Bilingualism: Language and Cognition

cambridge.org/bil

Crossing the cultural bridge: The role of inhibitory control during second language metaphor comprehension

Jiayan Chen, Junmei Lv and Baoguo Chen

Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Faculty of Psychology, Beijing Normal University, Beijing 100875, China

Research Article

Cite this article: Chen, J., Lv, J. and Chen, B. (2025). Crossing the cultural bridge: The role of inhibitory control during second language metaphor comprehension. *Bilingualism: Language and Cognition* **28**, 1393–1409. https://doi.org/10.1017/S1366728924001081

Received: 2 November 2023
Revised: 4 December 2024
Accepted: 4 December 2024
First published online: 19 February 2025

Keywords:
context; eye movements; inhibition; metaphor; predictive processing

Corresponding author:
Baoguo Chen;
Email: chenbg@bnu.edu.cn

Abstract

Previous research has found that metaphor comprehension is often more challenging in L2 than in L1 because of the prioritization of literal meanings, but the effect of cross-cultural conceptual differences and the role of inhibitory control during L2 metaphor processing remain uninvestigated. We explored these through a metaphor-induced lexical forgetting paradigm (Experiment 1), a metaphor interpretation task (Experiment 2), and an eye-tracking reading task (Experiment 3) to evaluate competing theories. Inhibitory control did not play a significant role during reading culturally congruent metaphors as it did for culturally incongruent ones. However, interpreting both kinds of L2 metaphors involved more inhibitory control than literals, even after explicit explanatory contexts. Although literal meanings (and culturally incongruent L1 metaphorical meanings) of L2 metaphors may always be activated, inhibition involvement depends on both task requirements and metaphor properties. These can be explained by the extended graded salience view and the predictive processing framework.

Highlights

- Reading culturally incongruent L2 metaphors seems to inhibit L1 metaphorical meanings.
- L2 metaphorical meanings are not directly accessed, even after explanatory contexts.
- Task requirements and metaphor properties moderate inhibition involvement.
- The predictive processing account accommodates variations in metaphor comprehension.

1. Introduction

Imagine your friend has just gotten off work and upon seeing you say, "My boss is a dragon". Now, no matter what cultural background you come from, you know those words are not to be taken literally, i.e., your friend does not really work for a legendary creature with reptilian traits. It is a metaphor, the use of language to describe one thing (the topic) in terms of something else that is conceptually very different (the vehicle). However, if you and your friend have always lived in European culture, you will probably take it that the boss is difficult to deal with, while if both of you have always lived in East Asia, you will see the boss as successful and praiseworthy. An intriguing situation arises if you come from a different cultural background than your friend. In that case, you may weigh your knowledge of dragons in two cultures (or the lack of it), your friend's facial expression and tone of voice, following remarks of the boss, etc., all of which will call for more cognitive effort. You may need to inhibit going straight to the metaphorical sense in your native linguistic culture to cross the cultural bridge, and you could go astray in the process.

Metaphors are prevalent in natural language, making up as much as 20% of discourse (Steen et al., 2010). They are used as not only figurative embellishments but also teaching devices (Littlemore et al., 2012; Ortony, 1975). A myriad of empirical studies have been conducted to investigate metaphor comprehension in native (L1) speakers, upon which several theories have been offered. For second language (L2) users, metaphor comprehension adds fuel of figurative understanding to the flames of being non-native, especially when an L2 metaphorical expression has a different metaphorical meaning in L1 culture. According to Littlemore et al. (2011), 40% of the spoken items that international students found difficult during university lectures were metaphors. However, psychological studies about metaphor have by and large ignored the linguistic cultural dimension, and bilingual metaphor processing remains underexplored. Meanwhile, it is reasonable to assume that inhibitory control, proved to be pivotal in both bilingual language processing (Bialystok & Martin, 2004; Liu et al., 2016; Luk et al., 2010) and L1 metaphor comprehension (Chiappe & Chiappe, 2007; Columbus et al., 2015; George & Wiley, 2016, 2019; Gernsbacher et al., 2001; Glucksberg et al., 2001; Rubio Fernández, 2007), will figure prominently

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

in L2 metaphor processing, though little attention has been given to its investigation. The lacunae left by these paucities motivated the present study. We aim to explore the role of individual differences in inhibitory control during comprehension of both culturally incongruent and culturally congruent L2 metaphors, based on which we try to evaluate and extend existing accounts to develop a more comprehensive account for L2 metaphor comprehension.

1.1. L1 metaphor comprehension and the role of inhibition

Traditionally, the Standard Pragmatic View argues that only when an obligatorily analysed literal meaning is inappropriate contextually (e.g., a boss cannot be an actual dragon) is it inhibited and triggers a nonliteral interpretation (Grice, 1975; Searle, 1979). This serial processing view has been extensively tested in L1 speakers and conclusively rejected (Harris, 1976; Keysar, 1989; Mcelree & Nordlie, 1999; Pollio et al., 1984). For example, Glucksberg et al. (1982) found that people were slower to respond "false" to sentences that were literally incongruous but had possible metaphorical interpretations (e.g., *Some surgeons are butchers.*) than to those without (e.g., *Some apples are oranges.*). This was called the Metaphor Interference Effect, indicating that extraction of metaphorical meaning can be immediate and obligatory, even when it interferes with literal processing as required by the context.

Alternatively, the Direct Access View suggests that listeners need not automatically analyse the complete literal meanings of nonliteral linguistic expressions either before or in parallel to accessing contextual pragmatic knowledge to figure out the intended figurative meanings (Gibbs, 2002; Ortony et al., 1978). A more recent version of this view is the Constraint Satisfaction Model (Katz, 2005; Katz & Ferretti, 2001, 2003; MacDonald & Seidenberg, 2006; Pexman, 2008; Pexman et al., 2000). It professes that comprehension is achieved through parallel satisfaction of multiple probabilistic constraints, including lexical and contextual ones. Under the direct access view, given strong prior contextual constraints, even the figurative meaning of an unfamiliar metaphor can be accessed directly and precisely, without invoking and then inhibiting the literal meaning.

Opposing the possible superiority of contextual influence over lexical one, the Graded Salience View (Giora, 2003, 2012) argues that upon seeing the stimulus, the more salient meaning encoded in the mental lexicon always reaches a sufficient level of activation faster and is accessed obligatorily before less salient ones, independent of the context. Salience depends on conventionality, frequency, familiarity, and prototypicality, regardless of figurativeness/literality. The predictive context may in parallel facilitate the construction of a different nonsalient meaning before or after the activation of the salient one. If the salient meaning turns out to be inappropriate in the local context, like the literal meaning of an unfamiliar metaphor, inhibition of it will take effect only when it interferes with constructing the appropriate contextual interpretation.

Tremendous effort has been poured into adjudicating between these different views, by figuring out different activation time points or processing efforts for literal and figurative meanings (Arzouan et al., 2007; Bambini et al., 2016; Blank, 1988; Blasko & Connine, 1993; Coulson & Van Petten, 2002; De Grauwe et al., 2010; Giora, 1997, 2003; Giora & Fein, 1999; Lai et al., 2009; McElree & Nordlie, 1999; Pynte et al., 1996; Weiland et al., 2014). However, to our knowledge, no study has ever tried to evaluate these theories by their implicitly postulated roles of inhibition, though it has been demonstrated that metaphor comprehension involves inhibitory control (Chiappe & Chiappe, 2007; Columbus et al., 2015; Lai et al., 2015; Rapp et al., 2012; Yang, 2014; Yoon et al., 2021). Several studies without explicitly measuring individual differences in inhibitory control, have investigated the inhibition of metaphor-irrelevant literal properties in L1 metaphor comprehension. Gernsbacher et al. (2001) and Glucksberg et al. (2001) using sentence-verification paradigms found that validation of literal target sentences took longer after reading metaphorical prime sentences. Rubio Fernández (2007) using a cross-modal lexical priming task, observed that both metaphor-relevant and irrelevant properties were active up to 400 ms from the offset of the prime, but after 1000 ms, only the relevant properties remained active. Using a metaphor-induced lexical forgetting paradigm, George & Wiley (2016, 2019) found that reading metaphors (e.g., *The lawyer for the defense is a shark*) led to reduced memory for previously studied literal associates of the metaphoric vehicles (e.g., *SHARK–swim*), and this inhibition of literal properties depended on metaphor familiarity, i.e., only happened for reading unfamiliar metaphors.

Although not discussed by the authors, we think the results of these studies shed light on the above theories. The inhibition of metaphor-irrelevant properties in Gernsbacher et al. (2001) and Glucksberg et al. (2001) seems explicable by both the standard pragmatic view and graded salience view, but the latter serves better accounting for George and Wiley (2016, 2019)'s finding that reading familiar metaphors didn't lead to inhibition of literal properties as did reading novel ones. The graded salience view can explain that the literal properties of familiar metaphors will not interfere with constructing the appropriate contextual interpretation as much as for novel metaphors. Therefore, they may not get inhibited. Also, Rubio Fernández (2007)'s finding that with prior paragraph contexts both metaphor-relevant and irrelevant properties got activated early on, opposes the standard pragmatic view and the direct access view, but lends support to the graded salience view. To further test these theories, however, we still need experiments with prior context manipulation and direct measurement of inhibition involvement. We plan to use the classic colour-word Stroop task as our inhibitory control measurement because it requires inhibiting a predominant response in favour of a contextually adaptive response, highly similar to the presumed process during (at least some) metaphor comprehension.

1.2. L2 metaphor comprehension and the role of inhibition

According to some estimates, more than half of the world's population is multilingual (Bialystok, 2017), but metaphorical language competence is generally not considered a core skill in L2 learning and teaching (Littlemore & Low, 2006). Hence, even highly proficient L2 speakers struggle to understand and use metaphorical expressions effectively in their L2, find these difficult even if they contain familiar words, and are often unaware of their misinterpretations (Littlemore et al., 2011). This pedagogical urgency only punctuates the regrettable paucity of comparable research into L2 metaphor processing in contrast to the abundance of L1 studies, though recent years have seen a growing interest in the mechanisms associated with bilingual figurative language processing.

On the one hand, metaphor understanding seems to present more challenges in L2 compared to L1 (e.g., Chen et al., 2013; Citron et al., 2020; Jankowiak et al., 2017; Mashal et al., 2015). Jankowiak et al. (2017) observed that even familiar metaphors were processed as novel in L2 as reflected by similarly reduced late positive component (LPC) amplitudes. Mashal et al. (2015) found that bilinguals demonstrated a left hemisphere advantage for processing conventional metaphoric expressions in L1, but a right hemisphere advantage for processing the same metaphors in L2,

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

mirroring how monolinguals process unfamiliar metaphors (Bohrn et al., 2012). In the case of L2 idioms, while L1 speakers could directly access the holistic idiomatic meaning, even proficient L2 speakers relied on the L1 conceptual system (Carrol & Conklin, 2014; Kecskés, 2000; Senaldi & Titone, 2022) and prioritized the verbatim literal meaning even when they knew its idiomatic meaning and the idiom was embedded in a figurative context (Cieślicka, 2006; Cieślicka & Heredia, 2011; Siyanova-Chanturia et al., 2011). The prioritization of literal meaning during L2 figurative comprehension has been encapsulated as the Literal-Salience Model by Cieślicka (2006), which is a specification of the graded salience view.

However, there is evidence that metaphor processing in L2 is not fundamentally more difficult than in L1 as L2 dominance or proficiency grows. Segal and Gollan (2018)'s balanced bilinguals showed the same left hemisphere advantage as natives in metaphor processing, in contrast to the right hemisphere advantage manifested by Mashal et al. (2015)'s unbalanced bilinguals. Vaid et al. (2015) found that L2-dominant and balanced bilinguals showed more than twice the amount of Metaphor Interference Effect in L2 than monolinguals, but this automatic activation of metaphorical meanings didn't appear when the L2 of the bilinguals was subdominant to L1. Heredia and Cieslicka (2016) also found that although bilinguals were more likely to access the literal meaning of metaphor first if their L2 was subdominant, they had early access to both the literal and figurative meanings if their L2 was dominant or at least balanced with L1. Research on L2 formulaic expressions has also suggested more native-like processing as L2 proficiency develops (Carrol et al., 2016; Cieślicka et al., 2014). This modulation of literal versus figurative meaning activation by L2 dominance or proficiency fits well with the graded salience view as in L1. Notably, most L2 metaphor studies did not explicitly manipulate context to test the direct access view. Supportive situational context has been found to boost L2 idiomatic language processing (Cieślicka et al., 2014; Cieślicka & Heredia, 2011), but it remains to be seen whether, with a strong prior context, bilinguals can directly tap the figurative meaning of an L2 metaphor without accessing the literal.

