The “Good is Up” Metaphoric Effects on Recognition: True for Source Guessing but False for Item Memory

Zixi Jin[1], Ulrich von Hecker[1], Nikoletta Symeonidou[2], Liu Yi[3] and Karl Christoph Klauer[4]

[1]School of Psychology, Cardiff University, UK.

[2]Department of Psychology, University of Mannheim, Germany

[3]Department of Psychology, School of Philosophy, Wuhan University, China

[4]Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, Germany.

Author Note:
This research is a part of Zixi Jin's PhD project funded by China Scholarship Council (CSC). Correspondence concerning this article should be addressed to Ulrich von Hecker, vonheckeru@cardiff.ac.uk.
We would like to thank the social lab group members from School of Psychology, Cardiff University for their suggestions in the early stage of this research. And thanks to Alina Kias, Lea Krumpholtz and Luise Metzger for their help with data collection. The data are available at:

https://osf.io/w39uq/?view_only=89da5b4b57914a15b92d9a7907e1f99f

The experiments were not preregistrated.

Abstract

The “good is up” metaphor which links valence and verticality was found to influence affective judgement and to direct attention (Meier & Robinson, 2004), but its effects on memory remain unclear with contradictory research findings. In order to provide a more accurate assessment of memory components involved in recognition, such as item memory and source guessing biases, a standard source monitoring paradigm (Johnson et al., 1993) was applied in this research. A series of three experiments, provided a conceptual replication and extension of Crawford et al’s (2014) Experiment 2 and yielded a consistent result pattern suggesting that the “good is up” metaphor biases participants’ guessing of source location. That is, when source memory failed, participants were more inclined to guess the “up” location versus “down” location for positive items (and vice versa for negative items). It did, however, not affect source memory or item memory for valenced stimuli learned from metaphor-congruent versus incongruent locations (i.e., no metaphor-(in)congruent effects in memory). We suggest that the “good is up” metaphor may affect cognitive processes in a more subtle way than originally suggested.

Keywords: metaphor, recognition memory, source monitoring, source guessing, latent-trait approach

Introduction

The “good is up” metaphor

Metaphors are fundamental part of people’s conceptual system (Kittay, 1990; Lakoff & Johnson, 2008), bridging the realms of abstract and concrete concepts. They can be used as a cognitive scaffold for people to rely on and thereby to learn about, reason with, and illustrate abstract concepts (Lakoff & Johnson, 1980, 1999). The “good is up” metaphor as a primary metaphor in people’s life (Lakoff & Johnson, 1999) indicates that concepts considered to represent something positive are commonly associated with physically high spatial locations, whereas those considered to represent something negative are commonly associated with physically low locations. Our daily life experiences intuitively provide evidence to support this metaphoric association. In colloquial English, people often use “I feel up/down today” to indicate their positive or negative moods. On the expressive side, when feeling confident or powerful, people’s body postures tend to be more upright, whereas people tend to be slouched when feeling depressed or upset. In competitive events like the Olympics, athletes spontaneously elevate the chest or raise their arms above the head to express pride of success, whereas they tend to display a hanging head or slump the shoulders to express shame of failure (Casasanto & Dijkstra, 2010; Riskind & Gotay, 1982; Stepper & Strack, 1993). Supportive evidence can also be found on the Internet, where users of social media or video websites use the “thumb up” or “thumb down” button as a simplified way to express positive or negative feedback. In fact, previous literature suggests that the “good is up” metaphor is not limited to the English culture or language. Its effects have also been demonstrated in Mandarin (Wu et al., 2019), German (Dudschig et al., 2015), Russian and French (Luodonpää-Manni & Viimaranta, 2010) contexts.

One cornerstone of the investigation on the “good is up” metaphor is research conducted by Meier and Robinson (2004), suggesting that *metaphor-congruency* (i.e., positive-up & negative-down, which are the congruent combinations under the “good is up” metaphor) facilitates participants’ affective judgement responses and that valenced word primes can direct participants’ attention to metaphor-congruent spatial locations. Moreover, fMRI studies have also provided confirming evidence of metaphor influence. Quadflieg and colleagues (2011) found similar neuronal states between discriminating the physical “up/down” connotation of a word stimulus (e.g., “attic” as having “up” connotation, “carpet” as having “down” connotation) and the “positive/negative” valence of a word stimulus (e.g., “party” as having positive valence and “accident” as having negative valence).

Researchers have also been investigating the effects of the “good is up” metaphor on memory. Overall, this literature shows mixed results as to whether metaphor-congruency or -incongruency may facilitate memory performance. One reason for the ambiguity may reside in the use of simple performance measures of memory in which different processes involved in memory-based judgements are confounded. Therefore, the present project aims at investigating metaphor-induced influences on memory using a source-monitoring paradigm and model (Bayen et al., 1996; Kuhlmann et al., 2021) that allows us to disentangle between the various processes contributing to memory-based judgements.

Diverging results in the literature

We started the series with an adaptation and modification of Experiment 2 from Crawford et al. (2014) in which they manipulated the location of the presented valenced (i.e., positive & negative) stimuli in the encoding phase and used a recognition test to show the facilitating effect of metaphor incongruency on recognition memory. In line with the Attention Elaboration Hypothesis (AEH, inconsistent materials get more attention), they found positive (vs. negative) words learned in the down (vs. up) location to be memorized better, indicating an incongruence effect in recognition memory. This stands in contrast to Palma et al. (2011) who found metaphor-congruent recall in an impression-formation task. Likewise, when asked to recall the locations on a map where positive and negative events occurred (e.g., “A family wins a trip to Disney World” vs. “A family is killed in a tragic car accident”, all virtual events learned in the learning phase), participants tended to recall the locations of positive events with an upward bias and to recall the locations of negative events with a downward bias (Brunyé et al., 2012). Note, however, that this work does not acknowledge that both memory and reconstructive processes (e.g., educated guessing) can contribute to participants’ responses in a memory test (Johnson et al., 1993; Kuhlmann et al., 2021). Crucially, the “good is up” metaphor might have distinct effects on these processes and failing to distinguish between them may contribute to the inconsistencies in findings mentioned earlier. In fact, previous studies have repeatedly shown that participants' semantic knowledge, such as schemas or stereotypes, have differential (sometimes even opposing) effects on memory versus reconstructive processes, emphasizing the necessity to separate and consider both (see next section). Assuming that metaphors, similar to schemas, are another example of semantic knowledge acquired outside the experimental environment, we borrowed from this schema-literature to derive our predictions.

Indirect effects of metaphors versus direct effects of stereotypes

As alluded, metaphors are conceptually close to schemas (or stereotypes) because they also tap into semantic knowledge. However, schema congruence or incongruence is arguably easier to detect than metaphor congruence or incongruence. Stereotype congruence versus incongruence is determined by directly comparing the encoded information (e.g., word "brutal") to an activated category (e.g., profession "nurse"). In contrast, the metaphor-congruence status of words such as “bliss” or “death”, although with clear valence implication, is not immediately established. Rather, the metaphor-congruence or incongruence relies on accessing extrinsic information, that is, perceptual constraints in the environment have to be taken into account. Specifically, for the good-is-up metaphor, one first needs to decode the item's valence status (positive or negative), then decode its pairing with a (salient) spatial differentiation (up or down), and finally decode the metaphoric implication of that location itself (positive or negative). Based on these considerations, we think it likely that congruence/incongruence effects due to a metaphor be weaker than effects due to a stereotype-based schema.

Results on schema congruency and predictions for metaphor-congruency

But what are the results on the effects of stereotype-based schema on memory? In the last two decades, this question has often been investigated in a source-monitoring framework that allows one to assess the differential contribution of memory and guessing processes to memory judgements that are usually confounded in overall assessments of memory performance.

Source monitoring. Source broadly refers to the origin of an information and thus encompasses (but is not limited to) the context in which an information is acquired, for example, when and where this information was perceived, or through whom or what media, etc. (Johnson et al., 1993). Source monitoring refers to a series of cognitive processes involved in the judgements about the source of information, such as remembering the source or reconstructing it based on plausibility or schemas (see Johnson et al., 1993 for the theoretical framework).

A source monitoring study typically consists of a learning phase in which items originating from (usually) two sources are presented (for example, words presented at an up or down location) and a testing phase in which participants are asked to identify a presented item as “old” or “new” (i.e., as having been presented in the learning phase or not). If participants classify the item as old, they are then asked to attribute the item to one of the two sources (i.e., they are asked to indicate whether the word was presented at either the up or down location in the learning phase).

In such a source monitoring paradigm, three main processes contribute to participants’ responses: *item memory*, which indicates the ability to recognise whether an item was studied or not; *source memory*, which indicates the ability to remember which source an item stemmed from; and *guessing biases*, which includes the tendency of guessing that an item was learned when not being able to recognise it (i.e., guessing an unrecognised item to be “old” or “new”), and the tendency of guessing that an item classified as “old” (no matter whether this judgement was due to memory or guessing) originates from a certain source when not being able to discriminate the source of this item.

The advantage of applying this standard source monitoring paradigm to the present research is that the metaphoric and stereotype effects on recognition memory can be investigated with fewer confounds. Previous research confounds the contribution of memory processes versus guessing processes when investigating the metaphoric effects on recognition memory, which makes it hard to disentangle effects on memory from effects on guessing biases. Applying the standard source monitoring paradigm allows us to use the two-high-threshold multinomial model of source monitoring (2HTSM; Bayen et al., 1996). This in turn makes it possible to look at metaphoric effects on “pure” item memory, corrected for guessing, which is innovative for research on the present topic. A sizeable literature of applications of source-monitoring models has provided evidence for the validity of the measures it provides for item memory, source memory, and item and source guessing (for an overview, see Kuhlmann et al., 2021; for selective-influence studies in which manipulations of these processes were found to selectively influence the appropriate parameter, see Bayen et al., 1996, and Bayen & Kuhlmann, 2011).

Predictions for metaphor-congruency effects. Based on the results for schema-congruency effects, we argue that an effect of metaphor congruency on item memory or on source memory is unlikely to occur. Although schema-(in)congruency effects on memory seem to unfold in free recall tests no such influences could be established for item memory (i.e., recognition tests, Bell et al., 2012; Ehrenberg & Klauer, 2005; Kroneisen & Bell, 2015; Küppers & Bayen, 2014). Secondly, with regards to such influences on source memory, it has been argued that schemas should primarily have an influence on source memory rather than item memory, as it is the source that makes the item congruent or incongruent (Ehrenberg & Klauer, 2005). However, positive evidence for such an influence was only found when source-item combinations strongly contradicted schematic expectations (e.g., an oven in the bathroom), but was not apparent when this contradiction was only weak (e.g., books in the bathroom, Bayen & Kuhlmann, 2011; Bayen et al., 2000; Kuhlmann et al., 2012). As argued above, metaphor-induced (in)congruency is established in a more complex and indirect way, making metaphor violations less salient and less likely to contradict existing expectations compared to stereotype-based schemas. Therefore, we predicted no metaphor-induced influences on source memory.

With regard to source guessing biases, recent research suggests, however, that schema-congruent source guessing might indeed be a psychological default mechanism. From this perspective, we now extend this expectation to the case of metaphor-induced congruency and its likely effects. In particular, in the schema literature, substantive evidence has accrued for schema-congruent source guessing (e.g., Bayen & Kuhlmann, 2011; Bayen et al., 2000; Bell et al., 2012; Ehrenberg & Klauer, 2005; Kroneisen & Bell, 2013, Schaper et al., 2019). If, therefore, in a source-monitoring situation, metaphors operate in a similar way as schemas, we would expect their effects to manifest in terms of metaphor-congruent source guessing, and not as congruence effects for item memory or source memory.

Other research in the schema literature also emphasizes the importance of congruency for metaphor-related memory. Sherman and Bessenoff (1999) demonstrated more misattribution of stereotypic than counterstereotypic behaviours to target persons in cases where retrieval of the true source information for these behaviours was difficult or disrupted. Thus, schema-congruent guessing was found to be a heuristic used to compensate source memory failure. Schema-congruent guessing also occurred when participants had to re-align their first impressions of faces with behavioural descriptions (Bell et al., 2015). In another study, illusory recollections were found to be congruent with stereotypic associations to instructed sources at retrieval (as either coming from a doctor or from a lawyer), supporting the idea that not only an existing memory trace mediated the responses, but also schematic information presented in the retrieval situation that would inform guessing in cases of insufficient source memory (Dodson et al., 2008).

Based on these findings, we hypothesize that the “good is up” metaphor can also create the expectation of congruency in the way schemas do and consequently bias source guessing toward a metaphor-congruent direction. Namely, when participants fail to retrieve the source location of a valenced stimulus, they will tend to guess a metaphor-congruent location rather than a metaphor-incongruent one. Item memory and source memory, however, should be unaffected by metaphoric influence.

Overview of Experiments

In Experiment 1, we examined memory for materials of positive or negative valence shown at the top or bottom of the screen: words (Exp. 1a) and emojis (Exp. 1b). In Experiments 2a and 2b, we replicated Experiments 1a and 1b, respectively, in a Mandarin speaking context. Finally, to corroborate the role of physical simulation as underlying the metaphoric mapping of “good” to “up”, we tested in the UK (Experiment 3a) and in China (Experiment 3b) whether concrete concepts with “up” and “down” vertical connotations can trigger the same effects as words with positive and negative valence, respectively.

Experiment 1

This and all other experiments in this research received Ethics approval from the relevant university committee.

Experiment 1 is a conceptual replication of Experiment 2 from Crawford et al (2014) in which participants were instructed to memorise positive and negative words randomly presented at the top or bottom of the screen. Instead of a simple recognition task, a source monitoring paradigm was applied. We used two different types of materials: Words (Experiment 1a) and emojis (Experiment 1b).

