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

Zixi Jin

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Supervisor: Ulrich von Hecker
School of Psychology, Cardiff University
September 2021

Summary

Embodied cognition is of great interest in the development of cognitive science which emphasizes the role of the body’s interaction with the world in cognition. Within this field, some researchers suggest that metaphoric associations of abstract concepts with concrete concepts can influence relevant cognitive processes. For example, the “good is up” metaphor which links valence and verticality concepts was found to facilitate affective judgement and attention directing processes when the stimuli were presented at metaphor-congruent locations (Meier & Robinson, 2004). Meanwhile, the metaphoric effects on memory performance remain unclear with contradictory research findings. The primary aim of this research was to investigate the metaphoric effects on recognition processes. The source monitoring paradigm was applied in order to provide a more accurate estimation of item memory performance, and to assess source guessing biases at the same time. Across the first 5 experiments, a consistent result pattern showed that the metaphor is effective to guide participants’ guessing of source location with metaphor-congruent biases when one is unable to recall source information. However, the metaphor is not effective to differentiate item memory performance of valenced stimuli learned from metaphor-congruent or incongruent locations. Two follow-up experiments were conducted which intended to replicate Meier & Robinson’s (2004) original findings of metaphor-congruent facilitation effects, but failed to demonstrate the same effects. Altogether, this thesis suggests that the “good is up” metaphor may be effective in cognition but in a more subtle way than one could expect based on previous literature. Despite the failed replications, the series of experiments is informative and provides useful insights for investigations on this metaphor. Future research on this metaphor topic is promising to altogether provide more reliable and conclusive suggestions.

Acknowledgments

There are so many people I would like to thank for their support and help during my PhD. First of all, I would like to thank my supervisor, Ulrich von Hecker, for providing his professional insights and guidance to supervise me, and for being patient to help me improve my English writing skills during the last four years. His passion for research and optimistic personality have influenced me significantly. Thanks to his help and encouragement, I have built up my confidence in not only this project, but also research in general, which will be so helpful in my future career.

Second, I would like to thank my colleagues and friends from academia. I would like to thank my colleagues from the Social Psychology lab and the COGDEV lab who provided a supportive environment for me to present my research as well as offered inspiring feedback. I also would like to express my gratitude to Nikoletta Symeonidou and Christoph Klauer for sharing their expertise on modelling with me, and for being so patient with the emails full of my detailed and sometimes stupid questions.

I also would like to thank my dearest friends from Cardiff, Olivia, Ling, Leo, Kevin, Matt, Neo, Alex, James, Will, Kelsey, Natalia, Aimee, and Charlie. Being so far away from home is never easy, but they have made Cardiff like a second home for me.

我还要感谢我亲爱的爸爸妈妈，金先生和虞女士，谢谢你们一直以来的包容与支持。以后我也会继续加油，希望我可以一直是你们的骄傲。

Finally, I would like to thank China Scholarship Council (CSC) for funding my PhD. Thanks for providing me with the opportunity for this wonderful journey. I also would like to thank iCog grant for partially funding some experiments in this thesis.

Table of contents

Chapter 1 General Introduction

In this chapter, I start by introducing embodied cognition and metaphors as a theoretical background. Then I explain what orientational metaphors are and subsequently focus on the “good is up” metaphor. Following that, I present previous research on the “good is up” metaphor, especially concentrating on the empirical evidence concerning the metaphoric effects on cognitive processes. Afterwards, I present previous findings with regards to the metaphoric effect of “good is up” on memory, in different research contexts, in order to demonstrate the need of a systematic investigation on this topic. At the end of this chapter, I introduce the methodology which is going to be applied in the next chapter, where empirical studies are reported.

Embodied cognition and metaphors

Embodied cognition is often considered as a new evolution in the development of cognitive science (Shapiro, 2019) which emphasizes the role of the body’s interaction with the world in cognitive processing. It is assumed that the brain captures and integrates experience-related information into a multimodal representation and stores this representation in memory (Barsalou, 1999, 2008). In later situations of retrieval, the multimodal representation is reactivated, thereby triggering simulations of the perceptual, motoric, and introspective states of the experience. In this way, the embodied experiences play a more effective role in cognition. Within embodiment, mental simulation is often given a prominent role as an explanatory concept.

In fulfilling a metaphorical function, the simulation of sensorimotor experiences can provide an individual with concrete counterparts of abstract concepts which can be used as a cognitive scaffold for people to rely on and thereby to learn, reason, and illustrate these abstract concepts (Lakoff & Johnson, 1980, 1999). For example, perceptions of physical height or elevation provide people with a direct impression of power and status and help people to learn about this abstract concept (see Schubert, 2005; von Hecker et al., 2013), which is expressed by the “power is up” metaphor. Researchers of linguistics and philosophy proposed that metaphors are not just poetic expressions, but also a fundamental part of the conceptual system in our brains (Kittay, 1990; Lakoff & Johnson, 2008) which bridges the realms of abstract and concrete concepts.

Following the above argument, the question was raised whether literal, and entirely language-based evidence can at all support the existence of conceptual metaphors in cognitive processes. Murphy (1996) suggests that the coherence between descriptions of abstract and concrete concepts may not reflect metaphoric representations, but rather reflect the structural similarity between these concepts (see also Gentner & Markman, 1997). Take the “love is a journey” metaphor as an example. The conceptual-metaphor view suggests that the concept of “love” is not explicitly represented but is instead understood through reference to the concept of “journey” in cognition. However, Murphy (1996) claimed the alternative explanation that the “love” concept has a representation independent of the “journey” concept, but the content is influenced by the metaphor such that the two concepts have the same structure. In other words, the metaphors influence the representations of the abstract concepts to be similar to the concrete concepts, but they still have separate representations, namely, the representations themselves are not metaphoric.

Since then, a growing body of research has been conducted to provide more empirical evidence concerning conceptual metaphors (Gibbs et al., 1997; McGlone & Harding, 1998). Specifically, research on non-linguistic metaphors provided more empirical evidence to argue against the doubts about the linguistic-only substantiation of metaphor-based cognitive processes. For example, investigations on spatiotemporal reasoning demonstrate that people build their understanding of the abstract domain of time on the concrete domain of space (Boroditsky & Ramscar, 2002), which means when primed with different spatial information (objects moving toward participants vs. participants moving toward objects) and then asked to think about time (“Next Wednesday’s meeting has been moved *forward* 2 days. What day is the rescheduled meeting?”), participants responded “Friday” after having been primed with themselves moving (in which case it was concluded that they perceived “moving forward” as moving to a later timepoint). In contrast, participants tended to respond “Monday” after being primed with objects moving toward them (in which case perceived “moving forward” as moving to an earlier timepoint). Findings from Fay and Maner's (2012) studies suggest that there is a physical grounding of the metaphoric link between warmth and social affiliation. Participants tended to perceive warm objects as physically closer than cold ones, and when primed with warmth they reported higher levels of social affiliative motivation. Conceptual metaphors have been more widely accepted with the broader range of supportive evidence.

Metaphors can be helpful not only to help understand abstract concepts, but also to illustrate and develop abstract thinking. After reviewing previous literature, Gibbs and colleagues (2004) presented a case study on the metaphor “desire is hunger” to demonstrate that people’s understandings of metaphorical expressions about human desires are motivated by their embodied experiences related to feeling hungry. As a result, they proposed that metaphorical thought and language are fundamentally grounded in embodied experiences. Furthermore, Jamrozik and colleagues (2016) suggest that repeated metaphoric use can serve as a useful tool to guide the development of new abstract representations. For example, Bowdle & Gentner (2005) suggest that there is a shift in the mode of alignment from comparison to categorization as novel metaphors are conventionalized, which means participants tend to shift their preferential use of expression from the simile form (e.g., “an obsession is like a tumor”) to the metaphor form (e.g., “an obsession is a tumor”) after repeatedly encountering novel similes using the same source domain (e.g., “doubt is like a tumor”, “a grudge is like a tumor”). The shift indicates that along with the familiarization, people align new targets with the metaphoric abstraction derived from the mappings, and an explicit comparison between the target and source domains becomes less necessary. Furthermore, metaphors are also influential to other cognitive processes. Crawford (2009) reviewed previous studies on metaphors of affect and showed that the metaphoric association between affect and physical dimensions can influence cognition processes such as shifting attention, judgement of brightness and memory of spatial information. In general, metaphors do not merely manifest the embodied cognition, but using metaphoric languages and thoughts also has a powerful impact on cognition.

O*rientational metaphor*, as one of the most commonly used kinds of metaphors, organises abstract concepts with respect to *spatial* orientations (Lakoff & Johnson, 1980, 2008), such as up-down, left-right, front-back etc. Various orientational metaphors have been shown to operate in cognition, such as “good is up” (Meier & Robinson, 2004), “divine is up” (Meier et al., 2007), “powerful is up” (Giessner & Schubert, 2007; Schubert, 2005; Zarzeczna et al., 2020), “high social status is up” (von Hecker et al., 2013), “dominant is left” in cultures with left-to-right reading/writing habits (and reverse direction in cultures with right-to-left reading/writing habits, see von Hecker et al., 2016), “future is in front of/behind ego” (direction depends on different cultural background, see Núñez & Sweetser, 2006) etc. The congruency between the physical present location of a stimulus (an abstract concept) and the spatial location within the mental simulation of the same stimulus can facilitate some cognitive processes, such as attention, comprehension and learning. Previous research showed that when primed with power-related concepts, participants automatically shifted their attention to the metaphor-congruent direction (i.e., powerful-up & powerless-down) (Schubert, 2005; Zanolie et al., 2012). Langston (2002) found that texts expressing content incongruent to orientational metaphors (e.g., “more is up”, “the future is forward”) were more difficult for participants to comprehend compared to texts that were congruent with the metaphors. Yasuda (2010) found that when learning idioms from a second language, participants who were explicitly taught about the related orientational metaphors showed better performance compared to those who learned in a non-metaphoric way. Orientational metaphors consist of an important part of our daily life and are unavoidable in the development of cognition.

“Good is up”

This research concentrates on one specific orientational metaphor, “good is up”. The “good is up” metaphor is a primary metaphor in people’s life (Lakoff & Johnson, 1999) which indicates that concepts considered to be good are commonly associated with up or high spatial locations, whereas those considered to be bad are commonly associated with down or low locations.

This argument is intuitively supported by evidence from linguistics and daily life. In oral English communication, people often use “I feel up/down today” to indicate their positive or negative moods. 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. For example, in competitive events like Olympics, athletes spontaneously elevate the chest or raise the arms above the head to express pride of their success, whereas they tend to hang the head or slump the shoulders to express shame of failure. Even congenitally blind people express pride and shame in this way with upward and downward postures and gestures, even they have never seen these behaviours modelled by other people (Tracy & Matsumoto, 2008). And supportive evidence can also be found on the Internet, where users of social media always press “thumb up” or “thumb down” button as a simplified way to express their positive or negative feedback.

Long before the theory of a “good is up” metaphor was proposed and widely investigated, there was evidence to suggest its existence. Wapner and colleagues (1957) found that participants who had just received a good grade on a test showed an upward bias on horizontal fixation, whereas those who had received a failed grade showed a downward bias. Stepper & Strack (1993) found that manipulated unobtrusive body posture influenced people’s feeling of pride when learned about their success, such that people feel prouder when sitting up straight compared to when slumping down. Though Stepper & Strack (1993) concentrated more on the proprioceptive determinants of emotional feelings instead of looking at the relationship between spatial location and affect in their research, it nevertheless provided support for the “good is up” metaphor.

Researchers have been encouraged to seek for more empirical evidence to support the “good is up” metaphor and its effects on cognition. One cornerstone is a research conducted by Meier & Robinson (2004) including three studies to demonstrate the metaphoric association between valence and vertical locations. Study 1 showed that participants responded faster judging the valence of the words (e.g., “polite” as positive, “rude” as negative) when they were presented at *metaphor-congruent* locations (i.e., positive words presented on top of the screen, or negative words presented at the bottom, which is congruent to the “good is up” metaphor) compared to when at *metaphor-incongruent* locations (i.e., negative words presented on top of the screen, or positive words presented at the bottom, which is incongruent to the metaphor). Study 2 showed that when primed with positive or negative words, participants shifted their attention towards metaphor-congruent locations automatically; in contrast, Study 3 showed that when primed with different vertical locations, participants showed similar latencies to distinguish the valence of the following central-vertically presented words, no matter metaphor-congruent or -incongruent to the prime location. Therefore, Meier & Robinson (2004) suggest that the association between valence and verticality is asymmetric. A potential embodied explanation is that people need to use concrete concepts to interpret abstract concepts in cognitive processes, however, concrete concepts by themselves can be comprehended and further processed in cognition without the need to activate relevant abstract concepts.

