

[cambridge.org/bil](https://cambridge.org/bil)Huilan Yang<sup>1</sup> , J. Nick Reid<sup>2</sup> and Yuru Mei<sup>1</sup><sup>1</sup>Zhejiang Gongshang University, Hangzhou, Zhejiang, China and <sup>2</sup>University of Manitoba, Winnipeg, Manitoba, Canada

## Research Article

**Cite this article:** Yang H, Reid JN, Mei Y (2023). Conceptual metaphor activation in Chinese–English bilinguals. *Bilingualism: Language and Cognition* **26**, 345–355. <https://doi.org/10.1017/S1366728922000530>

Received: 15 October 2021  
Revised: 14 July 2022  
Accepted: 18 July 2022  
First published online: 11 August 2022

### Keywords:

conceptual metaphor theory; false memory; bilingualism

**Address for correspondence:** Huilan Yang, Department of Foreign Languages, Zhejiang Gongshang University, Hangzhou, 310018, China (e-mail: [yanghuilan@mail.zjgsu.edu.cn](mailto:yanghuilan@mail.zjgsu.edu.cn)); or Nick Reid, Department of Psychology, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada (e-mail: [jeffreynicholas.reid@umanitoba.ca](mailto:jeffreynicholas.reid@umanitoba.ca)).

### Abstract

An episodic memory experiment was conducted to examine whether “conceptual metaphors” influence how metaphorical expressions are processed and encoded into memory. Forty Chinese–English bilinguals read lists of expressions in their L1 and L2. The data revealed that after reading a series of metaphorical expressions based on the same underlying conceptual metaphor, participants falsely recognized new sentences that instantiated the same conceptual metaphor mapping more often than control sentences that did not share this mapping. This false memory effect was robust in both participants’ L1 and L2, with the only difference between languages being that participants showed more memory errors for literal sentences related to the source domain of the conceptual metaphors when reading in their second language (i.e., English). These data suggest that although bilinguals can access the appropriate conceptual metaphors in their second language, they have difficulty inhibiting literal meaning when processing metaphorical statements.

### Introduction

Traditionally, metaphor has been regarded as a linguistic act rather than a cognitive way of thinking. So when processing a metaphor such as “*our marriage is a dead-end street*”, conventional comprehension models would suggest that the encoding depends on finding a shared similarity between the semantic concepts of “marriage” and “dead-end street” (e.g., Bowdle & Gentner, 2005) or treating “dead-end street” as an ad hoc category that the concept of “marriage” can be placed into, such as “things that are going nowhere” (e.g., Glucksberg & Keysar, 1990; see Holyoak & Stamenković, 2018, for a review). These models more or less assume that the concepts in a metaphor are already understood by the comprehender, and that metaphor processing depends on stretching the meaning of these words to find a point of comparison between the two concepts. In contrast to this view, Lakoff and Johnson (1980) argued that metaphor is not only a rhetorical device, but also a fundamental part of human thinking. This work has become well-known as Conceptual Metaphor Theory (CMT). This pioneering work opened the door to the study of metaphor systems in cognitive linguistics.

According to CMT, conceptual metaphors are encoded and represented through an underlying conceptual system. For instance, abstract concepts (e.g., arguments) are constructed and represented entirely through metaphor. The essence of metaphor is to understand and experience an abstract concept through a concrete concept that has structural similarities with the abstract one. For instance, in the conceptual metaphor, “ARGUMENT IS WAR”, the properties of the concrete concept “WAR” are mapped onto “ARGUMENTS”, which leads to a set of correspondences between the two domains: both involve an unpleasant relationship, two opposite enemies, careful strategizing, conflicts, etc. In this way, conceptual metaphors allow abstract concepts that cannot be touched or seen to be represented in terms of concrete concepts that are directly perceived or experienced. As such, CMT has made a major contribution to the area of embodied cognition, an area that argues cognition is related to perception, action and experience (e.g., Barsalou, 1999, 2008; Gibbs, 2006). A major challenge with embodied cognition is to explain how abstract concepts are represented in an embodied way, and CMT offers a feasible explanation.

Although CMT focuses on concept representation, semantic information, and language understanding, this theory has received only limited experimental examination in cognitive psychology, the area most concerned with experimentally investigating concept representation, human memory and language processing. There are some causes for this lack of examination (see Gibbs, 2009) – for instance, some have argued that the theory is underspecified for cognitive experiments and that the supportive evidence can be explained by more common mechanisms, such as accounts based on association (McGlone, 2011). The experiments that have been conducted on CMT have focused on discourse reading (Glucksberg, Brown, & McGlone; 1993; Gong & Ahrens, 2007; Keysar, Shen, Glucksberg, & Horton; 2000; Nayak

© The Author(s), 2022. Published by Cambridge University Press



& Gibbs, 1990; Thibodeau & Durgin, 2008) and the embodiment of abstract concepts (Casasanto, 2008; Casasanto & Boroditsky, 2008; Gibbs, 2013; Matthews & Matlock, 2011; Wilson & Gibbs, 2007; Yang et al., 2015, 2021; Zhong & Leonardelli, 2008; see Holyoak & Stamenković, 2018, for a review), but little research has examined whether conceptual metaphors play a role in other areas associated with cognition, such as learning, categorization, and memory.

### Conceptual metaphors and memory

A small group of studies conducted by Katz and colleagues have examined the psychological reality of conceptual metaphors through memory experiments. Firstly, Katz and Taylor (2008) employed several semantic and episodic memory tasks to show that the conceptual metaphor, "LIFE IS A JOURNEY", structured participants' semantic knowledge of typical life events. They found that when participants were asked to imagine the ideal life events of a 70-year-old, they tended to produce events in a chronological forward-temporal order and had highly consistent views on the age at which events would happen, the emotion associated with events, and whether the events would really happen or not. Katz and Taylor interpreted these findings to indicate that participants conceptualize typical lives as journeys with landmarks representing significant events in one's life, consistent with the LOVE IS A JOURNEY conceptual metaphor. Secondly, Katz and Law (2010) demonstrated in an episodic memory task that reading consecutive lists of metaphorical expressions based on the same conceptual metaphor elicited proactive interference analogous to what is observed for word lists based on taxonomic categories (see also Katz & Reid, 2020). Furthermore, a shift in the conceptual metaphor (e.g., switching from TIME IS MONEY to LOVE IS A JOURNEY) elicited a "release" from proactive interference akin to what happens when the taxonomic category is shifted between word lists (e.g., switching from exemplars of *birds* to *fruit*). This finding suggests that metaphorical expressions are organized according to conceptual metaphor categories in semantic memory, supporting their psychological reality.

Recently, Reid and Katz (2018) used a modified version of the Deese, Roediger and McDermott (DRM) memory paradigm to further test the psychological reality of conceptual metaphor theory (Deese, 1959; Roediger & McDermott, 1995). In a typical DRM task, participants read or hear a list of 15 words (e.g., jazz, horn, concert, orchestra, rhythm, piano, band, note, instrument, art, sound, symphony, radio, melody) and afterwards are asked to recall as many as possible (or recognize them from a list). Critically, each of the 15 words is associated with one non-presented word (i.e., the "critical lure"; in this case, the word *music*). Participants typically falsely remember this associated word at a very high rate even though it was not presented. This effect is also known as the "associative memory illusion," and a large number of experimental studies have been conducted over the years, using a variety of DRM paradigm variants and controlling many different experimental variables to investigate the factors affecting the illusion (see Gallo, 2006, 2010; Chang & Brainerd, 2021, for reviews).

Lakoff (2008) proposed that when a person encounters a metaphorical expression, the conceptual metaphor, which is the basis of this expression, is activated automatically. The theory is that when confronted with a phrase like "how did you spend the summer vacation?", its conceptual metaphor, "TIME IS MONEY", is automatically activated in people's minds. Moreover, metaphors

that people generally use, not only defined by Lakoff, can have the same effect. Thus, Reid and Katz (2018) hypothesized that after reading several metaphorical expressions based on one conceptual metaphor, people should mistakenly remember other metaphorical expressions based on the same conceptual metaphor. These non-presented expressions should be associated to the read expressions, and therefore, should be activated in memory just as words are activated by their list of associates in the DRM memory illusion. This hypothesis was confirmed as after reading a list of phrases based on a common conceptual metaphor, participants were more likely to falsely recognize new phrases based on the same conceptual metaphor mapping than control phrases that did not share this mapping from the source domain to the target domain. Reid and Katz (2022) recently replicated the experiment but included a divided attention condition that limited participants' conscious processing of the phrases. Even under divided attention, participants still displayed a robust false recognition effect, suggesting that conceptual metaphors are accessed automatically and effortlessly. These studies supply experimental evidence for the automatic activation of conceptual metaphors in English and validate the use of episodic memory tasks to explore the tenets of CMT.