Experiment 1a

In Crawford and colleagues' (2014) findings, a metaphor-incongruency effect on item memory occurred, that is, valenced words learned from metaphor-incongruent locations (e.g., “hostile” presented at the top) were better memorised than those from metaphor-congruent locations (e.g., “refreshing” presented at the top). However, as we argued above, such an effect is unlikely to show up for item memory or for source memory. We therefore do not expect such an effect to occur in the present experiment.

Crawford et al. (2014) also found a negativity advantage on item memory, which we have no reason to question and therefore include in our predictions. Different from Crawford et al.’s (2014) design, two instruction conditions were added. As explained in the Introduction, we think that metaphor-induced effects might be more indirect and weaker than stereotype-induced effects. It was therefore of interest whether different degrees of metaphor awareness would have an influence of the strength of such effects. Lebois and colleagues (2015) proposed that even obvious factors such as spatial locations or valence may only be effective if mentioned explicitly, otherwise participants would not pay attention to them. In this vein, the instructions of the first level of awareness mentioned that the stimuli would have positive or negative valence. The second (moderate) awareness condition in this experiment explicitly mentioned that the stimuli to be memorised would have positive or negative valence and were going to be presented at up or down vertical locations (e.g., “You will be presented with 40 different words, with either *positive* or *negative* valence, at either the *up* or *down* location on the screen in a random sequence.”). In the last and third awareness condition, participants were additionally encouraged to use this knowledge as a memory aid. Higher awareness of the metaphoric association was expected to increase the expectation of valence-verticality congruency and consequently to increase the differences in memory parameters between metaphor-congruent and incongruent presented words if any, as well as to increase the predicted metaphor-congruent source guessing biases.

Experiment 1b uses the same design with positive and negative emojis instead of words as materials. The purpose was to investigate whether metaphoric effects can be demonstrated in recognition when using stimuli with less semantic information but projecting affective valence more directly. This was expected to more efficiently address the role of valence per se, in terms of metaphoric effects.

Method

Participants. a: *N* = 103 / **b:** *N* = 103. English native speakers were recruited from Cardiff University, United Kingdom, to take part in this study, of which **a:** 21 / **b:** 17 were males, mean age **a:** *M* = 19.98 years (*SD* = 2.57) **b:** 19.61 years (*SD* = 2.79). Participants received 1 course credit or were paid £2 for their participation.

Design. Experiment 1 (**a** and **b**) used a mixed design with factors valence (positive vs. negative, within-subjects), verticality (up vs. down, within-subjects), and instruction (stimulus-only vs. stimulus-location vs. stimulus-location-metaphor, between-subjects). Participants were instructed to memorise positive and negative words which were randomly presented at the top or bottom of the screen. Contributing memory components were tested later by a source monitoring task.

Hypotheses. Hypothesis 1 (H1): There is no metaphor-incongruency effect on item memory; Hypothesis 2 (H2): There are metaphor-congruent source guessing biases; Hypothesis 3 (H3): Increasing awareness enlarges metaphoric effects. Hypothesis 4 (H4): There is a negativity advantage on item memory. Hypothesis 5 (H5): There is no metaphor-incongruency effect on source memory. H5 was tested in an additional analysis presented in the discussion section of the present experiments.

Materials. a: 40 positive words and 40 negative words were selected from the Affective

Norms for English Words (ANEW, Bradley & Lang, 1999) considering valence, arousal norms, and frequencies in daily English. The mean valence scores of positive words and negative words were 7.20 (*SD* = 0.37) and 2.86 (*SD* = 0.63), respectively, on a 9-point rating scale, which differed significantly, *t*(78) = 37.61, *p* < .001. No differences were found between positive and negative words on arousal scores, length, syllable length or frequency in *t* tests (see Table 1). The words were randomized to old and new status newly for each subject, and an equal number of congruent and incongruent trials were generated. For a list of all the selected materials see Appendix A1.

b: Eighty emojis were selected from the Lisbon Emoji and Emoticon Database (LEED, see Rodrigues et al., 2018), half of them with positive valence and the other half with negative valence. Subjective valence ratings were obtained in a pre-test (*N* = 20) on a scale ranging from -7 to +7 indicating extremely negative to extremely positive. Based on these ratings, 72 emojis were selected as materials in this experiment. The mean valence rating of positive and negative emojis were 3.73 (*SD* = 0.73) and -3.72 (*SD* = 0.72) respectively, which differed significantly, *t*(70) = 43.63, *p* < .001. See Appendix A2 for a full list of the selected emojis.

Table 1

Norms of positive and negative word stimuli (M±SD) in Experiment 1a

	positive (*M*±*SD*)	negative (*M*±*SD*)	*t*	*df*	*p*
Valence	7.20±0.37	2.86±0.63	37.61	78	.000***
Arousal	4.49±0.72	4.62±0.67	-0.86	78	.39
Length	6.10±1.72	6.20±1.74	-0.26	78	.80
Syllable	1.88±0.79	2.00±0.96	-0.64	78	.53
Frequency	26.74±39.04	22.16±47.81	0.44	78	.66

Note. “Arousal” stands for the arousal rating score from ANEW (Bradley & Lang, 1999), “Length” refers to the number of letters in a word, “Syllable” refers to the number of syllables in a word, “Frequency” stands for the word frequency score from ANEW (Bradley & Lang, 1999), with higher numbers indicating higher frequencies; *** p < .001.

Procedure. **a:** Participants were tested individually in laboratory rooms in the School of Psychology, Cardiff University, United Kingdom. Written consent was obtained from each participant before the experiment started. Participants were guided through all experimental procedures by a computer program written in PsychoPy (Peirce, 2009; Peirce et al., 2019) with instructions presented on the screen. In order to enlarge the difference between the up and down locations, a vertically set-up 24-inch screen was used in this experiment, which was about 12 inches wide and 20 inches tall. The top and bottom locations were approximately 8 inches above or below the midpoint of the screen, respectively.

The test was self-paced. For each participant, 40 words were presented in the learning phase, half of them positive and half negative ones. Half of the words from each valence were presented at metaphor-congruent locations, the other half at metaphor-incongruent locations. In other words, in the learning phase, 10 trials each presented positive words at the top of the screen, positive words at the bottom of the screen, negative words at the top of the screen and negative words at the bottom of the screen, in a random sequence. All words were presented in white font colour on a black background. In each learning trial, there was a fixation cue presented for 1000 ms at the location where the word was going to appear, followed by the word for 1000 ms. Then a blank screen was presented for 1000 ms as an inter-trial interval before the next trial began. In the following testing phase, the 40 learned words were randomly interspersed with 40 new words. In each test trial, one of these words appeared in the middle of the screen and participants were asked to respond whether it was a new (not studied) or old (studied) word by pressing the left or right arrow key (response mapping counterbalanced across participants). If the word was classified as old, participants were then asked to press the up or down arrow key to indicate whether the word had been presented at the up or down location in the learning phase. If the word was classified as new, the next trial began. There was a 500 ms inter-trial interval between test trials. After completing all 80 test trials, participants were debriefed, thanked and dismissed.

b: Except for the duration of presentation in the learning phase, the procedure was identical to Experiment 1a. In each learning trial, after 1000 ms of fixation cue, the emoji was presented for 2000 ms. Participants learned positive and negative emojis from either up or down vertical location in a random sequence, then were required to answer whether a presented emoji was learned or not and to subsequently discriminate where it was presented if classified as learned.

Results and Discussion

The data and code for data analysis of this and all other experiments in this report are available at:

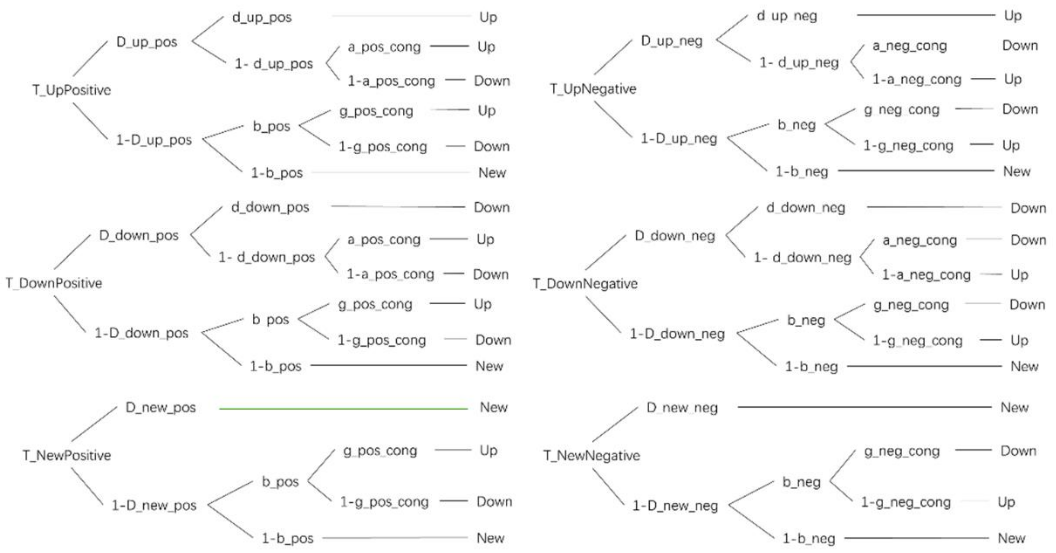
https://osf.io/w39uq/?view_only=89da5b4b57914a15b92d9a7907e1f99f

Modeling source-memory data. As commonly used in source monitoring research, the two-high-threshold multinomial processing tree model of source monitoring (2HTSM) was applied (e.g., Bayen & Kuhlmann, 2011; Buchner et al., 2009; Klauer & Meiser, 2000; Küppers, 2012; Küppers & Bayen, 2014; Meiser et al., 2007; Singmann et al., 2013). The model provides separate parameters for measuring item recognition, source discrimination, and guessing, based on the obtained frequency data of each source-response category (Bayen et al., 1996)[1]. Figure 1 illustrates an adapted 2HTSM model structure (so-called “multinomial tree”) with each pathway in the figure specifying a combination of the cognitive processes involved that might contribute to a certain response in a test trial.

Figure 1

Adapted 2HTSM Model of Experiment 1 with Six Multinomial Trees

[1] The source-monitoring model is based on a threshold model of recognition memory, postulating a discrete mediation of recognition judgments via an “all or none” process. Concerns regarding the use of threshold models have been voiced (e.g., Dubé & Rotello, 2012) by modelers preferring signal detection (SDT) models that postulate a mediation via a continuous memory signal. As pointed out by Kellen and Klauer (2018), these concerns are put into perspective by the fact that both SDT and threshold accounts almost invariably provide similar characterizations of performance in empirical studies. Moreover, in focused tests pitting threshold models and SDT models against each other, mixed results have been obtained (e.g., Kellen & Klauer, 2014; Province & Rouder, 2012). McAdoo et al. (2019) argue that task characteristics determine whether participants rely on graded or discretized memory signals, with SDT prevailing in tasks such as ranking tasks that require effortful comparisons between items, and threshold assumptions prevailing in tasks requiring “detect or don’t” decisions on single items (as in all of our experiments). Another reconciliation could arise from dual-process models in which it is assumed that recognition is mediated by a mixture of discrete and continuous processes (e.g., Brainerd et al., 2015).



Note. "T_" refers to the multinomial tree of each item type (e.g., up_pos = positive items presented at the "up" location). Explanations of parameters: *D* parameters for item memory: Probability of correctly detecting item status as new or old, with separate parameters for each item type; *d* parameters for source memory: Probability of correctly detecting the source (location at which they were presented), with separate parameters for each type of old items; *b* parameters for item-status guessing: Probability of guessing "old" when item status was not detected, with separate parameters for positive and negative items; a parameters for location guessing: Probability of guessing the metaphor-congruent location ("up" for positive items, "down" for negative locations) for items correctly detected as "old" without memory for source, with separate parameters for positive and negative items; *g* parameters for location guessing: Probability of guessing the metaphor-congruent location when item status could not be detected, with separate parameters for positive and negative items.

The root of each multinomial tree (left side of each tree in Figure 1) indicates the source of the test word. For example, "T_UpPositive" indicates a positive test word was presented at the up location in the learning phase. Each pathway leads to a possible response, "Up" or "Down" or "New", as listed on the right side of Figure 1. Obviously, more than one pathway can terminate in the same observed source-response category. The branches of the tree are labelled by the parameters of the model (*D*, *d*, *b*, *a*, *g*) or by their complements (1-*D*, 1-*d*, 1-*b*, 1-*a*, 1-*g*).

To illustrate, consider a positive word that originated from the up location as an example. When later presented in a test trial (see the first "tree" in Figure 1), the participant will recognise the word as “old” with probability *D*_up_pos (positive word, recognised). Further, with probability *d*_up_pos (location of the word is up), the recognised word will be correctly identified as stemming from the up location; with the complementary probability 1 - *d*_up_pos, the participant will not be able to identify the word as stemming from the up location and will, therefore, guess the source of the word. With probability *a*_pos_cong (guess that positive words go up), the participant will guess that the word was from the metaphor-congruent source, which is the up location; with the complementary probability 1 - *a*_pos_cong, the participant will guess the item was from the metaphor-incongruent source, namely the down location. If the word is not recognised (with probability 1 - *D*_up_pos) the participant is in a state of uncertainty. He or she will then, with probability *b*, guess that it is an old word. In this situation, the participant will then guess, with probability *g*, for location “up”, or with probability 1-*g*, for location “down”. If the participant guessed, with probability 1 – *b*_pos, that the item is new, the answer is “new”. For words that originated from other sources, the probabilities of the branches are to be understood in a similar way as above, just with different subscripts representing parameters. Notably, the model equations were formulated in such a way that the *a* and *g* parameters represent the probabilities of guessing the metaphor-*congruent* vertical location based on the valence of each stimulus irrespective of one particular source location (see Arnold et al., 2013; Kuhlmann et al., 2012).