Evidence from functional magnetic resonance imaging (fMRI) was also obtained to support the metaphorical association of valence and verticality. Quadflieg and colleagues (2011) found that neural activity corresponding to discriminating the physical “up” or “down” present location of a shape stimulus is similar to the activity when discriminating the “up/down” spatial connotation as well as the “positive/negative” valence of a word stimulus (e.g., “attic” as having “up” connotation, “carpet” as having “down” connotation; “party” as having “positive” valence and “accident” as having “negative” valence).

From the simulation perspective, the valence “good” is associated with not just the “up” location but also the relevant sensorimotor experiences which include upward motions. Brookshire and colleagues (2010) found that, when asked to judge the font colour of a positive or negative word by key-pressing, participants responded faster when the hand-movement towards the response key was metaphor-congruent to the valence of the presented word compared to when the hand-movement direction was metaphor-incongruent to the word valence, although valence was irrelevant to the task. Santana & de Vega (2011) found a similar result that when required to perform a speeded upward or downward hand-movement response while reading a positive or negative sentence, participants responded faster when the valence of sentence and hand motion were metaphorically congruent. In addition to the influence on reaction time, metaphor-congruency between valence and vertical motions was also found to affect subjective preferences. Gottwald and colleagues (2015) found that observing vertically moving actions can change participants' affective ratings of the involved object toward a metaphor-congruent direction, namely, participants generated more positive ratings after observing an upward movement and more negative ratings after observing a downward movement.

Interestingly, the "good is up" metaphor shows consistency across different languages and cultures, as evidence was also found in Mandarin (Wu et al., 2019), German (Dudschig et al., 2015), Russian and French (Luodonpää-Manni & Viimaranta, 2010) contexts, apart from English.

As an extension of the research interest on the "good is up" metaphor, researchers have been investigating its effect on more complex cognitive processes, such as memory.

Metaphoric effects on memory

Memory plays an important role in both people's daily life and cognitive science. Memory consists of three basic stages: encoding, storage, and retrieval (Melton, 1963, as cited in McDermott & Roediger, 2018). Encoding refers to the initial learning of information, storage means maintaining information over time and retrieval stands for the access to information when people need it (McDermott & Roediger, 2018). To investigate metaphoric impact on memory can enrich our understanding of how metaphors structure cognition (Crawford et al., 2014) in a more comprehensive way. However, previous research interestingly shows contradictory results, as to whether metaphor-congruency or -incongruency facilitates memory performance.

Some studies support facilitation of metaphor-congruency in memory tasks. Casasanto and Dijkstra (2010) found that when asked to recall positive or negative autobiographical memories, participants tended to respond faster when moving their hands in a metaphor-congruent direction simultaneously. When given neutral-valence instructions to recall autobiographical memories, participants tended to recall more positive memories with hands moving upward whereas they recalled more negative memories with hands moving downward. Similarly, Globig and colleagues (2019) found that while performing an upward or downward vertical head movement in a free recall task, participants retrieved more positive or negative words respectively, which were metaphor-congruent to their head movement. As a limitation, these two studies both manipulated valence and verticality only in the retrieval phase. Supportive evidence was also obtained when the two variables were manipulated in the encoding phase. Palma and colleagues (2011) asked participants to form an impression of a stereotypical character (e.g., a childcare professional as a positive character and a skinhead as a negative one) from reading stereotype-relevant behavioural information (e.g., a positive character helped an elderly person to use the ATM or a negative character intentionally ignored the phone calls from a friend) mixing with stereotype-irrelevant behavioural information (e.g., waited for the bus in that morning) randomly presented at the top or bottom of a screen. Afterwards, when unexpectedly asked to free recall the information, participants recalled more stereotype-relevant information that had been placed in metaphor-congruent than in metaphor-incongruent locations. They further replicated the metaphor-congruency advantage on free recall memory by manipulating vertical hand movements instead of locations of presentation during the impression formation procedure.

In contrast, some researchers suggest that metaphor-*in*congruency facilitates memory, based on the *Attention-Elaboration Hypothesis* (AEH) and have provided according evidence. The AEH argues that atypical information receives more attention, generates more conceptual elaboration and therefore is processed more deeply than typical information. This makes atypical information easier to retrieve and to recognise (Erdfelder & Bredenkamp, 1998; originally attributed to Bobrow & Norman, 1975). The AEH has been supported by empirical results from previous research on memory performance (Mäntylä & Bäckman, 1992; Vakil et al., 2003). Similarly, a large amount of research shows that information violating expectations can be better remembered than those consistent with expectations (behaviours congruent or incongruent to personal impression, Hastie & Kumar, 1979; schema-relevant information, Sakamoto & Love, 2004; see Rojahn & Pettigrew, 1992 for a meta-analysis). In line with this theory, Crawford and colleagues (2014) found that participants showed better memory of valenced words presented in metaphor-incongruent than in metaphor-congruent locations in the encoding phase, in both free recall and recognition tasks.

Different to the previous two views, other studies suggest no metaphoric effects on memory. McMullan (2016) used a similar paradigm as used by Palma and colleagues (2011), but found that vertical locations at which behavioural information was presented had no influence on participants' recall performance in an unexpected memory test. Moreover, in the abovementioned research conducted by Crawford and colleagues (2014) in which they found that metaphor-incongruent word location in encoding facilitates memory, they also found no metaphoric effects on memory of valenced words when manipulating the vertical position of participants' hands in the encoding phase.

After looking more closely at the conflicting results, I narrowed down the research interest to a more specific question, which is the metaphoric effects on memory when manipulating valence and verticality in the encoding phase, in a "pure" memorising task. When manipulating valence and verticality in the retrieval phase, the metaphor-congruency facilitating effect was substantially demonstrated by both Casasanto and Dijkstra (2010) and Globig et al (2019), with the hand/head movement providing direct bodily experience as a retrieval hint, in both autobiographical memory and semantic material memory contexts. It is the metaphoric effect in the encoding phase that remains rather unclear. Palma and colleagues (2011) and McMullan (2016) both required participants to complete an impression formation task to hide their interest of memory performance, then the metaphoric effect was measured by an unexpected memory test afterwards. On the contrary, the context Crawford and colleagues (2014) used was less complex and rather straightforward to investigate the effects in the encoding phase, with a direct instruction to require participants to memorise the valenced word stimuli and without any masking tasks such as impression formation.

Under these circumstances, I find the inconsistent memory performance results interesting, with respect to manipulating locations of stimulus presentation and manipulating the positions of participants' hands in the encoding phase, as reported by Crawford and colleagues (2014). According to the simulation theory, the vertical location of word presentation should serve the same purpose as hand position, and thus produce consistent effects on memory, which however was not the case in Crawford et al (2014). Consequently, in order to get a clearer picture on the metaphoric effects, the aim of this research is to systematically investigate whether the “good is up” metaphor can affect memory in the encoding phase. I started with a conceptual replication of Experiment 2 from Crawford et al (2014) in which they manipulated the location of the presented valenced stimuli (details see Chapter 2) and expected a metaphor-incongruency facilitating effect on memory based on their original findings. More than a mere replication of Crawford et al (2014), I applied a novel paradigm, *source monitoring*, to investigate the metaphoric effects on recognition processes.

Source monitoring

Source broadly refers to the context in which a piece of 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 is a series of cognitive processes involved in the judgements about the source of information (see Johnson et al., 1993 for the theoretical framework).

A source monitoring study typically consists of a learning phase where items originating from (usually) two sources are presented (for example, in Experiment 2 in Bayen et al., 2000, statements said by “a doctor” or “a lawyer” were used), and a testing phase in which participants are asked to identify a presented item as “old” or “new” (i.e., learned or not learned from the learning phase). If participants classify the item as previously learned, they are then asked to attribute the item to one of the two sources (i.e., either “a doctor” or “a lawyer” said this statement in the learning phase). Frequencies of all kinds of responses to items from different sources are calculated separately (i.e., it is counted how many statements from “a doctor” are classified correctly by participants as from “a doctor” or falsely classified as from “a lawyer” or as not learned respectively, same with statements originally presented from “a lawyer” and new statements).

The source monitoring paradigm includes three main parts, *item memory*, which indicates the ability to remember whether an item is learned or not; *source memory*, which indicates the ability to remember which source an item is from in the encoding phase; and *guessing biases*, which includes the tendency of guessing an item is learned when not being able to recognise (i.e., guessing an unrecognised item to be “old” or “new”), and the tendency of guessing an item that was classified as “old” (no matter whether this judgement was due to memory or guessing) as originating from one certain source when not being able to discriminate the source of this item.

The advantage of applying the source monitoring paradigm to the present research is that the metaphoric effects on recognition memory can be investigated with less noise. It is unavoidable in previous research on recognition memory that people may have guessed when required to generate a response while unable to recall, which makes it hard to distinguish the actual effect on memory against the effect on guessing biases. In the source monitoring paradigm instead, the effects on each step can be disentangled which makes it possible to look at the metaphoric effect on “pure” item memory. Additionally, metaphoric effects on other processes in recognition can be investigated at the same time, for example, the effect on source guessing biases, which is innovative for research on the present topic.

Metaphoric effects on source guessing

Apart from expecting that metaphor-incongruency might facilitate item memory, metaphor-congruent source guessing biases were expected.

Previous research suggests that the valence of stimuli can affect participants’ reproduction of their locations in the encoding phase, such that positive stimuli are biased upward compared to negative stimuli (Crawford et al., 2006). 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 from the learning phase), participants tended to recall the locations of positive events with an upward bias whereas to recall the locations of negative events with a downward bias (Brunyé et al., 2012). The demonstrated source recollection biases suggest that the valence may influence the responses of where the stimuli were presented.

In a typical source monitoring paradigm, source identification is required by a two-alternative forced choice response. Which means that when participants have classified an item to be “old”, they then have to indicate which source they think it was from in the learning phase, based on their memory or merely just guess. In this case, metaphoric effects on source identification actually consist of two parts, metaphoric effects on source memory and metaphoric effects on source guessing. Within this paradigm, the two kinds of possible effects can be investigated separately, and a growing body of research is more in favour of the effect on guessing biases than on source memory, as presented below.

Bayen and colleagues (2000) were the first to empirically demonstrate that prior knowledge influences source identification by biasing source guessing. In their research, expectation of source-information congruency was considered as an influence factor. Materials included information that was expected to come from one source or the other or equally from both. For example, again in their Experiment 2, “a doctor” and “a lawyer” were used as two sources for statements (e.g., “a doctor said…”). Statements which are typically expected from a doctor (e.g., “Are you taking any other medicine?”) or from a lawyer (e.g., “I ask that custody be given to the mother.”) or equally expected for either (“I have never liked spinach.”) were presented as half said by “a doctor” and the other half by “a lawyer” respectively. As a result, source identification was better for the statements coming from the sources congruent to the expected professions (e.g., statements expected from a doctor that said by “a doctor”). Model-based analyses revealed equally good source memory performance of statements from expected or unexpected source, but an expectation-based source guessing bias. Namely, when not be able to recall the information source, participants were biased to guess the expected source based on their prior knowledge of the professions. And participants showed unbiased guessing for statements that were equally expected from both sources. To conclude, Bayen and colleagues (2000) suggest that expectation-reliant source guessing, and not differential source memory, caused the better performance of the source identifications for statements from expected sources.