On the other hand, as language is deeply rooted in culture (Hadley, 1997), bilinguals' L1 cultural knowledge may hinder their comprehension of culturally incongruent L2 metaphors. Conceptual metaphor theory (Lakoff & Johnson, 2003) claims that metaphorical linguistic expressions manifest underlying conceptual patterns in thought that people use to conceive of the world, some of which inevitably vary across linguistic cultures due to different experiences, cognitive preferences and styles (Kövecses, 2005a, 2005b). From the viewpoint of embodied cognition, metaphor emerges from the interaction between body and culture (Yu, 2008). Primary metaphors derived directly from common human experience are more likely to be universal, whereas complex metaphors as combinations of primary metaphors and cultural beliefs tend to be culture-specific. Deignan et al. (1997) in a comparative analysis of conceptual metaphors across Polish and English identified four types of variation, one of which is the expression with equivalent linguistic forms, i.e., the same literal meaning across languages, but different conceptual metaphorical meanings. This kind of cross-cultural variation presented the largest difficulty for Malay-speaking learners of English body-part metaphors in Charteris-Black (2002), who resorted to the L1 conceptual basis, leading to misunderstanding. Türker (2016) also found L2 speakers had a better comprehension of those L2 metaphors with corresponding metaphorical interpretations in L1. During L2 formulaic processing, incongruent L2 formulaic expressions that have no L1 counterparts and can only be rephrased literally in L1 were found to suffer processing difficulty (Bortfeld, 2003; Carrol et al., 2016; Carrol & Conklin, 2014; Laufer, 2000; Titone et al., 2015; Wolter & Gyllstad, 2011, 2013). Besides the few scanty studies mentioned above, however, the effect of cross-cultural conceptual differences in L2 metaphor processing has rarely been investigated.

Based on existing literature on L2 metaphor processing, it seems that the literal meaning always gets activated, possibly requiring inhibition. In the case of culturally incongruent L2 metaphors where equivalent linguistic forms (i.e., the same literal meaning) correspond to different conceptual metaphorical meanings across languages, the presence of a conflicting L1 metaphorical meaning besides the literal one may demand even more inhibition. To our knowledge, only one study to date has examined the role of inhibitory control in L2 metaphor processing (Lü et al., 2019). They found participants with higher inhibitory control ability made significantly faster metaphorical judgements, but only for familiar L2 metaphors. While the theoretical views introduced above implicate different involvement of inhibition during metaphor processing, none of them have explicitly accommodated individual differences in inhibition. The Predictive Processing (PP) framework (Clark, 2016; Friston, 2010; Hohwy, 2013) can offer a more comprehensive theory for metaphor comprehension.

PP proposes that the brain is in a continuous effort to minimize bottom-up prediction error, which is part of incoming sensory evidence that remains unexplained by top-down predictions generated based on prior world knowledge and contextual expectations by a hierarchical model. Perception, cognition, and action are realized by minimization of prediction errors at multiple levels of the hierarchy across time. Prediction error minimization is a specific application of the free energy principle, according to which "any self-organizing system that is at equilibrium with its environment must minimize its free energy" (Friston, 2010). There are two complementary and mutually constraining ways to minimize prediction error. In perceptual inference, top-down predictions of sensory states are modified as a result of prediction error. For example, N400 and P600 have been respectively associated with semantic prediction error and syntactical and pragmatic prediction error pushing for an update of predictions (Fitz & Chang, 2019; Kuperberg & Jaeger, 2016; Rabovsky et al., 2018). In active inference, the causes of available sensory input are changed by action to bring about predicted sensory states. Saccadic eye movement has been argued as a ubiquitous way to minimize prediction error by actively selecting the array of available visual stimuli (Clark, 2016; Friston et al., 2012; Hohwy, 2013). Top-down predictions lead to saccades towards locations that promise to bring about a reduction of prediction error, including parts of linguistic stimulus that are either infrequent or unexpected given prior context and world knowledge.

PP has been applied to both lower-level language processes (e.g., Lupyan & Clark, 2015; Ylinen et al., 2017) and higher-level ones (Fabry, 2021; Fabry & Kukkonen, 2019; Lewis & Bastiaansen, 2015; Olkoniemi et al., 2022). Thinking about L2 metaphor comprehension in terms of prediction error minimization (Olkoniemi et al., 2022), the figurativeness of metaphors was supposed to trigger larger prediction errors than literal expressions, thus requiring longer reading time. Higher L2 proficiency and metaphor familiarity allowed better prediction, produced less prediction error, and correspondingly shortened reading time. From the perspective of PP, the Stroop incongruency has been conceptualized as a prediction error during the interplay between predictions originating from the high level of the hierarchical predictive model reflecting

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

long-term memory, and the stimulus and task structures at hand (Jiang et al., 2014; Wallentin et al., 2015). Thus, inhibitory ability as reflected by the Stroop task can be regarded as the efficiency in resolving prediction errors through more context-adaptive behaviour. Thus, individuals with higher inhibitory control can minimize prediction errors induced by metaphor more effectively.

1.3. Overview of the present study

To sum up, previous research has found that metaphor comprehension is generally more challenging in L2 than in L1, possibly because of the prioritization of literal meanings, though this difficulty seems to diminish as L2 proficiency grows. However, the effect of cross-cultural conceptual differences and the role of individual differences in inhibitory control during L2 metaphor processing have barely been investigated. Inhibitory control has proved important during L1 metaphor processing because the literal meaning may need to be suppressed, and presumably, it will figure more prominently during L2 metaphor comprehension, especially for culturally incongruent ones. Previous metaphor research has offered competing theories including the direct access view, the literal-salience model, and the graded salience view, all with different implications for the role of inhibition during L2 metaphor processing. However, these theories have not explicitly accounted for how individual differences in inhibition affect metaphor comprehension, while PP can do so. In the present study, through three experiments, we aim to nail down the role of individual differences in inhibitory control during comprehension of both culturally incongruent and culturally congruent L2 metaphors, based on which we will evaluate and extend existing accounts to develop a more comprehensive theory for L2 metaphor comprehension.

Specifically, in Experiment 1, we explored whether the comprehension of culturally incongruent L2 metaphors would involve the inhibition of conflicting L1 metaphorical meanings by using a metaphor-induced lexical forgetting paradigm (George & Wiley, 2016, 2019). In Experiment 2, we tried to figure out whether the presence of context would affect the involvement of inhibitory control during the interpretation of both culturally congruent and culturally incongruent L2 metaphors, in order to test present theories. In Experiment 3, we used eye-tracking to determine the potentially different time course of inhibition involvement during reading culturally congruent and culturally incongruent L2 metaphorical sentences, compared with literal ones.

2. Experiment 1

Experiment 1 extended the metaphor-induced lexical forgetting paradigm (George & Wiley, 2016, 2019) to test whether the comprehension of culturally incongruent L2 metaphors would involve the inhibition of conflicting L1 metaphorical meanings. George and Wiley (2016) found that reading metaphor sentences (e.g., *The lawyer for the defense is a shark*) led to reduced memory for previously studied literal associates of the metaphoric vehicles (e.g., *SHARK–swim*, where *SHARK* is the metaphoric vehicle and *swim* is its literal associate), but only for novel metaphors. This was interpreted as the inhibition of literal properties. A subsequent study (George & Wiley, 2019) provided further support for the inhibitory account for this forgetting effect over a blocking or cue-based interference account.

We reasoned that during the comprehension of culturally incongruent L2 metaphors, in addition to literal meanings, L1 metaphorical meanings would also get activated and then inhibited. We asked participants to learn word pairs where the cues were potential L2 metaphoric vehicles and the targets were associates of their culturally incongruent L1 metaphoric meanings (e.g., *dragon–success*). Then, participants read sentences containing half the cues used in their L2 metaphorical meanings (e.g., *My boss is a dragon because he often punishes us severely for small things.*). We predicted that subsequent forgetting of the L1 metaphorical associates in the final cued-recall test would be greater if the corresponding cues had appeared in L2 metaphorical sentences as the vehicles, manifesting as longer reaction times and lower recall accuracy. This could reflect the inhibition of L1 metaphorical meanings when processing L2 metaphorical meanings.

2.1. Method

Participants

Participants were 38 Chinese students (28 female) recruited from Beijing Normal University between the ages of 18–25 (M = 20.21, SD = 1.33). They were all native Chinese speakers learning English as L2 for at least 8 years. They had normal or corrected-to-normal vision and had no neurological or psychiatric disorder. They signed the written informed consent and got monetary compensation for participation. This study was approved by the ethics committee of Beijing Normal University.

The L2 proficiency requirement during recruitment was to have scored higher than 550 in CET-6 (College English Test Band 6) or 70 in TEM-4 (Test for English Majors-Band 4). Both tests are English examinations for college students conducted by the Ministry of Education of China. The CET-6 is for non-English majors and assesses listening, reading, and writing performance. The maximum score is 710, and the passing score is 425. The TEM-4 is for English majors and assesses listening, writing, grammar, vocabulary, and reading performance. The maximum score is 100, and the passing score is 60. Then participants were asked to self-rate their English proficiency (1 = low proficiency, 5 = high proficiency) in listening, speaking, reading, and writing on a 5-point scale. Furthermore, the Oxford Quick Placement Test (OQPT, 2001) was used to provide an objective assessment of English proficiency. It takes approximately 15 min and consists of 60 multiple-choice questions with increasing levels of difficulty, including discrete multiple-choice questions and multiple-choice cloze questions. The mean OPT score of our participants was 41.33 ± 2.04, which placed them at the upper intermediate level. Supplementary Table A1 in the Supplementary Materials shows participants' information on age, age of English acquisition, English proficiency ratings, and OQPT scores.

Materials

Based on the categorization of L2 metaphors by Deignan et al. (1997) and Charteris-Black (2002), we defined culturally congruent L2 metaphors as those with both equivalent linguistic expressions and metaphorical meanings across L1 and L2, while culturally incongruent L2 metaphors as those with equivalent linguistic forms, i.e., the same literal meaning, but different metaphorical meanings across L1 and L2. We chose 30 culturally incongruent metaphor vehicles from A Usage Dictionary of English and Chinese Conceptual Metaphors (Su, 2009), which is a Chinese version of the Collins CoBuild English Guides 7: Metaphor (Deignan, 1996). Thirty English metaphorical sentences were created, in which each vehicle was used in its L2 metaphorical sense followed by an explanatory clause to make sure participants correctly process the L2 metaphorical meaning (e.g., *My boss is a dragon because he often*

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

punishes us severely for small things.). For each metaphorical sentence, a statement was created to which participants had to judge true/false (e.g., *My boss often punishes us.*). Then 30 cue-target word pairs were created, in which the cues were the metaphor vehicles of the sentences, and the targets were associates of their L1 metaphorical meanings but not related to the L2 metaphorical meanings (e.g., *dragon–success*). All stimuli are presented in Table A2.

We asked 30 participants not involved in the formal experiment to judge the semantic relatedness between the cue's L1 metaphorical meaning and the target for each word pair (1 for highly unrelated, 5 for highly related). To assess the rate of recall of these word pairs, we conducted a norming experiment with another 14 participants who would not enter the formal experiment. They studied the 30 cue-target pairs on a sheet of paper for 5 min, followed by cue plus first-letter-stem recall on a sheet of paper for 5 min. The average percentage recall for these participants was 97.62% ± 3.91%. For each word pair, the average percentage recall was calculated across participants. These word pairs with corresponding metaphorical sentences were then split into two lists of 15 items that matched the targets on the base rate of recall, word frequency, number of letters, and number of syllables. Semantic relatedness of word pairs and length of metaphorical sentences were also matched (all $|t|$s < 1). The two lists were used to counterbalance materials across the two conditions, with equal numbers of participants in each condition.

Procedure

There were four phases for the experiment: study, initial cued-recall, metaphor comprehension, and final cued-recall. In the study phase, participants were given the 30 cue-target word pairs on a sheet of paper and instructed to study each word pair so that when given the cue word, they could recall the target. The order of items was randomized for each participant. They had 5 min to study this list before it was collected.