The aim of the current research was to employ the same false memory paradigm adopted by Reid and Katz (2018) to examine the automatic activation of conceptual metaphors in Chinese-English bilinguals. Reid and Katz only tested native English speakers; however, conceptual metaphors are generally considered to be cross-lingual and universal as they have been identified across many languages, including Chinese (Li, 2010; Lv & Zhang, 2012; Yu, 1995, 2003, 2008), Thai (Han, 2019), Dutch (Forceville, 2007), Tagalog (Palmer, 2003), Japanese (Berendt & Tanita, 2011; Nomura, 1996), Spanish (Soriano, 2003), and Brazilian Portuguese (Gibbs, Lima, & Francozo, 2004) to name a few. Although different languages do not always have the exact same conceptual metaphors or instantiations of these metaphors (Yu, 2008, 2017), these studies indicate that cross-domain mappings appear in many different languages and cultures. Furthermore, Türker (2016) found that bilinguals can better comprehend metaphorical expressions in their L2 when they are based on conceptual metaphors that also exist in their L1, suggesting that conceptual metaphor activation is important when processing metaphorical expressions in L2. As such, people should show evidence of conceptual metaphor activation when reading both Chinese and English sentences, which would support that not only their existence in language, but also their psychological reality, is universal. Before describing the experiments in detail, we will briefly review the literature on bilingual metaphor processing.

### Metaphor processing in bilinguals

Metaphor comprehension is also a popular topic in the bilingual literature, with studies indicating that there are important differences in processing between L1 and L2. Several studies indicate that conventional metaphors tend to be processed as novel metaphors in L2. For instance, Mashal, Borodkin, Maliniak, & Faust (2015) found that L2 speakers considered conventional metaphoric word pairs (e.g., *sweet revenge*) to be more novel than L1 speakers. While native speakers processed conventional metaphors quicker when presented to the left hemisphere (LH) than when presented to the right hemisphere (RH), L2 speakers processed conventional metaphors faster when presented to the RH

than the LH (Mashal et al., 2015). The RH is known to play a vital role in the processing of less prominent figurative expressions (e.g., novel metaphors) and more literal expressions (Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012; Forgács, Lukács, & Pléh, 2014; Kasparian, 2013). In an ERP study, late bilinguals were shown to respond similarly in their L2 to novel and conventional metaphorical expressions, demonstrating equal amplitude of a late positive component, whereas divergence between the two expressions was reported on the same component in their L1 (Jankowiak, Rataj, & Naskręcki, 2017). This suggests that L2 speakers engaged in continued effortful information retrieval for both conventional and novel metaphoric expressions, whereas L1 speakers only showed this continued effort for novel expressions. Additionally, Ikuta and Miwa (2021) presented L2 speakers forward and reversed metaphors (e.g., “some babies are angels” and “some angels are babies”, respectively) for various durations ranging from 500 ms to 8000 ms. Whereas previous research indicates that L1 speakers rate reversed metaphors as less comprehensible with increased presentation duration (Wolff & Gentner, 2011), Ikuta and Miwa found that L2 speakers instead rated reversed metaphors as more comprehensible with increased duration. This suggests that metaphor processing depends more on symmetrical comparison in L2. According to the Career of Metaphor theory (Bowdle & Gentner, 2005), novel metaphors are also initially processed via symmetrical comparison, whereas conventional metaphors are processed more asymmetrically. Therefore, this study also suggests that metaphors tend to be processed as more novel in L2 compared to L1.

Other research indicates that metaphors tend to activate more literal meaning when processed in bilinguals' L2. Heredia and Cieślicka (2016) examined metaphor processing in Spanish–English and English–Spanish bilingual readers. Participants read English passages biasing either a literal meaning or a metaphorical meaning. The results indicated that meaning activation was decided by language dominance as Spanish dominant bilinguals tended to activate the literal meaning whereas English dominant bilinguals and balanced bilinguals tended to access both the literal and metaphorical meanings. Moreover, Citron, Michaelis, and Goldberg (2020) found that metaphorical processing activity is higher in native speakers and that left amygdala activation rises as Metaphoricity increases. However, L2 speakers did not exhibit any significant activity beyond the caudate nucleus when Metaphoricity increased, suggesting that L2 speakers were less influenced by Metaphoricity than native speakers. These findings support the theory that metaphorical language is more engaging for native speakers but not always for L2 speakers, and that L2 speakers tend to process conventional metaphors more similarly to literal sentences than L1 speakers (Citron et al., 2020). Along similar lines, Chen, Peng and Zhao (2013) explored the neural mechanisms of bilinguals' metaphor processing and found that, for Chinese–English bilinguals, the amplitude of the N400 component was more negative for English metaphor expressions than for Chinese literal, English literal, and Chinese metaphor expressions, suggesting that these individuals had more difficulty rejecting the literal meaning in their second language. Behavioral evidence also indicates that inhibiting literal meaning is difficult for L2 speakers as they often erroneously interpret metaphorical statements as literal (Picken, 2005).

The finding that L2 speakers show difficulty inhibiting literal meaning when processing metaphors in their L2 is consistent with the graded salience hypothesis (Giora, 2003). This hypothesis posits that metaphoric meaning is activated depending on which meaning (literal or figurative) is more salient, with salient

meanings accessed more readily. Salience is influenced by factors like word frequency, familiarity, and conventionality. Non-salient meanings are less utilized, less familiar, and need more time to be triggered (Giora, 2003, p. 491). Although the graded salience hypothesis is not a bilingual theory, it suggests that metaphor comprehension in bilinguals could be influenced by bilingual factors, such as language proficiency and language dominance. So, for less proficient bilinguals, the literal meaning of a metaphor expression could be more dominant and salient in their second language. Along similar lines, the literal-salience resonant model (Cieślicka, 2006, 2015) posits that when bilinguals process idioms in their L2, the literal meanings of the idioms' constituents are more salient than their figurative meanings.

In sum, there is evidence that bilinguals show some differences in processing metaphor expressions in their first language versus their second language, particularly regarding the activation of literal meaning. Therefore, it is important to explore how conceptual metaphor activation may differ depending on whether bilinguals are reading in their first or second language.

### The current study

In this experiment, Chinese–English bilingual participants completed the episodic memory task used by Reid and Katz (2018, 2022). Participants read lists of metaphorical expressions (e.g. “*How did you spend the summer?*”) based on the same conceptual metaphor (e.g., TIME IS MONEY), and were then tested for recognition of new expressions that either used the same conceptual metaphor (e.g., “*that cost me a day*”) or did not (e.g., “*the weekend is approaching*”). If conceptual metaphors are activated automatically after reading a series of metaphor expressions, non-presented metaphor expressions based on the same conceptual metaphor should elicit more memory errors in the subsequent old/new sentence recognition test than control sentences.

Participants completed both Chinese and English versions of the episodic memory task, with each version being identical aside from the language. Through this bilingual design, this research can compare metaphor processing in native language and second language reading. According to the graded salience hypothesis, one could speculate that the first language has advantages in metaphor processing as the metaphorical meaning of the expressions might be more salient in participants' first language. Therefore, there may be more memory errors for non-presented expressions that use the same underlying conceptual metaphor when participants read in their L1. However, as outlined above, bilinguals tend to process conventional metaphors as novel in their L2. Conceptual metaphors typically have a larger influence on the processing of novel metaphors than conventional metaphors (Gentner, Bowdle, Wolff, & Boronat, 2001; Keysar et al., 2000; Thibodeau & Durgin, 2008). Therefore, there may be more memory errors for expressions using the same conceptual metaphor in L2 than L1 if conceptual metaphor activation depends on novelty.

The other major finding outlined above is that bilinguals have difficulty inhibiting literal meaning when processing metaphors in their L2. If the literal meaning is more salient for expressions in the bilinguals' second language, they may also show increased memory errors for literal control sentences associated with the source domain of the conceptual metaphor. That is, expressions based on the TIME IS MONEY conceptual metaphor may more strongly activate meanings related to the literal meaning of MONEY, and therefore, participants may show more memory

errors for literal sentences about money (e.g., “he makes biweekly payments”) on the recognition tests in their L2.