To achieve mathematical identifiability, additional constraints must be imposed on the parameters. The table in Appendix B1 shows the range of identifiable models, based on the submodels proposed by Bayen and colleagues (1996). We initially focused on models that include separate *D*_up_pos, *D*_down_pos, *D*_up_neg, and *D*_down_neg parameters in order to be able to replicate a possible effect of metaphor congruency on item memory as reported by Crawford et al. (2014). The analysis strategy was thus to leave the different *D* parameters free to assume different values in parameter estimation, then use hypothesis testing to compare *D* parameters for congruent items (i.e., items of types up_pos and down_neg) and those for incongruent items (of types up_neg and down_pos). We chose the most parsimonious submodels (with fewest separate parameters) allowing for different *D* parameters in order to avoid diluting the information in the data across overly many parameters and thus, we focused on submodels 5b and 5c (see Appendix B, Table B1). A model selection procedure as described in Appendix B was then conducted in order to settle on one of these two submodels for the hypotheses tests. This was followed by a goodness of fit test for the selected submodel to see whether it provided an adequate description of the data or whether more complex submodels would need to be considered (which – to foreshadow – was never the case). For parameter estimation in this and all subsequently reported experiments, we use a Bayesian hierarchical latent-trait approach (Klauer, 2010), for details see Appendix B. Group-level mean estimates of the posterior distribution for the parameters and corresponding 95% Bayesian credibility intervals (BCI) are reported in Appendix E. A 95%BCI of the differences between parameters in the posterior distribution that excludes zero was considered statistically substantial.

All finally selected models fitted well: Group 1 **a:** Model 5c: p_{T1} = .480 and p_{T2} = .443; **b:** Model 5b: p_{T1} = .389 and p_{T2} = .432; Group 2 **a:** Model 5b: p_{T1} = .322 and p_{T2} = .364; **b:** Model 5c: p_{T1} = .455 and p_{T2} = .382; Group 3 **a:** Model 5c: p_{T1} = .462 and p_{T2} = .544 **b:** Model 5b: p_{T1} = .576 and p_{T2} = .515. Group-level mean estimates of the posterior distribution for the relevant parameters and corresponding 95BCIs are reported in Appendix E.

In this and the following experiments, our hypotheses about metaphor-congruence and valence were assessed by means of parameter contrasts such that, pooling across both presentation locations (up vs. down), parameters indicating memory in congruent trials were subtracted from those indicating incongruent trials. For item memory, this was done by testing the parameter contrast *D*_incong, which is the parameter difference 1/2*(D_down_pos + D_up_neg - D_up_pos - D_down_neg), to reveal any hypothetical congruence-related effect in each study. Values of D_incong > 0 indicate an incongruency bias in item memory. Accordingly, negative values (D_incong < 0) would indicate a congruency bias in item memory. By pooling across „up“ and „down“ locations, this contrast is also not confounded with possible effects of location. Similarly, hypothetical effects of valence were tested using the parameter contrast D_valence, which is the parameter difference 1/2 * (D_down_neg + D_up_neg - D_down_pos - D_up_pos). Values of D_valence > 0 indicate a negativity bias in item memory while negative values would indicate a positivity bias. For congruence-related effects on source guessing we tested the parameter difference a_cong = a_pos_cong - (1 - a_neg_cong) which is an unconfounded estimate for the congruence effect in guessing. Congruence-related guessing corresponds to values of a_cong > 0.

Figure 2 shows the estimates of the guessing bias, a_cong, and Table 2 presents the parameter contrasts relevant to the hypotheses. Bayesian posterior tests are used to test one-tailed hypotheses, and a posterior probability *(pp)* of a null effect smaller than .05 is considered statistically substantial[2].

Table 2

Estimates for selected parameter contrasts (Group level) in experiments 1a, 1b and 3a

Instructions:	Stimulus only		Stimulus location		Stimulus location Metaphor	
parameter	*M*(*SD*)	95%BCI	*M*(*SD*)	95%BCI	*M*(*SD*)	95%BCI
			Experiment 1a			
D_incong	.08 (.06)	[-.03, .21]	.02 (.04)	[-.06, .11]	-.01 (.04)	[-.09, .08]
D_valence	.05 (.06)	[-.06, .15]	.04 (.04)	[-.04, .13]	.00 (.04)	[-.08, .09]
a_cong	.19 (.06)	[.06, .32]	.17 (.05)	[.06, .28]	.22 (.07)	[.07, .36]
			Experiment 1b			
D_incong	-.03 (.06)	[-.16, .07]	.03 (.06)	[-.08, .16]	-.02 (.06)	[-.15, .09]
D_valence	-.10 (.05)	[-.20, .01]	-.16 (.05)	[-.27, -.05]	-.05 (.06)	[-.16, .06]
a_cong	.11 (.04)	[.02, .20]	.16 (.05)	[.04, .27]	.22 (.05)	[.12, .32]

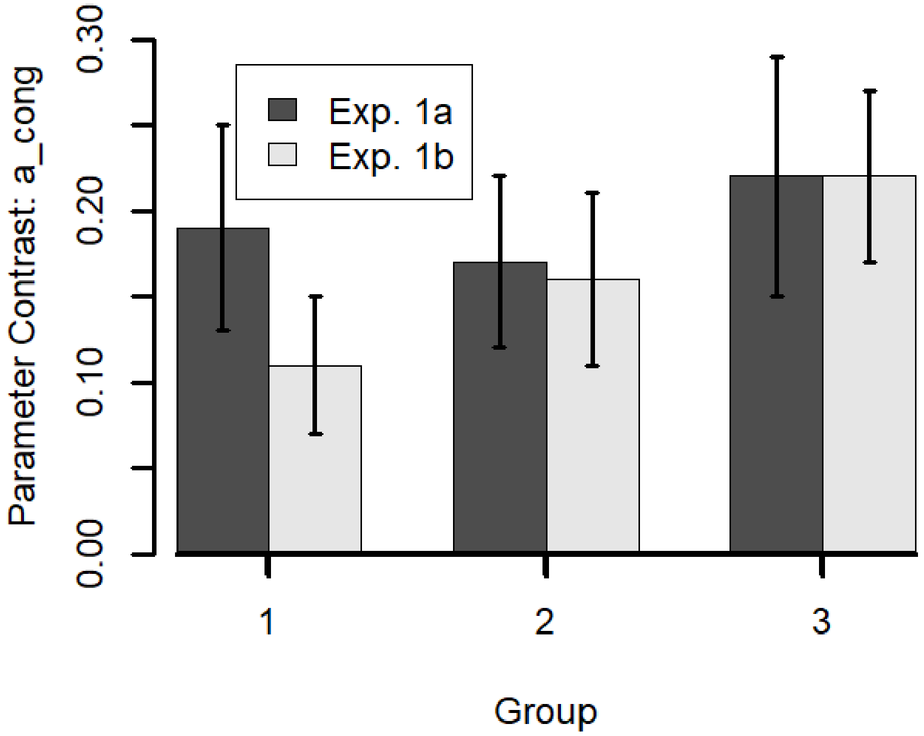
[2] We conduct hypothesis tests by means of the posterior distribution of the parameter contrast to be tested. This means that for a one-sided test of $H_0: \mu \leq 0$ against $H_1: \mu > 0,$ we use the posterior distribution of μ to compute the posterior p value, *pp* which is the posterior probability, given the data, of (parameter values consistent with) H_0, $pp = P(H_0) = 1 - P(H_1),$ and reject H_0 in favor of H_1 if *pp* is small (e.g., Casella & Berger, 1987), for example, if *pp* < .05. For a two-sided hypothesis $H_0: \mu = 0$ against $H_1: \mu \neq 0,$ this approach involves computing the 95% Bayesian credibility interval and rejecting H_0 if the value zero is not in the interval (Kruschke, 2011, chap. 12). Another common approach to hypothesis testing in Bayesian statistics is based on model selection via Bayes factors contrasting two models, one representing the H_0 and the other the H_1. There is a large literature discussing the relative merits of both approaches (see, e.g., Kruschke & Liddell, 2018, for a summary).

			Experiment 3a			
D_incong	-.01 (.05)	[-.12, .08]	-.04 (.05)	[-.16, .07]	.01 (.05)	[-.09, .10]
D_connotation	-.03 (.05)	[-.13, .06]	-.05 (.05)	[-.15, .06]	-.01 (.06)	[-.10, .07]
a_cong	.13 (.07)	[-.00, .27]	.26 (.06)	[.14, .38]	.15 (.06)	[.03, .27]

Note. The contrasts are defined as follows: D_incong: 1/2*(D_down_pos + D_up_neg - D_up_pos - D_down_neg); D_valence: 1/2 * (D_down_neg + D_up_neg - D_down_pos - D_up_pos); a_cong = a_pos_cong - (1 - a_neg_cong).

Figure 2

Estimates for source guessing parameters (group-level) from the three groups in Experiment 1a and 1b.



Note. Error bars indicate 95%BCI. Instruction in each group: Group 1: Stimulus only; Group 2: Stimulus + location; Group 3: Stimulus + location + Metaphor

Metaphor-incongruency effects on item memory were not found (confirming H1, **a:** Group 1: *pp* = .12; Group 2: *pp* = .30; Group 3: *pp* = .55, **b:** Group 1: *pp* = .72; Group 2: *pp* = .29; Group 3: *pp* = .63). An explanation could be that the limited attentional resources during encoding were allocated more to the spatial information and less to the valence of the stimuli. Consequently, the expectation-violation effect proposed by AEH was potentially reduced due to the lack of attention to valence information, which, as a result, would have diminished the differences in memory between metaphor-congruent and metaphor-incongruent presented words. Neither did we find evidence for negative words being remembered better than positive, in any group (*pp*'s in all groups > .05, both **a** and **b**).

H2 regarding metaphor-congruent source guessing biases received substantial support across all three groups (**a:** Group 1: *pp* = .002; Group 2: *pp* < .001; Group 3: *pp* = .002; **b:** Group 1: *pp* = .01; Group 2: *pp* = .002; Group 3: *pp* < .001). H3 regarding the influence of instructions on metaphoric effects was not supported in **a** (*pp* for the linear increase in the congruency bias in source guessing across groups = .39) whereas there was some support in **b** (*pp* = .04). Higher awareness, especially of both the spatial information *and* the metaphoric association, did not very strongly influence source guessing.

In general, the expected effects of metaphor-(in)congruency were only demonstrated on source guessing biases but not on item memory. Across the three groups of both experiments, the metaphor-congruent source guessing biases were generally consistent, in line with previous research and supporting the existence of metaphoric effects on cognition. In a situation in which participants cannot recall where a stimulus had been presented, the metaphoric association can serve as a guide for participants to generate the guessing responses. The lower estimates of item memory parameters *D* as found in Experiment 1**b** (as compared to Experiment 1**a**) indicate that participants found it harder to memorise emojis than words. A possible reason for this could lie in the greater difficulty of using verbal rehearsal in the case of emojis. Verbal rehearsal has been widely acknowledged as a useful strategy to improve memory performance (Dark & Loftus, 1976; Davachi et al., 2001; Forsberg et al., 2019; Woodward et al., 1973). Words, as descriptively richer stimuli may be easier to be silently (verbally) rehearsed during the encoding phase, which would facilitate memorisation. Emojis, in contrast, do not share this feature.

In this and all subsequent experiments we assessed the possibility we assessed the possibility that source memory varies dependent on the absolute location (i.e., up vs. down) and metaphor-congruency (i.e., congruent vs. incongruent location). However, the Mpt models used across all studies for evaluation of source guessing were of type 5b or 5c (see Appendices B). Models of this type do not allow us to test the possibility that the strength of source memory (parameter d) differs between the up and down location. Although we did not hypothesize that there would be such effects, we nevertheless wished to assess this possibility. We therefore, separately, fitted models of type 6c and 6d for all studies (see Appendices B and D), which allowed for an estimation of location-dependent source memory parameters *d* for positively versus negatively valenced stimuli (respectively, stimuli with “high” or “low” physical connotation in Experiment 4). In each of these models, we evaluated the difference between the *d* parameters found for congruent trials (source memory for stimuli presented up vs. down) minus the *d* parameters found for incongruent trials. In all cases and experiments reported in this manuscript, the credibility intervals included zero, meaning that on the basis of this additional modelling, there were no indications for credible differences, regardless of direction or valence, between source-memory estimates for congruent and incongruent trials. For a summary of all relevant source memory parameters across all experiments, and the critical evaluation, see Appendix D.

Experiment 2

The purpose of Experiment 2 was to provide further corroborating evidence on the hypotheses that substantial metaphoric effects can be found for source guessing biases, but not for item memory. Because of the COVID-19 pandemic, face-to-face experiments were banned in the U.K., instead of conducting online studies, a series of lab-based replication studies in China was conducted, in order to provide better control of the experimental environment than seemed possible via online studies. Specifically, Experiments 2a and 2b replicated Experiment 1a and 1b respectively, in a Mandarin speaking context.

On the basis of the previous experiments, we expected that metaphor-incongruency would have no effect on item memory, whereas source guessing biases with respect to metaphor-congruent locations were expected to replicate.

Another modification in this experiment was that only the Group 1 “stimulus-only” instruction condition was kept. The reason for this was that in Experiment 1 the instruction did not affect memory and guessing.

Method

Participants. a: *N* = 47, **b:** N = 43 Mandarin native speakers were recruited from Wuhan University to take part in this study, of which 25 (**a**) and 18 (**b**) were male, mean age **a:** 19.74 years (*SD* = 1.65); **b:** 19.35 years (SD = 1.12). Participants received 1 course credit or ¥10 for their participation.

