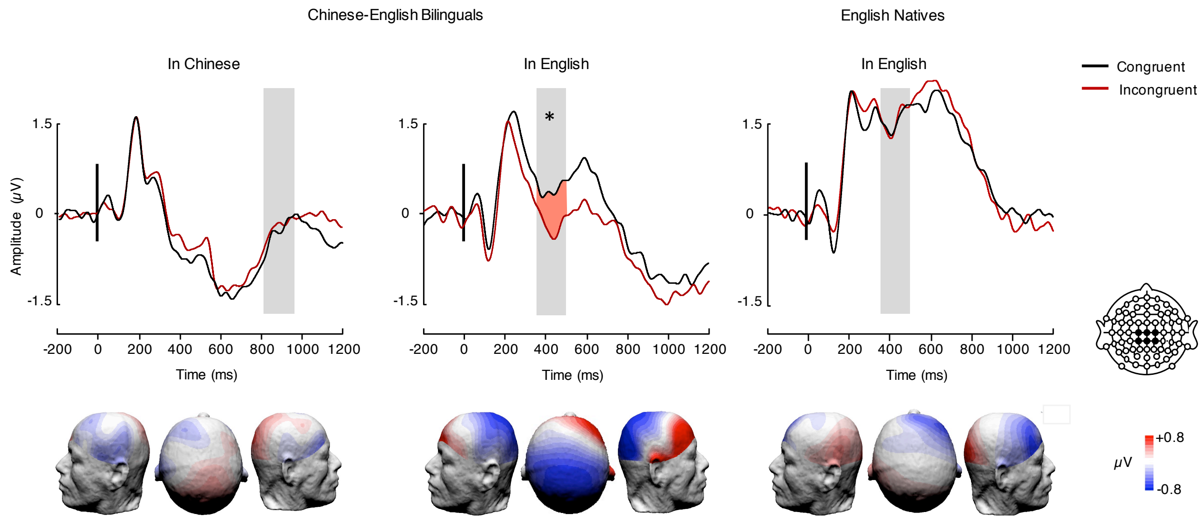

Figure 5. Event-related brain potentials elicited in experiment 1 (days). ERPs depict the linear derivation of 6 electrodes (C1, Cz, C2, CP1, CPz, CP2). Topographical maps show ERP activity across the 64-channel array in the following predictively determined time-windows. N400 amplitudes were computed between 350–500 ms based on previous literature, from the onset of the unique sound streams, irrespective of language or stimulus. In the case of Chinese stimuli, the interval of N400 amplitude extraction was 813–963 ms (since *xing-qi* – 'week' lasted 463 ms, see Methods) and in the case of English stimuli, the interval of N400 extraction was 350-500 ms, since day stimuli differed from one another from their onset. Topographies depict differences between incongruent and congruent conditions in all cases.

In order to further investigate the congruency effect found in bilinguals operating in English, we compared their results with that of English native participants. Accuracy was at ceiling in English native controls. No significant main effect of congruency (congruent, incongruent**;** $\boldsymbol{\chi^2_1 = 0.61, P = 0.44}$) or group (English, Chinese-English bilingual**;** $\boldsymbol{\chi^2_1 = 1.71, P = 0.19}$) was found on accuracy and there was no interaction ($\boldsymbol{\chi^2_1 = 0.01, P = 0.92}$**;** Fig. 4A). Regarding

RTs, Chinese-English bilinguals operating in English were significantly slower (β = 1056, *SE* = 48) than their English native peers (β = 855, *SE* = 50), as reflected by a significant main effect of group (F (1, 34.21) = 8.45, P < 0.001). **There was no significant main effect of congruency (F (1, 8.06) = 3.06, P = 0.12) and no interaction (F (1, 11.63) = 0.05, P = 0.83).**

A between-subjects repeated measures ANOVA, with congruency as within-subject factor and group (English, Chinese-English bilingual) as between-subject factor conducted on N400 mean amplitude revealed a significant main effect of group (F (1, 34) = 7.95, P = 0.01, η^2_p = 0.19) and a significant effect of congruency (F (1, 34) = 5.54, P = 0.02, η^2_p = 0.14). The interaction was also significant (F (1, 34) = 5.99, P = 0.02, η^2_p = 0.15). Planned comparisons showed that incongruent stimulus configurations elicited more negative N400 amplitudes than congruent configurations in bilingual participants (t (18) = 4.66, P < 0.001; Fig. 5), but not in their English peers (t (16) = -0.05, P = 0.48).