Klauer and Wegener (1998) also suggest that when people are unable to recall the source of a piece of information, they tend to guess according to acquired associations instead of making unbiased, random guesses. Our daily life experiences intuitively support this argument. For example, if we forget who told us about a new gym class, we may infer that it was from a close friend who goes to gym regularly instead of another friend who hates exercise. Source guessing is particularly biased when an existing expectation or schema implies congruency between a particular item and one particular source, supported by a large amount of empirical evidence with different kinds of contexts (social schemas: a reckless skinhead vs. a friendly social worker, Ehrenberg & Klauer, 2005; age stereotypes: younger vs. older adults, Kuhlmann et al., 2016; room objects schemas: bedroom vs. bathroom, Experiment 1 in Bayen et al., 2000) apart from the profession schemas used in Bayen et al's (2000) Experiment 2 (a lawyer vs. a doctor).

It was noticed that some evidence suggests better source memory for unexpected source than expected source when the item-source combination strongly violates the schema (Küppers & Bayen, 2014). As a result, the possibility of the metaphoric effects on source memory process was not excluded, details see the results section in each experiment in Chapter 2.

To sum up, it is proposed that the “good is up” metaphor can also create the expectation of congruency as the schemas do and consequently bias the source guessing toward a metaphor-congruent direction. Namely, when participants fail to retrieve the source location of a stimulus, they tend to guess a metaphor-congruent (i.e., positive-up, negative-down) rather than a metaphor-incongruent (i.e., positive-down, negative-up) location.

Methodology

To more accurately analyse the metaphoric effect on recognition memory, especially on different processes within the source monitoring paradigm, a Multinomial Processing Tree (MPT) modelling approach is going to be introduced. Specifically, a two-high-threshold source monitoring (2HTSM) model will be applied. In order to make the model mathematically identifiable, alternative submodels with different constraints between parameters will be presented and compared. On top of that, a recently developed Bayesian hierarchical approach on model estimation will be introduced in order to get more accurate parameter estimates by taking participants' heterogeneity into consideration.

Multinomial processing tree (MPT) modelling

Multinomial modelling was designed to estimate hypothetical parameters that represent the probabilities of unobserved processes. It has been applied in many fields, for example statistical genetics (e.g., Elandt-Johnson, 1971), before being introduced to cognitive psychology by Riefer and Batchelder (1988) who evaluated this methodology and its potentials for applications in the analysis of cognitive processes.

Multinomial processing tree (MPT) models in source monitoring are a class of models assuming that the responses made by participants emerge from a combination of discrete cognitive states that are attained with certain probabilities (Batchelder & Riefer, 1990; Bayen et al., 2000). The objective of the approach is to provide separate, theoretically based estimation of parameters representing the corresponding cognitive processes in generating the responses, including item recognition, source discrimination and guessing, based on the obtained frequency data of each source-response category.

Compared to general-purpose statistical analysis such as the analysis of variance (ANOVA), the advantage of MPT models is that they are theoretically motivated which means they make certain assumptions about the nature of the underlying cognitive processes (Riefer & Batchelder, 1988). Therefore, the influence of separate cognitive processes on behavioural outcomes can be measured more directly.

Bayen and colleagues (1996) presented three different kinds of MPT models and compared them against the empirical data from source monitoring, and then identified the two-high-threshold model to be the most valid one to provide accurate and independent measures of item memory, source memory and guessing biases. Since then, the *two-high-threshold source monitoring* (2HTSM) model has been widely accepted and used in empirical research (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).

Two-high-threshold source monitoring (2HTSM)

Bayen and colleagues (1996) proposed a 2HTSM model which assumes that there are two thresholds dividing the decision space into three discrete areas corresponding to undetected, detected as old and detected as new. If either threshold is crossed for a testing item, the item is classified as either old or new, depending on which threshold was crossed; if neither threshold is crossed, the item is in an undetected state and the participant responds either "old" or "new" depending on a guessing probability. *High*-threshold means that it is assumed that only old items can cross the detect-as-old threshold and only new items can cross the detect-as-new threshold. See Figure 1 for an illustration of the assumptions of item detection in a 2HTSM model.

Figure 1

Assumptions of Item Detection in a 2HTSM Model

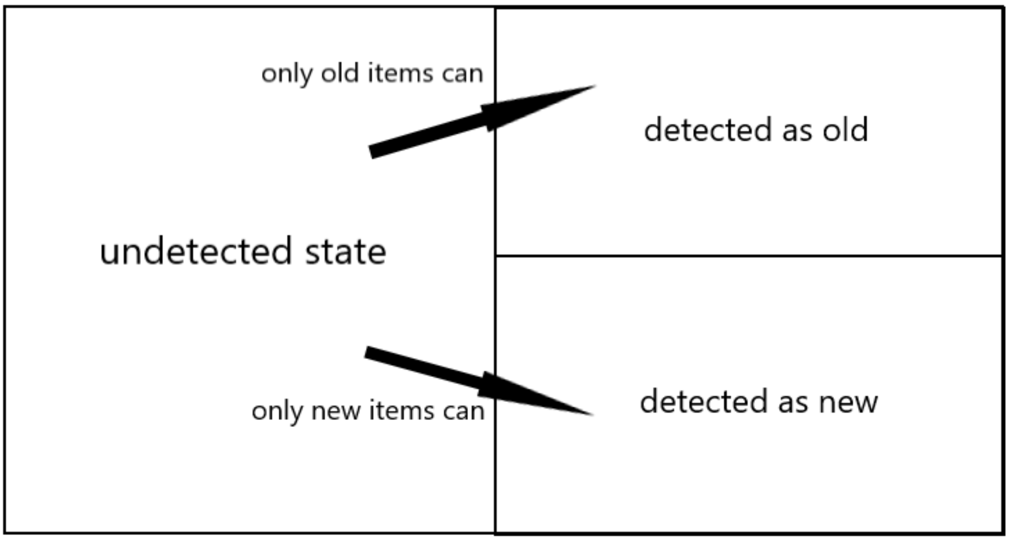
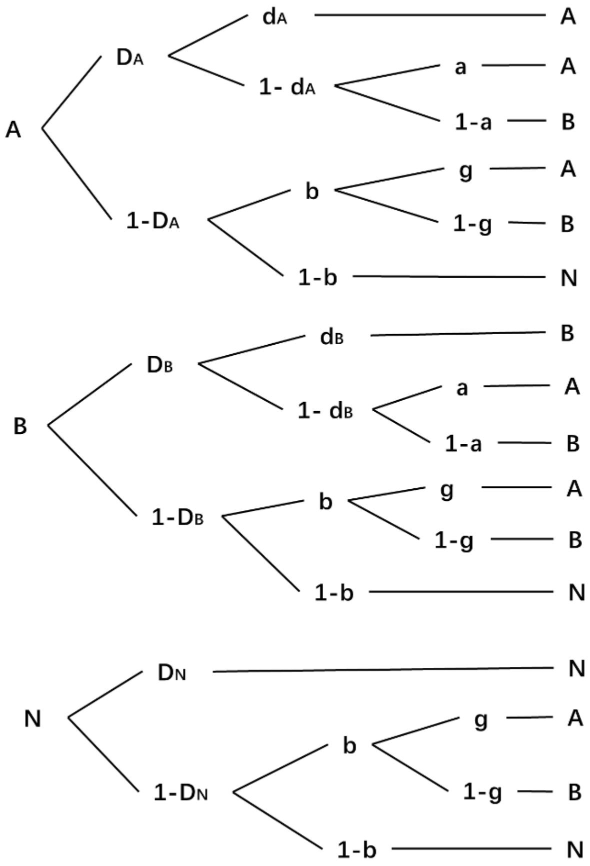


Figure 2 illustrates a standard 2HTSM model structure (so-called “multinomial trees”) with each pathway in the figure symbolizing a combination of involved cognitive processes that might contribute to a certain response in a testing trial. The root of each multinomial tree (left side in Figure 2, A or B or N) indicates whether the test item comes from source A (again, use Experiment 2 in Bayen et al., 2000 as an example, an item from source A represents a statement said by “a doctor” in the learning phase), or source B (“a lawyer”), or whether the item is new. Each pathway leads to a possible response, A or B or N, as listed on the right side of Figure 2. Obviously, more than one pathway can terminate in the same observed source-response category. Branches of the “trees” are the parameters of the model (D, d, b, a, g), which represent the probabilities of the involved cognitive processes. Take an item that originated from source A (i.e., a statement said by “a doctor” in learning) as an example. If it is presented in a testing trial, which refers to the first “tree” in Figure 2, and the participant will recognise the item as “old” with probability D_A. Then with probability d_A, the recognised item will be correctly identified as from source A; with the complementary probability $1 - d_A$, the participant will not be able to identify the item as from source A and will, therefore, guess the source of the item. With probability a, the participant will guess that the item was from source A (“a doctor”); with the complementary probability $1 - a$, the participant will guess the item was from source B (“a lawyer”). If the test item is not recognised, with probability $1 - D_A$, the participant will then, with probability b, guess that it is a learned item. Given that, the participant will then guess, with probability g, that the item was from source A, or with probability $1 - g$ to guess that the item was from source B. If the participant does not guess that unrecognised item is “old”, the item will be classified as “new”, with probability $1 - b$. For items that originated from source B and new items, the probabilities of the branches are to be understood in a similar way as above, just with different subscripts representing parameters, details see Table 1.

Figure 2

A Standard Two-high-threshold Source Monitoring (2HTSM) Model with Two Sources



Note. The left side indicates the item sources and the right side indicates responses. The model parameters are defined in Table 1. Adapted from “Source discrimination, item detection, and multinomial models of source monitoring” by U. J. Bayen, K. Murnane, and E. Erdfelder, 1996, *Journal of Experimental Psychology: Learning, Memory, and Cognitio*n, *22*, p. 202. Copyright 1996 by the American Psychological Association.

Table 1

Parameters Definition in the Standard 2HTSM Model in Figure 2

Parameter	Definition
D_A	Probability of correctly recognising an item from source A
D_B	Probability of correctly recognising an item from source B
D_N	Probability of correctly recognising a new item is new
d_A	Probability of correctly discriminating the source of an item from source A
d_B	Probability of correctly discriminating the source of an item from source B
b	Probability of guessing an unrecognised item is old
a	Probability of guessing a recognised item is from source A
g	Probability of guessing an unrecognised item is from source A

As illustrated, this model is able to disentangle the cognitive processes that are assumed to underlie participants' responses in the source monitoring task, by estimating the probabilities of each process. It is, as such, considered as a useful tool to investigate metaphoric effect on item memory and guessing biases separately.

Now the structure of the model has been introduced, the following question is how the abovementioned parameters within the 2HTSM model are mathematically estimated.

Mathematically identifiable submodels

In order to make the 2HTSM model mathematically identifiable, constraints between parameters are in need to reduce the degrees of freedom (Bayen et al., 1996). Table 2 presents all identifiable submodels of the standard 2HTSM model.

Table 2

All Identifiable Submodels of the Standard 2HTSM Model

Model 6a	Model 6b	Model 6c	Model 6d
D_B; $D_A = D_N$	D_A; $D_B = D_N$	D_B; $D_A = D_N$	D_A; $D_B = D_N$
$d_A = d_B$	$d_A = d_B$	d_A; d_B	d_A; d_B
a; g	a; g	$a = g$	$a = g$
b	b	b	b
Model 5a	Model 5b	Model 5c	Model 5d
$D_B = D_A = D_N$	D_B; $D_A = D_N$	D_A; $D_B = D_N$	$D_A = D_B = D_N$
$d_A = d_B$	$d_A = d_B$	$d_A = d_B$	d_A; d_B
a; g	$a = g$	$a = g$	$a = g$
b	b	b	b
Model 4			
$D_A = D_B = D_N$			
$d_A = d_B$			
$a = g$			
b			

Note. An equal sign indicates that two parameters are constrained to be equal, and a semicolon separates parameters that are allowed to vary in each submodel. The model parameters are the same as in Figure 2, defined in Table 1. Adapted from "Source discrimination, item detection, and multinomial models of source monitoring," by U. J. Bayen, K. Murnane, and E. Erdfelder, 1996, *Journal of Experimental Psychology: Learning, Memory, and Cognitio*n, *22*, p. 202. Copyright 1996 by the American Psychological Association.

The submodels are constructed by imposing equality constraints on parameters. For example, model 4 assumes that source memory for items from the two sources is equal ($d_A = d_B$), and also assumes source guessing probabilities to be equal for recognised and unrecognised items ($a = g$).