Design. Experiment 2 used a within-subjects design with factors valence (positive vs. negative) and verticality (up vs. down).

Hypotheses. H1: There is no metaphor-incongruency effect on item memory; H2: There are metaphor-congruent source guessing biases; H3: There is a negativity advantage on item memory.

Materials. a: 132 words were selected from Affective Norms for English Words (ANEW, Bradley & Lang, 1999). They were then translated to two-character Mandarin words for a pre-test (*N* = 38) to obtain subjective valence ratings from Chinese participants. Based on the pretest, 40 positive words and 40 negative ones were selected with a mean valence rating 7.21 (*SD* = 0.34) and 2.65 (*SD* = 0.31), respectively, on a scale ranging from 1 to 9 indicating negative to positive valence. The mean rating of the positive words differed from the mean rating of the negative ones, across all participants, *t*(78) = -62.4, *p* < .001. See Appendix A.3 for a full list of the selected materials.

b: The same 80 emojis from the Lisbon Emoji and Emoticon Database (LEED, see Rodrigues et al., 2018) as used in Experiment 1b were also used here and pre-tested (*N* = 38) again to obtain subjective valence ratings from Chinese participants. Considering the relatively low estimates of item memory in Experiment 1b, Experiment 2b reduced the number of materials in order to lower the difficulty level. 32 positive emojis and 32 negative ones were selected with mean valence ratings 6.97 (*SD* = 0.36) and 2.92 (*SD* = 0.27), respectively, on a scale ranging from 1 to 9 indicating negative to positive valence. The mean ratings of positive and negative emojis differed significantly, $t(62) = -50.72$, $p < .001$. See Appendix A.4 for a full list of the selected materials.

Procedure. Participants were tested individually in laboratory rooms at Wuhan University. The procedure was identical to Experiment 1a except that in the instructions at the beginning of the experiment, all participants were only told to memorise the stimuli (i.e., there was no manipulation of awareness).

Results. For model selection and estimation procedures see Appendix B, for estimates of the relevant parameter contrasts see Table 3. The final selected model fitted well: **a:** Model 5b: $p_{T1} = .535$ and $p_{T2} = .502$; **b:** 5b: $p_{T1} = .540$ and $p_{T2} = .577$. See Table 3 for the group-level mean estimates of the relevant parameter contrasts. Similar to previous experiments, 95%BCI and Bayesian posterior probability (*pp*) were used to report hypotheses tests.

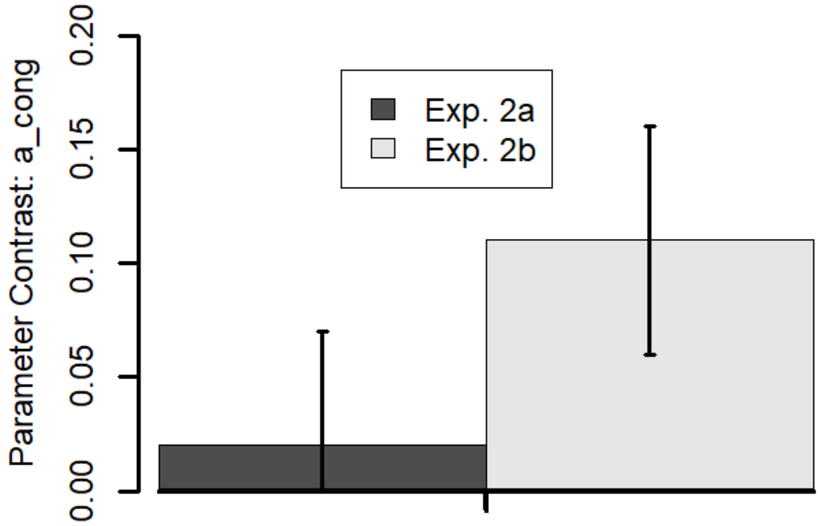
H1 regarding no metaphor-incongruency effects on item memory was supported, *pp* > .05 in both **a** and **b**. H2 regarding metaphor-congruent source guessing biases was not supported, *pp* = .32, in **a**, but was supported in **b**, *pp* = .01. H3 regarding a negativity advantage on item memory was supported, **a:** *pp* = .01; **b:** *pp* = .05. Memory for negative items was found to be better than for positive items. In contrast, a negativity advantage on item memory was not detected in Experiment 1b, Group 1. It appears that the experiments run in China (2a and 2b) were more susceptible to pick up a negativity effect than the ones run in Britain. Possible reasons for this await further investigation.

Table 3. *Estimates for Parameter Contrasts (Group-Level) in Experiments 2a, 2b and 3b*

parameter	*M*(*SD*)	95%BCI	*M*(*SD*)	95%BCI	*M*(*SD*)	95%BCI
	Experiment 2a		*Experiment 2b*		*Experiment 3b*	
D_incong	-.04 (.04)	[-.13, .03]	-.02 (.05)	[-.13, .08]	-.01 (.03)	[-.09, .05]
D_valence	.08 (.03)	[.01, .16]	.08 (.04)	[-.01, .17]	.02 (.03)	[-.04, .08]
a_cong	.02 (.05)	[-.08, .12]	.11 (.05)	[.01, .21]	.34 (.06)	[.20, .47]

Figure 3

Estimates for source guessing parameter contrast (group-level) from Experiments 2a and 2b.



Experiment 3

According to the simulation theory (Barsalou, 1999, 2008), if the “good is up” metaphor plays a role in processing valenced stimuli, it is supposed to function via the automatic activation of the related physical concept, verticality. To assess the plausibility of this proposed mechanism, Experiment 3 was designed to test, in the UK and in China, whether concrete concepts with “up” and “down” vertical connotations (e.g., “sky”, “cellar”) trigger the same effects as words with positive and negative valence, respectively.

Experiment 3a: United Kingdom

Method

Participants. *N* = 105 English native speakers were recruited from Cardiff University, United Kingdom, to take part in this study, of which 16 were male, mean age = 20.22 years (*SD* = 3.04). Participants took part in exchange of 1 course credit.

Design. We used a mixed design with factors connotation (high vs. low, within-subjects), verticality (up vs. down, within-subjects), and instruction (stimulus-only vs. stimulus-location vs. stimulus-location-association, between-subjects).

Hypotheses. Hypothesis 1 (H1): There is no connotation-incongruency effect on item memory. Hypothesis 2 (H2): There are connotation-congruent guessing biases; Hypothesis 3 (H3): Higher awareness of the spatial information and the connotation-verticality association was expected to increase the connotation-congruent source guessing biases. Considering that we changed from valence to connotation, we wanted to check whether instructions would make a difference in this case.

Materials. 80 words, half with high vertical connotation (e.g., “SKY”) and the other half with low vertical connotation (e.g., “PIT”), were selected from the materials that Lebois and colleagues (2015) used in their research on semantic processing, based on the provided norms on verticality (see Table 4). On a scale ranging from -9 to +9 indicating low to high vertical connotation, the mean ratings of high and low words were 5.75 (*SD* = 1.48) and -3.73 (*SD* = 1.03) respectively, which differed significantly, *t*(78) = 33.24, *p* <. 001. t-tests show no differences between high and low words on either length or syllable length, see Table 4. See Appendix A5 for a full list of the selected materials.

Table 4

Comparison of High and Low Connotation Words Stimuli (M±SD) in Experiment 3a

	high (*M*±*SD*)	low (*M*±*SD*)	*t*	*df*	*p*
Connotation	5.75±1.48	-3.73±1.03	33.24	78	.000***
Length	5.45±1.81	5.03±1.61	1.11	78	.27
Syllable	1.58±0.71	1.43 ±0.59	1.02	78	.31
Frequency M	2,037	2,029			
Frequency SD	2,660,414	2,615,766			

Note. “Length” = number of letters per word, “Syllable” = number of syllables per word; *** *p* < .001. The frequencies according to “Frequency English Web 2021” (absolute number of occurrences in that corpus during one year).

Procedure. This experiment was conducted online due to the COVID-19 pandemic when lab-based tests were not allowed in the UK. Informed consent was obtained from each participant before the experiment started. Moreover, due to the limitations of online data collection, the screen size and settings depended on participants' own end devices, so could not be set uniformly vertical (i.e., in portrait orientation) as was done in the previous, lab-based experiments. All other procedures were identical to Experiment 1a.

Results and Discussion

Due to concerns about data quality in online data-collection, the data were trimmed based on the average time taken for completing the test phase using the Tukey criterion. That is, the completion time needed to be within the range of [Q1 - 3*IQR, Q3 + 3*IQR], where Q1 (Q3) represent the lower (upper) quartiles, and IQR represents the interquartile range (see Tukey, 1977). Three participants were excluded because of very long times.

A similar adapted 2HTSM model as in Experiment 1a was used, except that positive/negative valence was substituted by high/low vertical connotation. Apart from this modification, the same model selection and parameter estimation procedures as in Experiment 1a were conducted (see Appendix B). All final selected models fitted well: Group 1 Model 5b: p_{T1} = .478 and p_{T2} = .298; Group 2 Model 5b: p_{T1} = .369 and p_{T2} = .216; Group 3 Model 5b: p_{T1} = .619 and p_{T2} = .286; for estimations of the relevant parameter contrasts see Table 2.

H1 regarding connotation-incongruency effects on item memory was not supported in any of the three groups, Group 1: *pp* = .88; Group 2: *pp* = .44; Group 3: *pp* = .62. H2 regarding connotation-congruent source guessing biases was substantially supported across the three groups (Group 1: *pp* = .02; Group 2: *pp* < .001; Group 3: *pp* = .006). H3 regarding the influence of instruction on item memory and source guessing was not supported (*pp* for the linear increase across groups = .36). Again, the absence of the predicted connotation-incongruency effect on item memory in all three groups implies that the expected enlarging effect from higher awareness was not found.

Taken together, when using concrete words with high and low connotations in the same recognition memory paradigm, the expected effects were demonstrated on source guessing biases, but not on item memory. The connotation-congruent guessing biases were quite consistent, analogous to the metaphor-congruent guessing tendencies demonstrated in Experiments 1a and 1b. In the current, entirely physically grounded context, the congruency between presentation location and stimulus' vertical connotation provides an expectation which participants can use to guess the source when their source memory fails[3]. The similarity of metaphor-congruent and connotation-congruent source guessing tendencies supports the argument that the metaphoric implication is mapped onto the physical dimension of verticality.

The connotation-incongruency did not show any influence on item memory. This lends some support to the assumption that metaphoric effects indeed do not affect item memory. According to the simulation theory (Barsalou, 1999, 2008), if the “good is up” metaphor plays a role in processing valenced stimuli, it is supposed to be mediated via the automatic activation of the related physical concept, verticality.

Experiment 3b: China

Participants. In a within-subjects design, N = 42 Mandarin native speakers were recruited from Wuhan University to take part in this study, of which 24 were male, mean age = 19.86 years (*SD* = 1.03). Participants received course credit or ¥10 for their participation.

Design. The design was the same as in Experiment 3a except only the instructions condition without mentioning of locations was used. Otherwise, hypotheses were the same as in Experiment 3a.

Materials. 96 words from Lebois and colleagues' (2015) research materials were directly translated (without revision) to two-character Mandarin words. Based on the pretest, eighty words, half with high verticality connotations, the other half with low verticality connotations, were selected with mean ratings 6.41 (*SD* = 0.59) and 3.16 (*SD* = 0.73), respectively, on a scale ranging from 1 to 9 indicating low to high connotation. The ratings differed significantly between words with low and high verticality connotation, *t*(78) = -21.94,

[3] These results have to be considered in the light of the up word subset having slightly greater verticality implication (see Materials section) than the down word subset.

p < .001. See Appendix A6 for a full list of the selected materials.

Procedure. The procedure was identical to Experiment 3a except that 3b was a lab-based study and the screen was set vertically as in Experiment 1-2 in order to increase the salience of the vertical dimension in terms of the difference between the up and down locations.

Results

Identical model selection and estimation procedure were conducted as in Experiment 3a (see Appendix B). The final selected model fitted well: Model 5b: p_{T1} = .442 and p_{T2} = .325. See Table 3 for the group-level mean estimates of the relevant parameter contrasts.

H1 regarding no connotation-incongruency effects on item memory was supported, *pp* = .41. Similarly, connotation-incongruency effects on item memory were not detected in Experiment 3a either. H2 regarding connotation-congruent source guessing biases was substantially supported, *pp* < .001.

Discussion

The hypothesis on the absence of metaphor-incongruency or connotation-incongruency effects on item memory could be maintained, and this conclusion also holds for words with physically vertical location connotations.

Source guessing biases were consistent across both sub-experiments (Experiment 3). As the N’s were small for all three experiments run in China (Experiments 2a, 2b, and 3b), we calculated post-hoc power analyses for each and found, for the critical hypothesis H2, effect sizes of d=.33 (E2a), d=2.2 (E2b), and d=5.66 (3b). The achieved power with the current N’s are $1\text{-}\beta$=.60 (2a), $1\text{-}\beta$=1.0 (2b), and $1\text{-}\beta$=1.0 (3b). Thus, whilst there remains concern about the reliability of the H2-related null effect in E2a, the support for H2 as demonstrated in the other two Chinese experiments (2b and 3b) appears robust.

Potentially, in Experiment 3, we observed the stronger and more consistent source guessing effects for physical connotation because the conceptual link between physical stimuli and spatial location is stronger compared to the link between valence and spatial location, with the former creating stronger expectations of congruency to guide source guessing.