Conventionality effects for one-day gaps affect behaviour but not ERP amplitudes

We first tested for effects of conventionality in Chinese-English bilinguals' mind. With regard to accuracy, **we found no significant main effect of language (Chinese, English; χ^2_1 < 0.01, P = 0.97) or conventionality (conventional, unconventional; χ^2_1 = 0.1, P = 0.75). However,** there was a significant interaction between language and conventionality (χ^2_1 = 3.88, P = 0.05). However, *post hoc* comparisons failed to show effects of conventionality in either Chinese (β = -0.68, *SE* = 0.45, z = -1.52, P = 0.13) or English (β = 0.49, SE = .40, z = 1.23, P = 0.22) considered separately. The effect of language in the conventional (z = -1.11, p = 0.27) and unconventional (z = 1.02, p = 0.31) conditions were not significant either. Regarding RTs, a significant main effect of language (F (1, 20.89) = 7.82, P = 0.01) showed that Chinese-English bilinguals were slower responding to English stimuli (β = 1043, *SE* = 60) than Chinese stimuli (β = 880, *SE* = 46; see Fig. 4B). The effect of conventionality was just significant (F (1, 66.96) = 3.88, P = 0.05), bilinguals being slower responding to unconventional (β = 972, *SE* = 45) than conventional stimuli (β = 951, *SE* = 45). **However, we found no interaction between language and conventionality on RT (F (1, 39.07) = 1.57, P = 0.22).** Amplitude analysis revealed no main effect of conventionality **(F (1, 18) = 0.75, P = 0.4, η^2_p = 0.04)** or language **(F (1, 18) = 1.94, P = 0.18, η^2_p = 0.1)** on N400 amplitude and no interaction **(F (1, 18) = 1.87, P = 0.19, η^2_p = 0.09**; Fig. 6).

As was the case in the bilingual group, English participants' accuracy was at ceiling in the one-day gap condition. Analysis comparing the Chinese-English bilinguals in English with the native English controls revealed no main effect of conventionality **($\chi^2_1 = 1$, $P = 0.32$)** or group **($\chi^2_1 = 2.92$, $P = 0.09$)** on accuracy and no interaction **($\chi^2_1 = 0.24$, $P = 0.62$)**. As regards RTs, a main effect of group ($F(1, 35.01) = 6.29$, $P = 0.02$) showed that Chinese-English bilinguals were slower responding to English stimuli ($\beta = 1051$, $SE = 66$) than their English native peers ($\beta = 888$, $SE = 56$). **There was no significant main effect of conventionality ($F(1, 6.73) = 0.59$, $P = 0.47$) and no interaction ($F(1, 8.29) = 1.84$, $P = 0.21$).** Amplitude analysis only revealed a significant main effect of group ($F(1, 34) = 6.75$, $P = 0.01$, $\eta^2_p = 0.17$) on N400 amplitude. **No significant main** effect of conventionality **($F(1, 34) = 0.02$, $P = 0.88$, $\eta^2_p < 0.01$) or no interaction ($F(1, 34) = 0.6$, $P = 0.45$, $\eta^2_p = 0.02$; Fig. 6) was detected.**