When applying the 2HTSM model approach to data analysis, it needs to be decided which submodel to use. The submodel selection process depends on specific contexts with different hypotheses. Details of the application process will be introduced in detail in Chapter 2 with actual datasets.

Bayesian hierarchical estimation approach

Traditionally, when using a 2HTSM model to analyse data, the parameters are calculated by maximum-likelihood estimation approach (Hu & Batchelder, 1994). In this approach, observed frequencies of each source-response category from all participants are aggregated together to obtain the fitted model and subsequently to get parameter estimates, based on the assumption that parameters are identical across participants.

However, this assumption is psychologically unrealistic because people differ in their cognitive abilities. For example, some people might have a generally better or worse item memory compared to others. To consider parameter heterogeneity across participants (sometimes across items as well), Bayesian hierarchical estimation approaches have been proposed more recently (Klauer, 2010; Matzke et al., 2015; Smith & Batchelder, 2010).

Bayesian hierarchical approaches estimate model parameters at the level of individuals while assuming a population-level continuous distribution of the parameters. In this thesis, the latent-trait approach proposed by Klauer (2010) is applied which uses a multivariate normal distribution of the probit-transformed model parameters as prior distribution with a mean and covariance matrix to be estimated from the data. Among all Bayesian hierarchical approaches, this approach is particularly advantageous because it also considers correlations among parameters. From a psychological point of view, it is plausible to assume that some cognitive processes might be correlated, for example, item memory and source memory might be both influenced by some covariates such as personality traits. This approach has been widely applied (Bott et al., 2020; Michalkiewicz & Erdfelder, 2016; Wulff & Kuhlmann, 2020) and validated (Arnold et al., 2015; Groß & Pachur, 2020) in recent years.

In the latent trait approach, the posterior distribution of parameters is sampled by means of the 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) will be reported. In this thesis, hypotheses concerning differences in parameters will be tested via Bayesian posterior probability (*pp*) or 95% BCI of the differences. Specifically, Bayesian posterior probability (*pp*) is used to test one-tailed hypotheses, representing the probability of one hypothesis occurring in the posterior samples (Gelman et al., 2013), and *pp* of null hypothesis smaller than .05 is considered statistically substantial (Rouder et al., 2007). A

95% BCI of the differences between parameters in the posterior distribution is used to test two-tailed hypotheses, and a 95% BCI excluding zero is considered statistically substantial (see Wulff & Kuhlmann, 2020 for an example). For details of analysis using this method, see Results sections in the experiments reported in Chapter 2.

Chapter 2 Metaphoric Effects on Recognition Memory

In the previous chapter, a general introduction of the theoretical rationale and the methodology was presented. The theoretical background of embodied cognition and metaphors was introduced, then the topic of the "good is up" and its effects on recognition processes were brought up. Specifically, a source monitoring paradigm was introduced in order to analyse metaphoric effects on recognition processes more accurately. To achieve this goal practically, a multinomial processing tree (MPT) modelling was introduced as a background, then more specifically, a Bayesian hierarchical estimation (latent-trait) approach on the basis of a two-high-threshold source monitoring (2HTSM) model was presented in detail.

The current chapter applies this methodology in empirical experiments. Following the previous research on the "good is up" metaphor, especially the conflicting results on metaphoric effects on memory as mentioned in the General Introduction, I would like to further investigate metaphoric effects on recognition processes using the model-based analysis, to estimate the effects on item memory and source guessing, respectively. Starting from a conceptual replication of the Experiment 2 from Crawford et al (2014) where a metaphor-incongruency effect on recognition memory was identified, this chapter presents a series of five experiments to further investigate the question.

Experiment 1

In this experiment, a conceptual replication of Experiment 2 from Crawford et al (2014) was conducted in which participants were instructed to memorise positive and negative words randomly presented at the top or bottom of the screen. Memory performance was tested by a subsequent recognition task. A source monitoring paradigm was applied to disentangle the metaphoric effects on each step of recognition processes.

As introduced in Chapter 1, in line with the Attention Elaboration Hypothesis (AEH), Crawford and colleagues (2014) found that valenced words were memorised better when presented at metaphor-incongruent locations than at metaphor-congruent locations. As a conceptual replication, the current experiment also expected the same metaphor-incongruency effect on item memory.

Apart from the metaphor-incongruency effect, a negativity advantage on item memory was also expected according to the original results from Crawford and colleagues (2014), which means negative stimuli were expected to be memorised better in general compared to positive stimuli. This was in line with a longstanding argument about a positive-negative asymmetry, that is, when equal degrees of good and bad factors are present, the psychological effects of the bad ones outweigh those of the good ones (Peeters & Czapinski, 1990; Rozin & Royzman, 2001). In general, negative information receives more processing and is better learned than positive information (Baumeister et al., 2001). Accordingly, in the current experiment, it was hypothesized that participants would memorise negative words better than positive ones.

Besides effects on item memory, metaphor-congruent source guessing biases were expected. As reviewed in Chapter 1, expectation-congruent guessing biases have been widely demonstrated in previous source-monitoring research (Bayen et al., 2000; Ehrenberg & Klauer, 2005; Kuhlmann et al., 2016). Similarly, it was hypothesised that the metaphoric association between valence and verticality encourages participants to guess positive stimuli presented more at up location compared to down location, and negative stimuli presented more at down location than at up location when unable to recall the present location.

Apart from manipulating valence and verticality variables, an awareness condition was added as another independent variable. When investigating the “good is up” metaphor, Sasaki and colleagues (2016) found that in order to activate sensorimotor simulations of vertical directions, conscious awareness of the emotional information was necessary. Similarly, 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, in the instructions of the current experiment, it was 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 *up* or *down* location on the screen in a random sequence.”). Moreover, to explore the expected metaphoric effects under different awareness conditions of the valence-verticality association, three groups were used in this variable, in which participants received different instructions, details see the Design section below. The raised awareness of the metaphoric association was expected to increase the expectation of valence-verticality congruency and consequently to increase the differences in memory performance between metaphor-congruent and incongruent items, as well as to increase the metaphor-congruent source guessing biases.

Methods

Design

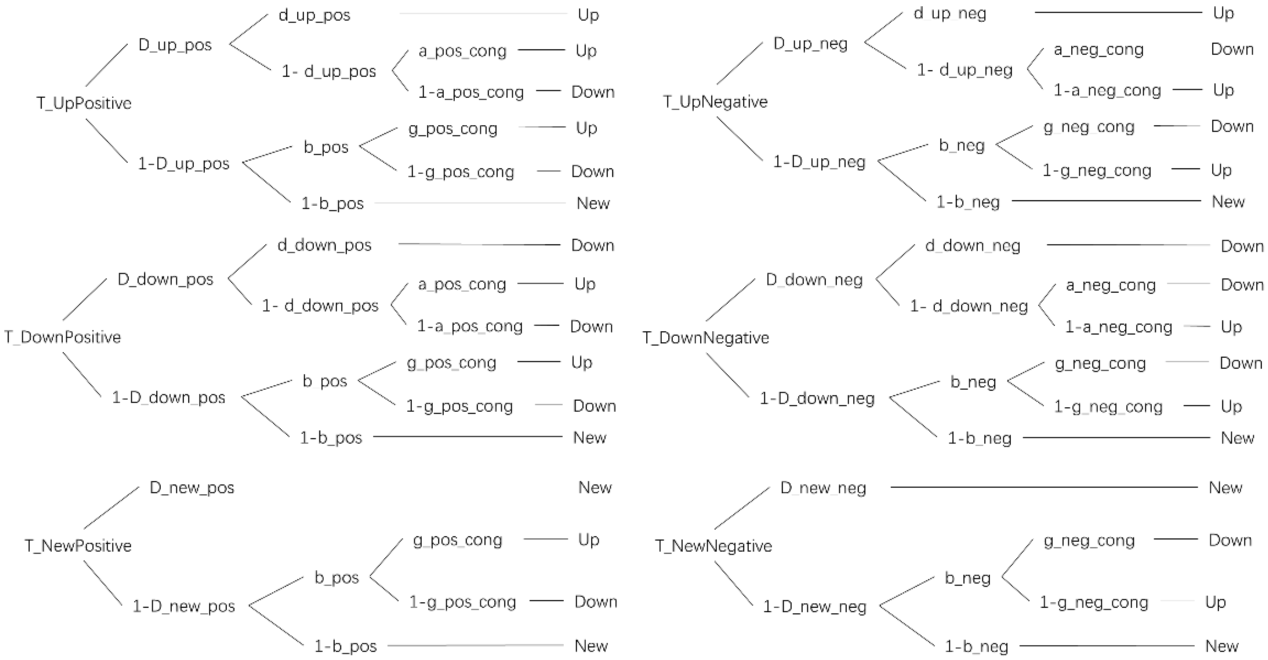
Experiment 1 used a mixed design of 2 valence (positive vs. negative, within-subjects) × 2 verticality (up vs. down, within-subjects) × 3 awareness (stimulus-only vs. stimulus-location vs. stimulus-location-metaphor, between-subjects). Participants were instructed to memorise positive and negative words which were randomly presented at the top or bottom of the screen. Performance was tested later by a recognition task. Specifically, participants in the 3 awareness conditions received different instructions at the beginning of the study. Group 1 participants (stimulus-only condition) were only instructed to memorise the words, while Group 2 participants (stimulus-location condition) were instructed to memorise the words and their present locations at the same time. Moreover, Group 3 participants (stimulus-location-metaphor condition) were instructed to memorise the words and their locations while being made aware of the metaphoric association between positive and up location as well as negative and down location at the beginning before the learning phase and told that they can use this association as a mnemonic strategy.

Adapted 2HTSM model

The standard 2HTSM model (see Figure 2) described in Chapter 1 was adapted in this experiment. In order to take the valence variable into consideration during modelling, six multinomial processing trees were constructed instead of the original three. A valence subscript was added besides the source subscript so that parameters specific to particular source-valence combinations can be estimated separately. For example, *D*_up_neg represents the probability of correctly recognising a negative item that had been presented at the up location during learning as an old item. Furthermore, 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 instead of one particular source location as traditionally do (see Arnold et al., 2013; Kuhlmann et al., 2012). For example, *a*_neg_cong represents the probability of guessing a recognised negative word was from the down location, whereas *a*_pos_cong represents the probability of guessing a recognised positive word was from the up location. See Figure 3 for the illustration, and details of all parameters in the adapted 2HTSM model are listed in Table 3.

Figure 3

Adapted 2HTSM Model of Experiment 1 with Six Multinomial Trees



Note. “T_” stands for the multinomial tree of each source-valence-combination.