General Discussion

This research presents six experiments to investigate the metaphoric effects of “good is up” in the context of recognition memory. Based on the extant studies, we investigated metaphoric effects on source guessing strategies and item discrimination. More specifically, we expected metaphor-congruent source guessing biases but no better discrimination for items presented at metaphor-incongruent locations. Substantial evidence was only obtained for source guessing biases as being linked to the metaphor, not for item discrimination. This pattern of results was obtained in our series of lab-based experiments as well as in an online experiment (3a). It also speaks to the robustness of our findings that the same general pattern was also obtained using different kinds of materials such as words, emojis and unfamiliar language characters, and different cultural backgrounds, such as the UK and China.

Summarising our argument with respect to the different memory components within the recognition situation, we, first, did not expect, and did not find, substantive metaphoric effects on item memory, in agreement with previous literature (Bell et al., 2012; Ehrenberg & Klauer, 2005; Kroneisen & Bell, 2015; Küppers & Bayen, 2014). Second, we did not find substantive metaphoric effects on source memory, which we explain with our assumption with regard to the strength of metaphoric influences: As congruence/incongruence with respect to metaphor is more indirect and complex than with respect to stereotypes (see Introduction), the strength of metaphoric influences as manifest in our paradigm is probably to be classified as “weak”, as in the context of the relevant literature (Bayen & Kuhlmann, 2011; Bayen et al., 2000; Kuhlmann et al., 2012). Thirdly, for guessing, we submit that similar to assuming schema-congruent source guessing to be the default (Schaper et al., 2019, 2023) we also assume an analogue default mechanism to hold for metaphors, as well. As a result, we consistently observe metaphor-congruent source guessing. These aspects will be discussed in order.

The “good is up” metaphor does not influence memory

Regarding item memory, our results did not favour an effect of metaphor congruency on item memory. Previous studies which reported no effects of the “good is up” metaphor on recognition memory all used paradigms other than that used in the current research, such as investigating person memory after an impression formation task (McMullan, 2016) or manipulating vertical locations of stimuli in the retrieval phase instead of the encoding phase (Experiment 3 from Crawford et al., 2014). The present research started with a conceptual replication of Experiment 2 from Crawford et al (2014), which suggests that the vertical location of valenced words in the encoding phase can influence recognition memory. Our conceptual replication did not support their argument. Assuming the expectation of metaphoric congruency between valence and verticality is not strong enough, the stimuli presented at metaphor-incongruent locations cannot trigger strong feelings of violation in the encoding phase, and consequently do not receive more attention and elaboration. As a result, there are no substantial differences in item memory between metaphor-congruent and incongruent stimuli. The question arises of why the strength of expectation was insufficient. Three possible explanations are (a) that the materials do not sufficiently express valence, (b) the verticality dimension is not sufficiently activated, or (c) the association between “valence” and “verticality” is a priori rather weak. The latter is implied by our earlier argument about this association being more complex and indirect in case of a metaphor, as compared to implications of social stereotypes (see Introduction). Note also that valence is associated with many other (possibly competing) concepts other than verticality, the same being true for verticality itself (see Schubert, 2005). The different results of metaphoric effects on item memory can be partially due to a motivational factor as well. At learning, the motivation for participants to encode the stimuli via the spatial metaphor might not be strong. In other words, the use of metaphoric information does not in any obvious way contribute to better memory performance, which might lower participants’ motivation to use that information when encoding stimuli. However, at test, the motivation to use the spatial metaphor as heuristic might be relatively elevated because of the requirement of an immediate response even when participants do not remember the stimulus’ location.

Moreover, the differences in analytic approaches might contribute to the discrepancies in results. An advantage of the 2HTSM model analysis is that it provides an independent estimation of multiple components within recognition: item discrimination per se, versus guessing components that actually form an integral part of memory in recognition paradigms. The current analysis allows for a more accurate estimate of these different memory components in a recognition situation while the conventional ANOVA measures used in Crawford et al (2014) fully neglect the contribution of these components.

In research investigating source memory, it was suggested that the inconsistency effect (on source memory) as predicted by the Attention Elaboration Hypothesis (AEH) only occurs when the expectation strength is high (Bayen & Kuhlmann, 2011; Bayen et al., 2000; Küppers & Bayen, 2014; Kuhlmann et al., 2012). This argument potentially explains the null metaphoric effects on source memory as observed here.

Guessing strategies

Metaphor-congruent source guessing biases suggest that, as hypothesized, an activated “good is up” metaphor creates the expectation of metaphor-congruency, expressing itself via source guessing responses when participants are unable to recall the presented location of a valenced stimulus. This is in line with previous research on schema-congruent guessing biases which suggests that source guessing is particularly biased when a congruency is implied by an existing schema connecting a particular item with a particular source (Bayen et al., 2000; Ehrenberg & Klauer, 2005; Kuhlmann et al., 2016; Wulff & Kuhlmann, 2020). Interestingly, the present research indicates that metaphors can function in the source guessing process in a similar way as schemas do, creating an expectation and guiding guessing responses in a direction congruent with the expectations. Some evidence for metaphor-congruent source guessing biases was obtained even when the metaphor was not mentioned explicitly as in the stimulus-only and stimulus-location awareness groups. This suggests that the association between valence and verticality expressed in metaphor-congruent source-guessing bias was implicitly represented in people’s minds. In essence, this is an interesting insight as we have argued above that establishing congruency/incongruency involving the meaning of a metaphor is likely to be more complex (and might have less of an impact) than when invoking a schema based on a social stereotype.

Combined with the connotation-congruent source guessing biases demonstrated in Experiments 3a and 3b, the results support the argument that valence, as an abstract concept, is mapped onto the vertical dimension in way similar to the concept of physical height that underlies vertical connotations in the processing of material objects (e.g., “sky” and “up” vs. “pit” and “down”). It is possible that when processing a valence concept, the simulation of the metaphor-congruent vertical location is activated automatically, as suggested by embodied cognition theory (Barsalou, 2008; Lakoff & Johnson, 1980, 2008). Presumably, although the “good is up” metaphor creates the association between valence and verticality, it is still not as strong and obvious as the link between physically grounded concepts and the corresponding spatial locations. As the physical connotation is more concrete and direct, and the psychological establishment of congruency/incongruency is more indirect in the case of a metaphor, thinking about the word “sky” may immediately and automatically activate the simulation of sky which includes the bodily experience of looking up. In contrast, valence, as an abstract concept per se, may require additional mental mediation to be mapped onto vertical locations (see Introduction).

Negativity advantage on item memory is inconsistent

The hypothesis of a negativity advantage was supported in two of the Experiments conducted in China, that is, Experiment 2a and 2b, but in none of the other experiments. In general, it seems to be easier to demonstrate a negativity advantage on item memory when using word stimuli as compared to emojis. As discussed in Experiment 1b, this might be due to the somewhat comical feature of emojis, which may interfere with the impression of negativity and as such may hinder the deeper cognitive elaboration, often attributed to negativity. Another potential reason is the generally poorer item memory for emojis compared to word materials. It may be harder to identify differences as a function of valence if item memory is poor to begin with[4].

[4] As a reviewer pointed out, these problems are even exacerbated when considering

Conclusion

The “good is up” metaphor does not only play a role in our daily language use, but also plays a role in recognition memory. It provides a schema to guide people’s source guessing when they cannot recall the source of a piece of valenced information. The metaphor-congruent source guessing biases reveal the use of heuristics based on the valence-verticality association. However, previous findings of schema-related effects on item memory or source memory were not replicated in the current research, indicating that the metaphor may affect cognition in a more subtle way than expected.

neutrality in order to localize the effect: It may be that each of positivity and negativity actually have an effect, but it is also possible that the overall effect is mainly driven by just one of them, the other being close to neutrality.

References

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*, 480-500.

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source–item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language, 64*, 1–17. https://doi.org/10.1016/j.jml.2010.09.001

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. L. (2000). The use of schematic knowledge about sources in source monitoring. Memory & Cognition, 28(3), 480–500. https://doi.org/10.3758/BF03198562

Bell, R., Mieth, L., & Buchner, A. (2015). Appearance-based first impressions and person memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 456.

Bell, R., Buchner, A., Kroneisen, M., & Giang, T. (2012). On the flexibility of social source memory: a test of the emotional incongruity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1512.

Bott, F. M., Heck, D. W., & Meiser, T. (2020). Parameter validation in hierarchical MPT models by functional dissociation with continuous covariates: An application to contingency inference. Journal of Mathematical Psychology, 98, 102388. https://doi.org/10.1016/j.jmp.2020.102388

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report C-1, the center for research in psychophysiology, University of Florida.

Brainerd, C. J., Nakamura, K., Chang, M., & Bialer, D. M. (2019). Verbatim editing: A general model of recollection rejection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 1776.

Brunyé, T. T., Gardony, A., Mahoney, C. R., & Taylor, H. A. (2012). Body-specific representations of spatial location. Cognition, 123(2), 229–239. https://doi.org/10.1016/j.cognition.2011.07.013

Buchner, A., Bell, R., Mehl, B., & Musch, J. (2009). No enhanced recognition memory, but better source memory for faces of cheaters. Evolution and Human Behavior, 30(3), 212–224. https://doi.org/10.1016/j.evolhumbehav.2009.01.004

Casella, G., & Berger, R. L. (1987). Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem. *Journal of the American Statistical Association, 82*, 106–111. https://doi.org/10.2307/2289130

Casasanto, D., & Dijkstra, K. (2010). Motor action and emotional memory. Cognition, 115(1), 179–185. https://doi.org/10.1016/j.cognition.2009.11.002

Cook, G. I., Hicks, J. L., & Marsh, R. L. (2007). Source monitoring is not always enhanced for valenced material. Memory & Cognition, 35, 222-230.

Crawford, E. L., Cohn, S. M., & Kim, A. B. (2014). “Good is up” is not always better: A memory advantage for words in metaphor-incompatible locations. PLoS ONE, 9(9), e108269. https://doi.org/10.1371/journal.pone.0108269

Crawford, E. L., Margolies, S. M., Drake, J. T., & Murphy, M. E. (2006). Affect biases memory of location: Evidence for the spatial representation of affect. Cognition and Emotion, 20(8), 1153–1169. https://doi.org/10.1080/02699930500347794

Dark, V. J., & Loftus, G. R. (1976). The role of rehearsal in long-term memory performance. Journal of Verbal Learning and Verbal Behavior, 15(4), 479–490. https://doi.org/10.1016/S0022-5371(76)90043-8

Davachi, L., Maril, A., & Wagner, A. D. (2001). When keeping in mind supports later bringing to mind: Neural markers of phonological rehearsal predict subsequent remembering. Journal of Cognitive Neuroscience, 13(8), 1059–1070. https://doi.org/10.1162/089892901753294356

Dodson, C. S., Darragh, J., & Williams, A. (2008). Stereotypes and retrieval-provoked illusory source recollections. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(3), 460–477. https://doi.org/10.1037/0278-7393.34.3.460

Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. Emotion, 1, 5.

Dube, C., & Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 130.

Dudschig, C., de la Vega, I., & Kaup, B. (2015). What’s up? Emotion-specific activation of vertical space during language processing. Acta Psychologica, 156, 143–155. https://doi.org/10.1016/j.actpsy.2014.09.015

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. Journal of Experimental Social Psychology, 41(4), 369–387. https://doi.org/10.1016/j.jesp.2004.08.001

Fischer, E. (2014). Philosophical intuitions, heuristics, and metaphors. Synthese, 191(3), 569–606. https://doi.org/10.1007/s11229-013-0292-2

Forsberg, A., Johnson, W., & Logie, R. H. (2019). Aging and feature-binding in visual working memory: The role of verbal rehearsal. Psychology and Aging, 34(7), 933–953. https://doi.org/10.1037/pag0000391

Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology*, *49*, 280-286.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequence. Statistical Science, 7(4), 457–472. https://doi.org/10.1214/ss/1177011136Globig, L. K., Hartmann, M., & Martarelli, C. S. (2019). Vertical head movements influence memory performance for words with emotional content. Frontiers in Psychology, 10, 672. https://doi.org/10.3389/fpsyg.2019.00672

Groß, J., & Pachur, T. (2020). Parameter estimation approaches for multinomial processing tree models: A comparison for models of memory and judgment. Journal of Mathematical Psychology, 98, 102402. https://doi.org/10.1016/j.jmp.2020.102402

Guttentag, R. E., & Carroll, D. (1994). Identifying the basis for the word frequency effect in recognition memory. Memory, 2(3), 255–273. https://doi.org/10.1080/09658219408258948.

Guttentag, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. Journal of Memory and Language, 37(4), 502–516. https://doi.org/10.1006/jmla.1997.2532.

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS : An R package for hierarchical multinomial-processing-tree modeling. Behavior Research Methods, 50(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7

Jeffreys, H. (1961). *Theory of probability (3rd ed.)*. New York, NY: Oxford University Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. Psychological Bulletin, 114(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Kellen, D., & Klauer, K. C. (2014). Discrete-state and continuous models of recognition memory: testing core properties under minimal assumptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1795.

Kellen, D., & Klauer, K. C. (2018). Elementary signal detection and threshold theory. *Stevens' handbook of experimental psychology and cognitive neuroscience*, *5*, 1-39.

Kittay, E. F. (1990). Metaphor: Its cognitive force and linguistic structure. Oxford University Press.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. Psychometrika, 75(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. Personality and Social Psychology Bulletin, 26(9), 1074–1093. https://doi.org/10.1177/01461672002611005

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the “Who said what?” paradigm. Journal of Personality and Social Psychology, 75(5), 1155–1178.

https://doi.org/10.1037/0022-3514.75.5.1155

Kroneisen, M., & Bell, R. (2013). Sex, cheating, and disgust: Enhanced source memory for trait information that violates gender stereotypes. *Memory*, *21*, 167-181.

Kruschke, J. K. (2011). Doing Bayesian Data Analysis. Academic Press.

Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic bulletin & review*, *25*, 155-177.