Immediately following the study period, participants received a cued-recall test of all 30 cue-target pairs via E-Prime 2.0 (Psychology Software Tools, Schneider et al., 2012). This test was presented using a PC (Lenovo) with an LCD monitor (Philips 220V4, 22 inches, screen area: 1440 by 900 pixels; refresh rate: 16.67 ms). Participants sat in a comfortable chair with a viewing distance of 80 cm in a well-lit, sound-attenuated room. Stimuli were centred as black words on a grey background in lowercase Times New Roman with a font size of 40 points. These basic experiment set-ups were the same throughout our study unless stated otherwise. The order of items was pseudo-randomized for each participant. Each trial began with the presentation of a fixation cross for 500 ms, followed by the presentation of the cue word, along with the first letter of the target before a blank response box, for up to 6 s. Participants were instructed to fill in the target by typing into the response box as quickly as possible. Recall RT was calculated from the cue onset to the first keystroke response. After 5 s or after participants pressed the enter key, the next trial began. We familiarized participants with 3 example trials before the formal cued-recall test.

Following the initial recall, the metaphor comprehension task was presented also using E-Prime 2.0. Participants were presented with one of the two lists of 15 metaphorical sentences in which the vehicle was a cue from the previously studied word pairs (metaphor condition). The other half of the cue words did not appear in any sentences (no-metaphor condition). The sentences were presented in pseudo-random order one at a time in 20-point Times New Roman font. Each trial began with the presentation of a fixation cross for 500 ms, followed by the presentation of a sentence at the centre of the screen for up to 20 s. Participants were instructed to press the space bar after they finished reading. Then, to make sure participants process the L2 metaphor, a statement related to the previous metaphorical sentence appeared on the screen to which they had to judge true or false with keys F/J counterbalanced.

Following the metaphor comprehension task, participants received a final cued-recall test of all 30 cue-target pairs in the same way as the first cued-recall test.

2.2. Results

All participants judged the metaphor-related statements with an accuracy above 90%. Leniency was given to spelling errors and plurals in the cued-recall tests. One participant was excluded for low recall accuracy (0.5) on the second test. Recall time was log-transformed due to positive skewness. Average recall accuracies and correct RTs under metaphor condition and non-metaphor condition for the two cued-recall tests were presented in Table A3. Data analyses were conducted with the lme4 package (Bates et al., 2015) in R (Version 4.0.1; R Core Team, 2020). For two recall tests collectively and separately, linear mixed-effects models (LMM) were built for RTs, and generalized linear mixed-effects models (GLMM) with a binomial link function were built for accuracy. Condition (metaphor versus no-metaphor) was fitted to each model as a treatment-contrast coded fixed effect variable, while recall time (first versus second) and its interaction with condition were also fitted to the grand models including two recalls. Random effects were modelled as supported by the data and determined using backward selection (Matuschek et al., 2017). That is, we started with the maximal random structure allowing convergence, and reduced the random structure until a further reduction would imply a significant loss in the goodness-of-fit (χ^2 test criterion: $\alpha = .2$). Two-tailed probability values and degrees of freedom associated with each statistic were determined using the Satterthwaite approximation implemented in lmerTest (Kuznetsova et al., 2017).

Neither the model for initial recall accuracy nor initial recall RT revealed any significant effect of condition ($p = .26$ for the former, $p = .66$ for the latter). Condition still did not manifest any significant effect in the models for either final recall accuracy ($p = .92$) or final recall RT ($p = .13$). However, in the grand model for recall RT, the interaction between recall time and condition approached significance ($p = .06$, please see Figure 1), and the main effect of recall time was also significant ($p < .001$). This suggested that targets whose cue words had been read in their L2 metaphorical meanings in sentences did not enjoy as much shortening of recall time on the second test as others. The grand model for recall accuracy did not reveal any significant effects.

2.3. Discussion

Compared with cue words not read in the metaphor comprehension task, those used in their L2 metaphorical meanings in sentences led to slower recall of target words semantically associated with their L1 metaphorical meanings. This confirmed our expectation, suggested that comprehension of culturally incongruent L2 metaphors did involve the inhibition of conflicting L1 metaphorical meanings, and extended previous findings about inhibition of irrelevant literal information in L1 metaphor processing (George & Wiley, 2016, 2019; Gernsbacher et al., 2001; Glucksberg et al., 2001; Rubio Fernández, 2007).

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

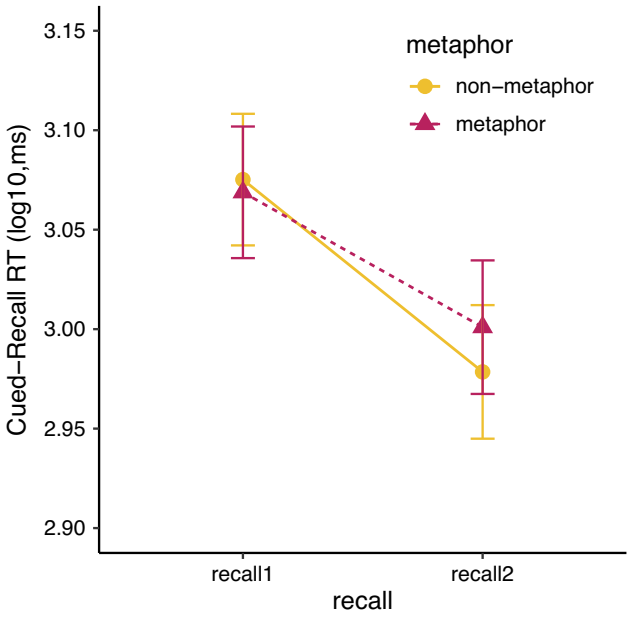


Figure 1. Interaction between recall time and condition in Experiment 1.
Notes: This graph shows the estimated effects of recall timepoint on the RTs in two conditions based on the grand linear mixed – effects model. Error bars represent the standard error of the mean.

Both the literal-salience model and the graded salience view could be extended to account for the results. For an extension of the literal-salience model, it would be that after the initial prioritization of literal meaning, the effortful figurative computation, unfortunately, goes further astray towards the L1 metaphorical meaning which doesn't fit with the context, and requires inhibition during the final computation of the L2 metaphorical meaning. For an extension of the graded salience view, it seems that the L1 metaphorical meaning of the culturally incongruent L2 metaphor enjoys higher salience than the L2 metaphorical meaning, interferes with constructing the appropriate contextual interpretation, and thus gets inhibited. However, as we did not explicitly manipulate context to see whether a strong prior context would allow the L2 metaphorical meaning to be precisely tapped without activation and inhibition of the L1 metaphorical meaning, we could not refute the direct access view here. This remained to be tackled in experiment 2. Besides, inhibition's involvement was indirectly inferred from the slower cued recall of target words after reading L2 metaphors. In the following experiments, we directly tested the relationship between individual inhibitory control and performance in L2 metaphor comprehension.

3. Experiment 2

In experiment 2, we tried to evaluate and extend existing accounts by their implicitly postulated roles of inhibition during well-contextualized L2 metaphor processing. Specifically, we tested whether the presence of context would affect the involvement of inhibitory control during the processing of both culturally congruent and culturally incongruent L2 metaphors.

We asked participants to read and interpret target sentences including literal sentences, culturally congruent metaphorical sentences, and culturally incongruent metaphorical sentences. Participants were then presented with explanatory sentences to contextualize former target sentences and asked to interpret them again. Comprehension time and interpretation quality of target sentences before and after explanatory contexts were examined. Individual differences in inhibitory control were measured by the Stroop task. It would be used as a predictive variable when modelling (re)comprehension time and (re)interpretation quality of target sentences to reflect inhibition involvement.

The direct access view predicted that after the explanatory contexts were provided, processing of neither kind of metaphor would require more inhibitory control than literals because suitable L2 metaphorical meanings would already be precisely activated. Statistically, this implied that when modelling the recomprehension time and reinterpretation quality, the two treatment-contrast coded variables contrasting two kinds of metaphor processing with literal processing (see the result section) would not interact with the inhibitory control variable. In contrast, the extended literal-salience model predicted that metaphor comprehension especially culturally incongruent one would always require more inhibition than literal processing, even after explanatory contexts. Statistically, this implied that when modelling both the first and the second comprehension time and interpretation quality, the two treatment-contrast coded variables contrasting two kinds of metaphor processing with literal processing (see the result section) would interact with the inhibitory control variable. For culturally congruent metaphors, the extended graded salience view predicted that whether their processing would need more inhibition than literal expressions would depend on whether their literal meanings were salient, irrespective of context. This remained to be seen. For culturally incongruent metaphors, based on findings from experiment 1, this view predicted that both before and after contexts were provided, more inhibition would be needed than literal processing, at least to suppress L1 metaphorical meanings. Statistically, this implied that when modelling both the first and the second comprehension time and interpretation quality, the treatment-contrast coded variable contrasting culturally incongruent metaphor processing with literal processing would interact with the inhibitory control variable.

3.1. Method

Participants

Participants were 34 Chinese students (26 female) recruited from Beijing Normal University between the ages of 18–23 (M = 19.78, SD = 1.05). None of them participated in Experiment 1. All the recruitment standards and participation requirements were the same as in Experiment 1. They signed the written informed consent and got monetary compensation for participation. This study was approved by the ethics committee of Beijing Normal University. The mean OPT score of these participants was 41.45 ± 2.14, which placed them at the upper intermediate level. Table A4 shows participants' information on age, age of English acquisition, English proficiency ratings, and OQPT scores.

Materials

There were 3 conditions for target sentences, i.e., literal sentences, culturally congruent metaphors, and culturally incongruent metaphors. Examples are given in Table 1. Each condition initially contained 30 potential items. They were mostly in the syntactic form of "A be B" where B referred to a noun, with a few exceptions of prepositional phrases or adjectives. The definitions of culturally congruent and culturally incongruent metaphors were the same as in experiment 1. Culturally incongruent metaphors were mostly taken from experiment 1. Culturally congruent metaphors and literal sentences were taken from previous research (Chiappe & Chiappe, 2007; Glucksberg et al., 1997). Metaphor vehicles were

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

Table 1. Examples of target sentences and explanatory sentences in Experiment 2

Conditions	Target sentences	Explanatory sentences
Culturally congruent metaphor	Love is a journey.	We will encounter different landscapes in love, sometimes smooth, and sometimes not.
Culturally incongruent metaphor	My boss is a dragon.	He often gets angry at the slightest thing and punishes us severely.
Literal sentence	His teacher is an expert.	His teacher has made outstanding contributions in his specialization and has strong communication skills.

4-11 letters in length and were high-frequency words (van Heuven et al., 2014). To make sure our manipulation of metaphoricity was effective, 30 judges with similar backgrounds as formal participants were presented with 90 target sentences and asked to decide whether each one was literally plausible, metaphorically plausible, or not plausible. Only expressions rated by at least 80% of the judges as either metaphorically or literally plausible would be selected as formal materials (Gold et al., 2010; Goldstein et al., 2012). Another group of 30 students with similar backgrounds as formal participants were asked to rate each target sentence's familiarity (1 = very unfamiliar, 5 = very familiar) and difficulty (1 = very difficult, 5 = very easy). We finally chose 20 formal target sentences out of 30 for each condition. They were 3–6 words in length without significant difference across conditions. Despite our effort to control, there remained significant differences in familiarity ($F(2, 57) = 28.41$, $p < .001$) and difficulty ($F(2, 57) = 16.64$, $P < .001$) across conditions. Thus, familiarity and difficulty would be added as covariates during statistical analyses.

For each target sentence, we compiled an explanatory sentence to be used as a context. Examples are given in Table 1. To make sure these explanatory sentences would provide effective contexts, 20 students with similar backgrounds as formal participants were asked to rate whether each explanatory sentence could help with understanding the target sentence, on a scale of 5 (1 = is not helpful at all, 5 = is very helpful). Averaged ratings for every sentence achieved above 4. Explanatory sentences were 10–19 words in length, without significant differences across conditions. All stimuli are presented in Table A5.

Procedure

Sentence Interpretation task. Stimulus presentation was based on E-Prime 2.0 (Psychology Software Tools, Schneider et al., 2012) with the same experimental setup as experiment 1. Sixty target sentences were pseudo-randomized for each participant and separated into 4 blocks balanced across conditions. All the sentences appeared at the centre of the screen in 20-point Times New Roman font. Each trial began with a fixation cross in the centre of the screen for 500 ms. Then a target sentence appeared, which participants were required to press the spacebar as soon as they comprehended. Reaction time was calculated from target onset and named as the Time for First Comprehension of The Target Sentence. Then an input box appeared, into which participants were instructed to type down their best interpretation in Chinese as soon as possible and then press the enter key. After a blank screen for 1000 ms, an explanatory sentence appeared, based on which participants were asked to re-comprehend the previous target sentence and then press the spacebar as soon as possible. Reaction time was calculated from explanatory sentence onset and named as the Total Time for Second Comprehension. Then an input box appeared again, and participants were to type down their reinterpretation of the target sentence in Chinese as soon as possible, and then press the enter key to trigger the next trial. There were 10 practice trials before the formal experiment.