Table 3

Parameters in the Adapted 2HTSM Model in Figure 3

Parameter	Definition
*D*_up_pos	Probability of correctly recognising a positive word from up location
*D*_down_pos	Probability of correctly recognising a positive word from down location
*D*_new_pos	Probability of correctly recognising a new positive word is new
*D*_up_neg	Probability of correctly recognising a negative word from up location
*D*_down_neg	Probability of correctly recognising a negative word from down location
*D*_new_neg	Probability of correctly recognising a new negative word is new
*d*_up_pos	Probability of correctly discriminating a positive word is from up location
*d*_down_pos	Probability of correctly discriminating a positive word is from down location
*d*_up_neg	Probability of correctly discriminating a negative word is from up location
*d*_down_neg	Probability of correctly discriminating a negative word is from down location
*b*_pos	Probability of guessing an unrecognised positive word is old
*b*_neg	Probability of guessing an unrecognised negative word is old
*a*_pos_cong	Probability of guessing a recognised positive word is from up location
*a*_neg_cong	Probability of guessing a recognised negative word is from down location
*g*_pos_cong	Probability of guessing an unrecognised positive word is from up location
*g*_neg_cong	Probability of guessing an unrecognised negative word is from down location

Hypotheses

Hypothesis 1 (H1): Metaphor-incongruency effect on item memory, *D*_up_neg > *D*_down_neg; *D*_down_pos > *D*_up_pos;

Hypothesis 2 (H2): Negativity advantage on item memory, *D*_up_neg > *D*_up_pos; *D*_down_neg > *D*_down_pos;

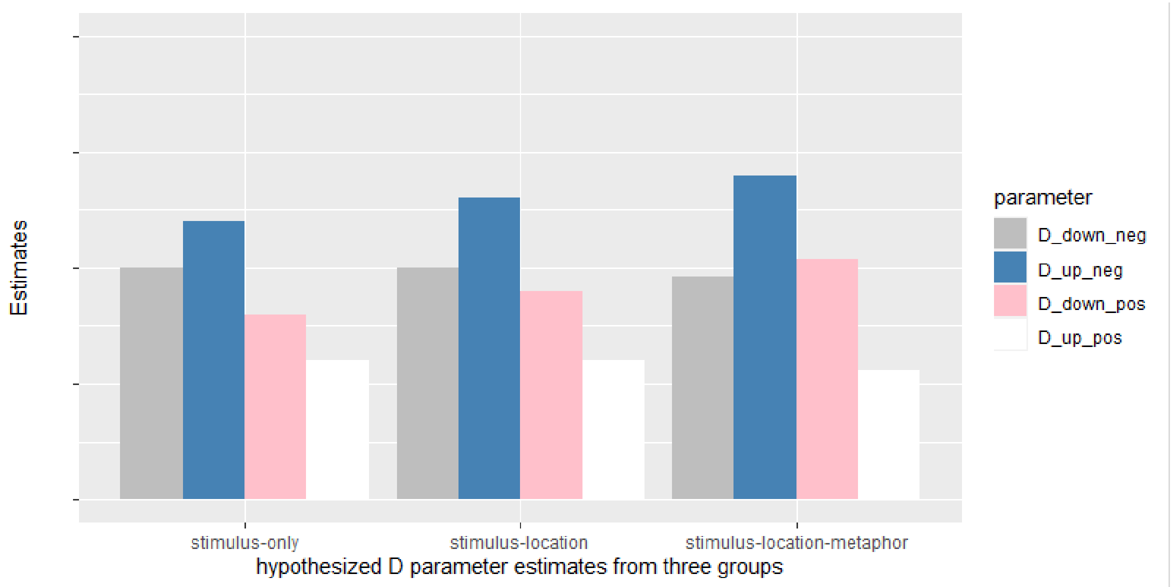
Hypothesis 3 (H3): Metaphor-congruent source guessing biases, *a*_pos_cong > 0.5; *g*_pos_cong > 0.5; *a*_neg_cong > 0.5; *g*_neg_cong > 0.5;

Hypothesis 4 (H4): Awareness was expected to increase metaphoric effects such that higher awareness of the spatial information and the metaphoric association was expected to increase metaphoric effects on item memory and to increase metaphor-congruent source guessing biases.

See Figure 4 and 5 for a visual illustration of the hypotheses.

Figure 4

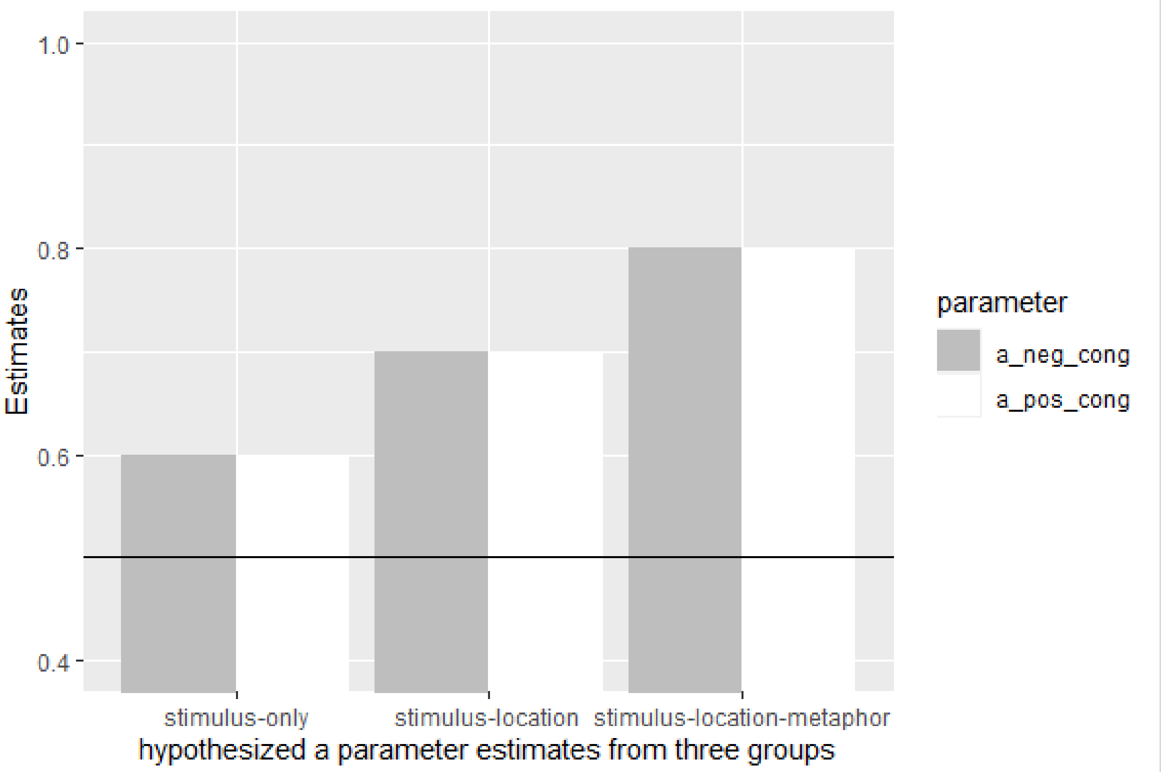
Hypothesised Estimates for Parameter D in Experiment 1



Note. D parameters represent the probabilities of correctly recognising a positive/negative word that had been presented at the up/down location during learning as an old item, with the subscripts indicate different valence-verticality conditions.

Figure 5

Hypothesised Estimates for Parameter a in Experiment 1



Note. *a* and *g* were hypothesised equal so are not listed individually. Details are explained in the Results section. *a* parameters represent the probabilities of guessing the metaphor-congruent vertical location as the source of a present positive/negative word.

Materials

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, $t(78) = 37.61$, $p < .001$. T-tests showed no differences between positive and negative words on arousal scores, length, syllable length or frequency (see Table 4). A full list of all the selected materials see Appendix A.1.

Table 4

Norms of Positive and Negative Words Stimuli (M±SD) in Experiment 1

	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 arousal rating score from ANEW (Bradley & Lang, 1999), “Length” stands for how many letters in each word, “Syllable” stands for how many syllables in each word, “Frequency” stands for word frequency score from ANEW (Bradley & Lang, 1999), with higher numbers indicating higher frequencies; “***” indicates $p < .001$.

Participants

Previous research using source monitoring paradigm recruited 24-30 participants in each between-subject condition. For example, Bayen and colleagues (1996) recruited 24 participants in each group, Ehrenberg & Klauer (2005) used 25 participants per condition, and Meiser and colleagues (2007) allocated 26-30 participants in each condition of each study. Consequently, more than 30 participants were recruited in each group in the current experiment and all the subsequent experiments in the current chapter which used source monitoring paradigm.

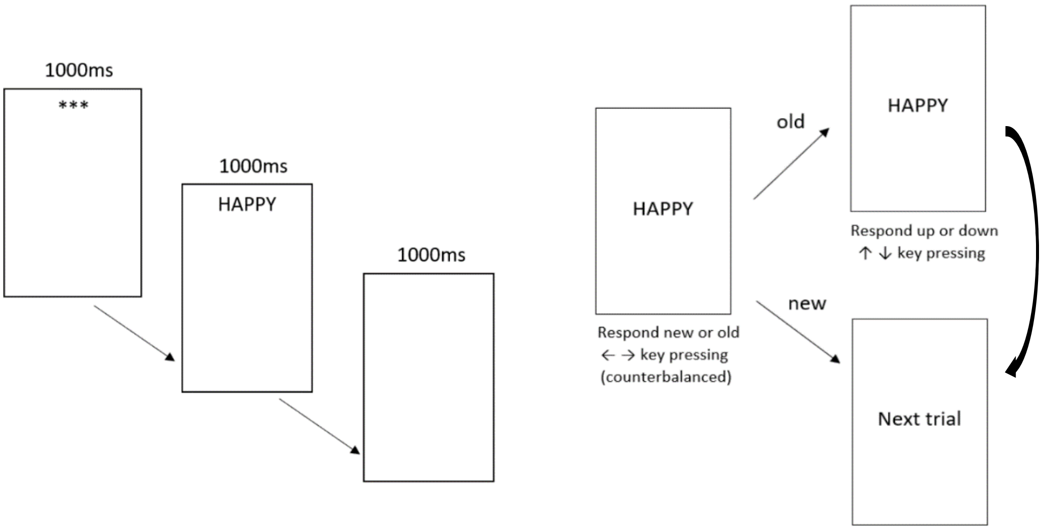
N = 103 English native speakers (N = 35, 34, 34, respectively in each group) were recruited from Cardiff University via flyers and online university-wide Experimental Management System to take part in this study, of which 21 were males, mean age M = 19.98 years (SD = 2.57). Participants got paid by 1 course credit, or £2 cash for their participation.

Procedure

Participants were tested individually in laboratory rooms in the School of Psychology, Cardiff University. 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 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. Top or bottom location was approximate 8 inches above or below the middle point of the screen respectively. For each participant, there were 40 words presented in the learning phase, half of them were positive words and the other half were negative ones. Half of the words from each valence were presented at metaphor-congruent locations, the other half were presented 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 at the location where the word was going to appear for 1000 ms, 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, 40 learned words were randomly interspersed with 40 new words. In each testing trial, one of these words appeared in the middle of the screen and participants were asked to respond whether it was a new (not-learned) or old (learned) word by pressing left or right arrow key (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 testing trials. After completing all 80 testing trials, participants were then debriefed, thanked and dismissed. Figure 6 illustrates an example of a learning trial and a testing trial.

Figure 6

Illustration of a Learning Trial and a Testing Trial in Experiment 1



Note. The left is an example of a learning trial; the right is an example of a testing trial; "HAPPY" represents an example word.

Results and Discussion

An appropriate modelling procedure is necessary to get accurate estimates of parameters in the 2HTSM model. Firstly, alternative models have to be mathematically identifiable. Additional restrictions of parameters are necessary to achieve this, see Table 2 in Chapter 1 for all identifiable submodels of the standard 2HTSM model. Accordingly, the equality constraints were applied within each valence separately in the adapted 2HTSM model of the current experiment, details see Table 5. Notably, similar to the modifications applied on guessing bias parameters (i.e., *a* and *g* represent probability of guessing metaphor-congruent location instead of one specific source), D_A was adapted to represent item memory of metaphor-congruent source presented stimuli (i.e., *D*_up_pos, *D*_down_neg) whereas D_B was adapted to represent item memory of metaphor-incongruent source presented stimuli (i.e., *D*_up_neg, *D*_down_pos).

Table 5

Constraints of the Identifiable Submodels Applied on Experiment 1 Adapted 2HTSM

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

Secondly, among the identifiable submodels, the most appropriate alternative one(s) needs to be selected. The basic rule is to select the models with the capability to test the hypotheses while being as parsimonious as possible. In the current experiment, due to expecting item memory of words presented at different vertical locations to be different, Model 4, 5a and 5d were the first to be excluded. Then, source guessing biases of recognised items (*a*) and unrecognised items (*g*) were constrained to be equal, which was empirically supported and widely used in previous 2HTSM research (Bayen et al., 2000; Wulff & Kuhlmann, 2020). As a result, Model 6a and 6b were excluded. Previous research also provided evidence to support equal source memory performance for both sources when applying 2HTSM models (Bayen et al., 2000; Wulff & Kuhlmann, 2020). Also due to the parsimony reason, Model 5b and 5c were preferred than Model 6c and 6d[1].