Kuhlmann, B. G., & Touron, D. R. (2011). Older adults' use of metacognitive knowledge in source monitoring: Spared monitoring but impaired control. Psychology and Aging, 26(1), 143–149. https://doi.org/10.1037/a0021055

Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. Psychology and Aging, 31(8), 875–889. https://doi.org/10.1037/pag0000140

Kuhlmann, B. G., Symeonidou, N., Tanyas, H., & Wulff, L. (2021). Remembering and reconstructing episodic context: An overview of source monitoring methods and behavioral findings. In K. D. Federmeier & L. Sahakyan (Eds.), Psychology of Learning and Motivation. The Context of Cognition: Emerging Perspectives (Vol. 75, pp. 79–124). Academic Press. https://doi.org/10.1016/bs.plm.2021.06.002

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(5), 1365–1376. https://doi.org/10.1037/a0028147

Küppers, V. (2012). Source Monitoring: Source-Memory Effects and Guessing Strategies. [Doctoral dissertation, Universitäts-und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf].

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. Quarterly Journal of Experimental Psychology, 67(10), 2042–2059. https://doi.org/10.1080/17470218.2014.904914

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. Cognitive Science, 4(2), 195–208. https://doi.org/10.1016/S0364-0213(80)80017-6

Lakoff, G., & Johnson, M. (1999). Philosophy in the flesh: The embodied mind and its challenge to western thought. Basic Books.

Lakoff, G., & Johnson, M. (2008). Metaphors we live by. University of Chicago Press.

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. Cognitive Science, 39(8), 1764–1801. https://doi.org/10.1111/cogs.12174

Luodonpää-Manni, M., & Viimaranta, J. (2010). Metaphoric expressions on vertical axis revisited: An empirical study of Russian and French material. Metaphor and Symbol, 25(2), 74–92. https://doi.org/10.1080/10926481003715994

McAdoo, R. M., Key, K. N., & Gronlund, S. D. (2019). Task effects determine whether recognition memory is mediated discretely or continuously. *Memory & Cognition*, *47*, 683-695.

McMullan, R. (2016). The embodiment of “good” and “bad” via vertical space: An investigation of conceptual metaphor in impression formation. [Doctoral dissertation, Western Sydney University].

Meier, B. P., & Robinson, M. D. (2004). Why the Sunny Side Is Up: Associations Between Affect and Vertical Position. Psychological Science, 15(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Meiser, T., Sattler, C., & von Hecker, U. (2007). Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. Quarterly Journal of Experimental Psychology, 60(7), 1015–1040. https://doi.org/10.1080/17470210600875215

Michalkiewicz, M., & Erdfelder, E. (2016). Individual differences in use of the recognition heuristic are stable across time, choice objects, domains, and presentation formats. Memory & Cognition, 44(3), 454–468. https://doi.org/10.3758/s13421-015-0567-6

Millar, R. B. (2009). Comparison of hierarchical bayesian models for overdispersed count data using DIC and Bayes’ Factors. Biometrics, 65(3), 962–969. https://doi.org/10.1111/j.1541-0420.2008.01162.x

Palma, T. A., Garrido, M. V, & Semin, G. R. (2011). Grounding person memory in space: Does spatial anchoring of behaviors improve recall? European Journal of Social Psychology, 41(3), 275–280. https://doi.org/10.1002/ejsp.795

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The Adaptive Decision Maker. Cambridge University Press.

Peirce, J. (2009). Generating stimuli for neuroscience using PsychoPy. Frontiers in Neuroinformatics, 2(10), 1–8. https://doi.org/10.3389/neuro.11.010.2008

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. Behavior Research Methods, 51(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Province, J. M., & Rouder, J. N. (2012). Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences*, *109*, 14357-14362.

Quadflieg, S., Etzel, J. A., Gazzola, V., Keysers, C., Schubert, T. W., Waiter, G. D., & Macrae, C. N. (2011). Puddles, parties, and professors: Linking word categorization to neural patterns of visuospatial coding. Journal of Cognitive Neuroscience, 23(10), 2636–2649. https://doi.org/10.1162/jocn.2011.21628

Quent, J. A., Henson, R. N., & Greve, A. (2021). A predictive account of how novelty influences declarative memory. *Neurobiology of Learning and Memory*, *179*, 107382.

R Core Team. (2013). R: A language and environment for statistical computing.

Riskind, J. H., & Gotay, C. C. (1982). Physical posture: Could it have regulatory or feedback effects on motivation and emotion?. Motivation and emotion, 6, 273-298.

Rodrigues, D., Prada, M., Gaspar, R., Garrido, M. V, & Lopes, D. (2018). Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions. Behavior Research Methods, 50(1), 392–405. https://doi.org/10.3758/s13428-017-0878-6

Sasaki, K., Yamada, Y., & Miura, K. (2016). Emotion biases voluntary vertical action only with visible cues. Acta Psychologica, 163, 97–106. https://doi.org/10.1016/j.actpsy.2015.11.003

Schaper, M. L., Bayen, U. J., & Hey, C. V. (2023). Remedying the Metamemory Expectancy

Illusion in Source Monitoring: Are there Effects on Restudy Choices and Source Memory?. *Metacognition and Learning*, *18*, 55-80.

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metamemory expectancy illusion and schema-consistent guessing in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(3), 470.

Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. Journal of personality and social psychology, 89, 1.

Shapiro, L. (2019). Embodied cognition. Routledge.

Sherman, J. W., & Bessenoff, G. R. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science*, *10*(2), 106-110.

Singmann, H., Kellen, D., & Klauer, K. C. (2013). Investigating the other-race effect of Germans towards Turks and Arabs using multinomial processing tree models. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), Proceedings of the 35th annual conference of the cognitive science society (pp. 1330–1335). Cognitive Science Society.

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(4), 631–651. https://doi.org/10.1037/0278-7393.28.4.631

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 64, 583–616. https://doi.org/10.1111/1467-9868.00353

Stepper, S., & Strack, F. (1993). Proprioceptive determinants of emotional and nonemotional feelings. Journal of Personality and Social Psychology, 64, 211–220. https://doi.org/10.1037/0022-3514.64.2.211

Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley.

Wapner, S., Werner, H., & Krus, D. M. (1957). The effect of success and failure on space location. Journal of Personality, 25, 752–756. https://doi.org/10.1111/j.1467-6494.1957.tb01563.x

Wickham, H. (2016). Programming with ggplot2. In ggplot2. Use R! (pp. 241–253). Springer. https://doi.org/10.1007/978-3-319-24277-4_12

Woodward, A. E., Bjork, R. A., & Jongeward, R. H. (1973). Recall and recognition as a function of primary rehearsal. Journal of Verbal Learning and Verbal Behavior, 12(6), 608–617. https://doi.org/10.1016/S0022-5371(73)80040-4

Wu, Q., Kidd, E., & Goodhew, S. C. (2019). The spatial mapping of concepts in English and Mandarin. Journal of Cognitive Psychology, 31(7), 703–724. https://doi.org/10.1080/20445911.2019.1663354

Wulff, L., & Kuhlmann, B. G. (2020). Is knowledge reliance in source guessing a cognitive trait? Examining stability across time and domain. Memory & Cognition, 48(2), 256–276. https://doi.org/10.3758/s13421-019-01008-1

Appendices

Appendix A. Materials of all the experiments

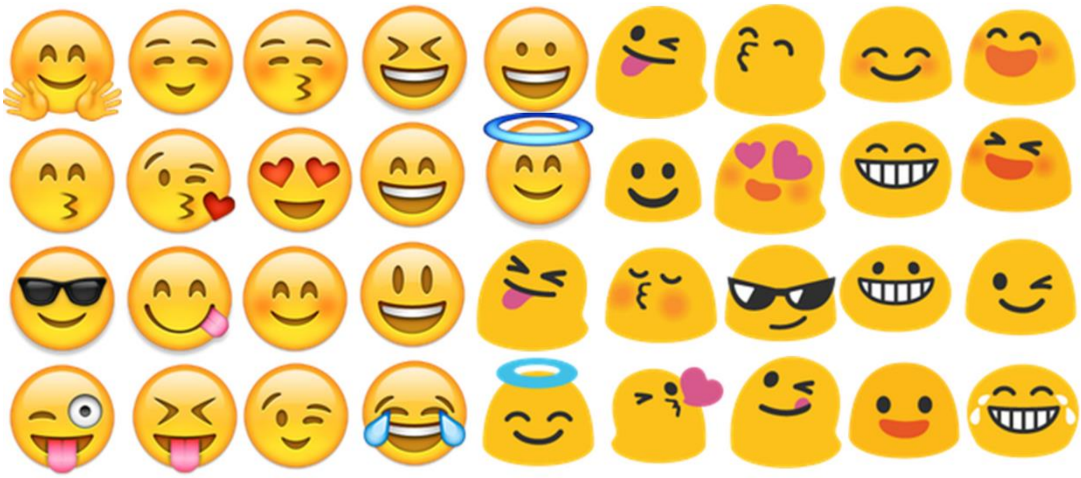
A1 Positive and negative words in Experiment 1a

Positive words: BEAUTY, BLESS, BLISS, BUNNY, CAKE, CAREFREE, COMFORT, COZY, CROWN, CUDDLE, DIGNIFIED, DOVE, DREAM, EASYGOING, FLOWER, GENTLE, GLAMOUR, GRATEFUL, HONEY, HOPE, INNOCENT, INSPIRE, INTELLECT, JEWEL, LEISURELY, POLITENESS, RESPECT, REWARD, SAFE, SCHOLAR, SECURE, SNUGGLE, TALENT, TENDER, TRIUMPH, TRUTH, WARMTH, WISE, WIT, ZEST;

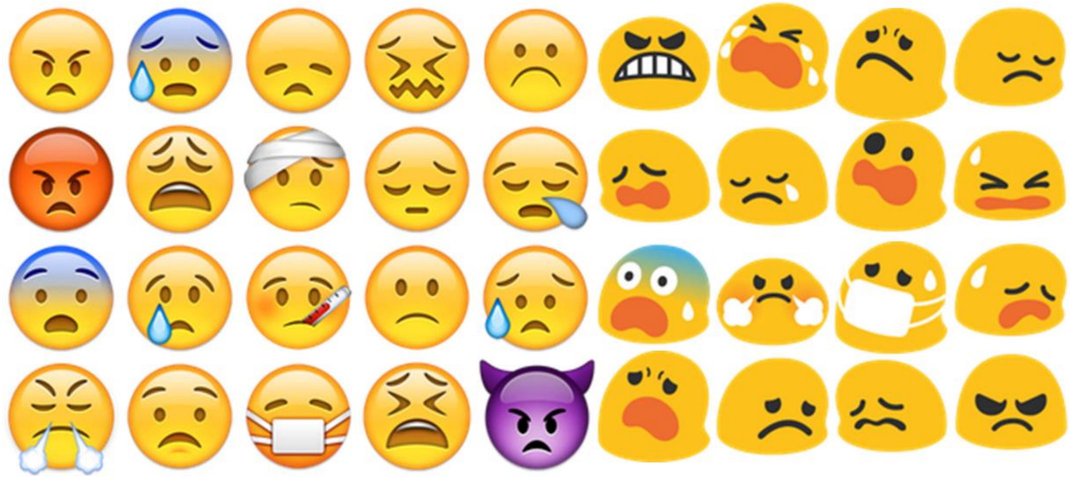
Negative words: BLISTER, BLOODY, BORED, COFFIN, CORPSE, COWARD, CRIMINAL, DEATH, DEFORMED, DETACHED, DISAPPOINT, DISCOURAGED, DUMP, FLABBY, FOUL, FRAUD, FRIGID, FUNERAL, GERMS, GREED, GRIEF, GRIME, HANDICAP, HINDER, HOOKER, IMMATURE, IMMORAL, INFECTION, INFERIOR, MOLD, MUTATION, NEGLECT, OBESITY, SCAR, SCUM, SLUM, TOMB, UNHAPPY, WASTE, WOUNDS.