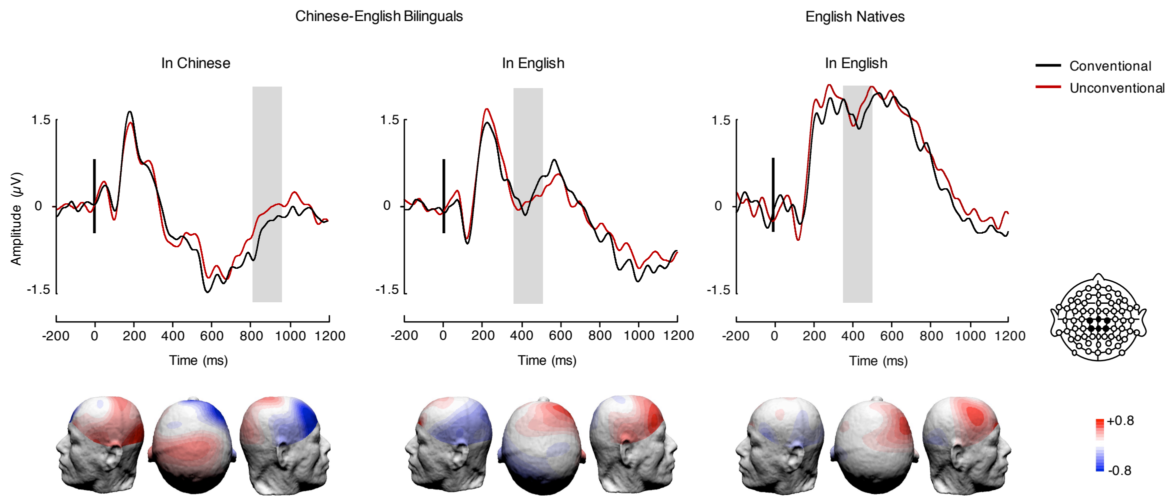

Figure 6. Event-related brain potentials elicited in experiment 1. ERPs depict the linear derivation of 6 electrodes (C1, Cz, C2, CP1, CPz, CP2). Topographical maps show ERP activity across the 64-channel array in the following predictively determined time windows: 813–963 ms after Chinese stimulus onset and between 350–500 ms after English stimulus onset. Topographies depict differences between **unconventional and conventional** conditions in all cases.

Replication of the spatiotemporal metaphor effect with year stimuli

In experiment 2, as was the case for days, Chinese-English bilinguals were at ceiling in the interval calculation task with two-year gap stimuli in both the congruent and the incongruent conditions and in both their languages (Fig. 7A). Results revealed no significant main effect of language (Chinese, English; **$\chi^2_1 = 0.33$, $P = 0.57$)** or congruency (congruent, incongruent**; $\chi^2_1 = 2.55$, $P = 0.11$)** on accuracy and no interaction **($\chi^2_1 = 0.21$, $P = 0.64$)**. We found no effect of language of operation **($F(1, 2.61) < 0.01$, $P = 0.98$)** or congruency **($F(1, 2.1) = 0.41$, $P = 0.59$)** on RTs and no interaction **($F(1, 2.29) = 0.26$, $P = 0.66$)**.

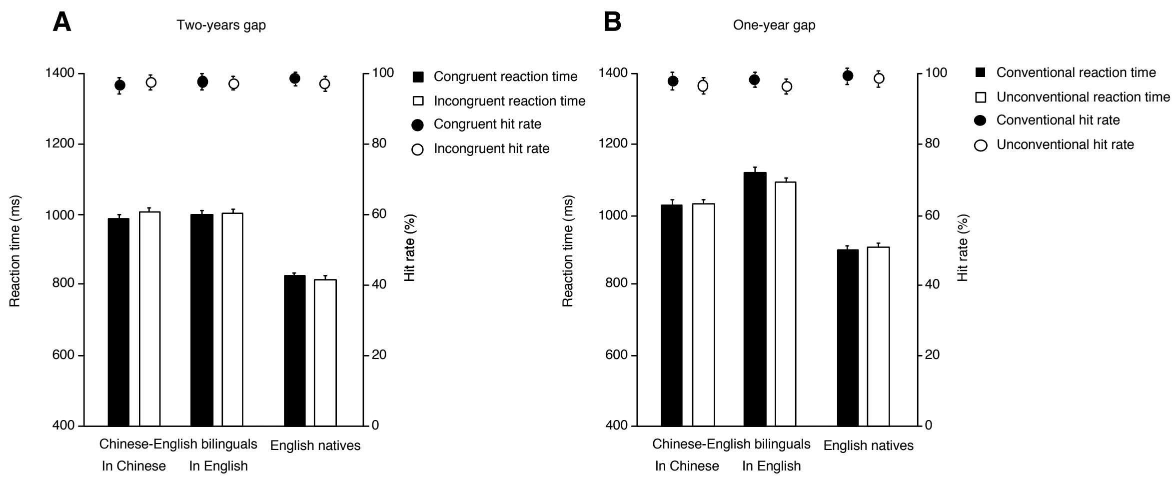

Figure 7. Behavioural results. *(A)* Two-years gap. *(B)* One-year gap. Bars represent reaction times and bullets represent accuracy. Error bars depict s.e.m.

The within-subject repeated measures ANOVA of ERP data revealed a main effect of congruency on mean N400 amplitude in bilingual participants ($F(1,18) = 6.96$, $P = 0.02$, $\eta^2_p = 0.28$) and a significant interaction between congruency and language ($F(1,18) = 4.6$, $P = 0.05$, $\eta^2_p = 0.2$). **The main effect of language was not significant ($F(1, 18) = 0.04$, $P = 0.85$, $\eta^2_p < 0.01$).** Replicating the pattern found for 2-day gap calculations, planned comparisons showed that N400 amplitude was significantly greater for incongruent than congruent stimulus configurations when bilinguals operated in English ($t(18) = 3.89$, $P < 0.001$; Fig. 8) but not when they operated in Chinese ($t(18) = 0.31$, $P = 0.38$).