At last, it came to the last two alternatives, Model 5b and 5c, with the only difference of whether item memory of distractors was equal to metaphor-congruent presented learned items or metaphor-incongruent ones (i.e., *D*_new_pos = *D*_up_pos & *D*_new_neg = *D*_down_neg vs. *D*_new_pos = *D*_down_pos & *D*_new_neg = *D*_up_neg). In principle, D_N was set to be limited by the lower bound of item memory between the two, following previous research (Kuhlmann & Touron, 2011; Meiser et al., 2007), which is *D*_congruent (i.e., *D*_up_pos and *D*_down_neg) in the current experiment due to the hypothesised metaphor-incongruency effect on item memory. Therefore, Model 5b was supposed to be preferred. But in previous research, when using this constraint, it did not mean that the other alternative submodel was not considered at all. Meiser and colleagues (2007) reported that setting the D_N equal to the *D* parameter with superior value did not change any of the results regarding their hypotheses. Kuhlmann & Touron (2011) reported that the deviance information criterion (DIC, a criterion combining a Bayesian measure of model fit and a

---

[1] Considering the potential inclusion of Model 6c and 6d to investigate the possibility of the metaphoric effects on source memory process, the deviance information criterion (DIC, Spiegelhalter et al., 2002) was calculated for 4 alternatives (Model 5b, 5c, 6c and 6d) in each group data, with the smaller DIC value indicating better model fit. Results either showed preference of Model 5b/5c ($DIC_{6c/6d} - DIC_{5b/5c} > 6$), or did not show any preferences (all $\Delta DICs \leq 6$), details see Appendix B. Taking parsimony into consideration at the same time, Model 5b/5c was preferred due to smaller degrees of freedom. The situation remained the same for the following experiments within this chapter. As a result, only Model 5b and 5c were considered and reported in this thesis.

measure of model complexity, with smaller value indicating model preference, see Spiegelhalter et al., 2002) showed preferences of the selected models ($D_{\mathrm{N}} = D_{\mathrm{old}}$ with lower value) over the other alternative ($D_{\mathrm{N}} = D_{\mathrm{old}}$ with higher value). However, support of Model 5b from DIC value was not always obtained in the current experiment. For example, in Group 1, Model 5c was preferred rather than 5b ($\mathrm{DIC}_{5b} = 1375$, $\mathrm{DIC}_{5c} = 1368$, $\Delta\mathrm{DIC} = 7$), judged by a popular rule of thumb for model comparison that a difference in excess of 6 provides strong support (Millar, 2009; Spiegelhalter et al., 2002). This situation made it less convincing to constantly prefer Model 5b than 5c in the current research.

As a solution, both Model 5b and 5c were considered for each dataset and DICs were calculated for both models, when there was a strong support to prefer Model 5c ($\mathrm{DIC}_{5b} - \mathrm{DIC}_{5c} > 6$), it was selected for further analysis of parameter estimates, otherwise, Model 5b was selected. The model-selection process described above remained the same for the following experiments in this chapter, and details of DIC value of the alternative models are presented in Appendix B.

As mentioned in Chapter 1, to take heterogeneity of participants into consideration, a latent trait approach (Klauer, 2010) was applied to the current analysis, using the TreeBUGS package (Heck et al., 2018) in R (R Core Team, 2013). Samples from the posterior distribution of parameters are drawn with the Monte Carlo–Markov chain (MCMC) algorithm. 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 $\hat{R} \leq 1.1$[2] (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). All final selected models

---

[2] except for a few ρ parameters which indicate the correlations between parameters, all parameters actually met the criterion of $\hat{R} \leq 1.05$, see Bott et al (2020).

fitted well: Group 1 Model 5c: $p_{T1} = .480$ and $p_{T2} = .443$; Group 2 Model 5b: $p_{T1} = .322$ and $p_{T2} = .364$; Group 3 Model 5c: $p_{T1} = .462$ and $p_{T2} = .544$.

After the final models were selected, parameter estimates and hypotheses tests were performed. Figures of parameter estimates were created with the R package ggplot2 (Wickham, 2016). Group-level mean estimates of the posterior distribution for the parameters and corresponding 95% Bayesian credibility intervals (BCI) are reported in Table 7. Figure 7 and 8 illustrate and compare item memory and guessing biases from three groups. Bayesian posterior *probability* (*pp*) is used to test one-tailed hypotheses and *pp* of null hypothesis smaller than .05 is considered statistically substantial.

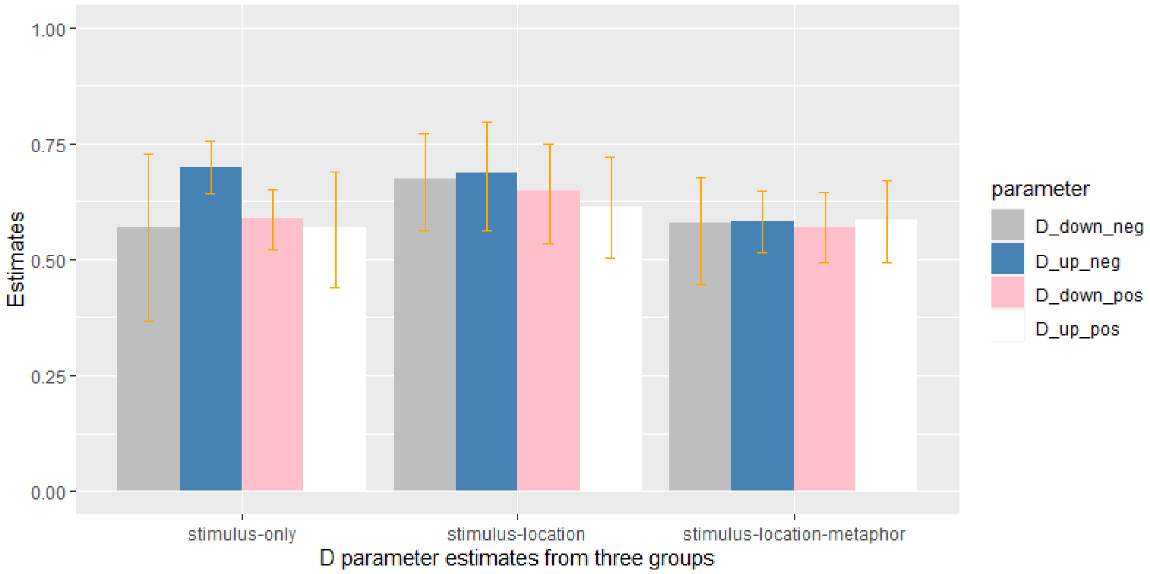
Table 6

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

parameter	*M*(*SD*)	95%BCI	Group	*M*(*SD*)	95%BCI	Group	*M*(*SD*)	95%BCI	Group
*D*_up_neg	.70 (.03)	[.64, .75]	1	.69 (.06)	[.56, .80]	2	.58 (.03)	[.51, .65]	3
*D*_down_neg	.57 (.09)	[.37, .73]	1	.67 (.05)	[.56, .77]	2	.58 (.06)	[.45, .68]	3
*D*_up_pos	.57 (.06)	[.44, .69]	1	.61 (.06)	[.50, .72]	2	.59 (.05)	[.49, .67]	3
*D*_down_pos	.59 (.03)	[.52, .65]	1	.65 (.05)	[.54, .75]	2	.57 (.04)	[.49, .65]	3
*d*_neg	.44 (.06)	[.33, .55]	1	.44 (.09)	[.25, .61]	2	.44 (.09)	[.25, .61]	3
*d*_pos	.36 (.08)	[.19, .50]	1	.39 (.08)	[.21, .54]	2	.45 (.07)	[.32, .58]	3
*a*_neg_cong	.63 (.03)	[.56, .69]	1	.56 (.04)	[.49, .63]	2	.55 (.05)	[.46, .65]	3
*a*_pos_cong	.58 (.05)	[.48, .67]	1	.61 (.04)	[.54, .69]	2	.68 (.05)	[.58, .77]	3
*b*_neg	.36 (.06)	[.24, .49]	1	.36 (.06)	[.24, .48]	2	.38 (.06)	[.27, .50]	3
*b*_pos	.26 (.05)	[.18, .36]	1	.29 (.05)	[.19, .40]	2	.38 (.04)	[.29, .46]	3

Figure 7

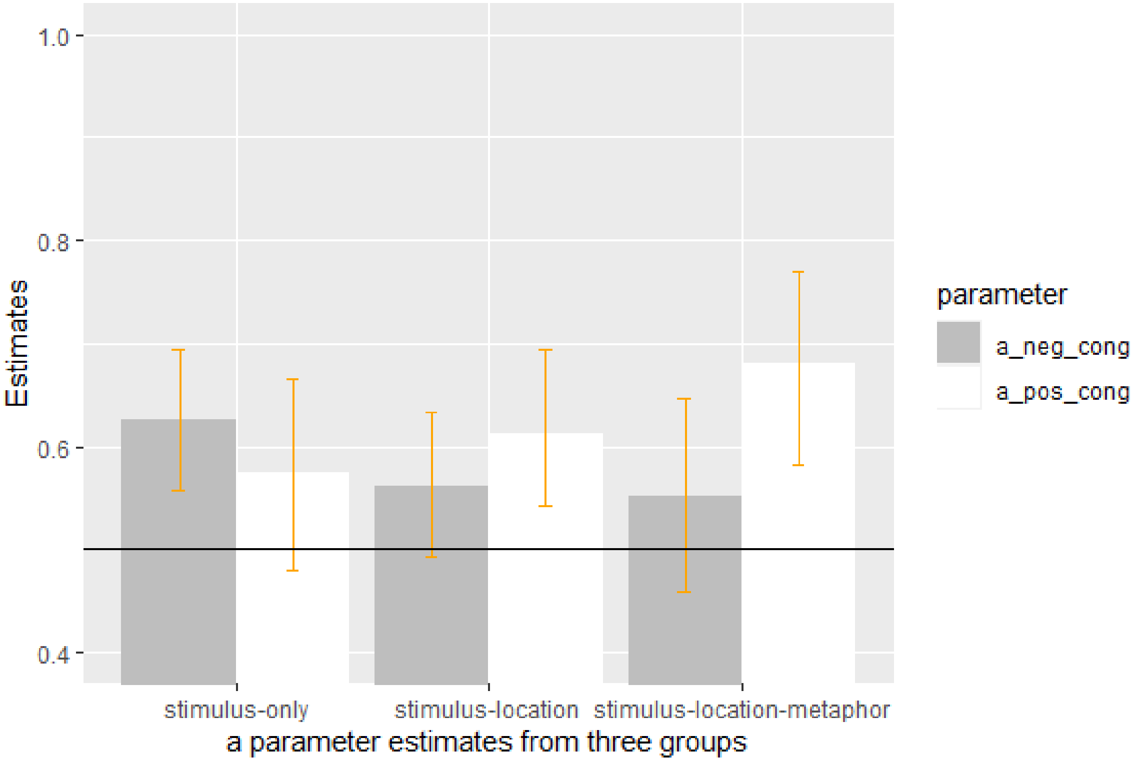
Estimates for Parameter D (Group-Level) from Three Groups in Experiment 1



Note. Error bars indicate 95%BCI. *D* parameters represent the probabilities of correctly recognising a positive/negative word that had been presented at the up/down location during learning as an old item, with the subscripts indicate different valence-verticality conditions.

Figure 8

Estimates for Parameter a (Group-Level) from Three Groups in Experiment 1



Note. Error bars indicate 95%BCI. *a* parameters represent the probabilities of guessing the metaphor-congruent vertical location as the source of a present positive/negative word.

H1 regarding metaphor-incongruency effect on item memory was not substantially supported. Negative words were only marginally better remembered in Group 1 when presented at up location compared to when presented at down location ($pp_{D\_up\_neg<D\_down\_neg}$ = .08). The effect was not demonstrated on positive stimuli in Group 1 ($pp_{D\_down\_pos<D\_up\_pos}$ = .40). Moreover, the predicted effect was neither identified in Group 2 ($pp_{D\_up\_neg<D\_down\_neg}$ = .40, $pp_{D\_down\_pos<D\_up\_pos}$ = .30) nor in Group 3 ($pp_{D\_up\_neg<D\_down\_neg}$ = .49, $pp_{D\_down\_pos<D\_up\_pos}$ = .61).