## Method

### Participants

Forty Chinese–English bilinguals (sample age: mean = 19,  $SD = 1.73$ ) from Zhejiang Gongshang University participated in this experiment in return for a small gift. Of the forty participants, 36 were female and 4 were male. All participants indicated that they were proficient in reading Simplified Chinese as well as in reading English words, and that Chinese was their first and dominant language whereas English was their second language. All participants were born in China and lived there at the time of the experiment. The language learning order of all participants was firstly Chinese and then English, and their dominant language was Chinese at the time of the experiment. Sixteen participants had passed the Test for English Majors-Band 4 (TEM 4), 16 participants had completed the College English Test Band 6 (CET 6) with a mean score of 562, 19 participants had completed the College English Test Band 4 (CET 4) with a mean score of 605, and 3 participants had completed the IELTS with a mean score of 7. All had normal or corrected-to-normal vision and no reading disorders. More details about participants’ language backgrounds are displayed in Table 1.

## Materials

### Study lists

Eight study lists (four Chinese and four English) were created following the procedures outlined in Reid and Katz (2018). Each study list contained 16 metaphorical phrases based on 1 specific conceptual metaphor. The lists were based on four well-known conceptual metaphors that appear in both Chinese and English: “LOVE IS A JOURNEY,” “TIME IS MONEY,” “ECONOMY IS A HUMAN BEING,” and “ARGUMENT IS WAR”. Therefore, the same conceptual metaphors were used in both languages, although the expressions in these lists differed as metaphorical expressions vary across languages and cultures, even when the underlying mapping is the same (Yu, 2008). For the expressions in the lists, each expression framed a concept from the target domain of the conceptual metaphor in terms of a concept from

the source domain of the metaphor. For example, for the “TIME IS MONEY” list, English study expressions included “How did you *spend* the *summer break*?” and “The *diversion* should *buy* him a *few minutes*.” Here, “spend” and “buy” are associated with the source domain (i.e., MONEY) whereas “summer break” and “a few minutes” are associated with the target domain (i.e., TIME). Chinese expressions included “你把暑假的时间都花在哪里了(English translation: How did you spend the summer break?)” and “他投资了一年的时间做实验(He invested a year in the experiment.)”. Here, “花(English translation: spend)” and “投资(English translation: invested)” are associated with the source domain whereas “暑假(English translation: summer break)” and “一年(English translation: a year)” are associated with the target domain.<sup>1</sup> Each participant read only 3 of the 4 study lists in each language for the purposes of the recognition test (described below). This yielded 4 versions of the experiment, each of which included a different non-presented study list. Participants were randomly assigned to the different versions. The study lists were presented in blocks according to language, and the order of language (i.e., Chinese lists first or English lists first) was counterbalanced across participants.

### Distractor task

After each study list, there was a short distractor task consisting of 10 simple mathematical questions that required participants to follow the proper order of operations (e.g.,  $5 + 6 \div (9 - 3)$ ). The purpose of this task was to prevent participants from rehearsing the study list expressions in short term memory, which would affect the subsequent results of the recognition tests. Participants were asked to complete these math questions mentally without using any tools, such as a pen and paper or a calculator. There was no time limit for the math questions, but participants were asked to answer them as quickly and accurately as possible.

### Recognition tests

After completing each study list and distractor task, participants completed a recognition task on the study list they had just finished. There was no time limit. Each recognition test consisted of 26 or 27 phrases, divided into 14 old items and 12–13 new items, or “lures”. The old items included all the phrases in the previous study list except the phrases in the second and the fourteenth serial positions. The new items included four types: 3 critical consistent lures, 3 control metaphor lures, 3 control literal lures and 3 to 4 unrelated lures.

The 3 critical consistent lures were expressions based on the same conceptual metaphor as the study list expressions but were not presented previously in the study list. The expression “That cost me a day” is an example of a critical consistent lure, which frames a TIME concept (“day”) in terms of a MONEY concept (“cost”).

The 3 control metaphor lures were phrases that framed the same target domain but with different source domains. For example, for the “LOVE IS A JOURNEY” list, one of the metaphor control lures was based on the “LOVE IS SWEET” conceptual metaphor. The purpose of the control metaphor lures was to eliminate the possibility that the participants would simply encode the target domain of the study list instead of activating a conceptual mapping from the source domain to the target domain. For instance, this ruled out that participants simply memorized that all the phrases were metaphorical expressions

**Table 1:** Language experience for Chinese-English bilingual participants

	Chinese	English	Chinese	English
<i>% of daily usage for each language</i>				
Daily usage	70.50	23.18		
<i>Self-rating of current skill (out of 10)</i>			<i>Average starting age of each language skill</i>	
Understanding	9.33	7.00	2	8.5
Speaking	9.15	6.80	3	9
Reading	9.33	7.95	4	9
Writing	9.1	7.25	5	9

about “LOVE” rather than activating the specific metaphor mapping, “LOVE IS A JOURNEY”.

The 3 control literal lures were related to the source domain. For example, for the “LOVE IS A JOURNEY” list, a control literal lure was “This pathway is a short walk”, which is a literal statement about a JOURNEY. The purpose of the control literal lures was to eliminate the possibility that the participants simply encoded the source domain of the study list. For instance, this ruled out that participants simply attended to the fact that all of the sentences in the “LOVE IS A JOURNEY” list mentioned something about journeys, and used this criterion to make recognition decisions.

Finally, there were 3 to 4 unrelated lures that did not overlap with the study list in terms of either the target or source domain. These were the 3 critical consistent lures, 3 control metaphor lures, 3 control literal lures, and the conceptual metaphor label associated with the non-presented list (e.g., “argument is war”). These 10 phrases were split into 3 parts with 3–4 phrases each and randomly assigned to different study lists to serve as the unrelated lures. We expected false recognition to be quite low for these items as they had little overlap with the studied items.

To ensure that the different types of lure sentences did not differ in terms of factors that could influence memory, we asked 12 Chinese–English bilingual volunteers to rate the affective valence and familiarity of the expressions. The items were rated on two seven-point scales, one for familiarity, and one for affective valence (1 = very negative/very unfamiliar, 7 = very positive/very familiar). We then compared the sentence length, valence, and familiarity values between the critical consistent, control metaphor, control literal and unrelated lures. For the Chinese stimuli, there were no sentence length differences between those four conditions,  $F = 0.891$ ,  $p = 0.450$  (with mean lengths of 11, 9, 10, and 10, respectively) nor differences in emotional valence,  $F = 0.418$ ,  $p = 0.740$  (with mean emotional valences of 3.5, 3.3, 3.8, and 3.6, respectively). There were also no sentence familiarity differences between the four conditions,  $F = 0.064$ ,  $p = 0.979$  (with mean familiarities of 4.8, 4.7, 4.9, and 4.8, respectively).

For the English stimuli, there were no length differences between the four conditions,  $F = 0.386$ ,  $p = 0.763$  (with mean lengths of 7, 6, 6, and 6, respectively). There were also no emotional valence differences between those four conditions,  $F = 1.554$ ,  $p = 0.208$  (with mean emotional valences of 3.3, 3.3, 4.1, and 3.6, respectively), nor differences in sentence familiarity,  $F = 0.932$ ,  $p = 0.430$  (with mean familiarities of 4.6, 5.0, 4.7, and 4.8, respectively). All in all, both the Chinese and English sentences were matched well in terms of sentence length, emotional valence and familiarity across the four main comparison conditions.

### Procedure

The procedure was similar to that used in Reid and Katz (2018). After participants arrived at the laboratory, they read the letter of information about this experiment, and verbal consent was obtained from each participant before they began. Participants were told that this was an experiment on language, memory and mental mathematics ability. The purpose of this was to encourage participants to pay more attention to the distractor tasks to exert the interference effect as much as possible.