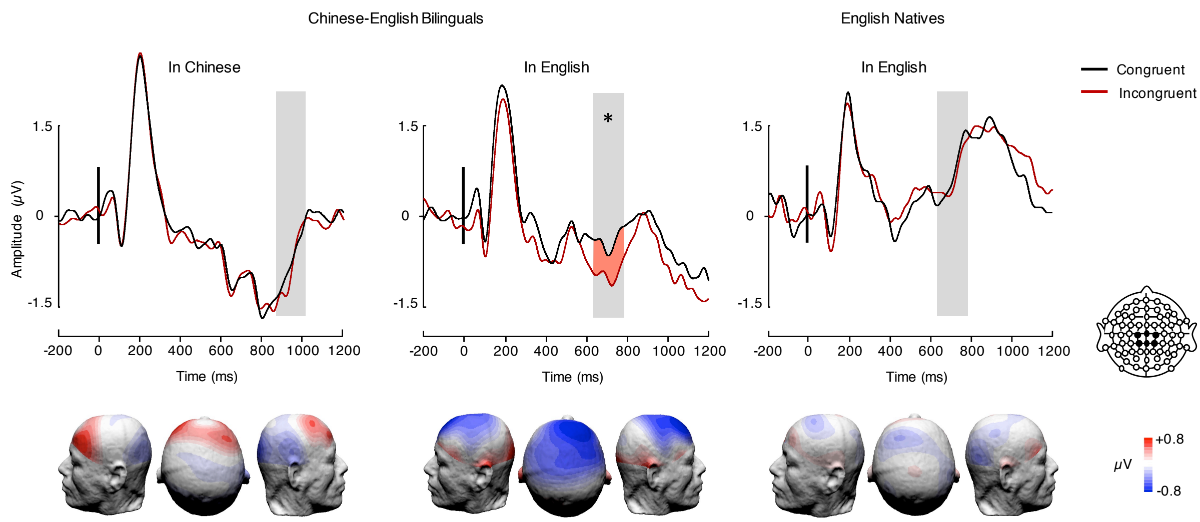

Figure 8. Event-related potentials elicited in experiment 2. ERPs depict the linear derivation of 6 electrodes (C1, Cz, C2, CP1, CPz, CP2). Topographical maps show ERP activity across the 64-channel array between 869-1019 ms after Chinese stimulus onset and 630-780 ms after English stimulus onset. The predicted time-window of the N400 for Chinese stimuli was between 869–1019 ms after stimulus onset, given that the er-ling – 'twenty' portion

of the auditory stream lasted for 519 ms. In the case of English stimuli, the N400 time analysis window was 630–780 ms (since 'twenty' lasted 280 ms). Topographies depict differences between incongruent and congruent conditions in all cases.

As in Experiment 1, we sought to further characterise the congruency effect found for the English condition in bilinguals by comparing their results with that of native English speakers. English participants' accuracy was at ceiling. No significant main effects **(group: $\chi^2_1 = 0.12$, $P = 0.73$; congruency: $\chi^2_1 = 1.59$, $P = 0.21$)** or interaction between congruency and group **($\chi^2_1 = 0.24$, $P = 0.62$)** was detected. Regarding RTs, Chinese-English bilinguals operating in English were significantly slower ($\beta = 1002$, $SE = 74$) than English native participants ($\beta = 819$, $SE = 61$), as shown by a main effect of group ($F\,(1, 28.13) = 6.96$, $P = 0.01$). No **significant main effect of congruency ($F\,(1, 2.23) = 0.3$, $P = 0.64$) or interaction ($F\,(1, 1.84) = 0.12$, $P = 0.76$) was detected.**

A between-subject repeated measures ANOVA on N400 mean amplitudes showed a significant main effect of group ($F\,(1, 34) = 4.13$, $P = 0.05$, $\eta^2_p = 0.11$) and a significant main effect of congruency ($F\,(1, 34) = 7.21$, $P = 0.01$, $\eta^2_p = 0.18$). The interaction between group and congruency was also significant ($F\,(1, 34) = 4.51$, $P = 0.04$, $\eta^2_p = 0.12$). Planned comparisons showed that incongruent stimulus configurations elicited greater N400 amplitudes than congruent ones in bilingual participants ($t\,(18) = 3.89$, $P < 0.001$; Fig. 8), but not in English controls ($t\,(16) = 0.35$, $P = 0.37$).