H2 regarding negativity advantage on item memory was not substantially supported either. Some evidence was found in Group 1 that negative words presented at the up location were better remembered compared to positive words presented at up location ($pp_{D\_up\_neg<D\_up\_pos}$ = .02). But also in Group 1, when presented at down location, the

difference was not found ($pp_{\text{D\_down\_neg<D\_down\_pos}}$ = .55). The effect was neither identified in Group 2 ($pp_{\text{D\_up\_neg<D\_up\_pos}}$ = .81, $pp_{\text{D\_down\_neg<D\_down\_pos}}$ = .84) nor in Group 3 ($pp_{\text{D\_up\_neg<D\_up\_pos}}$ = .53, $pp_{\text{D\_down\_neg<D\_down\_pos}}$ = .44).

H3 regarding metaphor-congruent source guessing biases, on the contrary, was substantially supported across three groups. In Group 1, guessing biases of words presented at metaphor-congruent locations was strongly supported with negative words ($pp_{\text{a\_neg\_cong<0.5}}$ = .00) and marginally supported with positive words ($pp_{\text{a\_pos\_cong<0.5}}$ = .06). Group 2 showed evidence to support the effect within both valence contexts ($pp_{\text{a\_pos\_cong<0.5}}$ = .00; $pp_{\text{a\_neg\_cong<0.5}}$ = .04). Group 3, however, only showed the bias with positive words as stimuli ($pp_{\text{a\_pos\_cong<0.5}}$ = .00) but not with negative ones ($pp_{\text{a\_neg\_cong<0.5}}$ = .14).

H4 regarding the influence of awareness on metaphoric effects was only partially supported. Higher awareness of the spatial information and the metaphoric association was expected to increase both metaphor-incongruent effects on item memory and metaphor-congruent source guessing biases. The metaphor-incongruent effect on item memory was only marginally detected in Group 1 and not in either Group 2 or 3, which consequently did not support the enlarging effect of awareness. On the contrary, the predicted effect of awareness on source guessing biases was partially demonstrated. The rise of valence-verticality awareness led to an increase of metaphor-congruent source guessing for positive stimuli, however a decrease for negative stimuli, see Figure 8 for an illustration[3].

In general, the expected effects on recognition processes were only demonstrated on source guessing biases but not on item memory performance.

Across the three groups, the metaphor-congruent source guessing biases were generally consistent, which was in line with previous research as mentioned in the introduction, supporting the existence of metaphoric effect on cognition. In a situation where participants cannot recall where a stimulus had been presented, the metaphoric association can serve as a guidance to rely on for participants to generate the guessing responses.

---

[3] In the TreeBUGS package, the comparison of one parameter across models is limited to between two models only. Consequently, there is no way to implement a linear analysis of the parameters across three groups. As a result, the arguments regarding the awareness effects are based on observations from the figures of parameter estimates.

Moreover, when made aware of the vertical location or the valence-verticality association before encoding, the bias was more pronounced for positive stimuli as expected, but not for negative ones. This partially supported the hypothesis that stronger awareness can emphasize the metaphor-congruent expectations and result in higher metaphor-congruent source guessing biases, but for some unknown reasons, this was only true for positive words rather than negative ones.

On the contrary, there was just a marginal metaphor-incongruency effect on item memory which only showed in Group 1 with negative words (*D*_up_neg > *D*_down_neg). Similarly, the predicted negativity advantage on item memory was only partially demonstrated in Group 1 on words presented at up location (*D*_up_neg > *D*_up_pos). Also, contradicting hypothesis (H4), with the rise of awareness of verticality and metaphoric association, the metaphoric effects on item memory were not enlarged at all, but diminished instead. And the negativity advantage disappeared in Group 2 and 3 as well. An explanation for the lacking metaphoric effect on item memory in Group 2 and 3 could be that the limited attentional resources during encoding were allocated more on the spatial information and less on 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, diminished the differences in memory performance between metaphor-congruent and metaphor-incongruent presented words, and the differences in memory performance between negative words and positive words as well. Further investigation is in need to support the findings and the speculative explanation.

Experiment 2

Experiment 2 was designed to serve as a comparison to Experiment 1 and altogether to emphasize the important role that valence played in the metaphoric effects demonstrated in Experiment 1, mainly on the source guessing biases, but also in a more subtle way on item memory in Group 1. It was expected that the differences of spatial location in the encoding phase can elicit neither source guessing bias nor item memory differences when valence is absent. Experiment 2 used the same paradigm as Experiment 1, except that valence-neutral stimuli were used as memory materials.

Methods

Design

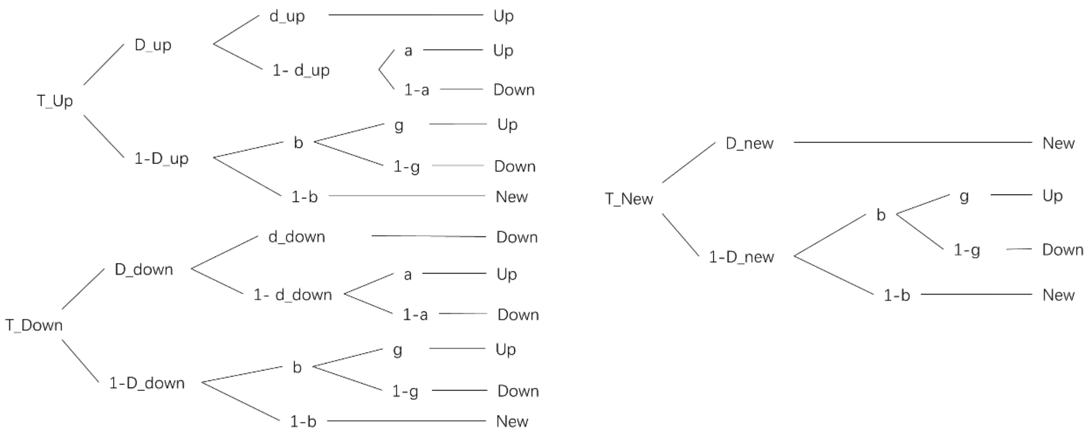
A mixed design of 2 verticality (up vs. down, within-subjects) × 2 awareness conditions (stimulus-only vs. stimulus-location, between-subjects) design was used in this experiment. Due to the absence of valence, there were no predictions regarding metaphoric association of valence and verticality, resulting in only 2 awareness conditions. Participants in different awareness groups received instructions at the beginning of the experiment, to either memorise the stimuli only, or alternatively, to memorise the stimuli along with the present locations.

Adapted 2HTSM model

Similar to Experiment 1, an adapted 2HTSM model was constructed in the current experiment, see Figure 9 for details. The model was simpler compared to Experiment 1 due to the absence of valence.

Figure 9

Adapted 2HTSM Model of Experiment 2 with Three Multinomial Trees



Note. “T_” stands for multinomial tree of each source.

Hypotheses

Hypothesis 1 (H1): No substantial differences of item memory of stimuli presented at the top or bottom of the screen in encoding: $D\_up = D\_down$.

Hypothesis 2 (H2): No guessing biases towards either vertical location, that is, when unable to recall the original location of a word, participants should equally guess the up or down location: a (= g) = 0.5.

Materials

Tigrinya characters were selected to create valence-neutral stimuli, assuming unfamiliarity of this language to the student participants from the U.K. Subjective preference ratings were obtained in a pre-test for 48 Tigrinya characters (N = 15), on a scale ranging from -7 to +7, indicating extreme disliking to extreme liking. Three characters were excluded due to their average pre-test ratings being smaller than -3 or larger than 3. One more character with a relatively extreme rating (M = 2.87) was also excluded to make the number of stimuli even, so that the formal test could have equal number of materials been allocated to up and down locations. After exclusion, the mean preference rating of the selected 44 Tigrinya characters was 0.73 (SD = 0.85). Materials were made into pictures with black font colour on a white background in size of 200×200 pixels, to achieve a consistent presentation during the procedure. See Appendix A.2 for a full list of the selected materials.

Participants

N = 70 fluent English-speaking participants (N = 35 in each group) were recruited from Cardiff University to take part in this study, of which 11 were males, mean age = 19.33 years (SD = 1.66). Participants got paid by 1 course credit or alternatively, £2 cash for their participation.

Procedure

The procedure was identical to Experiment 1, except for the presentation duration in the learning phase. Due to the participants' unfamiliarity with the stimuli, in each learning trial, after 1000 ms of fixation cue, the character was presented for 2000 ms instead of 1000 ms. Participants learned 22 randomly selected Tigrinya characters from either up or down vertical location in a random sequence. Then the total 44 Tigrinya characters were presented in the testing phase in a random sequence. In each testing trial, participants were required to answer whether a presented character was learned or not from the learning phase and to subsequently discriminate where it was presented if classified a character as learned. At the end of the experiment, participants were asked whether they had knowledge of the Tigrinya language, in order to make sure the stimuli were equally unfamiliar to all participants.

Results and Discussion

Similar model selection procedure as in Experiment 1 was conducted. Notably, although the null hypothesis was expected regarding item memory, namely, no differences in item memory for stimuli presented at different locations were expected (D_up = D_down), submodel 5b and 5c were again selected as to compare with Experiment 1. The idea was to leave them free for estimation, then use hypothesis testing to compare their values, which was the identical procedure as in Experiment 1.

Convergence checks of the selected final models were conducted in an identical way as in Experiment 1. All final selected models[4] fitted well: Group 1 Model 5b: $p_{T1} = .579$ and $p_{T2} = .705$; Group 2 Model 5b: $p_{T1} = .417$ and $p_{T2} = .372$.

Group-level mean estimates of the posterior distribution for the parameters and corresponding 95%BCI are reported in Table 7. Figure 10 and 11 illustrate item memory and guessing biases from the two groups. A 95%BCI of the differences between parameters in the posterior distribution that excludes zero was considered statistically substantial.

---

[4] Unlike in Experiment 1, there was no predicted lower bound of Ds. As a result, the model with smaller DIC was selected but if the selected model was not substantially better than the other (ΔDIC ≤ 6), the hypothesis test was conducted in the other alternative model as well to assure the reliability of results, as in Group 2. Otherwise (ΔDIC > 6), only the selected model results were considered, as in Group 1. DICs of all alternative models are reported in Appendix B.

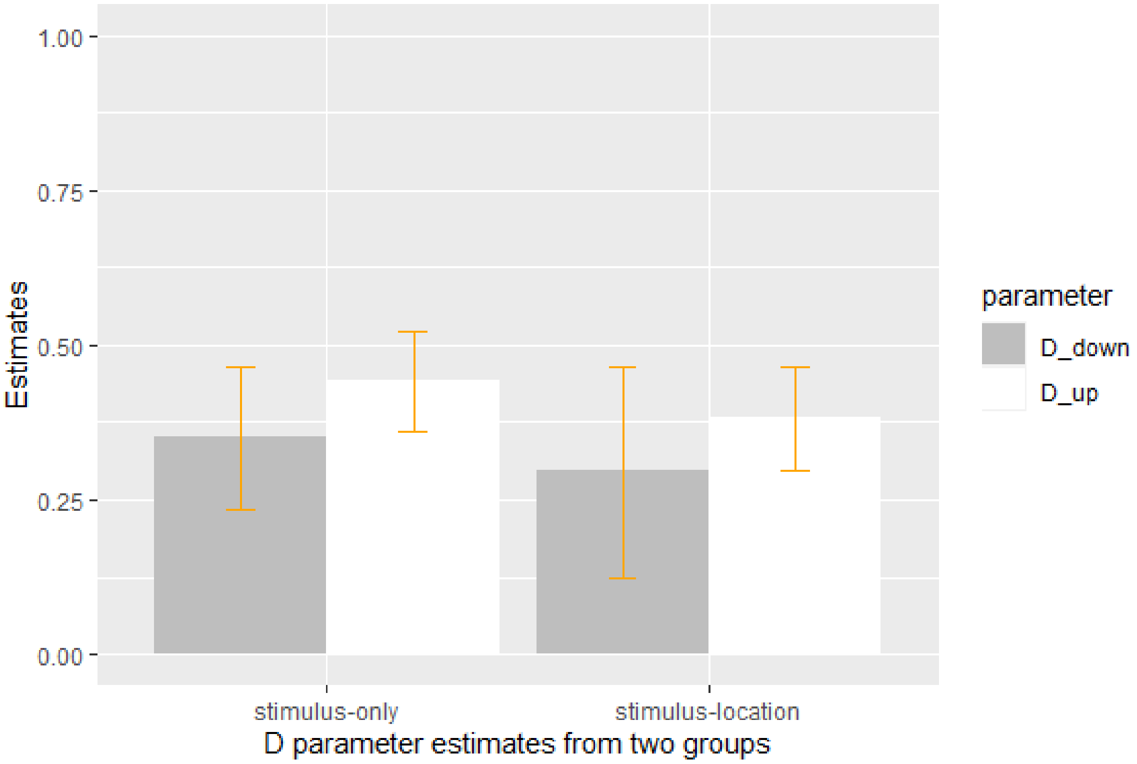
Table 7

Estimates for Parameters (Group-Level) from the Selected Models in Experiment 2

parameter	*M*(*SD*)	95%BCI	Group	*M*(*SD*)	95%BCI	Group
*D*_up	.44 (.04)	[.36, .52]	1	.38 (.04)	[.30, .47]	2
*D*_down	.35 (.06)	[.24, .46]	1	.30 (.09)	[.12, .46]	2
d	.71 (.16)	[.36, .97]	1	.82 (.17)	[.33, 1.00]	2
a	.51 (.03)	[.46, .58]	1	.50 (.03)	[.45, .55]	2
b	.52 (.03)	[.46, .58]	1	.59 (.03)	[.52, .65]	2