After the explanation, the experimenter opened the experiment on the computer. E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA; see Schneider et al., 2002) was used for data collection. After the participants read the

experiment descriptions that were displayed on the computer, they pressed the space bar and started the test. The Chinese and English versions of the experiment were counterbalanced across different participants. Each version was divided into practice and formal experiment trials. The practice trial was designed to familiarize participants with the task and provided feedback with the correct answers to the math questions and recognition items after participants responded. Moreover, the practice trial items were metaphor and literal phrases created by the experimenter, and none of these items appeared later in the formal experiment trials. Each version contained one practice trial and three formal trials. The only difference between these two parts was that the practice trial displayed correct answers to the distractor task and recognition test whereas the formal trials did not. Participants firstly saw the study list, then finished the math distractor, and finally answered the recognition test for each study list they had just read. This process was repeated for each of the presented study lists. Therefore, in contrast to some DRM experiments wherein all study lists are presented first and are followed by a large recognition test drawing from all lists (e.g., Gallo, Roediger, & McDermott, 2001; Soro et al., 2017), due to the difficulty of remembering full sentences, we tested recognition following each list (as done in Reid & Katz, 2018; see also Kawasaki & Yama, 2006).

Each phrase in each study list was preceded by a fixation cross “+” presented for 500 milliseconds in the centre of the screen, and then the phrase itself was presented for 3 seconds. After the study list was presented, the 10 math questions in the distractor task followed. Finally, the recognition test took place with each phrase presented one at a time in the center of the screen. Participants were instructed to identify phrases as old or new by pressing the “O” or “N” buttons, respectively, on the keyboard. Similar to the distractor task, the recognition task did not have a time limit.

The entire memory task took about 25–40 minutes to complete. In order to understand participants’ Chinese and English language proficiency and provide a reference basis for later data analysis, participants also completed a language experience questionnaire after the memory task, which took about 5–10 minutes. Therefore, the whole process took about 30–50 minutes. This research was approved by the Foreign Languages Department of Zhejiang Gongshang University.

### Results

Generalized Linear mixed-effects (GLMM) models from the *lme4* package in R were used to analyze false recognition (Bates, Mächler, Bolker, & Walker, 2015; Lo & Andrews, 2015; R Core Team, 2015), treating subjects and items as random effects and treating Lure Type and Language as fixed effects (Baayen, 2008; Baayen, Davidson, & Bates, 2008). The function *Anova* in the *car* package (Fox & Weisberg, 2016) was used to determine *p*-values from Wald chi-square tests for fixed items with more than two levels. For the false recognition analysis, the model was  $\text{Accuracy} = \text{glmer}(\text{accuracy} \sim \text{Lure Type} * \text{Language} + (1 | \text{subject}) + (1 | \text{item}), \text{family} = \text{"binomial"}, \text{control} = \text{glmerControl}(\text{optimizer} = \text{"bobyqa"}))$ .

For completeness, we also examined response latencies for the lure sentences when they were correctly identified as “new” items (as will be seen, there were too few falsely recognized items in some categories to do a meaningful analysis of response latencies for false recognition). In false recognition studies with word lists,

participants typically take longer to identify new words as “new” when they are related to the studied words because these words seem more familiar and are more difficult to reject as “old” (see Gallo, 2006, for a review). Therefore, if critical consistent lures have longer response latencies when they are correctly identified as “new”, this would suggest that they seemed more familiar to participants, supporting the hypothesis that their corresponding conceptual metaphors were activated at study.

Response latencies less than 250 ms or more than 3 standard deviations from the participant’s mean response times, as well as false recognitions (14.3% of the data) were excluded from the latency analyses. Before running the model, R-default treatment contrasts were altered to sum-to-zero contrasts (Levy, 2014; Singmann & Kellen, 2017). The R code for the model was:  $RT = \text{glmer}(RT \sim \text{Lure Type} * \text{Language} + (1 | \text{subject}) + (1 | \text{item}), \text{family} = \text{Gamma}(\text{link} = \text{“identity”}), \text{control} = \text{glmerControl}(\text{optimizer} = \text{“bobyqa”}))$ .

More complex models that included all relevant random structures were used in our initial analyses, but we used the simpler models noted above due to convergence failures with the more complex random slope models (Barr, 2013). Different link functions were used for the false recognition and latency analyses because gamma link is appropriate for positively skewed continuous data that are characteristic of response latencies, and binomial link is suitable for recognition data, which is binomial data (each trial is a dichotomous datapoint, that is, the item is categorized as either “old” or “new”). More details about data analysis in GLMM models can be found in Fox and Weisberg (2018). The percentage of false recognition and mean response latencies from a subject-based analysis for the four types of lures are shown in Table 2.

## Analyses

### Preliminary analysis

We first conducted a preliminary analysis on the critical consistent, control metaphor, and control literal lures when they were UNRELATED to the study lists. Although the sentences were matched well on length, familiarity, and affective valence, this analysis was conducted to directly examine the likelihood of false recognition for these items when they were not related to the presented items (in other words, to ensure that some lures were not more likely than others to be falsely recognized for reasons other than conceptual metaphor activation). A 2 (Language: Chinese or English) by 3 (Unrelated Lure Type: critical consistent, control metaphor, or control literal) within-subject GLMM analysis was conducted. In the analysis for false recognition, there were no main effects of Lure Type ( $\chi^2 = 2.008, p = .366$ ) or Language ( $\chi^2 = 2.383, p = .123$ ), nor was the interaction between Language

and Lure Type significant,  $\chi^2 = 2.178, p = .337$ . We also conducted an analysis of response latencies, which revealed no significant main effect of Lure Type ( $\chi^2 = 4.567, p = .102$ ), but a significant main effect of Language ( $\chi^2 = 241.517, p < .001$ ), suggesting that correct rejections of unrelated lures were faster overall in Chinese (1068ms) than in English (1964ms). This was not surprising as it was expected that participants would take longer to process and make recognition decisions for sentences in their L2. Critically, there was no significant interaction between Language and Lure Type,  $\chi^2 = 1.786, p = .406$ . Therefore, any differences between the lure types in the subsequent analyses cannot be attributed to pre-existing differences in the lure items.

### Main analysis

Participants finished a recognition test following each study list. The data presented in Table 2 are based on participants’ false recognition across all study lists. The crucial interest was the false recognition rate contrast between the four lure types. We conducted a 2 (Language: Chinese and English) by 4 (Lure Type: critical consistent, control metaphor, control literal, and unrelated) repeated measurements GLMM analysis on false recognition and response latencies for lures correctly identified as new.

### False recognition analysis

The false recognition analysis revealed a significant main effect of Lure Type ( $\chi^2 = 115.723, p < .001$ ), indicating that mean false recognition varied significantly between the different lure types. Moreover, there was a significant interaction between Language and Lure Type,  $\chi^2 = 13.419, p = .004$ . Lastly, the main effect of Language did not approach significance,  $\chi^2 = 1.065, p = .302$ .

In order to determine which lure types differed statistically, we conducted post-hoc comparisons between the false recognition rates for the different lure types within each language. For Chinese, the false recognition rate of the critical consistent lures was significantly higher than for control metaphor lures ( $z = -4.003, p < .001$ ), control literal lures ( $z = -6.212, p < .001$ ) and unrelated lures ( $z = -7.881, p < .001$ ), replicating the findings of Reid and Katz (2018) in Chinese. The false recognition rates also differed significantly between the control metaphor lures and the control literal lures ( $z = -3.061, p = .012$ ), and between the control metaphor lures and unrelated lures ( $z = -4.209, p < .001$ ). For English, the false recognition rate of the critical consistent lures was also significantly higher than for the control metaphor lures ( $z = -3.289, p = .006$ ), control literal lures ( $z = -3.313, p = .005$ ) and unrelated lures ( $z = -7.230, p < .001$ ),

**Table 2.** Percentage of items categorized as old for actual old items (correct recognition) and for the four lure types (false recognition). Mean latencies also displayed for lure types when lure sentences were correctly identified as “new”.

Lure type	Number of items per participant (per language)	Percent recognized for Chinese lists (SD in brackets)	Percent recognized for English lists (SD in brackets)	Mean latencies for Chinese lists (SD in brackets)	Mean latencies for English lists (SD in brackets)
Old	42	86.7 (6)	84.0 (9)		
Critical consistent	9	33.6 (21)	27.2 (24)	1598(580)	2422(1005)
Control metaphor	9	9.7 (14)	13.6 (17)	1297(490)	2108(581)
Control literal	9	1.9 (6)	10.3 (12)	1139(305)	1986(480)
Unrelated lures	10	1.3 (3)	3.2 (6)	1057(267)	1920(484)

indicating that the conceptual metaphor false memory effect is robust even in participants' second language. The false recognition rates also differed significantly between the control metaphor lures and the unrelated lures ( $z = 4.056, p < .001$ ). Unlike in Chinese however, the false recognition rates did not differ significantly between the control metaphor lures and the control literal lures ( $z = 0.123, p = .999$ ).