The Total Time for Second Comprehension included both the time for comprehending the explanatory sentence and the time for re-comprehending the target sentence. To dissociate them, we conducted an adapted version of the above sentence interpretation task in an additional group of 30 participants with similar backgrounds as those in the formal experiment. The main difference with the previous version was that these participants were not asked to re-comprehend and reinterpret target sentences. Each trial began with a fixation cross in the centre of the screen for 500 ms. Then a target sentence appeared, which participants were required to press the spacebar as soon as they comprehended. After a blank screen for 1000 ms, an explanatory sentence appeared, in which participants were simply required to press the spacebar as soon as they comprehended. Reaction time was calculated from the explanatory sentence onset and named as the time for comprehending the explanatory sentence. For 25% of all trials, a statement related to the explanatory sentence followed, which participants were to judge as true or false. This was to make sure of participants' involvement. There were also 10 practice trials before this task. For each explanatory sentence, comprehension time averaged across these participants would be used as a covariate when modelling the Total Time for Second Comprehension.

Assessment of individual inhibitory control. Individual differences in inhibitory control were measured with a Stroop task (Heidlmayr et al., 2014; Lü et al., 2019; Qiu et al., 2006). There were three conditions: neutral, congruent, and incongruent. The congruent stimuli consisted of three Chinese characters for colours 红, 黄, 蓝 (meaning "red, yellow, blue") presented in the same colors as the meanings of the characters themselves (e.g., 红 "red" presented in red ink). The incongruent stimuli consisted of the same three characters presented in the two colors not matching their meanings (e.g., 蓝 "blue" presented in red or yellow ink). In the neutral condition, non-colour characters 笔, 球, 表 (meaning "pen, ball, watch") were presented in each one of the three colors. There was a total of 120 trials, including 90 congruent trials, 15 incongruent trials, and 15 neutral trials (Chiappe & Chiappe, 2007). Individual inhibitory control was measured by subtracting the neutral condition from the incongruent condition (i.e., the Inhibition effect, Coderre et al., 2011; Heidlmayr et al., 2014; Lu et al., 2017). The smaller the inhibition effect, the greater the inhibitory control ability of the participant. The order of presentation was pseudo-randomized to avoid characters of the same experimental condition appearing four times in succession and to avoid the same character or the same colour appearing consecutively. The pseudo-randomized order of trials was created using the program Conan (Nowagk, 1998).

Stimulus presentation was based on E-Prime 2.0 (Psychology Software Tools, Schneider et al., 2012) with the same experimental setup as experiment 1. Participants were asked to judge the colour of the stimulus character as quickly and accurately as possible by pressing the number key 1 for red, 2 for yellow, and 3 for blue, with the index, middle, and ring fingers of their dominant hand. They were instructed to rest their fingers on the keyboard during the task. For each trial, a fixation cross was first presented in the centre of the

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

screen for 500–1000 ms. Then a stimulus character, in SimSun font and 36 font size, was presented against a black background in the centre of the screen. It stayed until one of the three colour response keys was pressed or for 1500 ms maximally if no key was pressed. Finally, a blank screen was presented for 500 ms. Response time (RT) was defined as the interval between the stimulus onset and pressing a response key. To allow the participants to learn the colour-key correspondences, a practice session was presented before the formal task. It consisted of 24 trials, 3 in neutral conditions, 3 in incongruent conditions, and 18 in congruent conditions.

3.2. Results

We examined the time for first comprehension and re-comprehension of target sentences, and the qualities of their first interpretation and reinterpretation, to explore the role of inhibitory control during the processing of both culturally congruent and culturally incongruent L2 metaphors, before and after context. For the time for First Comprehension of Target Sentences and the Total Time for Second Comprehension, trials with RT less than 200 ms or more than 3 standard deviations from individual means were excluded (both <1%), then both RTs were log-transformed due to positive skewness. The qualities of first interpretation and reinterpretation of target sentences were judged by 3 doctors of English on a 3-point scale. For example, for the metaphor "Some divorces are earthquakes.", a detailed and precise interpretation like "Divorces can destroy parts of life and bring about plenty of changes" (in Chinese) would be given 2 points, a related but vague interpretation like "Divorces have negative consequences" would be given 1 point, while a tangent one like "Some marriage are failures", wrong one or empty response would get no point. Scorers did not know these interpretations included first interpretations and reinterpretations from the same participants, i.e., they judged each interpretation independently. Interrater reliability as reflected by Cronbach's alpha coefficient was .92, higher than the required .7 (Stemler, 2004). Scores for first and second interpretations for each item were averaged over 3 raters. Basic statistics for comprehension RTs and interpretation scores are displayed in Table A6. A paired t-test revealed that scores for reinterpretation were significantly higher than that for first interpretation ($p < .001$), attesting to the effectiveness of the explanatory context.

Data analyses were conducted with the lme4 package (Bates et al., 2015) in R (Version 4.0.1; R Core Team, 2020). Linear mixed-effects models (LMM) were built for the Time for First Comprehension of Target Sentences, the Total Time for Second Comprehension, and scores of first interpretation and reinterpretation of target sentences, separately. We used two kinds of treatment-contrast coding schemes. The first used two contrast variables, Congruent Metaphor, and Incongruent Metaphor, to contrast the processing of culturally congruent and culturally incongruent L2 metaphors with literal sentences (baseline), respectively. The second scheme set culturally congruent L2 metaphors as the baseline and used Literal and Metaphor to contrast literal and culturally incongruent L2 metaphors with the congruent ones, respectively. We only report the model results of the second coding scheme if there were significant effects concerning Metaphors, i.e., the difference between congruent versus incongruent metaphors, because all the other model information was the same across the two schemes. For each model, core fixed effect variables included two contrast variables, Inhibitory Control, and their interactions. Covariates included familiarity and difficulty of target sentences. Random effects were modelled as supported by the data and determined using backward selection (Matuschek et al., 2017). Two-tailed probability values and degrees of freedom associated with each statistic were determined using the Satterthwaite approximation implemented in lmerTest (Kuznetsova et al., 2017).

In the model for the time for first comprehension of target sentences based on the first coding scheme (Table A7), the main effects of Inhibitory Control ($p = .007$) and Congruent Metaphor ($p = .012$) and their interaction ($p = .041$) were significant, suggesting that lower inhibitory control (larger Stroop effect) took a heavier toll on culturally congruent metaphors than literal sentences concerning first comprehension time. When using the second coding scheme, we observed a significant interaction between Inhibitory Control and Metaphor ($p = .01$), suggesting that lower inhibitory control took a heavier toll on culturally congruent metaphors than incongruent ones. Please see Figure 2 for a demonstration of these interactions.

The model for scores of first interpretation (Table A8) revealed a marginally significant interaction between Incongruent Metaphor and Inhibitory Control ($p = .06$), suggesting that worse inhibitory control had a more negative influence on culturally incongruent metaphors than literal sentences regarding first interpretation quality. None of the other concerned variables reached significance.

The Total time for Second Comprehension included both the time for comprehending the explanatory sentence and the time for re-comprehending the target sentence. As we were concerned with the time for re-comprehending the target sentence and how the core variables affected it, we added the time for explanatory sentence comprehension as a covariate when modelling the Total Time for Second Comprehension. To avoid model convergence failure due to huge scale differences among three covariates, we standardized them and named them Z_explanatory, Z_familiarity, and Z_difficulty. Model results (Table A9) based on the first coding scheme revealed (marginally) significant interactions between Congruent Metaphor and Inhibitory Control ($p = .04$), and between Incongruent Metaphor and Inhibitory Control ($p = .07$), in addition to two significant main effects of Congruent Metaphor ($p < 0.001$) and Incongruent Metaphor ($p < 0.001$). These suggested that worse inhibitory control took a heavier toll on both culturally congruent and culturally incongruent metaphors than literal sentences, leading to longer recomprehension time. Please see Figure 2 for a demonstration of these interactions. There were no significant differences found between the two metaphor conditions based on the second coding scheme.

The model for scores of reinterpretations (Table A10) revealed a significant main effect for Incongruent Metaphor ($p = .04$), suggesting that reinterpretation qualities of culturally incongruent metaphors were better than literal sentences. None of the other concerned variables reached significance.

3.3. Discussion

In experiment 2, we found that compared with literal sentences, better inhibitory control saved more comprehension time for culturally congruent metaphors and improved interpretation quality more for culturally incongruent metaphors, before contexts were provided. Crucially, after explanatory sentences were provided to contextualize understanding, inhibitory control still had a larger impact on recomprehension time for both kinds of metaphors than literal sentences. These results did not support the direct access view, because after the explanatory contexts were provided which supposedly had activated suitable L2 metaphorical meanings, comprehension of both kinds of metaphor still required more inhibition

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

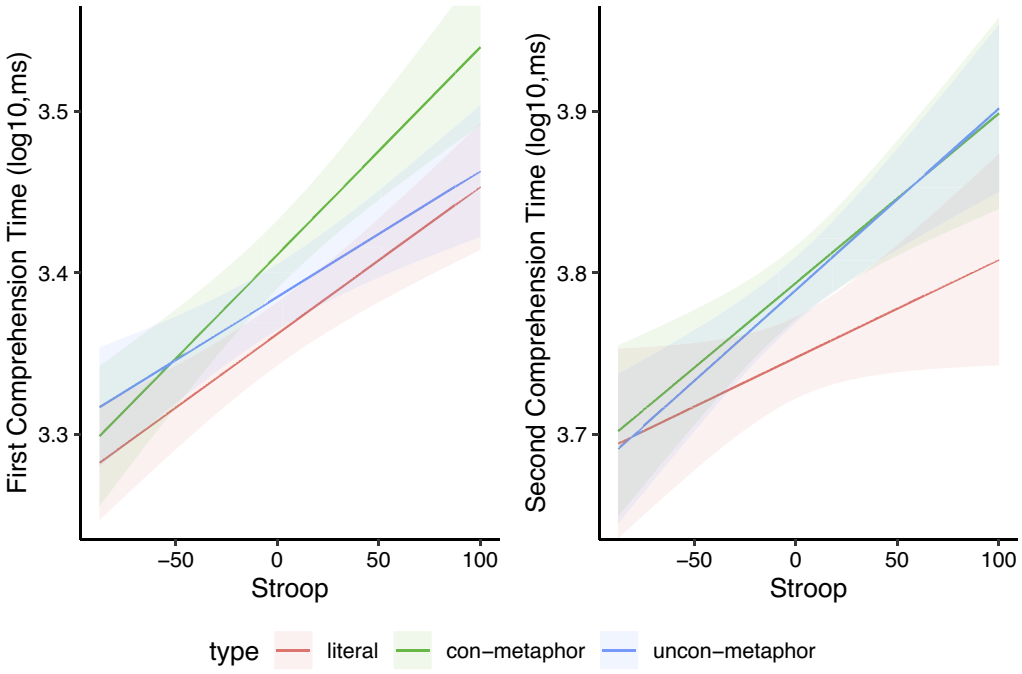


Figure 2. Interactions between Stroop effect and conditions in Experiment 2.
Notes: Stroop effect values were mean-centred, with a higher value indicating worse inhibitory control. This graph shows the estimated effects of the Stroop effect on the first and second comprehension time under three conditions based on their linear mixed-effects models. Shaded areas represent the standard error of the mean.

than literals. However, it was predicted by the extended literal-salience model. According to this model, literal meanings are always prioritized during the processing of culturally congruent metaphors, which demand inhibition. For culturally incongruent metaphors, further activation of conflicting L1 metaphorical meanings requires more inhibition. The extended graded salience view can also accommodate the results by specifying saliency for literal meanings of these metaphors, which is just the literal saliency model.

By measuring individual differences in inhibitory control with the Stroop inhibition effect, this experiment more directly pointed at the role of inhibition during L2 metaphor processing. Though the extended literal-salience model could explain the involvement of inhibition, it did not explicitly account for its individual differences. In contrast, PP has a natural explanation for individual differences. From this perspective, across the L2 situations for our participants with upper intermediate level English, literal language use has gained higher expectations than figurative one. Consequently, the use of a figurative expression results in larger prediction errors than a literal sentence, which then requires a corrective process to form a more suitable interpretation. Individual inhibitory control reflects the efficiency in resolving prediction errors, which explains its effects on comprehension time and interpretation quality.