No measurable effect of conventionality in the case of 1-year gaps

As previously, we first compared bilingual participants' performance in English and Chinese using within-subject analyses. No significant main effect of language **(Accuracy: $\chi^2_1 = 0.45$, $P = 0.5$; RT: $F\,(1, 8.07) = 2.69$, $P = 0.14$)** or conventionality **(Accuracy: $\chi^2_1 = 1.8$, $P = 0.18$; RT: $F\,(1, 2.11) = 0.27$, $P = 0.65$)** on either accuracy or RT and no interaction were detected **(Accuracy: $\chi^2_1 < 0.01$, $P = 0.96$; RT: $F\,(1, 2.69) = 0.78$, $P = 0.45$).** The analysis conducted on mean N400 amplitude showed no significant main effect **(language: $F\,(1, 18) = 0.14$, $P = 0.71$, $\eta^2_p = 0.01$; conventionality: $F\,(1, 18) = 3.44$, $P = 0.08$, $\eta^2_p = 0.16$)** or interaction **($F\,(1, 18) = 0.06$, $P = 0.82$, $\eta^2_p = 0.003$;** Fig. 9).

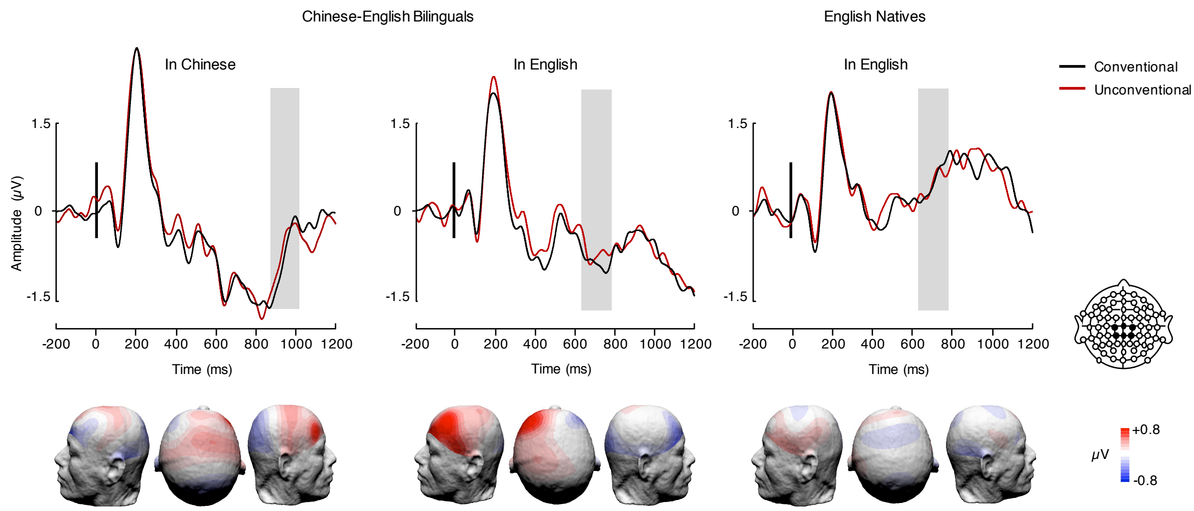

Figure 9. Event-related potentials elicited in experiment 2. ERPs depict the linear derivation of 6 electrodes (C1, Cz, C2, CP1, CPz, CP2). Topographical maps show ERP activity across the 64-channel array between 869-1019 ms after Chinese stimulus onset and 630-780 ms after English stimulus onset. Topographies depict differences between **unconventional and conventional** conditions in all cases.

Finally, we compared Chinese-English bilinguals in English with their English native peers. **No significant main effect of group ($\chi^2_1 = 0.01$, $P = 0.94$) or conventionality ($\chi^2_1 = 0.06$, $P = 0.8$) on accuracy was detected.** We found a significant interaction between group and conventionality on accuracy ($\chi^2_1 = 7.14$, $P < 0.01$). However, *post hoc* comparisons showed that there was no effect of conventionality in either Chinese-English bilinguals ($\beta = 0.46$, $SE = 0.27$, $z = 1.68$, $P = 0.09$) or English natives ($\beta = -0.39$, $SE = 0.28$, $z = -1.43$, $P = 0.15$). As regards RTs, there was a significant effect of group ($F(1, 32.92) = 6.04$, $P = 0.02$), bilingual participants ($\beta = 1109$, $SE = 55$) being slower responding to English stimuli than English native participants ($\beta = 911$, $SE = 58$; Fig. 7B). **There was no significant main effect of conventionality ($F(1, 1.1) = 0.66$, $P = 0.56$) and no interaction ($F(1, 1.67) = 0.54$, $P = 0.55$).** As regards the ERP analysis, we only found a significant main effect of group on mean N400 amplitude ($F(1, 34) = 4.56$, $P = 0.04$, $\eta^2_p = 0.12$; Fig. 9). **There was no significant main effect of conventionality ($F(1, 34) = 0.13$, $P = 0.73$, $\eta^2_p < 0.01$) and there was no interaction ($F(1, 34) = 0.34$, $P = 0.56$, $\eta^2_p = 0.01$).**