As mentioned above, there was also a significant interaction between Language and Lure Type. To explore this interaction, we conducted a series of paired *t*-tests to compare each Chinese lure type to its English counterpart. Alpha was adjusted using the Bonferroni correction, yielding an adjusted value of .013. The only contrast that reached significance was between control literal lures,  $t(39) = -4.22, p < .001$ , *Cohan's d* =  $-0.89$  (all other *p*'s > .025). Therefore, the interaction between Language and Lure Type was driven mainly by increased false recognition for literal sentences about the source domain in English, the participants' L2, relative to their first language of Chinese.

To examine the testing order effect statistically, we also conducted a mixed-design GLMM analysis with Testing Order (first block vs. second block), Language (Chinese vs. English) and Lure Type (critical consistent, control metaphor, control literal and unrelated lures) as fixed factors. The results revealed that the main effect of Lure Type and interaction between Language and Lure Type were still significant (*p*'s < .05), and that the main effect of Language remained nonsignificant ( $p = .369$ ). The interaction between Language and Testing Order was significant,  $\chi^2 = 11.702, p = .001$ , suggesting that Chinese-English bilinguals had fewer false recognitions when reading English metaphors than when reading Chinese metaphors in the first block, whereas the pattern was reversed in the second block. Critically however, the interaction between Lure Type and Testing Order was not significant,  $\chi^2 = 1.095, p = .778$ , nor was the three-way interaction between Testing Order, Language and Lure Type,  $\chi^2 = 1.043, p = .791$ , indicating that the Testing Order did not impact the pattern of false recognition between the different lure types across the two languages.

### Latency analysis

The latency analysis on lures correctly identified as "new" revealed significant main effects of Lure Type ( $\chi^2 = 706.608, p < .001$ ) and Language ( $\chi^2 = 1842.498, p < .001$ ), indicating that recognition latencies varied significantly between the different lure types and languages. There was also a significant interaction between Language and Lure Type,  $\chi^2 = 34.161, p < .001$ . The recognition latencies (and standard deviations) for the different lure types are also displayed in Table 2.

Post-hoc comparisons were conducted to examine whether the recognition latencies differed between specific pairs of lure types. For Chinese, the mean latency for critical consistent expressions was significantly slower than for control metaphor lures ( $z = -4.508, p < .001$ ), control literal lures ( $z = -7.045, p < .001$ ) and unrelated lures ( $z = -7.263, p < .001$ ). The mean latency for the control metaphor lures was also significantly slower than for the control literal lures ( $z = -4.945, p < .001$ ) and unrelated lures ( $z = -7.339, p < .001$ ). For English, the mean latency of the critical consistent lures was also significantly slower than for the control metaphor lures ( $z = -2.876, p = .021$ ), control literal lures ( $z = -9.980, p < .001$ ) and unrelated lures ( $z = -8.548, p < .001$ ), indicating a similar pattern of conceptual metaphor activation in participants' second language. Mean latency was

also significantly slower for the control metaphor lures than the control literal lures ( $z = -7.010, p < .001$ ) and unrelated lures ( $z = -7.018, p < .001$ ). These data suggest that even when critical consistent lures were not falsely recognized, they took longer to identify as "new" compared to the other lure types in both Chinese and English. This was likely because the critical consistent lures were more familiar, and therefore, harder to reject as "old" (see Gallo, 2006). As such, this is consistent with the false recognition results and further suggests that the conceptual metaphors were activated at study.

In sum, across both languages, non-presented expressions that were based on the same underlying conceptual metaphor as the study list items were falsely recognized more often than controls. This replicates the memory effect found by Reid and Katz (2018) both in a completely different language (Simplified Chinese), and in English with Chinese-English bilinguals who speak English as a second language. Metaphorically consistent expressions also took longer to identify as "new" when they were not falsely recognized, suggesting that they were more familiar to participants, and thus, harder to reject as "old". The major difference between the languages was that control literal sentences about the source domains of the conceptual metaphors were falsely recognized more often in Chinese-English bilinguals' second language, English.

### General discussion and conclusion

The present study shows three major findings. Firstly, the results suggest that conceptual metaphors are activated automatically and immediately during sentence reading, which influences how such sentences are encoded into memory. After reading lists of metaphorical phrases based on the same conceptual metaphor, participants were more likely to falsely recognize non-presented phrases based on the same conceptual metaphor mapping (i.e., "critical consistent" lures), replicating the results obtained in English speakers (Reid & Katz, 2018). We interpret this finding to suggest that the underlying conceptual metaphor was activated while participants read the study list expressions, and as a result, participants found new expressions that used this conceptual metaphor to be more familiar at recognition. This could be due to a processing fluency advantage for critical consistent expressions. That is, because the conceptual mapping needed to process these items was already activated from the study list, these expressions were processed more fluently during recognition, whereas for the control expressions, a new conceptual mapping would need to be activated, yielding relatively slower processing. Processing fluency is a heuristic for judging familiarity in recognition tasks and has been proposed as a key mechanism for false memory effects (Doss, Bluestone, & Gallo, 2016; Gallo & Roediger, 2003; Whittlesea, 2002).<sup>2</sup> Alternatively, from the perspective of fuzzy-trace theory (Brainerd & Reyna, 2002; Reyna et al., 2016), the conceptual metaphor may represent the overall theme, or "gist", of each list; and new expressions that use the same metaphor mapping may be deemed by participants as being more consistent with this gist, yielding higher false recognition. Regardless of the specific memory theory, the fact that the critical consistent expressions were falsely recognized significantly more often than controls suggests that conceptual metaphors influence how expressions are encoded into memory, supporting their psychological reality.

Secondly, our results also support the universality of conceptual metaphors as the memory effect is observed not only in

English speakers, but also in Chinese speakers. This suggests that the effect found by Reid and Katz (2018) was not just due to the particular stimuli used, but that it generalizes to an entirely new set of expressions from a different language. We also employed two lists based on new conceptual metaphors, ARGUMENT IS WAR and THE ECONOMY IS A HUMAN BEING, that were not explored by Reid and Katz, suggesting the memory effect is robust across different conceptual mappings as well.

Thirdly, the memory effect was also observed in participants' second language of English. The robustness of the effect in participants' L2 is somewhat surprising as metaphor comprehension often poses problems for second language learners (see Nacey, 2016, for a review). Nonetheless, the Chinese–English bilinguals in this study showed a significant memory effect for critical consistent lures in their L2, and after correcting for multiple comparisons, the false recognition rate for these lures did not differ between L1 and L2. This demonstrates that participants were able to activate the appropriate conceptual mappings even in their L2.

Although the conceptual metaphor false memory effect was observed in both participants' L1 and L2, the major difference that emerged between the languages was for the control literal lures, in which there was a higher false recognition rate in participants' L2. Bilinguals often show difficulty inhibiting the literal meaning of metaphorical expressions in their second language (Chen *et al.*, 2013) and sometimes interpret metaphorical expressions literally when there is a lack of context (Picken, 2005). As such, we interpret the difference for control literal lures as indicating that the study list expressions activated a literal representation of the source domain to a greater extent in participants' second language. Because the literal source domain was activated during reading, when participants subsequently recognized items, they were more likely to make memory errors for literal sentences about the source domain. In contrast, when participants read the study lists in their L1, they more easily inhibited the literal representation of the source domain to focus on its metaphorical aspects, resulting in very few memory errors for the control literal lures in their L1. This is also consistent with the graded salience model (Giora, 2003), which posits that less salient meanings take longer to be activated. We hypothesize that metaphorical meanings are more salient in participants' L1 due to their expertise in the language, but in L2, these meanings may be less salient, leaving more room for literal representations to activate.

Although we have discussed our results mostly in terms of CMT, it is important to consider other metaphor theories that could possibly explain these effects. Another dominant approach to metaphor is Gentner and colleagues' CAREER OF METAPHOR THEORY (Bowdle & Gentner, 2005; Gentner *et al.*, 2001), which posits that novel metaphors are processed through structural alignment, much like analogies. Over time, if a metaphor becomes highly conventionalized, the metaphoric meaning associated with the source term is abstracted and becomes a secondary, lexicalized meaning of the source term. At this point, this conventionalized meaning resembles a metaphoric category, consistent with the attributive category theory of metaphor (Glucksberg & Keysar, 1990). For instance, "goldmine" can refer literally to a place where gold is mined, or to an abstract metaphor category of "things that are valuable".