In addition, we observed that the role of inhibitory control before and after context was related to the cultural congruency of metaphors. For culturally congruent metaphors, both before and after context, low inhibitory control induced more costs on comprehension time than literals, but it did not influence interpretation quality. However, for culturally incongruent metaphors, low inhibitory control took its toll on interpretation quality before context but only affected recomprehension time after context. Counterintuitively, the first comprehension time of culturally incongruent metaphors did not suffer from low inhibitory control as much as culturally congruent ones. A qualitative review of participants' interpretation of culturally incongruent metaphors showed that during the first interpretation, they tended to take the road more travelled by of literal meanings with less inhibitory control under time pressure, saving first comprehension time at the expense of first interpretation quality, especially for individuals with lower inhibitory control. The explanatory contexts helped participants access L2 cultural knowledge, and they revised their answers to cross the cultural bridge during reinterpretation. Eventually, inhibitory control also only affected the recomprehension time of culturally incongruent metaphors as for culturally congruent ones.

Although there was no significant difference in inhibitory control's effect on recomprehension time between 2 metaphor conditions based on the second coding scheme, we believed a more nuanced analysis of online processing during the natural reading would reveal more inhibition involvement for culturally incongruent metaphors than congruent ones during the later stage of processing. This was our aim for experiment 3.

4. Experiment 3

In experiment 3, we used eye-tracking to determine the potentially different time course of inhibition involvement during reading culturally congruent and culturally incongruent L2 metaphorical sentences, compared with literal ones. Metaphorical sentences were constituted by L2 metaphors in the form of "A is B" followed by explanatory clauses. Based on the results of experiments 1 and 2, the extended literal-salience model predicted that inhibition would manifest its importance from the early stage of metaphorical understanding to suppress the activated literal meanings, irrespective of cultural congruency. But inhibition would play a stronger role in the late stage of processing for culturally incongruent metaphorical sentences than congruent ones, especially for their explanatory clauses, where conflicts between L1 and L2 metaphorical meanings

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

would need to be resolved. PP offered the same predictions albeit with a different logic. It predicted that both culturally congruent metaphors and incongruent ones would trigger larger prediction errors than literals early on, which could benefit from better inhibitory control. Top-down predictions would lead to saccades towards locations that promise to bring about a reduction of prediction error as active inference. This meant the eye would linger longer on the explanatory clauses for metaphorical sentences than literal ones, especially for culturally incongruent ones, because the culturally incongruent L2 metaphorical meaning would be less predicted than the L1 metaphorical meaning. Better inhibitory control would allow quicker resolution of this prediction error.

4.1. Method

Participants

Participants were 36 Chinese students (24 female) recruited from Beijing Normal University between the ages of 18–25 (M = 22.19, SD = 1.96). None of them participated in experiments 1 or 2. All the recruitment standards and participation requirements were the same as in Experiment 1. They signed the written informed consent and got monetary compensation for participation. This study was approved by the ethics committee of Beijing Normal University. One participant's Stoop task was omitted, leaving 35 participants for final analysis. The mean OPT score of these participants was 40.28 ± 2.37, which placed them at the upper intermediate level. Table A11 shows participants' information on age, age of English acquisition, English proficiency ratings, and OQPT scores.

Materials

Forty-five target sentences were adapted from the material of experiment 2, 15 sentences for each of the three conditions. The principle was to avoid those with low first interpretation scores to make sure participants knew the meanings. They were constituted by main clauses in the form of "A is B" where B referred to a noun, followed by explanatory clauses. Examples are given in Table 2, where two regions of interest (ROIs) are underlined, i.e., the vehicle of the metaphor and the corresponding part in the literal sentence, and the explanation region (Ashby et al., 2018; Olkoniemi et al., 2022).

Main clauses were 3-6 words in length, and explanatory clauses were 4-11 words in length, with no significant difference across conditions. Word frequency and word length of the objects of main clauses also did not differ across conditions ($F(2, 42) = .64$, $p = .53$; $F(2, 42) = .60$, $p = .55$). Another group of 30 students with similar backgrounds as formal participants rated each target sentence's familiarity (1 = very unfamiliar, 5 = very familiar) and difficulty (1 = very difficult, 5 = very easy). As in experiment 2, they would be added as covariates during statistical analyses. In addition, we created 45 filler sentences with varying syntactic forms to prevent participants from building up the expectancy of syntax during reading. To make sure of participant involvement, each sentence was followed by a related question requiring a true or false judgement (see Table 2). All stimuli were presented in Table A12.

Table 2. Examples of reading materials and questions in Experiment 3

Conditions	Stimulus sentences	Questions
Culturally congruent metaphor	Love is a journey because it is sometimes smooth, but sometimes not.	There are also difficulties in love.
Culturally incongruent metaphor	My boss is a dragon because he often punishes us severely for small things.	My boss often punishes us.
Literal sentence	His teacher is an expert because he is very skilled in his field.	His teacher is not professional.
Filler sentence	Our school is three times larger than my grandma's garden.	Our school is smaller than my garden.

Notes: The underlined regions in each stimulus sentence are two regions of interest (ROIs), i.e., the vehicle of the metaphor and the corresponding part in the literal sentence, and the explanation region. Underlines are used here for illustration and did not appear during the experiment.

Procedure

Right eye movements were recorded monocularly using EyeLink 1000 (SR Research Ltd., Ontario, Canada) at a 1000 Hz sampling frequency. Stimuli were presented on a 19-inch CRT monitor with a refresh rate of 150 Hz and a resolution of 1024 × 768 pixels. Participants were seated 60 cm from the screen, and a chin-and-forehead rest was used to stabilize the head. Participants were first calibrated using a 3-point calibration before the experiment started and calibration accuracy was rechecked by drift correction before reading each sentence. Recalibration was performed after the rests between blocks and whenever necessary during the experiment.

Ninety sentences were pseudo-randomized for each participant and separated into 3 blocks balanced for conditions. Each trial started with a black fixation point appearing on the left of the screen, where the first word of the sentence would appear. Participants were instructed to focus their eyes on the fixation point and the experimenter would click the mouse to trigger the sentence display after proper drift correction. The whole sentence appeared at once as black words on a grey background in 14-point Courier New font, and each letter subtended about 0.3° of visual angle. Participants were instructed to read each sentence at their own pace and press the spacebar on the keyboard to move on. Then a true/false comprehension question should be answered by pressing designated buttons on a keyboard, which would trigger the next trial. There were 6 practice trials before the formal experiment. After the reading task, an assessment of individual inhibitory control was conducted in the same way as in experiment 2.

4.2. Results

All participants answered the comprehension questions with an accuracy above 90%. Predetermined cutoffs were used to trim the data (Rayner, 1998). Fixations shorter than 80 ms and longer than 800 ms were eliminated (2% of fixations). Trials in which there was a blink or track loss during reading were removed before analysis. Finally, trials with any dependent measure 3 standard deviations beyond the individual means were also eliminated (Cong & Chen, 2022). The two resulted in exclusions of <7% of trials.

The critical values measured were gaze duration (also called first-pass reading time, the sum of the duration of all fixations made on the ROI before exiting), regression path duration (also go-past time, the sum of all fixation durations, which starts with the first fixation on the ROI up to the time the eyes fixate to the right of the ROI), and total reading time (the sum of all fixation durations made on the ROI including re-reading). We analysed these three values for the vehicle region of the metaphors and the corresponding part in the literal sentences and the explanation region in both metaphors and literal sentences. All RTs were log-transformed due to

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

positive skewness. For the vehicle region of metaphor, the gaze duration informs about the early stage of metaphorical processing, the regression path duration informs about how well the vehicle can be integrated with the topic before explanation, whereas total reading time reflects the late stage of whole sentence integration. For the explanation region, the gaze duration informs about whether the previous metaphor affects the early stage of explanation processing, whereas the regression path duration and total reading time reflect how well the previous understanding of the metaphor can be integrated with the explicit explanation into a whole. The means and standard deviations of these eye movement measures are presented in Table A13.

Data analyses were conducted with the lme4 package (Bates et al., 2015) in R (Version 4.0.1; R Core Team, 2020). Separate linear mixed-effects models (LMM) were built for each eye movement measure for each region of interest. We used two kinds of treatment-contrast coding schemes. The first used two contrast variables, Congruent Metaphor, and Incongruent Metaphor, to contrast the processing of culturally congruent and culturally incongruent L2 metaphors with literal sentences (baseline), respectively. The second scheme set culturally congruent L2 metaphors as the baseline and used Literal and Metaphor to contrast literal and culturally incongruent L2 metaphors with the congruent ones, respectively. We only report the model results of the second coding scheme if there were significant effects concerning Metaphors, i.e., the difference between congruent versus incongruent metaphors, because all the other model information was the same across the two schemes. For each model, core fixed effect variables included two contrast variables, Inhibitory Control, and their interactions. Covariates included familiarity and difficulty of sentences. Random effects were modelled as supported by the data and determined using backward selection (Matuschek et al., 2017). Two-tailed probability values and degrees of freedom associated with each statistic were determined using the Satterthwaite approximation implemented in lmerTest (Kuznetsova et al., 2017).

The models for gaze duration on the vehicle region did not show any significant effects under either coding scheme, suggesting that the three conditions didn't differ from each other at this early stage and inhibition didn't manifest a special role for metaphors at this time. Results of the model for regression path duration on the vehicle region based on the first coding scheme are presented in Table A14, revealing a marginally significant interaction between Incongruent Metaphor and Inhibitory Control (p = .07). A significant interaction between Metaphor and Inhibitory Control (p = .03) was revealed by the second coding scheme. These suggest that compared with both literal and culturally congruent metaphors, inhibitory control had a larger impact on culturally incongruent metaphors on the integration of the vehicle part with the previous topic before the explanation. Please see Figure 3 (left) for a demonstration of these interactions. Results of the model for total reading time on the vehicle region based on the first coding scheme are presented in Table A15, which revealed a marginally significant interaction between Incongruent Metaphor and Inhibitory Control (p = .06). The interaction between Metaphor and Inhibitory Control in the second coding scheme showed a trend towards significance (p = .09). These suggest that compared with both literals and culturally congruent metaphors, inhibitory control had a larger impact for culturally incongruent metaphor on the late stage of whole sentence integration. Please see Figure 3 (middle) for a demonstration of these interactions.

The models for gaze duration and total reading time on the explanation region did not show any significant effects of interest under either coding scheme. The absence of significant effects in the gaze duration model suggests that the previous metaphor did not affect the early stage of explanation processing. For the late stage of explanation processing however, the model for its

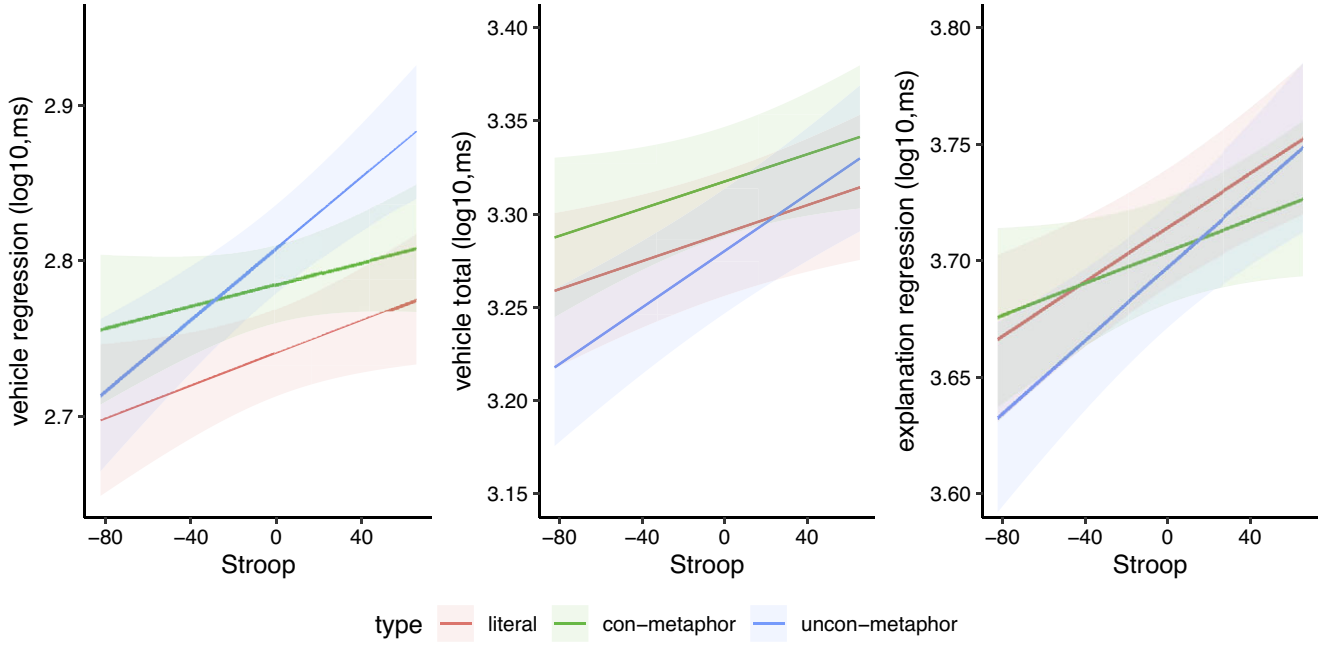


Figure 3. Interactions between Stroop effect and conditions in Experiment 3.
Notes: Stroop effect values were mean-centred, with a higher value indicating worse inhibitory control. This graph shows the estimated effects of the Stroop effect on the regression path duration and the total reading time for the vehicle region and regression path duration for the explanation region under three conditions based on their linear mixed-effects models. Shaded areas represent the standard error of the mean.