In addition, we ran a direct comparison between the congruency effects detected in Experiment 1 and Experiment 2. Paired sample t-test suggested that the difference waves were not statistically different across experiments ($t(18) = 0.24$, $P = 0.81$). A Bayesian paired sample t-test confirmed that the null hypothesis (i.e., no difference in effect magnitude between experiments) was around 4 times more likely than the alternative ($BF_{01} = 4.1$).

Discussion

Here we investigated a potential effect of native spatiotemporal metaphors on time conceptualization in Chinese-English bilinguals operating in their native or their second language. When tested in Chinese, participants did not display congruency effects predicted by spatiotemporal metaphors. Strikingly, however, when they were presented with English stimuli, native language representations interfered with time conceptualization as indicated by more negative N400 amplitudes in the incongruent conditions. Importantly, this pattern of result was mostly replicated using years instead of days as auditory stimuli. In contrast, conventionality effects only appeared as subtle behavioural variations in the case of 1-day intervals and did not entail any N400 amplitude modulation.

First, our results are consistent with previous studies that have established unconscious language non-selective access in bilinguals, and particularly Chinese-English bilinguals operating in English (Thierry and Wu, 2007). Indeed, and despite recent attempts to provide an alternative account for this mechanism (Costa, Pannunzi, Deco, and Pickering, 2017; Oppenheim, Wu, and Thierry, 2018), Chinese-English bilinguals appear to automatically access Chinese when processing input in English, because otherwise it would be difficult to account for the interference effects observed here. The results thus expand our understanding of language non-selective lexical activation mechanisms in different script bilinguals (Thierry and Wu, 2007; Wu and Thierry, 2010, 2012) by showing unconscious activation of spatiotemporal metaphor representations of Chinese when participants hear English words.

Our findings are partly compatible with results from previous behavioural studies suggesting that spatiotemporal metaphors can influence individuals' conceptualization of time (Boroditsky, 2001; Casasanto et al., 2004; Fuhrman et al., 2011; Lai and Boroditsky, 2013; Núñez and Sweetser, 2006, but see Chen, 2007; January and Kako, 2007; Tse and Altarriba, 2008). Critically, however, our data establish the locus of interference between language specific expression and time representation at a *conceptual level* in the absence of participants' awareness, since congruency effects were detected in N400 amplitude modulations rather than behavioural measurements and in conditions where time orientation was irrelevant. Indeed, at debriefing, detailed questioning of the participants revealed no explicit knowledge of hidden manipulations relating to spatiotemporal metaphors. All participants reported having interpreted the task as a simple arithmetic problem, that is, computing an interval of 1or 2 days, or 1 or 2 years, irrespective of future or past temporal reference. Even when directly confronted with the actual construction of the experiment, none of the participants recognised that the future or past reference afforded by the stimuli should conflict with the location of the speakers through which these stimuli were presented, or having resorted consciously to labelling 2-day and 2-year gaps as “front/back-day” or “front/back-year” in Chinese.

It may be considered a surprise, however, that bilingual participants experienced the spatiotemporal metaphor interference effect when performing the task in English rather than Chinese, given that the metaphors belong to Chinese. But this result is in fact compatible with the frequent observation that verbal interference tends to cancel effects of language on conceptualisation (Drivonikou et al., 2007; Gilbert, Regier, Kay, and Ivry, 2006; Roberson and Davidoff, 2000). When stimuli are presented in Chinese, participants suffer within-language competition, such that they cannot verbally recode information because accessing the labels for days and years and engaging in arithmetic computations in Chinese directly compete for selection with metaphoric lexical representations. However, this is arguably not the case when participants operate in English, since no direct within-language competition applies: Metaphors in Chinese can be accessed through cross-language activation. Then, and only then, can interference take place. This mechanistic explanation is consistent with selective interference effects previously shown in bilinguals switching back and forth between their first and second language, whilst making non-verbal decisions on motion events (Athanasopoulos et al., 2015).