A2 Positive and negative emojis in Experiment 1b

Positive emojis:



Negative emojis:



A3 Positive and negative Mandarin words in Experiment 2a

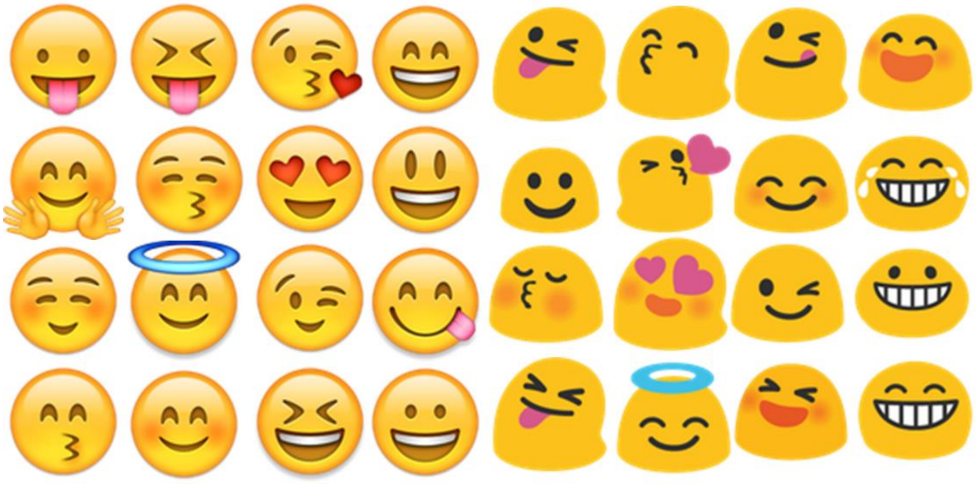
Positive words: 冠军，胜利，希望，可爱，乐观，智慧，自由，祝福，温暖，欣喜，温柔，美丽，明智，热情，耀眼，和平，奖励，浪漫，福气，感激，友善，贴心，尊重，迷人，钦佩，富有，魅力，拥抱，舒适，天堂，春天，安全，真理，礼貌，礼物，名望，天赋，鼓掌，接纳，启发;

Negative words: 虐待，背叛，残暴，罪犯，窒息，绑架，尸体，死亡，缺德，恶魔，恐慌，作弊，血腥，痛苦，肮脏，劣等，战争，贪婪，可怕，恐吓，坟墓，懦夫，疾病，埋

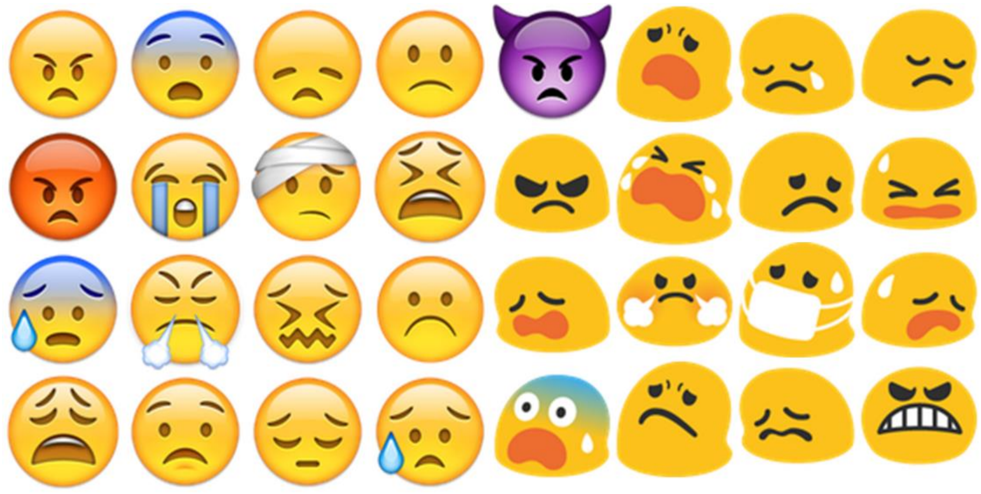
葬, 葬礼, 小偷, 沮丧, 轻蔑, 处决, 烦人, 棺材, 蟑螂, 悲哀, 污垢, 残障, 愤怒, 垃圾, 发霉, 失败, 杀手.

A4 Positive and negative emojis in Experiment 2b

Positive emojis:



Negative emojis:



A5 High and low connotation words in Experiment 3a

High connotation words: UMBRELLA, HAIR, HEAD, FOUNTAIN, PULPIT, TIP, FLOAT, BRANCH, CROWN, HILL, LARK, SPIRE, ARCH, ROOF, STEEPLE, CEILING, SMOKE, BALLOON, LIFT, MONUMENT, TORNADO, KITE, MISSILE, TOWER, FLY, AIRSHIP, EAGLE, HAWK, APEX, VOLCANO, LIGHTNING, MOUNTAIN, CLOUD, SKYSCRAPER, PLANE, ROCKET, SUN, MOON, SKY, STAR;

Low connotation words: SUBMARINE, DUNGEON, CHASM, ANCHOR, PIT, SEWER, ROOT, VALLEY, FALL, DIVE, BURROW, SUNSET, BASEMENT, FOUNDATION, CELLAR, SINK, DITCH, EARTH, GROUND, CAVE, HOLE, BASE, BASIN, DRAIN, SOIL, WORM, LAND, RUG, CARPET, TUMBLE, DIP, MAT, ASPHALT, FEET, HEEL, FLOOR, TOE, PUDDLE, SAND, SHOE.

A6 High and low connotation Mandarin words in Experiment 3b

High connotation words: 太阳，天空，顶点，星星，月亮，尖顶，尖端，火箭，飞机，老鹰，云朵，飞船，尖塔，飞翔，王冠，风筝，屋顶，导弹，气球，闪电，塔楼，云雀，钟楼，举起，火山，山脉，头盔，山丘，帽子，前额，领子，漂浮，喷泉，树枝，头发，烟雾，讲坛，台风，雨伞,假发;

Low connotation words: 深渊，沉没，坠落，潜艇，地窖，地洞，地牢，阴沟，陷阱，潜水，坟墓，树根，地基，棺材，地垫，盆地，洞窟，沟渠，峡谷，脚踏，脚跟，基座，排水，脚踝，漏洞，脚趾，双脚，袜子，土壤，鞋子，靴子，跌倒，地毯，日落，船锚，地板，拖鞋，尘土，地面，沙地.

Appendix B. Table of identifiable models, model selection and parameter estimation strategies

Table B1. Constraints of the Identifiable Submodels

	Positive	Negative
Model 6a	*D*_down_pos; *D*_up_pos = *D*_new_pos *d*_up_pos = *d*_down_pos *a*_pos_cong; *g*_pos_cong *b*_pos	*D*_up_neg; *D*_down_neg = *D*_new_neg *d*_up_neg = *d*_down_neg *a*_neg_cong; *g*_neg_cong *b*_neg
Model 6b	*D*_up_pos; *D*_down_pos = *D*_new_pos *d*_up_pos = *d*_down_pos *a*_pos_cong; *g*_pos_cong *b*_pos	*D*_down_neg; *D*_up_neg = *D*_new_neg *d*_up_neg = *d*_down_neg *a*_neg_cong; *g*_neg_cong *b*_neg
Model 6c	*D*_down_pos; *D*_up_pos = *D*_new_pos *d*_up_pos; *d*_down_pos *a*_pos_cong = *g*_pos_cong *b*_pos	*D*_up_neg; *D*_down_neg = *D*_new_neg *d*_up_neg; *d*_down_neg *a*_neg_cong = *g*_neg_cong *b*_neg
Model 6d	*D*_up_pos; *D*_down_pos = *D*_new_pos *d*_up_pos; *d*_down_pos *a*_pos_cong = *g*_pos_cong *b*_pos	*D*_down_neg; *D*_up_neg = *D*_new_neg *d*_up_neg; *d*_down_neg *a*_neg_cong = *g*_neg_cong *b*_neg
Model 5a	*D*_up_pos = *D*_down_pos = *D*_new_pos *d*_up_pos = *d*_down_pos *a*_pos_cong; *g*_pos_cong *b*_pos	*D*_up_neg = *D*_down_neg = *D*_new_neg *d*_up_neg = *d*_down_neg *a*_neg_cong; *g*_neg_cong *b*_neg
Model 5b	*D*_down_pos; *D*_up_pos = *D*_new_pos *d*_up_pos = *d*_down_pos *a*_pos_cong = *g*_pos_cong *b*_pos	*D*_up_neg; *D*_down_neg = *D*_new_neg *d*_up_neg = *d*_down_neg *a*_neg_cong = *g*_neg_cong *b*_neg
Model 5c	*D*_up_pos; *D*_down_pos = *D*_new_pos *d*_up_pos = *d*_down_pos *a*_pos_cong = *g*_pos_cong *b*_pos	*D*_down_neg; *D*_up_neg = *D*_new_neg *d*_up_neg = *d*_down_neg *a*_neg_cong = *g*_neg_cong *b*_neg
Model 5d	*D*_up_pos = *D*_down_pos = *D*_new_pos *d*_up_pos; *d*_down_pos *a*_pos_cong = *g*_pos_cong *b*_pos	*D*_up_neg = *D*_down_neg = *D*_new_neg *d*_up_neg; *d*_down_neg *a*_neg_cong = *g*_neg_cong *b*_neg
Model 4	*D*_up_pos = *D*_down_pos = *D*_new_pos *d*_up_pos = *d*_down_pos *a*_pos_cong = *g*_pos_cong	*D*_up_neg = *D*_down_neg = *D*_new_neg *d*_up_neg = *d*_down_neg *a*_neg_cong = *g*_neg_cong

*b*_pos	*b*_neg

Note. When only one item memory parameter is set equal to *D*_new_pos or *D_new_neg (Models 6a, 6b, 6c, 6d, 5b, 5c),* then it is for both valences either the parameter for metaphor-congruent stimuli (up_pos and down_neg) or for metaphor-incongruent stimuli (up_neg and down_pos). Models 6c and 6d differ in the same way as do Models 5b and 5c (see text).

B2. Model selection

For our analyses, we focused on the most parsimonious submodels that still permit us to test the above-stated hypotheses and thus, in particular, permit item memory to vary as a function of vertical location as postulated by our hypothesis of a methaphor-incongruency effect in item memory. This latter requirement excludes Models 4, 5a, and 5d. Among the remaining models, Models 5b and 5c were the most parsimonious.

In choosing between Models 5b and 5c, we have a preference for Model 5b as discussed next which we, however, allow to be overturned by strong empirical evidence in favour of Model 5c. Models 5b and 5c differ in which item memory parameter is set equal to *D*_new_pos and *D*_new_neg. In previous research, *D*_new, the ability to detect that a new item is in fact new was often set equal to the smaller of the item memory parameters for the old items (Kuhlmann & Touron, 2011; Meiser et al., 2007).

Based on the literature reviewed in the introduction, we do not expect a metaphor-congruency effect in item memory; if anything, a null effect or a small metaphor-incongruency effect seemed more likely a priori. And thus, we expect the item memory parameters for congruent stimuli (*D*_up_pos and *D*_down_neg) to be equal to or at best smaller than those for incongruent stimuli (*D*_up_neg and *D*_down_pos). Therefore, Model 5b was the preferred model. We also computed the deviance information criterion (DIC, a criterion combining a Bayesian measure of model fit and a measure of model complexity, with smaller value indicating model preference, see Spiegelhalter et al., 2002) for both models, however, and allowed this preference to be overturned in favour of Model 5c in case there was substantial evidence in this model-selection index favouring Model 5c (i.e., when DIC favoured Model 5c over Model 5b by differences exceeding six). A popular rule of thumb for model comparison is that a difference in excess of six provides substantial support; Millar, 2009; Spiegelhalter et al., 2002). Model fit checks conducted for the selected model in addition ensured that the assumptions embodied by the model (the equality constraints shown in Table B1) can be maintained. The model-selection process described above remained the same for all experiments in this research, and the DIC values of the different models are presented in Appendix C.

In terms of assessing the possibility of space-dependent source memory (see Results section in Experiment 1), our model selection strategy for this alternative modelling followed the one explained above. In particular, and analogous to our basic strategy, Model 6c was the preferred model. We also computed the deviance information criterion (DIC) for both models 6c and 6d and allowed this preference to be overturned in favour of Model 6d in case there was substantial evidence in this model-selection index favouring Model 6d (i.e., when DIC favoured Model 6d over Model 6c by differences exceeding six). The latter situation, however, did never occur.

B3. Parameter estimation.

A Bayesian hierarchical latent-trait approach (Klauer, 2010), was adopted for parameter estimation in this research, in order to take heterogeneity of participants into consideration. The latent-trait approach estimates model parameters at the level of individuals while assuming that they are constrained by a continuous population-level distribution of the parameters. The approach has been widely applied (Bott et al., 2020; Michalkiewicz & Erdfelder, 2016; Wulff & Kuhlmann, 2020) and validated in recent source monitoring research (Arnold et al., 2015; Groß & Pachur, 2020). In the analysis, the posterior distribution of parameters is sampled by means of a Monte Carlo - Markov chain (MCMC) algorithm, and the parameter estimation and inferences are all based on the posterior distribution given the obtained data. Group-level mean estimates and corresponding 95% Bayesian credibility intervals (BCI) are reported. The analysis was conducted using the TreeBUGS package (Heck et al., 2018) in R (R Core Team, 2013).

Convergence of all selected models were assured by visual inspection of the trace plots. Parameter convergence was assured by means of Gelman-Rubin statistics of $\hat{R}$ ≤ 1.1 (Gelman & Rubin, 1992). Model fit was assessed graphically and by posterior predictive *p* (PPP) values. PPP values were based on the means and covariances of the observed frequencies across participants by using the T1 and T2 test statistics (Klauer, 2010), respectively. T1 quantifies the discrepancy between the observed and the expected means of response frequencies, whereas T2 quantifies the discrepancy between the observed and the expected covariances of individual response frequencies. Close to zero posterior predictive p_{T1} and p_{T2} values indicate model misfit (PPPs < .05).

Appendix C. DICs of alternative models in Experiment 1-3

Two alternative models were considered for each dataset, were Model 6c and 6d. The deviance information criterion (DIC, Spiegelhalter et al., 2002) was calculated for the 4 alternatives and the model preference was judged by a popular rule of thumb for model comparison that a difference in excess of 6 provides strong evidence in favour of the model with smaller DIC (Millar, 2009; Spiegelhalter et al., 2002). Regarding the comparison between Model 6c and 6d, if Model 6d was preferred by the smaller DIC ($DIC_{6c} - DIC_{6d} > 6$), Model 6d was selected, otherwise Model 6c was selected. The DICs of the 2 alternative models for each dataset are provided in the tables below.