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

regression path duration based on the second coding scheme (Table A16) revealed a significant interaction between Metaphor and Inhibitory Control (p = .03). This suggests that compared with culturally congruent metaphors, inhibitory control had a larger impact for culturally incongruent metaphors on the late stage of whole sentence processing when understanding of the previous main clause got integrated with the explicit explanation. Please see Figure 3 (right) for a demonstration of this interaction. None of the other concerned effects under either coding scheme reached significance.

4.3. Discussion

Compared with both literal and culturally congruent metaphors, inhibitory control had a larger impact on culturally incongruent metaphorical sentences during the integration of the vehicle part with the previous topic before explanation, and on the late stage of explanation processing and whole sentence integration, when understanding the previous main clause got integrated with the explicit explanation into a whole. That is to say, inhibitory control contributed significantly during the late stage of reading culturally incongruent metaphors, especially for their explanatory clauses, which confirms our prediction. However, contrary to our predictions, early stage of reading neither culturally congruent nor culturally incongruent metaphors showed more inhibition involvement than literal sentences, and culturally congruent metaphors didn't require more inhibition than literal throughout reading. Granted that the activated literal meanings did not get inhibited immediately during the early phase, the absence of its inhibition even until the late stage for culturally congruent metaphors seems to contradict the findings of experiment 2. It seems to suggest the possibility of directly accessing culturally congruent L2 metaphorical meanings, even before explanatory contexts were given, and challenge the literal salience model. We discuss this in the general discussion, combined with the results of experiments 1 & 2 and previous findings.

5. General discussion

In the present study, we explore the role of individual differences in inhibitory control during comprehension of both culturally incongruent and culturally congruent L2 metaphors, based on which we try to evaluate and extend existing accounts to develop a more comprehensive theory for L2 metaphor comprehension. In experiment 1 using a metaphor-induced lexical forgetting paradigm, we observed reduced memory for the L1 metaphorical meaning associates, which could be best explained as the activation and inhibition of L1 metaphorical meanings during comprehension of culturally incongruent L2 metaphors. This extended previous findings about the inhibition of irrelevant literal information in L1 metaphor processing. In experiment 2, we found that after explanatory sentences were provided to contextualize understanding, inhibitory control still had a larger effect for both culturally congruent and culturally incongruent L2 metaphors interpretation than literal sentences. This did not support the direct access view but could be explained by the extended literal-salience model, the extended graded salience view, and PP. Using eye-tracking, experiment 3 nailed down the significant role of inhibitory control during the late stage of reading culturally incongruent metaphorical sentences. However, culturally congruent metaphors did not require more inhibition than literals throughout their reading, which seems to contradict the findings of experiment 2 and challenges the literal salience model.

5.1. Task requirements and metaphor properties affect inhibition involvement

Notably, experiment 3 asked participants to simply read sentences and answer true/false questions, all in L2, while in experiment 2 they had to explicitly produce detailed interpretations in L1 for target sentences. Explicit interpretation may demand deeper processing than reading. Besides, language switching could lead participants to further process the metaphor in L1 superfluously, adding to inhibition demand. Interestingly, in George and Wiley's (2016) metaphor-induced lexical forgetting study, subsequent forgetting of the literal associates was observed for both familiar and novel metaphors when participants were instructed to interpret the metaphors (Experiment 2) but was observed for only novel metaphors when participants were asked to simply read the metaphors (Experiment 3). The parallel pattern between their results and ours suggests that inhibition involvement during metaphor processing depends on both task requirements and metaphor properties. It is not likely that our culturally congruent metaphors got direct access to metaphorical meanings during reading in experiment 3, without activation of literal meanings as in experiment 2. It is more plausible that their activated literal meanings did not need to be inhibited during reading comprehension as they did during interpretation.

People's experience of metaphorical meaning can be crudely distinguished along a temporal continuum of processing that includes comprehension, recognition, interpretation, and appreciation (Gibbs, 1994; Gibbs & Colston, 2012). People process a metaphor only to the extent that it offers enough information to draw a relevant contextual pragmatic message (Gibbs, 2023). In fact, the graded salience view has stated that late processes such as retention of relevant and irrelevant information and suppression of contextually inappropriate outputs are more attentive to global discourse considerations such as global coherence, than to local ones (Giora, 2003; Giora et al., 2007). Thus, even if the literal meaning of a culturally congruent L2 metaphor seemed irrelevant in a local reading context, it might be retained because it was perceived as instrumental in constructing the appropriate metaphoric meaning, as conducive to reading the next expression in line, or simply because it didn't interfere reading. The continual activation of literal conceptual properties along with metaphorical interpretation gives rise to a certain level of metaphor awareness (Recanati, 2001). But for the interpretation task, the literal meaning ultimately has to be inhibited to allow precise production of the metaphorical meaning. Viewed in this light, it also makes sense that for culturally incongruent L2 metaphors, activation of conflicting L1 metaphorical meanings in addition to literal meanings inflicted stronger interference and necessitated inhibition involvement even for a reading task.

In addition, PP can still account for these results. Though the bottom-up stimuli were the same, the interpretation task would have required more specific top-down predictions down to the motor production part of the hierarchy, which could have led to larger prediction errors. To produce the metaphorical explanation precisely, the prediction error would have to be resolved in time. Thus, reading culturally congruent metaphors may indeed have triggered larger prediction errors than literals, but this prediction error may not be as large enough and in urgent need of resolving to manifest the effect of inhibitory control, as for culturally incongruent metaphors or as in the interpretation task.

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

5.2. Towards a dynamic view of metaphor comprehension

Gibbs and Colston (2012, 2020) reviewed the vast literature on metaphor understanding and showed that the data one obtains in psycholinguistic and cognitive neuroscience experiments can be influenced by four broad, interacting factors: (1) the participants, (2) the specific language and utterance encountered in context, (3) the specific understanding task, and (4) the method by which the data are analysed to assess language comprehension. Both linguists and psychologists should not assume that a single theoretical model will account for the complex empirical findings, but an alternative meta-theoretical perspective, best illustrated by the dynamical systems theory, can help us make sense of the stabilities and variations in real-life metaphor use (Gibbs, 2013). Thus, although we argued that the graded salience view could be extended to account for the results of our study, while the direct access view was not supported, we admit that generalization should be cautious. In fact, we observed seemingly contradictory results between experiment 3 and experiment 2, where a change of task also affected inhibition involvement during the processing of the same culturally congruent metaphors. It is conceivable that for highly proficient L2 speakers, within a long paragraph context, the reading of culturally incongruent L2 metaphors may even manifest no behavioural difference from literals. It is also likely that ERPs may reveal subtle differences between culturally congruent metaphors and literals during reading that were not revealed by eye-movement measures in our study. Neither case will be readily accountable by the graded salience view.

Our application of the PP framework is an exploration along the lines of a more dynamic view on metaphor comprehension against traditional modular views like the graded salience view. From the perspective of PP, the figurativeness of metaphors triggers larger prediction errors than literal expressions, which could be exacerbated by the conflict between cultural-linguistic knowledge of different languages. Different processing tasks of the same stimuli will induce different amounts of prediction error and different necessities to resolve them in time. Individual differences in inhibitory control can be regarded as the efficiency in resolving prediction errors through more context-adaptive behaviour. Thus, individuals with higher inhibitory control can minimize prediction errors induced by metaphor more effectively. High language proficiency should lead to smaller prediction errors when encountering metaphors, as a proficient reader has probably encountered them more often and therefore assigns its figurative meaning higher probability in the first place. Compared with the graded salience view, PP, with its emphasis on the situated cognizing agent, is more able to account for how variations in metaphor stimuli, task requirements, and individual differences of participants jointly bring about specific observations in an experiment. However, the comprehensiveness and flexibility of this relatively new framework come with the cost of less specified predictions, which should be addressed by further research and its theoretical development.

5.3. L2 metaphor comprehension as embodied enculturation

Our study demonstrates that L2 metaphor constitutes a peculiar challenge for L2 speakers, not only because it may require more executive control to process than literals, but also because L2 users may lack the specific cultural-linguistic experiences to realize the conflict between L1 and L2 metaphorical meanings. A strong embodied view of metaphor claims that sensorimotor activation is necessary during initial metaphor acquisition, but L2 speakers lack those culturally specific sensorimotor experiences, or even the same sensorimotor experiences may have different connotations across cultures (Kövecses, 2005a, 2005b; Lakoff & Johnson, 1999, 2003; Yu, 2008). Our study explored the cultural-linguistic dimension of L2 metaphor understanding by providing culturally incongruent knowledge in the form of an explanatory clause and found that it did demand more involvement of inhibitory control compared to a culturally congruent situation. This effortful bottom-up construction of the culturally incongruent L2 metaphorical meaning could be regarded as a first step in an embodied enculturation process described in the PP framework (Fabry, 2018). As the L2 metaphorical meaning becomes familiarized and lexicalized through repeated use and memory consolidation (Chen & Chen, 2023), the original vehicle will gradually become a polysemy (Bowdle & Gentner, 1999). Diachronically, the process of meaning construction involved in novel L2 metaphor comprehension may evolve through automatization into a process of meaning selection and ambiguity resolution, while inhibition of contextually inappropriate information changes from a later attentional process to an early automatic one (Rubio Fernández, 2007). This happens both for individuals (Cardillo et al., 2012; Goldstein et al., 2012) and for a population of speakers of a language (Zharikov & Gentner, 2019), constituting language evolution on a microscopic level.

5.4. Future research

There are several limitations of our study. First, we mainly investigated nominal metaphors based on nouns. Whether our conclusion can be generalized to other syntactic forms of metaphor, including predicate metaphors (e.g., *The flower purred in the sunshine*, based on a verb), attributive metaphors (e.g., *the weary mountain*, based on an adjective), and conventionalized locative expressions (e.g., *He's feeling up today*, based on a preposition), awaits further research. Second, our participants had an upper intermediate level of L2 proficiency. Previous studies have proved that L2 proficiency has a significant impact on metaphor processing (Heredia & Cieślicka, 2016; Segal & Gollan, 2018; Vaid et al., 2015). Intuitively, as proficiency grows, L2 metaphor processing will require less inhibitory control, which remains a hypothesis to be tested. Thirdly, behavioural measures including eye movements may not be sensitive enough to subtle differences between metaphors and literals during early processing to reveal the role of inhibition, which could be amended by the ERP technique in future studies. Fourthly, the semantic relatedness between L2 metaphorical meanings and L1 metaphorical meanings for culturally incongruent metaphors, and the semantic relatedness between metaphorical meanings and literal meanings for both culturally incongruent and culturally congruent metaphors will most likely moderate inhibition involvement during processing, which is worth further exploration. Last but not least, considering the sample sizes of our experiments are rather small, these results await replications with larger samples in terms of both participants and items.

6. Conclusion

Through three experiments, we found that inhibitory control had a larger effect during the interpretation of both culturally congruent and culturally incongruent L2 metaphors than literal sentences, even after explanatory contexts were provided, opposing the direct access view. However, it did not manifest this significant effect

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

during the reading of culturally congruent metaphors as culturally incongruent ones. Although literal meanings (and culturally incongruent L1 metaphorical meanings) of L2 metaphors may always be activated, inhibition of them depends on both task requirements and metaphor properties. The conflict between L1 metaphorical meanings with incongruent L2 metaphorical meanings, or a more demanding interpretation task that requires precise production of L2 metaphorical meanings, may be factors necessitating inhibition involvement. Together these results can be accounted for by the extended graded salience view, while we propose that a more dynamical situated view, i.e., the predictive processing framework, can better explain how variations in metaphor stimuli, task requirements, and individual differences of participants jointly bring about the specific observations in metaphor comprehension studies.