Some aspects of our data are consistent with Career of Metaphor theory. The finding that participants have more false recognitions in L2 for literal sentences associated with the source domains of the metaphors is consistent with Career of Metaphor

because these metaphors should presumably be more novel in participants' L2. According to Career of Metaphor, novel metaphors involve structurally aligning the concrete and literal representations of the target and source, which means the literal representation of the source should be activated for novel metaphors. In contrast, for highly conventional metaphors, the metaphoric meaning is lexicalized and can be accessed directly. Therefore, it makes sense that literal representations should be less activated in participants' L1 with metaphors that are highly familiar as they should be able to directly access the metaphorical meaning of the source terms used in these metaphors. Therefore, the increased rate of false recognition for literal control sentences in participants' L2 is consistent with Career of Metaphor theory.

From our perspective, it is more difficult to explain the pattern of false recognition in L1 in terms of Career of Metaphor theory. According to this theory, highly conventional metaphors, such as those used in this study, can be processed by categorization. Proponents of the attributive category theory of metaphor have rejected the notion that conceptual metaphors play a role in processing conventional metaphor statements (Glucksberg, Brown, & McGlone, 1993; Keysar, Shen, Glucksberg, & Horton, 2000; McGlone, 1996, 2007, 2011), arguing instead that people access stereotypical meanings of the source term that could be attributed to the target term. Furthermore, Gentner *et al.* (2001) argue that category-based approaches are "localist" in that they posit that metaphors highlight categorical relations between only the specific terms used in their statements, and therefore, do not have a mechanism to explain extended metaphorical mappings. If conceptual metaphors are not activated when processing conventional metaphors, we may expect a false memory effect in L2 where the expressions are novel, but there is no reason to expect that critical consistent lures should be falsely recognized more often than control metaphor lures in L1 where the expressions are conventional, unless false recognition was caused by another factor such as overall word similarity. However, Reid and Katz (2018) demonstrated that word similarity alone cannot account for false recognition effects based on conceptual metaphors (see also Katz & Reid, 2020). Therefore, although the increased rate of false recognition for literal control sentences in L2 aligns with Career of Metaphor theory, the pattern of false recognition in L1 is inconsistent with the theory. Nonetheless, we do not reject the notion that structural alignment is an important mechanism in metaphor comprehension, and it is likely important for how conceptual metaphors are initially learned (Gentner *et al.*, 2001; Holyoak & Stamenković, 2018).

There are some limitations in the present study. For instance, we only recruited medium proficiency Chinese–English bilinguals, but according to the graded salience hypothesis, language proficiency should also influence metaphor processing, so future research should compare conceptual metaphor activation in bilinguals with different levels of proficiency. Furthermore, we extended the evidence of conceptual metaphor activation only to Chinese–English bilinguals, but it may be interesting to explore the activation of conceptual metaphors in different languages to confirm the universality of Conceptual Metaphor Theory. Finally, another avenue for future research may be to compare English monolinguals reading English metaphoric expressions to Chinese monolinguals reading Chinese expressions to explore cultural differences in conceptual metaphor processing. Chinese culture emphasizes holistic processing, which involves a focus on how elements are interconnected, whereas Western culture prefers analytical processing, focusing on the properties of



individual objects (Li, Masuda, Hamamura, & Ishii, 2018). As such, Chinese readers may focus more on how the study list expressions are connected through an underlying metaphor compared to English readers, who may focus more on the aspects of each expression that discriminate it from the others. Therefore, Chinese participants may show larger memory effects for critical consistent expressions than English participants. Furthermore, recent research has found that culture influences the DRM memory illusion (Wang et al., 2021). Therefore, it may be interesting to explore the role of culture in false memories elicited from conceptual metaphor activation.

In conclusion, through a false memory task conducted with Chinese-English bilinguals, we provide evidence for the activation of conceptual metaphors during reading in the Chinese language, and for individuals reading expressions in their second language, English. This replicates and extends the findings of Reid and Katz (2018), generalizing the effect to both another language (Simplified Chinese) and another language group (second language learners). The data also suggest that participants had difficulty inhibiting literal meaning in their L2, supporting a bilingual extension of the graded salience hypothesis (Giora, 2003). Future research should consider other languages, as well as the role culture plays in memory for language based on conceptual metaphors.

**Acknowledgements.** The data and material are available upon request from the first author. This research was supported by the Zhejiang Provincial Philosophy and Social Science Planning Project (22NDQN238YB) and the 2021 Zhejiang Gongshang University Provincial platform teaching project (1070XJ0520111-31) to Huilan Yang, and the 2022 Zhejiang Gongshang University school-level graduate education reform project (YJG2022104) to Sumin Zhang.

**Conflicts of interest.** None.

## Notes

<sup>1</sup> The labels for the conceptual metaphors themselves (e.g., “love is a journey”) were also included in each study list. These are mnemonics traditionally used in the cognitive linguistics literature to name the mappings (Lakoff, 1993); however, they are also linguistic exemplars of the mappings. That is, the expression “love is a journey” is itself an instantiation of the LOVE IS A JOURNEY conceptual metaphor (Lakoff, 1993). We considered examining recognition for these items separately; however, the rate of recognition for these items did not differ significantly from the other old items. For simplicity, we report correct recognition collapsed across all old items.

<sup>2</sup> Note that recognition latencies are not a direct measure of processing fluency as a recognition decision still needs to be made after the item is processed. Critical consistent lures correctly identified as “new” may be processed more fluently, but this may result in a more difficult (and therefore, longer) recognition decision due to the feeling of familiarity attributed to this fluency.

## References

- Baayen RH (2008) Analyzing linguistic data: A practical introduction to statistics using R. In Processing. Retrieved from <http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenCUPstats.pdf>
- Baayen RH, Davidson DJ and Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390–412. doi: <https://doi.org/10.1016/j.jml.2007.12.005>
- Barr DJ (2013) Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology* 4, 328. doi: <https://doi.org/10.3389/fpsyg.2013.00328>
- Barsalou LW (1999) Perceptual symbol systems. *Behavioral and Brain Sciences* 22, 577–660. doi: 10.1017/S0140525X99002149
- Barsalou LW (2008) Grounded cognition. *Annual Review of Psychology* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639
- Bates D, Mächler M, Bolker BM and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. doi: <https://doi.org/10.18637/jss.v067.i01>
- Berendt EA and Tanita K (2011) The ‘heart’ of things: A conceptual metaphor analysis of heart and related body parts in Thai, Japanese and English. *Intercultural Communication Studies* 20, 65–78. doi: <https://web.uri.edu/iaics/files/06ErichABerendtKeikoTanita.pdf>
- Bowdle BF and Gentner D (2005) The career of metaphor. *Psychological Review* 112, 193–216. doi:10.1037/0033-295X.112.1.193
- Brainerd CJ and Reyna VF (2002) Fuzzy-trace theory and false memory. *Current Directions in Psychological Science* 11, 164–169. doi:10.1111/1467-8721.00192
- Cardillo ER, Watson CE, Schmidt GL, Kranjec A and Chatterjee A (2012) From novel to familiar: tuning the brain for metaphors. *Neuroimage* 59(4), 3212–3221. doi: <https://doi.org/10.1016/j.neuroimage.2011.11.079>
- Casasanto D (2008) Similarity and proximity: When does close in space mean close in mind? *Memory & Cognition* 36, 1047–1056. doi:10.3758/MC.36.6.1047
- Casasanto D and Boroditsky L (2008) Time in the mind: Using space to think about time. *Cognition* 106, 579–593. doi: 10.1016/j.cognition.2007.03.004
- Chang M and Brainerd CJ (2021) Semantic and phonological false memory: A review of theory and data. *Journal of Memory and Language* 119, 104210. doi: 10.1016/j.jml.2020.104210
- Chen H, Peng X and Zhao Y (2013) An ERP study on metaphor comprehension in the bilingual brain. *Chinese Journal of Applied Linguistics* 36, 505–517. doi: 10.1515/cjal-2013-0034
- Cieślicka A (2006) Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research* 22, 115–144. doi: 10.1191/0267658306sr2630a
- Cieślicka AB (2015) Idiom acquisition and processing by second/foreign language learners. In RR Heredia and AB Cieślicka (Eds.), *Bilingual Figurative Language Processing*. New York, NY: Cambridge Press, pp. 208–244.
- Citron FMM, Michaelis N and Goldberg AE (2020) Metaphorical language processing and amygdala activation in L1 and L2. *Neuropsychologia* 140, 107381. doi: <https://doi.org/10.1016/j.neuropsychologia.2020.107381>
- Deese J (1959) On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology* 58, 17–22. doi:10.1037/h0046671
- Doss MK, Bluestone MR and Gallo DA (2016) Two mechanisms of constructive recollection: Perceptual recombination and conceptual fluency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42, 1747–1758. doi:10.1037/xlm0000273
- Forceville C (2007) Multimodal metaphor in ten Dutch TV commercials. *Public Journal of Semiotics* 1, 15–34. doi: <http://www.semiotics.ca/issues/pjos-1-1.pdf>
- Forgács B, Lukács Á and Pléh C (2014) Lateralized processing of novel metaphors: disentangling figurativeness and novelty. *Neuropsychologia* 56, 101–109. doi: 10.1016/j.neuropsychologia.2014.01.003
- Fox J and Weisberg S (2016) car: Companion to applied regression. R Package Version 2.1-2. <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>
- Fox J and Weisberg S (2018) An R companion to applied regression. Sage publications. Retrieved from [https://books.google.com.tw/books?hl=zh-CN&lr=&id=uPNrDwAAQBAJ&oi=fnd&pg=PP1&dq=a+r+comparison+to+applied+regression&ots=MwH2dA4y10&sig=\\_wXu6S70kQAsathORUHL1JzxAc&redir\\_esc=y#v=onepage&q=a%20r%20comparison%20to%20applied%20regression&f=false](https://books.google.com.tw/books?hl=zh-CN&lr=&id=uPNrDwAAQBAJ&oi=fnd&pg=PP1&dq=a+r+comparison+to+applied+regression&ots=MwH2dA4y10&sig=_wXu6S70kQAsathORUHL1JzxAc&redir_esc=y#v=onepage&q=a%20r%20comparison%20to%20applied%20regression&f=false)
- Gallo DA (2006) *Associative illusions of memory: False memory research in DRM and related tasks*. New York, NY: Psychology Press.
- Gallo DA (2010) False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition* 38, 833–848. doi:10.3758/MC.38.7.833
- Gallo DA and Roediger HL (2003) The effects of associations and aging on illusory recollection. *Memory & Cognition* 31, 1036–1044. doi:10.3758/BF03196124
- Gallo DA, Roediger HL and McDermott KB (2001) Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review* 8, 579–586. doi: 10.3758/BF03196194