Table 1

DICs of Alternative Models in Experiment 1a

	Group 1	Group 2	Group 3
Model 6c	1372*	1309*	1420*
Model 6d	1378	1306	1414

Note. “*” represents the selected model in each Group

Table 2

DICs of Alternative Models in Experiment 1b

	Group 1	Group 2	Group 3
Model 6c	1463*	1426*	1436*
Model 6d	1461	1420	1443

Note. “*” represents the selected model in each Group

Table 3

DICs of Alternative Models in Experiment 2a,b and 3b

	E2a	E2b	E3b
Model 6c	1875*	1776*	1623*
Model 6d	1878	1771	1621

Note. “” represents the selected model in each Group*

Table 4

DICs of Alternative Models in Experiment 3a

	Group 1	Group 2	Group 3
Model 6c	1334*	1149*	1690*
Model 6d	1334	1151	1686

Note. “*” represents the selected model in each Group

Appendix D. Source memory parameters as estimated in models of type 6c in Experiments 1-3

Parameter	*d_up_pos*		*d_up_neg*		*d_down_pos*		*d_down_neg*		*d_cong*	
	M(*SD*)	95%BCI	*M*(*SD*)	95%BCI	*M*(*SD*)	95%BCI	M(SD)	95%BCI	M(SD)	95%BCI
Experiment 1a, Group 1	.34 (.16)	[.04, .62]	.34 (.10)	[.11, .52]	.34 (.19)	[.03, .76]	.46 (.23)	[.05, .88]	.06 (.20)	[-.33, .44]
Experiment 1a, Group 2	.60 (.19)	[.18, .93]	.44 (.16)	[.09, .73]	.22 (.13)	[.01, .49]	.38 (.19)	[.03, .75]	.16 (.19)	[-.22, .53]
Experiment 1a, Group 3	.61 (.18)	[.14, .92]	.48 (.22)	[.06, .91]	.35 (.16)	[.05, .72]	.34 (.18)	[.02, .73]	.05 (.21)	[-.37, .46]
Experiment 1b, Group 1	.35 (.18)	[.03, .74]	.75 (.18)	[.27, .99]	.12 (.10)	[.00, .37]	.54 (.21)	[.09, .92]	.00 (.19)	[-.36, .39]
Experiment 1b, Group 2	.42 (.17)	[.06, .75]	.69 (.22)	[.15, .98]	.37 (.18)	[.04, .81]	.33 (.20)	[.02, .80]	-.15 (.21)	[-.55, .28]
Experiment 1b, Group 3	.55 (.22)	[.08, .94]	.71 (.21)	[.20, .98]	.58 (.24)	[.08, .97]	.57 (.26)	[.06, .97]	-.08 (.24)	[-.54, .39]
Experiment 2a	.42 (.18)	[.04, .75]	.40 (.14)	[.09, .66]	.26 (.14)	[.02, .56]	.28 (.16)	[.02, .60]	.02 (.18)	[-.33, .37]
Experiment 2b	.56 (.25)	[.07, .97]	.58 (.20)	[.14, .94]	.71 (.21)	[.21, .99]	.36 (.21)	[.02, .83]	-.20 (.23)	[-.64, .26]
	.47 (.20)	[.05, .85]	.23 (.18)	[.01, .70]	.41 (.19)	[.05, .82]				

Experiment 3a *Group 1*							.28 (.21)	[.01, .78]	.05 (.21)	[-.36, .46]
Experiment 3a *Group 2*	.54 (.21)	[.08, .92]	.63 (.19)	[.19, .95]	.24 (.15)	[.01, .59]	.38 (.22)	[.02, .85]	.03 (.21)	[-.38, .43]
Experiment 3a *Group 3*	.52 (.23)	[.07, .92]	.45 (.18)	[.08, .79]	.69 (.09)	[.48, .86]	.41 (.18)	[.05, .78]	-.10 (.20)	[-.49, .28]
Experiment 3b	.38 (.19)	[.03, .75]	. 31 (.17)	[.02, .67]	.47 (.11)	[.22, .67]	.31 (.17)	[.02, .67]	-.08 (.18)	[-.43, .29]

Note. In Experiment 3, the estimated d parameters for location memory (=source memory) refer to stimulus presentations at the top or at the bottom of the screen, that have physical connotations “high” or “low”, see text. The two columns “d_cong” show the congruence effect (d_up_pos + d_down_neg - d_up_neg - d_down_pos)/2 and its corresponding BCI.

As the three experiments conducted in China had relatively small N’s (E2a / N= 47; E2b / N= 43; E3b / N= 42), we conducted post-hoc power analyses and found that, amongst these three experiments, we were able to detect the found effect sizes for the *d_cong* parameters in Experiment 2b (found: d=.75, detectable d=.38), and 3b (found: d=.41, detectable d=.39), but not in Experiment 2a (found: d=.13, detectable d=.36)

Appendix E. Parameter estimations (in detail) for each experiment.

The parameter notation follows the explanations given in Experiment 1. In all experiments, to test our congruence-related hypotheses, several parameter contrasts are introduced. *D*_incong is a contrast of *D* parameters, *D_incong* = 1/2*(*D_down_pos* + *D_up_neg* - *D_up_pos* - *D_down*), contrasting memory for incongruent versus congruent valence-location combinations. As such it quantifies the metaphor congruency effect signed so that positive values indicate a memory advantage for metaphor-incongruent cases. *D_valence*, defined as 1/2 * (*D_down_neg* + *D_up_neg* - *D_down_pos* - *D_up_pos*), reflects the memory bias favoring negative items, *a_neg_cong* and *a_pos_cong* reflect a possible guessing bias (values greater than .50 indicate a tendency to respond in a congruent manner) for negative and positive items, separately, whereas *a_cong* reflects the overall tendency to guess in a metaphor-congruent way (with values of *a_cong* larger than zero indicating a congruency effect; see text). The names of the parameter contrasts used to test our hypotheses are set in bold in the tables.

Estimates for Parameters (Group-Level) from the Selected Models in Experiment 1a

	Group 1		Group 2		Group 3	
parameter	*M*(*SD*)	95%BCI	*M*(*SD*)	95%BCI	*M*(*SD*)	95%BCI
*D*_up_neg	.70 (.03)	[.64, .75]	.69 (.06)	[.56, .80]	.58 (.03)	[.51, .65]
*D*_down_neg	.57 (.09)	[.37, .73]	.67 (.05)	[.56, .77]	.58 (.06)	[.45, .68]
*D*_up_pos	.57 (.06)	[.44, .69]	.61 (.06)	[.50, .72]	.59 (.05)	[.49, .67]
*D*_down_pos	.59 (.03)	[.52, .65]	.65 (.05)	[.54, .75]	.57 (.04)	[.49, .65]
D_incong	.08 (.06)	[-.03, .21]	.02 (.04)	[-.06, .11]	-.01 (.04)	[-.09, .08]
D_valence	.05 (-.06)	[-.06, .15]	.04 (.04)	[-.04, .13]	.00 (.04)	[-.08, .09]

*d*_neg	.44 (.06)	[.33, .55]	.44 (.09)	[.25, .61]	.44 (.09)	[.25, .61]
*d*_pos	.36 (.08)	[.19, .50]	.39 (.08)	[.21, .54]	.45 (.07)	[.32, .58]
*a*_neg_cong	.63 (.03)	[.56, .69]	.56 (.04)	[.49, .63]	.55 (.05)	[.46, .65]
*a*_pos_cong	.58 (.05)	[.48, .67]	.61 (.04)	[.54, .69]	.68 (.05)	[.58, .77]
a_cong	.19 (.06)	[.06, .32]	.17 (.05)	[.06, .28]	.22 (.07)	[.07, .36]
*b*_neg	.36 (.06)	[.24, .49]	.36 (.06)	[.24, .48]	.38 (.06)	[.27, .50]
*b*_pos	.26 (.05)	[.18, .36]	.29 (.05)	[.19, .40]	.38 (.04)	[.29, .46]

Estimates for Parameters (Group-Level) from the Selected Models in Experiment 1b,

	Group 1		Group 2		Group 3	
parameter	*M*(*SD*)	95%BCI	*M*(*SD*)	95%BCI	*M*(*SD*)	95%BCI
*D*_up_neg	.39 (.07)	[.25, .51]	.42 (.04)	[.34, .51]	.40 (.08)	[.24, .54]
*D*_down_neg	.43 (.03)	[.36, .50]	.37 (.08)	[.19, .50]	.31 (.06)	[.19, .42]
*D*_up_pos	.53 (.03)	[.47, .60]	.55 (.08)	[.39, .69]	.47 (.04)	[.38, .56]
*D*_down_pos	.50 (.08)	[.31, .64]	.56 (.04)	[.49, .63]	.34 (.07)	[.19, .47]
D_incong	-.03 (.06)	[-.16, .07]	.03 (.06)	[-.08, .16]	-.02 (.06)	[-.15, .09]
D_valence	-.10 (.05)	[-.20, .01]	-.16 (.05)	[-.27, -.05]	-.05 (.06)	[-.16, .06]

*d*_neg	.67 (.10)	[.49, .88]	.64 (.19)	[.20, .95]	.73 (.18)	[.29, .98]
*d*_pos	.22 (.10)	[.03, .41]	.38 (.07)	[.23, .52]	.60 (.16)	[.28, .93]
*a*_neg_cong	.54 (.04)	[.47, .61]	.57 (.04)	[.48, .65]	.63 (.04)	[.55, .70]
*a*_pos_cong	.57 (.03)	[.52, .62]	.60 (.04)	[.53, .67]	.60 (.03)	[.54, .66]
a_cong	.11 (.04)	[.02, .20]	.16 (.05)	[.04, .27]	.22 (.05)	[.12, .32]
*b*_neg	.46 (.03)	[.40, .53]	.46 (.05)	[.37, .55]	.47 (.04)	[.40, .54]
*b*_pos	.55 (.05)	[.46, .65]	.56 (.04)	[.48, .65]	.57 (.05)	[.47, .67]

Estimates for Parameters (Group-Level) from the Selected Model in Experiment 2a,

parameter	*M*(*SD*)	95%BCI
*D*_up_neg	.63 (.06)	[.51, .74]
*D*_down_neg	.70 (.03)	[.64, .75]
*D*_up_pos	.59 (.04)	[.51, .67]
*D*_down_pos	.56 (.04)	[.47, .64]
D_incong	-.04 (.04)	[-.13, .03]
D_valence	.08 (.03)	[.01, .16]
*d*_neg	.40 (.07)	[.25, .53]
*d*_pos	.38 (.08)	[.21, .53]
*a*_neg_cong	.47 (.03)	[.41, .53]
*a*_pos_cong	.56 (.04)	[.48, .64]
a_cong	.02 (.05)	[-.08, .12]
*b*_neg	.52 (.05)	[.42, .63]
*b*_pos	.35 (.05)	[.25, .45]

Estimates for Parameters (Group-Level) from the Selected Model in Experiment 2b,

parameter	*M*(*SD*)	95%BCI
*D*_up_neg	.41 (.06)	[.29, .52]
*D*_down_neg	.38 (.04)	[.31, .45]
*D*_up_pos	.36 (.03)	[.29, .42]
*D*_down_pos	.28 (.07)	[.13, .41]
***D*_incong**	-.02 (.05)	[-.13, .08]
***D*_valence**	.08 (.04)	[-.01, .17]
*d*_neg	.49 (.10)	[.30, .68]
*d*_pos	.70 (.19)	[.27, .98]
*a*_neg_cong	.55 (.04)	[.48, .63]
*a*_pos_cong	.56 (.04)	[.48, .63]
***a*_cong**	.11 (.05)	[.01, .21]
*b*_neg	.49 (.04)	[.41, .57]
*b*_pos	.63 (.03)	[.57, .70]

Estimates for Parameters (Group-Level) from the Selected Models in Experiment 3a,

D_connotation is calculated in an analogous way as D_valence in the previous experiments.

	Group 1		Group 2		Group 3	
parameter	M(SD)	95%BCI	M(SD)	95%BCI	M(SD)	95%BCI
D_up_low	.52 (.07)	[.38, .65]	.39 (.08)	[.23, .54]	.55 (.07)	[.40, .69]
D_down_low	.49 (.05)	[.39, .59]	.50 (.06)	[.37, .62]	.52 (.03)	[.46, .59]
D_up_high	.57 (.04)	[.49, .65]	.49 (.06)	[.36, .61]	.56 (.04)	[.47, .64]
D_down_high	.51 (.06)	[.38, .61]	.50 (.07)	[.35, .62]	.55 (.06)	[.42, .67]
D_incong	-.01 (.05)	[-.12, .08]	-.04 (.05)	[-.16, .07]	.01 (.05)	[-.09, .10]
D_connotation	-.03 (.05)	[-.13, .06]	-.05 (.05)	[-.15, .06]	-.01 (.06)	[-.10, .07]
d_low	.29 (.18)	[.03, .73]	.53 (.12)	[.27, .77]	.47 (.08)	[.30, .62]
d_high	.50 (.09)	[.31, .68]	.31 (.14)	[.06, .57]	.64 (.07)	[.49, .78]
a_low_cong	.52 (.04)	[.44, .61]	.61 (.05)	[.51, .69]	.59 (.04)	[.51, .66]
a_high_cong	.61 (.05)	[.51, .71]	.66 (.04)	[.59, .73]	.57 (.04)	[.49, .65]
a_cong	.13 (.07)	[-.00, .27]	.26 (.06)	[.14, .38]	.15 (.06)	[.03, .27]
b_low	.40 (.06)	[.29, .51]	.42 (.07)	[.29, .56]	.48 (.06)	[.36, .61]
b_high	.32 (.05)	[.23, .42]	.26 (.06)	[.15, .38]	.40 (.06)	[.28, .53]

Estimates for Parameters (Group-Level) from the Selected Model in Experiment 3b,

parameter	*M*(*SD*)	95%BCI
*D*_up_low	.70 (.04)	[.61, .78]
*D*_down_low	.67 (.03)	[.62, .73]
*D*_up_high	.69 (.03)	[.63, .76]
*D*_down_high	.64 (.04)	[.55, .72]
***D*_incong**	-.01 (.03)	[-.09, .05]
***D*_valence**	.02 (.03)	[-.04, .08]
*d*_low	.40 (.08)	[.24, .54]
*d*_high	.47 (.05)	[.36, .58]
*a*_low_cong	.63 (.05)	[.54, .72]
*a*_high_cong	.71 (.05)	[.62, .80]
***a*_cong**	.34 (.06)	[.20, .47]
*b*_low	.31 (.05)	[.21, .42]
*b*_high	.33 (.06)	[.22, .44]

*