Supplementary material. To view supplementary material for this article, please visit http://doi.org/10.1017/S1366728924001081.

Data availability statement. The data are available upon request. Materials are all in the supplementary material document.

Acknowledgements. This work was supported by funding from the National Natural Science Foundation of China (31970976) to Baoguo Chen.

Competing interests. The authors declare none.

References

Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research*, **1160**(1), 69–81. https://doi.org/10.1016/J.BRAINRES.2007.05.034

Ashby, J., Roncero, C., de Almeida, R. G., & Agauas, S. J. (2018). The early processing of metaphors and similes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, **71**(1), 161–168. https://doi.org/10.1080/17470218.2016.1278456

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling metaphor from context: An ERP study. *Frontiers in Psychology*, **7**(MAY), 171722. https://doi.org/10.3389/FPSYG.2016.00559/BIBTEX

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, **143**(3), 233–262. https://doi.org/10.1037/bul0000099

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science*, **7**(3), 325–339. https://doi.org/10.1111/J.1467-7687.2004.00351.X

Blank, G. (1988). Metaphors in the lexicon. *Metaphor and Symbolic Activity*, **3**(3), 21–36. https://doi.org/10.1207/s15327868ms0301_2

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**(2), 295–308. https://doi.org/10.1037/0278-7393.19.2.295

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language—A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, **50**(11), 2669–2683. https://doi.org/10.1016/j.neuropsychologia.2012.07.021

Bortfeld, H. (2003). Comprehending idioms cross-linguistically. *Experimental Psychology*, **50**(3), 217–230. https://doi.org/10.1027/1618-3169.50.3.217

Bowdle, B. F., & Gentner, D. (1999). Metaphor Comprehension: From Comparison to Categorization. *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*, 90–95. https://doi.org/10.4324/9781410603494-21

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *NeuroImage*, **59**(4), 3212–3221. https://doi.org/10.1016/j.neuroimage.2011.11.079

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism*, **17**(4), 784–797. https://doi.org/10.1017/S1366728913000795

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation. *Studies in Second Language Acquisition*, **38**(3), 403–443. https://doi.org/10.1017/S0272263115000492

Charteris-Black, J. (2002). Second language figurative proficiency: A comparative study of Malay and English. In *Applied Linguistics* (Vol. **23**, Issue 1, pp. 104–133).. https://doi.org/10.1093/applin/23.1.104

Chen, H., Peng, X., & Zhao, Y. (2013). An ERP study on metaphor comprehension in the bilingual brain. *Chinese Journal of Applied Linguistics*, **36**(4), 505–517. https://doi.org/10.1515/cjal-2013-0034

Chen, J., & Chen, B. (2023). Learning new meanings for known L2 words: Long-term semantic representation is updated to integrate new information after consolidation. *Psychophysiology*, **60**(5), e14228. https://doi.org/10.1111/psyp.14228

Chiappe, D. L., & Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of Memory and Language*, **56**(2), 172–188. https://doi.org/10.1016/j.jml.2006.11.006

Cieślicka, A. B. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, **22**(2), 115–144. https://doi.org/10.1191/0267658306sr263oa

Cieślicka, A. B., & Heredia, R. R. (2011). Hemispheric asymmetries in processing L1 and L2 idioms: Effects of salience and context☆. *Brain and Language*, **116**(3), 136–150. https://doi.org/10.1016/j.bandl.2010.09.007

Cieślicka, A. B., Heredia, R. R., & Olivares, M. (2014). It's all in the eyes: How language dominance, salience, and context affect eye movements during idiomatic language processing. In *Second language learning and teaching* (Vol. **21**, pp. 21–41). Springer Verlag. https://doi.org/10.1007/978-3-319-01414-2_2

Citron, F. M. M., Michaelis, N., & Goldberg, A. E. (2020). Metaphorical language processing and amygdala activation in L1 and L2. *Neuropsychologia*, **140**, 107381. https://doi.org/10.1016/j.neuropsychologia.2020.107381

Clark, A. (2016). Surfing Uncertainty. In *TLS – The times literary supplement* (Vols. **2016**-Novem, Issue 5928). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001

Coderre, E., Conklin, K., & Van Heuven, W. J. B. (2011). Electrophysiological measures of conflict detection and resolution in the Stroop task. *Brain Research*, **1413**, 51–59. https://doi.org/10.1016/J.BRAINRES.2011.07.017

Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: An eye movement study of sentence reading. *Frontiers in Human Neuroscience*, **8**(JAN), 1057. https://doi.org/10.3389/fnhum.2014.01057

Cong, F., & Chen, B. (2022). Parafoveal orthographic processing in bilingual reading. *International Journal of Bilingual Education and Bilingualism*, **25**(10), 3698–3710. https://doi.org/10.1080/13670050.2022.2073786

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory and Cognition*, **30**(6), 958–968. https://doi.org/10.3758/BF03195780/METRICS

De Grauwe, S., Swain, A., Holcomb, P. J., Ditman, T., & Kuperberg, G. R. (2010). Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia*, **48**(7), 1965–1984. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2010.03.017

Deignan, A. (1996). *Collins Cobuild English Guides 7. Metaphor.: Helping learners with real English*. Hueber Max GmbH + Company KG. https://books.google.co.jp/books?id=cn9dNAAACAAJ

Deignan, A., Gabryś, D., & Solska, A., (1997). Teaching English metaphors using cross-linguistic awareness-raising activities. *ELT Journal*, **51**(4), 352–360. https://doi.org/10.1093/elt/51.4.352

Fabry, R. E. (2018). Betwixt and between: the enculturated predictive processing approach to cognition. *Synthese*, **195**(6), 2483–2518. https://doi.org/10.1007/S11229-017-1334-Y/METRICS

Fabry, R. E. (2021). Getting it: A predictive processing approach to irony comprehension. *Synthese*, **198**(7), 6455–6489. https://doi.org/10.1007/s11229-019-02470-9

Fabry, R. E., & Kukkonen, K. (2019). Reconsidering the mind-wandering reader: Predictive processing, probability designs, and enculturation. *Frontiers in Psychology*, **9**(JAN). https://doi.org/10.3389/FPSYG.2018.02648/FULL

Fitz, H., & **Chang, F.** (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, **111**, 15–52. https://doi.org/10.1016/j.cogpsych.2019.03.002

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, **11**(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., **Adams, R. A.**, **Perrinet, L.**, & **Breakspear, M.** (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, **3**. https://doi.org/10.3389/fpsyg.2012.00151

George, T., & **Wiley, J.** (2016). Forgetting the literal: The role of inhibition in metaphor comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **42**(8), 1324–1330. https://doi.org/10.1037/xlm0000237

George, T., & **Wiley, J.** (2019). When is literal meaning inhibited? Evidence from nonsense in the metaphor-induced lexical forgetting paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, **45**(5), 869–880. https://doi.org/10.1037/xlm0000621

Gernsbacher, M. A., **Keysar, B.**, **Robertson, R. R. W.**, & **Werner, N. K.** (2001). The role of suppression and enhancement in understanding metaphors. *Journal of Memory and Language*, **45**(3), 433–450. https://doi.org/10.1006/jmla.2000.2782

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.

Gibbs, R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, **34**(4), 457–486. https://doi.org/10.1016/S0378-2166(01)00046-7

Gibbs, R. W. (2013). The real complexities of psycholinguistic research on metaphor. *Language Sciences*, **40**, 45–52. https://doi.org/10.1016/j.langsci.2013.03.001

Gibbs, R. W. (2023). Pragmatic complexity in metaphor interpretation. In *Cognition*, **237**, 105455. https://doi.org/10.1016/j.cognition.2023.105455

Gibbs, R. W., & **Colston, H. L.** (2012). Interpreting figurative meaning. In *Interpreting vigurative meaning*. Cambridge University Press. https://doi.org/10.1017/CBO9781139168779

Gibbs, R. W., & **Colston, H. L.** (2020). Pragmatics always matters: An expanded vision of experimental Pragmatics. *Frontiers in Psychology*, **11**, 543709. https://doi.org/10.3389/FPSYG.2020.01619/BIBTEX

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistic*, **8**(3), 183–206. https://doi.org/10.1515/cogl.1997.8.3.183

Giora, R. (2003). *On our MindSalience, context, and figurative language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195136166.001.0001

Giora, R. (2012). 9. Happy new war – The role of salient meanings and salience-based interpretations in processing utterances. In *Cognitive pragmatics* (pp. 233–260). De Gruyter. https://doi.org/10.1515/9783110214215.233

Giora, R., & **Fein, O.** (1999). Irony: Context and salience. *Metaphor and Symbol*, **14**(4), 241–257. https://doi.org/10.1207/S15327868MS1404_1

Giora, R., **Fein, O.**, **Aschkenazi, K.**, & **Alkabets-zlozover, I.** (2007). Negation in context: A functional approach to suppression. *Discourse Processes*, **43**(2), 153–172. https://doi.org/10.1080/01638530709336896

Glucksberg, S., **Gildea, P.**, & **Bookin, H. B.** (1982). On understanding non-literal speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, **21**(1), 85–98. https://doi.org/10.1016/S0022-5371(82)90467-4

Glucksberg, S., **McGlone, M. S.**, & **Manfredi, D.** (1997). Property attribution in metaphor comprehension. *Journal of Memory and Language*, **36**(1), 50–67. https://doi.org/10.1006/jmla.1996.2479

Glucksberg, S., **Newsome, M. R.**, & **Goldvarg, Y.** (2001). Inhibition of the literal: Filtering metaphor-irrelevant information during metaphor comprehension. *Metaphor and Symbol*, **16**(3–4), 277–298. https://doi.org/10.1080/10926488.2001.9678898

Gold, R., **Faust, M.**, & **Goldstein, A.** (2010). Semantic integration during metaphor comprehension in Asperger syndrome. *Brain and Language*, **113**(3), 124–134. https://doi.org/10.1016/J.BANDL.2010.03.002

Goldstein, A., **Arzouan, Y.**, & **Faust, M.** (2012). Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization. *Brain and Language*, **123**(2), 137–142. https://doi.org/10.1016/j.bandl.2012.09.008

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). BRILL. https://doi.org/10.1163/9789004368811_003

Hadley, G. (1997). Lexis and Culture: Bound and Determined? *Journal of Psycholinguistic Research*, **26**(4), 483–496. https://doi.org/10.1023/A:1025086204994/METRICS

Harris, R. J. (1976). Comprehension of metaphors: A test of the two-stage processing model. *Bulletin of the Psychonomic Society*, **8**(4), 312–314. https://doi.org/10.3758/BF03335150

Heidlmayr, K., **Moutier, S.**, **Hemforth, B.**, **Courtin, C.**, **Tanzmeister, R.**, & **Isel, F.** (2014). Successive bilingualism and executive functions: The effect of second language use on inhibitory control in a behavioural Stroop Colour Word task*. *Bilingualism: Language and Cognition*, **17**(3), 630–645. https://doi.org/10.1017/S1366728913000539

Heredia, R. R., & **Cieślicka, A. B.** (2016). Metaphoric reference: An eye movement analysis of Spanish–English and English–Spanish bilingual readers. *Frontiers in Psychology*, **7**(MAR), 439. https://doi.org/10.3389/fpsyg.2016.00439

Hohwy, J. (2013). *The predictive mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Jankowiak, K., **Rataj, K.**, & **Naskr□cki, R.** (2017). To electrify bilingualism: Electrophysiological insights into bilingual metaphor comprehension. *PLoS One*, **12**(4), e0175578. https://doi.org/10.1371/journal.pone.0175578

Jiang, J., **Heller, K.**, & **Egner, T.** (2014). Bayesian modeling of flexible cognitive control. In *Neuroscience and Biobehavioral Reviews* (Vol. **46**, Issue P1, pp. 30–43). https://doi.org/10.1016/j.neubiorev.2014.06.001

Katz, A. N. (2005). Discourse and sociological factors in understanding non-literal language. In H. L. Colston & A. N. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (pp. 183–207). Routledge. https://doi.org/10.4324/9781410611628

Katz, A. N., & **Ferretti, T. R.** (2001). Moment-by-moment reading of proverbs in literal and nonliteral contexts. *Metaphor and Symbol*, **16**(3–4), 193–221. https://doi.org/10.1080/10926488.2001.9678895

Katz, A. N., & **Ferretti, T. R.** (2003). Reading proverbs in context:The role of explicit markers. *Discourse Processes*, **36**(1), 19–46. https://doi.org/10.1207/S15326950DP3601_2

Kecskés, I. (2000). A cognitive-pragmatic approach to situation-bound utterances. *Journal of Pragmatics*, **32**(5), 605–625. https://doi.org/10.1016/S0378-2166(99)00063-6

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, **28**(4), 375–385. https://doi.org/10.1016/0749-596X(89)90017-X

Kövecses, Z. (2005a). Causes of variation in metaphor. In *Metaphor in culture* (pp. 231–258). Cambridge University Press. https://doi.org/10.1017/CBO9780511614408.012

Kövecses, Z. (2005b). *Metaphor in culture*. Cambridge University Press. https://doi.org/10.1017/CBO9780511614408

Kuperberg, G. R., & **Jaeger, T. F.** (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, **31**(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kuznetsova, A., **Brockhoff, P. B.**, & **Christensen, R. H. B.** (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13), 1–26. https://doi.org/10.18637/JSS.V082.I13

Lai, V. T., **Curran, T.**, & **Menn, L.** (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, **1284**, 145–155. https://doi.org/10.1016/J.BRAINRES.2009.05.088

Lai, V. T., **Van Dam, W.**, **Conant, L. L.**, **Binder, J. R.**, & **Desai, R. H.** (2015). Familiarity differentially affects right hemisphere contributions to processing metaphors and literals. *Frontiers in Human Neuroscience*, **9**(FEB), 108552. https://doi.org/10.3389/FNHUM.2015.00044/BIBTEX

Lakoff, G., & **Johnson, M.** (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*.