In other words, we contend that only when participants heard temporal references in English, they accessed conceptually related expressions specific to their native language. For instance, when a participant tested on a Wednesday heard the English word “Monday”, they would have activated *qian-tian* (literally translated as “front–day”), given that Monday was the day before yesterday relative to the day of testing. This would arguably not have happened when the same participant was tested in the native language Chinese because of the within-language competition effects described above. Alternatively, this would not have happened because days and years in Chinese contain a digit enabling direct gap calculation (with the exception of Sunday). For instance, *xing-qi yi* – ‘Monday’ literally translates into “week-1” in English and *er-ling yi-wu* – ‘2015’ literally translates into “two-zero-one-five”. Thus, calculating intervals is straightforward in Chinese but not in English, given the previously noted difficulty of bilinguals to compute operation in the second language (Salillas and Wicha, 2012).

As expected, we found a difference between conventional and unconventional control conditions in the case of 1-day gaps in the absence of any metaphorical interference, presumably due to there being no spatiotemporal metaphor for tomorrow and yesterday in either English or Chinese. Indeed, in Chinese, tomorrow is *ming-tian* (literally, "bright-day"), yesterday is *zuo-tian* ("past-day"), next year is *ming-nian* (bright-year), and last year is *qu-nian* ("gone-year"), thus any effect of orientation for one day/year gaps could only relate to effects of spatial orientation conventions for time. Conventionality had an effect in experiment 1 (days) but not experiment 2 (years). We contend that this was the case because time conventionality effects weaken as the size of time chunks increases, i.e., it is more difficult to conceptualise the year ahead as in front than the day ahead as in front (Hellström and Rammsayer, 2004；Lewis and Miall, 2003). Furthermore, conventionality did not affect ERP amplitude as metaphor congruency did. Here the argument would be that interference between convention and time representation does not occur at a semantic level but rather in terms of direct mapping between stimulus and response. Spatiotemporal metaphors rely exclusively on language and thus result in a semantic interference effect to start with (here resulting in a measurable N400 modulation). In other words, spatiotemporal metaphors are resolved at a pre-response, semantic level, whereas conventionality effects do not come into play during semantic access but rather interfere directly with the task at hand (particularly in the case of days).

To conclude, the present study provides the first electrophysiological evidence for a deep, unconscious, and pervasive influence of native spatiotemporal metaphors on time conceptualization in bilinguals. These findings not only bridge unconscious language non-selective access in bilinguals with predictions from linguistic relativity theory but also demonstrate the staggering level of interactivity involved. After all, our Chinese-English bilingual participants suffered semantic interference when the English label of the day after tomorrow was played through loudspeakers located in front of them, as compared to when the same label was played in their back. Given that this did not happen when they listened to the label of tomorrow, or any label in Chinese, and that it generalised to year labels, our study demonstrates that abstract concepts such as that of time are highly permeable to linguistic representations specific of the native language even when bilinguals operate in their second language.

Author contributions

Y.L. and G.T. conceived the experiment; Y.L. collected the data; Y.L., A.C., and G.T. analysed the data; Y.L., A.C., Y.J.W., and G.T interpreted the data. Y.L. and G.T. wrote the first draft of the paper. Y.L., A.C., Y.J.W., and G.T revised the manuscript until final.

Acknowledgements

The authors wish to thank E. Devienne for her assistance with data collection and also P. Athanasopoulos, C. Barbet, and M. Jones, and M. Vihman for comments.

All materials used and data collected in this study are available upon request to the corresponding author.

References

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.

Barr, D. J. 2013. Random effects structure for testing interactions in linear mixed-effects models. *Front Psychol, 4*, 328-328.

Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Bates, D., Maechler, M., Dai, B. 2008. lme4: Linear mixed effects models using s4 classes (Version R package version 0.999375-28).

Boroditsky, L. 2001. Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43*(1), 1-22.

Boroditsky, L., Fuhrman, O., McCormick, K. 2011. Do English and Mandarin speakers think about time differently? *Cognition, 118*(1), 123-129.

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. Paper presented at the *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Chen, J-Y. 2007. Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition, 104*(2), 427-436.

Costa, A., Pannunzi, M., Deco, G., Pickering, M. J. 2017. Do Bilinguals Automatically Activate Their Native Language When They Are Not Using It? *Cogn Sci, 41*(6), 1629-1644.

de la Fuente, J., Santiago, J., Román, A., Dumitrache, C., Casasanto, D. 2014. When you think about it, your past is in front of you: how culture shapes spatial conceptions of time. *Psychol Sci, 25*(9), 1682-1690.

Dijkstra, T., van Heuven, W. J. B. 2002. The Architecture of the Bilingual Word Recognition System: From Identification to Decision. *Bilingualism: Language and Cognition*, 5, 175-197.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., et al. 2007. Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proc Natl Acad Sci USA, 104*(3), 1097-1102.

Fox, J., Weisberg, S. 2014. Package car.