- Gentner D, Bowdle BF, Wolff P and Boronat C** (2001) Metaphor is like analogy. In D Gentner, KJ Holyoak and BN Kokinov (Eds.), *The Analogical Mind: Perspectives from Cognitive Science*. Cambridge, MA: MIT Press, pp. 199–253. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.5863&rep=rep1&type=pdf>
- Gibbs RW** (2006) *Embodiment and cognitive science*. New York, NY: Cambridge University Press.
- Gibbs RW** (2009). Why do some people dislike conceptual metaphor theory? *The Journal of Cognitive Semiotics* 5, 14–36. doi: 10.1515/cogsem.2013.5.12.14
- Gibbs RW** (2013) Walking the walk while thinking about the talk: Embodied interpretation of metaphorical narratives. *Journal of Psycholinguistic Research* 42, 363–378. doi: 10.1007/s10936-012-9222-6
- Gibbs RW, Lima PLC and Francozo E** (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics* 36, 1189–1210. doi: 10.1016/j.pragma.2003.10.009
- Giora R** (2003) *On our mind: Saliency, context, and figurative language*. New York, NY: Oxford University Press.
- Glucksberg S, Brown M and McGlone MS** (1993) Conceptual metaphors are not automatically accessed during idiom comprehension. *Memory & Cognition* 21, 711–719. doi:10.3758/BF03197201
- Glucksberg S and Keysar B** (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review* 97, 3–18. doi:10.1037/0033-295X.97.1.3
- Gong S-P and Ahrens K** (2007) Processing conceptual metaphors in on-going discourse. *Metaphor and Symbol* 22, 313–330. doi: 10.1080/10926480701528121
- Han J** (2019) The study of Thai elephant culture based on the “elephant metaphors” in Thai idioms. *Comparative Literature: East & West* 3, 148–162. doi:10.1080/25723618.2019.1701306
- Heredia RR and Cieślicka AB** (2016) Metaphoric reference: An eye movement analysis of Spanish–English and English–Spanish bilingual readers. *Frontiers in Psychology* 7, 439. doi: 10.3389/fpsyg.2016.00439
- Holyoak KJ and Stamenković D** (2018) Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin* 144, 641–671. doi:10.1037/bul0000145
- Ikuta M and Miwa K** (2021) Structure mapping in second-language metaphor processing. *Metaphor and Symbol* 36, 288–310. doi:10.1080/10926488.2021.1941971
- Jankowiak K, Rataj K and Naskręcki R** (2017) To electrify bilingualism: Electrophysiological insights into bilingual metaphor comprehension. *PLoS One* 12(4), e0175578. <https://doi.org/10.1371/journal.pone.0175578>
- Kasparian K** (2013) Hemispheric differences in figurative language processing: contributions of neuroimaging methods and challenges in reconciling current empirical findings. *Journal of Neurolinguistics* 26(1), 1–21. doi: <https://doi.org/10.1016/j.jneuroling.2012.07.001>
- Katz AN and Law A** (2010) Experimental support for conceptual metaphors with an episodic memory task. *Metaphor and Symbol* 25, 263–270. doi:10.1080/10926488.2010.510930
- Katz AN and Reid JN** (2020) Tests of Conceptual Metaphor Theory with episodic memory tests. *Cognitive Semantics* 6, 56–82. doi:10.1163/23526416-00601003
- Katz AN and Taylor TE** (2008) The journeys of life: Examining a conceptual metaphor with semantic and episodic memory recall. *Metaphor and Symbol* 23, 148–173. doi:10.1080/10926480802223051
- Kawasaki Y and Yama H** (2006) The difference between implicit and explicit associative processes at study in creating false memory in the DRM paradigm. *Memory* 14(1), 68–78. doi:10.1080/09658210444000520
- Keysar B, Shen Y, Glucksberg S and Horton WS** (2000) Conventional Language: How Metaphorical Is It? *Journal of Memory and Language* 43, 576–593. doi: 10.1006/jmla.2000.2711
- Lakoff G** (1993) The contemporary theory of metaphor. In A Ortony (Ed.), *Metaphor and thought: Vol. 2*. New York, NY: Cambridge University Press, pp. 202–251.
- Lakoff G** (2008) The neural theory of metaphor. In RW Gibbs (Ed.), *Cambridge handbook of metaphor and thought*. New York, NY: Cambridge University Press, pp. 17–38.
- Lakoff G and Johnson M** (1980) *Metaphors we live by*. Chicago, IL: University of Chicago Press.
- Levy R** (2014) Using R formulae to test for main effects in the presence of higher-order interactions. ArXiv Preprint ArXiv:1405.2094. Retrieved from <https://arxiv.org/pdf/1405.2094.pdf>
- Li X** (2010) A Comparative Analysis of English and Chinese Idioms-From the Perspective of Conceptual Metaphor of “Happiness”. *Journal of Language Teaching and Research* 1(4): 473–476. doi: 10.4304/jltr.1.4.473-476
- Li LMW, Masuda T, Hamamura T and Ishii K** (2018) Culture and decision making: Influence of analytic versus holistic thinking style on resource allocation in a fort game. *Journal of Cross-Cultural Psychology* 49, 1066–1080. doi:10.1177/0022022118778337
- Lo S and Andrews S** (2015) To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology* 30(6), 1171. doi: <https://doi.org/10.3389/fpsyg.2015.01171>
- Lv Z and Zhang Y** (2012) Universality and variation of conceptual metaphor of love in Chinese and English. *Theory and practice in language studies* 2(2): 355–359. doi: 10.4304/tpls.2.2.355-359.
- Mashal N, Borodkin K, Maliniak O and Faust M** (2015) Hemispheric involvement in native and non-native comprehension of conventional metaphors. *Journal of Neurolinguistics* 35: 96–108. doi: <https://doi.org/10.1016/j.jneuroling.2015.04.001>
- Matthews JL and Matlock T** (2011) Understanding the link between spatial distance and social distance. *Social Psychology* 42, 185–192. doi:10.1027/1864-9335/a000062
- McGlone MS** (1996) Conceptual metaphors and figurative language interpretation: Food for thought? *Journal of Memory and Language* 35, 544–565. doi: <https://doi.org/10.1006/jmla.1996.0029>
- McGlone MS** (2007) What is the explanatory value of a conceptual metaphor? *Language & Communication* 27, 109–126. doi:10.1016/j.langcom.2006.02.016.
- McGlone MS** (2011) Hyperbole, homunculi, and hindsight bias: An alternative evaluation of conceptual metaphor theory. *Discourse Processes* 48(8), 563–574. doi:10.1080/0163853X.2011.606104
- Nacey S** (2016) Metaphor comprehension and production in a second language. In e bSemino and Z Demjén (Eds.), *The Routledge handbook of metaphor and language*. London: Routledge, pp. 503–515.
- Nayak NP and Gibbs RW** (1990) Conceptual knowledge in the interpretation of idioms. *Journal of Experimental Psychology: General* 119, 315–330. doi:10.1037/0096-3445.119.3.315
- Nomura M** (1996) The ubiquity of the fluid metaphor in Japanese: A case study. *Poetica* 46, 41–75. [https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/68211/1/Nomura%20\(1996\).pdf](https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/68211/1/Nomura%20(1996).pdf)
- Palmer G** (2003) Talking about thinking in Tagalog. *Cognitive Linguistics* 14, 251–280. doi: 10.1515/cogl.2003.010
- Picken JD** (2005) Helping foreign language learners to make sense of literature with metaphor awareness-raising. *Language Awareness* 14, 142–152. doi: 10.1080/09658410508668830
- R Core Team** (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Retrieved from <https://cran.r-project.org/doc/manuals/r-release/R-ints.html>
- Reid JN and Katz AN** (2018) Something false about conceptual metaphors. *Metaphor and Symbol* 33, 36–47. doi: 10.1080/10926488.2018.1407994
- Reid JN and Katz AN** (2022) Conceptual metaphors influence memory automatically: Evidence from a divided attention false memory task. *Memory & Cognition*. Advance online publication. doi: 10.3758/s13421-022-01322-1
- Reyna VF, Corbin JC, Weldon RB and Brainerd CJ** (2016) How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition* 5, 1–9. doi:10.1016/j.jarmac.2015.12.003
- Roediger HL and McDermott KB** (1995) Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Schneider W, Eschman A and Zuccolotto A** (2002) *E-Prime: User's guide*. Pittsburgh, PA: Psychology Software Tools.
- Singmann H and Kellen D** (2017) An Introduction to Mixed Models for Experimental Psychology. In D Spieler and E Schumacher (Eds.), *New Methods in Neuroscience and Cognitive Psychology* (pp. 1–39). Retrieved from <https://cran.r-project.org/web/packages/afex/vignettes/introduction-mixed-models.pdf>