Lakoff, G., & **Johnson, M.** (2003). *Metaphors we live by*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226470993.001.0001

Laufer, B. (2000). Avoidance of idioms in a second language: The effect of L1–L2 degree of similarity. *Studia Linguistica*, **54**(2), 186–196. https://doi.org/10.1111/1467-9582.00059

Lewis, A. G., & **Bastiaansen, M.** (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, **68**, 155–168. https://doi.org/10.1016/j.cortex.2015.02.014

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

Littlemore, J., **Chen, P. T.**, **Koester, A.**, & **Barnden, J.** (2011). Difficulties in metaphor comprehension faced by international students whose first language is not english. *Applied Linguistics*, **32**(4), 408–429. https://doi.org/10.1093/applin/amr009

Littlemore, J., & **Low, G.** (2006). Metaphoric competence, second language learning, and communicative language ability. *Applied Linguistics*, **27**(2), 268–294. https://doi.org/10.1093/applin/aml004

Littlemore, J., **MacArthur, F.**, **Cienki, A.**, & **Holloway, J.** (2012). How to make yourself understood by international students: The role of metaphor in academic tutorials. *ELT Research Papers: British Council Publications*, **12**(6), 1–27.

Liu, H., **Liang, L.**, **Dunlap, S.**, **Fan, N.**, & **Chen, B.** (2016). The effect of domain-general inhibition-related training on language switching: An ERP study. *Cognition*, **146**, 264–276. https://doi.org/10.1016/j.cognition.2015.10.004

Lü, J., **Liang, L.**, & **Chen, B.** (2019). The effect of executive control ability on the comprehension of second language metaphor. *International Journal of Bilingualism*, **23**(1), 87–101. https://doi.org/10.1177/1367006917709096

Lu, Y., **Wu, J.**, **Dunlap, S.**, & **Chen, B.** (2017). The inhibitory mechanism in learning ambiguous words in a second language. *Frontiers in Psychology*, **8**(APR), 636. https://doi.org/10.3389/fpsyg.2017.00636

Luk, G., **Anderson, J. A. E.**, **Craik, F. I. M.**, **Grady, C.**, & **Bialystok, E.** (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and Cognition*, **74**(3), 347–357. https://doi.org/10.1016/J.BANDC.2010.09.004

Lupyan, G., & **Clark, A.** (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, **24**(4), 279–284. https://doi.org/10.1177/0963721415570732

MacDonald, M. C., & **Seidenberg, M. S.** (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In *Handbook of psycholinguistics* (pp. 581–611). Elsevier.

Mashal, N., **Borodkin, K.**, **Maliniak, O.**, & **Faust, M.** (2015). Hemispheric involvement in native and non-native comprehension of conventional metaphors. *Journal of Neurolinguistics*, **35**, 96–108. https://doi.org/10.1016/j.jneuroling.2015.04.001

Matuschek, H., **Kliegl, R.**, **Vasishth, S.**, **Baayen, H.**, & **Bates, D.** (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, **94**, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mcelree, B., & **Nordlie, J.** (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, **6**(3), 486–494. https://doi.org/10.3758/BF03210839

Nowagk, R. (1998). *Conan: A barbarian tool for constrained randomization (Version 1.9)*. Leipzig: Max-Planck-Institute of Cognitive Neuroscience Leipzig.

Olkoniemi, H., **Bertram, R.**, & **Kaakinen, J. K.** (2022). Knowledge is a river and education is like a stairway: An eye movement study on how L2 speakers process metaphors and similes. *Bilingualism: Language and Cognition*, **25**(2), 307–320. https://doi.org/10.1017/S1366728921000869

Ortony, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory*, **25**(1), 45–53. https://doi.org/10.1111/J.1741-5446.1975.TB00666.X

Ortony, A., **Schallert, D. L.**, **Reynolds, R. E.**, & **Antos, S. J.** (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, **17**(4), 465–477. https://doi.org/10.1016/S0022-5371(78)90283-9

Oxford Quick Placement Test. (2001). Oxford University Press. http://elzbietapodolska.pl/wp-content/uploads/2018/04/epodolska-elzbieta-podolska-test-p01.pdf

Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science*, **17**(4), 286–290. https://doi.org/10.1111/j.1467-8721.2008.00591.x

Pexman, P. M., **Ferretti, T. R.**, & **Katz, A. N.** (2000). Discourse factors that influence online reading of metaphor and irony. *Discourse Processes*, **29**(3), 201–222. https://doi.org/10.1207/S15326950dp2903_2

Pollio, H. R., **Fabrizi, M. S.**, **Sills, A.**, & **Smith, M. K.** (1984). Need metaphoric comprehension take longer than literal comprehension? *Journal of Psycholinguistic Research*, **13**(3), 195–214. https://doi.org/10.1007/BF01068463

Pynte, J., **Besson, M.**, **Robichon, F. H.**, & **Poli, J.** (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language*, **55**(3), 293–316. https://doi.org/10.1006/BRLN.1996.0107

Qiu, J., **Luo, Y.**, **Wang, Q.**, **Zhang, F.**, & **Zhang, Q.** (2006). Brain mechanism of Stroop interference effect in Chinese characters. *Brain Research*, **1072**(1), 186–193. https://doi.org/10.1016/j.brainres.2005.12.029

Rabovsky, M., **Hansen, S. S.**, & **McClelland, J. L.** (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, **2**(9), 693–705. https://doi.org/10.1038/s41562-018-0406-4

Rapp, A. M., **Mutschler, D. E.**, & **Erb, M.** (2012). Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *NeuroImage*, **63**(1), 600–610. https://doi.org/10.1016/j.neuroimage.2012.06.022

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**(3), 372–422. https://doi.org/10.1037//0033-2909.124.3.372

Recanati, F. (2001). Literal/nonliteral. *Midwest Studies in Philosophy*, **25**(1), 264–274. https://doi.org/10.1111/1475-4975.00048

Rubio Fernández, P. (2007). Suppression in metaphor interpretation: Differences between meaning selection and meaning construction. *Journal of Semantics*, **24**(4), 345–371. https://doi.org/10.1093/jos/ffm006

Schneider, W., **Eschman, A.**, & **Zuccolotto, A.** (2012). *E-Prime 2.0, Psychology Software Tools*. Psychology Software Tools Inc..

Searle, J. R. (1979). *Expression and meaning*. Cambridge University Press. https://doi.org/10.1017/CBO9780511609213

Segal, D., & **Gollan, T. H.** (2018). What's left for balanced bilinguals? Language proficiency and item familiarity affect left-hemisphere specialization in metaphor processing. *Neuropsychology*, **32**(7), 866–879. https://doi.org/10.1037/neu0000467

Senaldi, M. S. G., & **Titone, D. A.** (2022). Less direct, more analytical: Eye-movement measures of L2 idiom reading. *Languages, Vol.* **7**(2), 91. https://doi.org/10.3390/LANGUAGES7020091

Siyanova-Chanturia, A., **Conklin, K.**, & **Schmitt, N.** (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, **27**(2), 251–272. https://doi.org/10.1177/0267658310382068

Steen, G. J., **Dorst, A. G.**, **Herrmann, J. B.**, **Kaal, A.**, **Krennmayr, T.**, & **Pasma, T.** (2010). A method for linguistic metaphor identification. In *Converging evidence in language and communication research* (Vol. **14**). John Benjamins Publishing Company. https://doi.org/10.1075/celcr.14

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement a comparison of consensus, consistency, and measurement approaches to estimating interrater reliability approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation*, **9**, 4. https://doi.org/10.7275/96jp-xz07

Su, L. (2009). *A usage dictionary of English and Chinese conceptual metaphors*. Nankai University.

Team, R.-C. (2020). R: A language and environment for statistical programming. In R foundation for sstatistical computing. *(4.0.1)*, Vienna, Austria. https://www.R-project.org/

Titone, D., **Columbus, G.**, **Whitford, V.**, **Mercier, J.**, & **Libben, M.** (2015). Contrasting bilingual and monolingual idiom processing. *Bilingual figurative language processing*, 171–207. https://doi.org/10.1017/CBO9781139342100.011

Türker, E. (2016). The role of L1 conceptual and linguistic knowledge and frequency in the acquisition of L2 metaphorical expressions. *Second Language Research*, **32**(1), 25–48. https://doi.org/10.1177/0267658315593336

Vaid, J., **López, B. G.**, & **Martínez, F. E.** (2015). Linking the figurative to the creative: Bilinguals' comprehension of metaphors, jokes, and remote associates. *Bilingual figurative language processing*, 53–86.

van Heuven, W. J. B., **Mandera, P.**, **Keuleers, E.**, & **Brysbaert, M.** (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, **67**(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Wallentin, M., **Gravholt, C. H.**, & **Skakkebæk, A.** (2015). Broca's region and Visual Word Form Area activation differ during a predictive Stroop task. *Cortex*, **73**, 257–270. https://doi.org/10.1016/j.cortex.2015.08.023

Weiland, H., **Bambini, V.**, & **Schumacher, P. B.** (2014). The role of literal meaning in figurative language comprehension: Evidence from masked

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press

priming ERP. *Frontiers in Human Neuroscience*, **8**(AUG), 93185. https://doi.org/10.3389/FNHUM.2014.00583/ABSTRACT

Wolter, B., & **Gyllstad, H.** (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, **32**(4), 430–449. https://doi.org/10.1093/APPLIN/AMR011

Wolter, B., & **Gyllstad, H.** (2013). Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition*, **35**(3), 451–482. https://doi.org/10.1017/S0272263113000107

Yang, J. (2014). The role of the right hemisphere in metaphor comprehension: A meta-analysis of functional magnetic resonance imaging studies. *Human Brain Mapping*, **35**(1), 107–122. https://doi.org/10.1002/HBM.22160

Ylinen, S., **Bosseler, A.**, **Junttila, K.**, & **Huotilainen, M.** (2017). Predictive coding accelerates word recognition and learning in the early stages of language development. *Developmental Science*, **20**(6), e12472. https://doi.org/10.1111/desc.12472

Yoon, H. D., **Shin, M.**, & **Jeon, H. A.** (2021). The critical role of interference control in metaphor comprehension evidenced by the drift–diffusion model. *Scientific Reports*, **11**(1), 1–16. https://doi.org/10.1038/s41598-021-98351-8

Yu, N. (2008). Metaphor from Body and Culture. In *The cambridge handbook of metaphor and Thought* (pp. 247–261). Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.016

Zharikov, S. S., & **Gentner, D.** (2019). Why do metaphors seem deeper than similes? In *Proceedings of the twenty-fourth annual conference of the cognitive science society* (pp. 976–981). Routledge. https://doi.org/10.4324/9781315782379-203

https://doi.org/10.1017/S1366728924001081 Published online by Cambridge University Press