Fuhrman, O., McCormick, K., Chen, E., Jiang, H., Shu, D., Mao, S., et al. 2011. How linguistic and cultural forces shape conceptions of time: English and Mandarin time in 3D. *Cogn Sci, 35*(7), 1305-1328.

Gilbert, A. L., Regier, T., Kay, P., Ivry, R. B. 2006. Whorf hypothesis is supported in the right visual field but not the left. *Proc Natl Acad Sci USA, 103*(2), 489-494.

Gratton, G., Coles, M. G. H., Donchin, E. 1983. A new method for off-line removal of ocular atficat. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468-484.

Hellström Å Rammsayer T.H. 2004. Effects of time-order, interstimulus interval, and feedback in duration discrimination of noise bursts in the 50- and 1000-ms ranges. *Acta* Psychologica, *116*(1),1-20.

Hoshino, N., Thierry, G. 2012. Do Spanish–English bilinguals have their fingers in two pies–or is it their toes? An electrophysiological investigation of semantic access in bilinguals. *Frontiers in psychology, 3*, 9.

Jaeger, T. F. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446.

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

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

Kutas, M., Hillyard, S. A. 1984. Event-related brain potentials (ERPs) elicited by novel stimuli during sentence processing. *Annals of the New York Academy of Sciences, 425*(Jun), 236-241.

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

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software. 1(13),*12787.

Lai, V. T., Boroditsky, L. 2013. The immediate and chronic influence of spatio-temporal metaphors on the mental representations of time in english, mandarin, and mandarin-english speakers. *Front Psychol, 4*, 142.

Lee, B.; Meade, G.; Midgley, K.J.; Holcomb, P.J.; Emmorey, K. 2019) ERP Evidence for Co-Activation of English Words during Recognition of American Sign Language Signs. *Brain Sci*. 9, 148.

Lewis, P.A., Miall, R.C. 2003. Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Current Opinion in Neurobiology*. *13*(2), 250-255.

Lenth, R.V. 2016. lsmeans: Least-Squares Means (Version R package version 2.22.).

Lupyan, G. 2012. Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front Psychol, 3*, 54.

Meade, G.; Midgley, K.J.; Sevcikova Sehyr, Z.; Holcomb, P.J.; Emmorey, K. 2017. Implicit co-activation of American Sign Language in deaf readers: An ERP study. *Brain Lang*. *170*, 50–61.

Núñez, R. E., Sweetser, E. 2006. With the future behind them: convergent evidence from aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cogn Sci, 30*(3), 401-450.

Oppenheim, G., Wu, Y. J., Thierry, G. 2018. Found in translation: Late bilinguals do automatically activate their native language when they are not using it. *Cogn Sci, 42: 1700-1713.*

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Roberson, D., Davidoff, J. 2000. The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory Cognition, 28*(6), 977-986.

Salillas, E., Wicha, N. Y. Y. 2012. Early Learning Shapes the Memory Networks for Arithmetic: Evidence From Brain Potentials in Bilinguals. *Psychol Sci, 23*(7), 745-755.

Slobin, D. 1996. From 'thought and language' to 'thinking for speaking'. In J. Gumperz S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge: Cambridge University Press.

Thierry, G. 2016. Neurolinguistic Relativity: How Language Flexes Human Perception and Cognition. *Lang Learn, 66*(3), 690-713.

Thierry, G., Wu, Y. J. 2004. Electrophysiological evidence for language interference in late bilinguals. *NeuroReport, 15*(10), 1555-1558.

Thierry, G., Wu, Y. J. 2007. Brain potentials reveal unconscious translation during foreign-language comprehension. *Proc Natl Acad Sci USA, 104*(30), 12530-12535.

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

van Heuven, W. J. B., Dijkstra, T. 2010. Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews, 64*(1), 104-122.

van Heuven, W. J. B., Dijkstra, T., Grainger, J. 1998. Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*(3), 458-483.

Wen, Y., Filik, R. and van Heuven, W.J.B 2018. Electrophysiological dynamics of Chinese phonology during visual word recognition in Chinese-English bilinguals *Scientific Reports*. 6869.

Whorf, B. 1956. *Language, thought, and reality*. Cambridge, MA: MIT Press.

Wu, Y. J., Thierry, G. 2010. Chinese-English Bilinguals Reading English Hear Chinese. *Journal of Neuroscience, 30*(22), 7646-7651.

Wu, Y. J., Thierry, G. 2012. Unconscious translation during incidental foreign language processing. *Neuroimage, 59*(4), 3468-3473.

***Data and Code Availability Statement**

All materials used and data collected in this study are available upon request to the corresponding author.