- Soriano C** (2003) Some anger metaphors in Spanish and English: A contrastive review. *International Journal of English Studies* 3, 107–122. doi: <https://revistas.um.es/ijes/article/view/48381>
- Soro JC, Ferreira MB, Semin GR, Mata A and Carneiro P** (2017) Ad hoc categories and false memories: Memory illusions for categories created on-the-spot. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43, 1779–1792. doi:10.1037/xlm0000401
- Thibodeau P and Durgin FH** (2008) Productive figurative communication: Conventional metaphors facilitate the comprehension of related novel metaphors. *Journal of Memory and Language* 58, 521–540. doi: 10.1016/j.jml.2007.05.001
- Türker E** (2016) The role of L1 conceptual and linguistic knowledge and frequency in the acquisition of L2 metaphorical expressions. *Second Language Research* 32, 25–48. doi: 10.1177/0267658315593336
- Wang J, Otgaar H, Santtila P, Shen X and Zhou C** (2021) How culture shapes constructive false memory. *Journal of Applied Research in Memory and Cognition* 10, 24–32. doi: 10.1016/j.jarmac.2020.12.002
- Whittlesea BWA** (2002) False memory and the discrepancy-attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology: General* 131, 96–115. doi:10.1037/0096-3445.131.1.96
- Wilson NL and Gibbs RW** (2007) Real and imagined body movement primes metaphor comprehension. *Cognitive Science* 31, 721–731. doi:10.1080/15326900701399962
- Wolff P and Gentner D** (2011) Structure-mapping in metaphor comprehension. *Cognitive Science* 35, 1456–1488. doi:10.1111/j.1551-6709.2011.01194.x
- Yang H, He X, Zhao X and Zhang W** (2015) Multiple metaphorical representations of power: Evidence from size and color. *Acta Psychologica Sinica*, 47 (7), 939–949. doi: 10.3724/SP.J.1041.2015.00939
- Yang H, Reid JN, Katz AN and Li D** (2021) The embodiment of power as forward/backward movement in Chinese and English speakers. *Metaphor and Symbol* 36, 181–193. doi: 10.1080/10926488.2021.1907185
- Yu N** (1995) Metaphorical expressions of anger and happiness in English and Chinese. *Metaphor and Symbol* 10, 59–92. doi: [https://doi.org/10.1207/s15327868ms1002\\_1](https://doi.org/10.1207/s15327868ms1002_1)
- Yu N** (2003) Metaphor, body, and culture: The Chinese understanding of gall-bladder and courage. *Metaphor and symbol* 18(1): 13–31. doi: 10.1207/S15327868MS1801\_2
- Yu N** (2008) Metaphor from body and culture. In RW Gibbs (Ed.), *Cambridge handbook of metaphor and thought*. New York, NY: Cambridge University Press, pp. 247–261.
- Yu N** (2017) LIFE AS OPERA: A cultural metaphor in Chinese. In F Sharifian (Ed.), *Advances in Cultural Linguistics*. Singapore: Springer, pp. 65–87. doi: [https://doi.org/10.1007/978-981-10-4056-6\\_4](https://doi.org/10.1007/978-981-10-4056-6_4)
- Zhong C-B and Leonardelli GJ** (2008) Cold and lonely: Does social exclusion literally feel cold? *Psychological Science* 19, 838–842. doi:10.1111/j.1467-9280.2008.02165.x

## Appendix

Examples of Expressions under different condition used in Chinese experiment

时间就是金钱 (TIME IS MONEY study list)  
 你把暑假的时间都花在哪里了  
 我还有足够的时间完成这项工作  
 计划好你的时间  
 周末很宝贵  
 我没有时间花在这件事上  
 给你三个月的时间  
 老人剩下的时间不多了  
 他投资了一年的时间做实验  
 为此留出几天时间  
 你能花一个下午做这个吗  
 时间都被浪费了  
 游戏规定的时间快用完了  
 免费时间很宝贵  
 在他身上花那么多时间不值得  
 这将为我省很多时间

### Recognition test list

(Critical CM)  
 时间就是金钱

(Critical consistent lures)  
 借给我几分钟  
 那花了我一天的时间  
 不要浪费你的时间

(Control metaphor lures)  
 时间好像静止了  
 时间给他开了个玩笑  
 截止日期快到了

(Control literal lures)  
 你每月房租多少  
 他每两周付款一次  
 她借了低息贷款

(Unrelated items)  
 他们之间的谈话总是唇枪舌剑的  
 我们进行了非常生动的讨论  
 战争和疾病使得人口减少

Examples of Expressions under different condition used in English experiment

### TIME IS MONEY study list

How did you spend the summer break  
 I have some days off banked from last month  
 Budget your hours  
 Weekends are precious  
 I don't have the hours for this  
 I'll give you a minute  
 Is that worth your while  
 Years are invested  
 Put aside a few days for this  
 Can you spare an afternoon  
 Hours are wasted  
 How many minutes do I have left  
 Free hours are valuable  
 The diversion should buy him a few minutes  
 This will save me many hours

### Recognition test list

(Critical CM)  
 Time is money

(Critical consistent lures)  
 Lend me a few minutes  
 That cost me a day  
 You don't use your hours profitably

(Control metaphor lures)  
 The weekend seems so far away  
 The years have not been kind to him  
 The deadline is approaching

(Control literal lures)  
 How much is your rent per month  
 He makes biweekly payments  
 She took out a low-interest loan

(Unrelated items)  
 Your claims are indefensible  
 We had a very lively discussion  
 We need to protect our allies