Figure 10

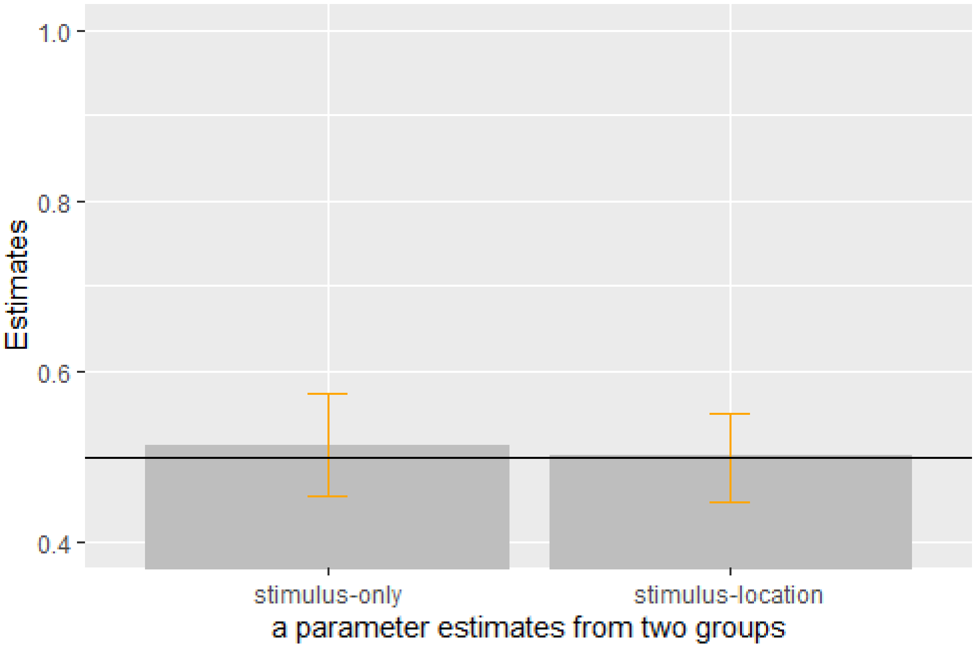
Estimates for Parameter D (Group-Level) from Two Groups in Experiment 2



Note. Error bars indicate 95%BCI. *D* parameters represent the probabilities of correctly recognising a stimulus that had been presented at the up/down location during learning as an old item.

Figure 11

Estimates for Parameter a (Group-Level) from Two Groups in Experiment 2



Note. Error bars indicate 95%BCI. *a* parameters represent the probabilities of guessing the up location as the source of a present stimulus.

H1 regarding no substantial differences of item memory of stimuli presented at the top or bottom of the screen in encoding was supported in both groups, Group 1: Δ*D* = (*D*_up - *D*_down) = .09 [-.05, .24]; Group 2: Δ*D* = (*D*_up - *D*_down) = .09 [-.08, .27][5].

H2 regarding no guessing biases towards either vertical location was supported in both groups as well, Group 1: Δ*a* = (*a* - .50) = .01 [-.04, .08]; Group 2: Δ*a* = (*a* - .50) = .00 [-.05, .05][6].

As predicted, there were no spatial effects on item memory or source guessing biases. With the absence of valence, no item memory differences were identified due to vertical location itself. Moreover, unsurprisingly, there were no guessing biases of either vertical location for valence-neutral stimuli. This was in line with previous research on source guessing which suggest that if participants do not have any previous expectations or

---

[5] In the other alternative model 5c, Δ*D* = (*D*_up - *D*_down) = .10 [-.04, .25].

[6] In the other alternative model 5c, Δ*a* = (*a* - .50) = .00 [-.05, .05], identical to in model 5b.

knowledge on the item-source relations, their guessing biases mimic the item-source contingency in the learning phase, namely, $a = .50$ if items were equally divided across sources as in the current experiment (probability-matching theory, proposed by Spaniol & Bayen, 2002; see also Arnold et al., 2013; Bayen & Kuhlmann, 2011).

And this pattern was consistent across two awareness groups, which means that, whether participants paid attention to spatial information or not did not affect their memory performance or guessing biases. Comparing with Experiment 1, the results altogether highlight the importance of valence in metaphoric effects, such that the spatial locations of stimuli only make a difference when valence is present (Experiment 1), and make no differences when valence is absent (Experiment 2).

Experiment 3

Experiment 3 was designed to be a parallel study to Experiment 1, using positive and negative emojis instead of words as materials. The purpose was to investigate whether metaphoric effects can be demonstrated in recognition processes when using stimuli with less semantic information but more direct on affective valence. This was expected to more directly address the role of valence per se, in terms of metaphoric effects.

Methods

Design

The design was identical to Experiment 1 using a mixed design of 2 valence (positive vs. negative, within-subjects) × 2 verticality (up vs. down, within-subjects) × 3 awareness conditions (stimulus-only vs. stimulus-location vs. stimulus-location-metaphor, between-subjects).

An identical adapted 2HTSM model as in Experiment 1 was used, see Figure 3 and Table 3 in Experiment 1 section for a reminder of details. Hypotheses were the same as in Experiment 1 as well, and are briefly described below:

Hypothesis 1 (H1): Metaphor-incongruency effect on item memory: D_up_neg > D_down_neg; D_down_pos > D_up_pos; Hypothesis 2 (H2): Negativity advantage on item memory: *D*_up_neg > *D*_up_pos; *D*_down_neg > *D*_down_pos; Hypothesis 3 (H3): Metaphor-congruent source guessing biases: *a*_pos_cong (= *g*_pos_cong) > 0.5; *a*_neg_cong (= *g*_neg_cong) > 0.5; Hypothesis 4 (H4): Higher awareness of the spatial information and the metaphoric association was expected to increase metaphoric effects on item memory and metaphor-congruent source guessing biases. See Figure 4 and 5 in Experiment 1 for a visual illustration of the hypotheses.

Materials

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. In order to ensure a general consensus on the valence of each emoji, the 4 emojis with more than 2 unmatched valence ratings (i.e., more than 2 participants out of the total 20 in the pre-test rated a positive stimulus as having negative valence or vice versa) were excluded from the materials. Another 4 stimuli with mean ratings relatively closer to 0 than the rest were excluded in order to equalise the number of positive and negative emojis while keeping the valence of materials polarised. 72 emojis were left to be used as items 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. The selected emojis were all made into pictures of 72×68 pixels in size, to achieve a consistent presentation. See Appendix A.3 for a full list of the selected emojis.

Participants

N = 103 fluent English-speaking participants (*N* = 35, 34, 34, respectively in each group) were recruited from Cardiff University to take part in this study, of which 17 were males, mean age = 19.61 years (*SD* = 2.79). Participants got paid by 1 course credit or £2 cash for their participation.

Procedure

Except for the duration of presentation in the learning phase, the procedure was identical to Experiment 1. In each learning trial, after 1000 ms of fixation cue, the emoji was presented for 2000 ms instead of 1000 ms in Experiment 1. 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 from the learning phase and to subsequently discriminate where it was presented if classified an emoji as learned.

Results and Discussion

Identical model selection and estimation procedure as in Experiment 1 was conducted. All final selected models fitted well: Group 1 Model 5b: p_{T1} = .389 and p_{T2} = .432; Group 2 Model 5c: p_{T1} = .455 and p_{T2} = .382; Group 3 Model 5b: p_{T1} = .576 and p_{T2} = .515.

Group-level mean estimates of the posterior distribution for the parameters and corresponding 95%BCI are reported in Table 8. Figure 12 and 13 illustrate item memory and guessing biases parameter estimates from the three groups.

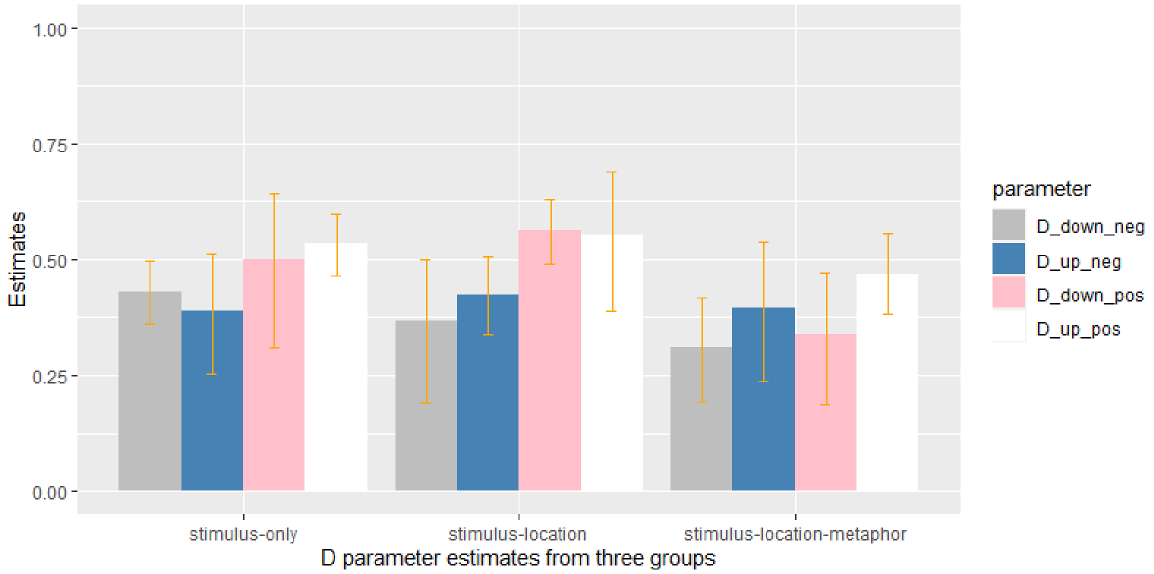
Table 8

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

parameter	*M*(*SD*)	95%BCI	Group	*M*(*SD*)	95%BCI	Group	*M*(*SD*)	95%BCI	Group
*D*_up_neg	.39 (.07)	[.25, .51]	1	.42 (.04)	[.34, .51]	2	.40 (.08)	[.24, .54]	3
*D*_down_neg	.43 (.03)	[.36, .50]	1	.37 (.08)	[.19, .50]	2	.31 (.06)	[.19, .42]	3
*D*_up_pos	.53 (.03)	[.47, .60]	1	.55 (.08)	[.39, .69]	2	.47 (.04)	[.38, .56]	3
*D*_down_pos	.50 (.08)	[.31, .64]	1	.56 (.04)	[.49, .63]	2	.34 (.07)	[.19, .47]	3
*d*_neg	.67 (.10)	[.49, .88]	1	.64 (.19)	[.20, .95]	2	.73 (.18)	[.29, .98]	3
*d*_pos	.22 (.10)	[.03, .41]	1	.38 (.07)	[.23, .52]	2	.60 (.16)	[.28, .93]	3
*a*_neg_cong	.54 (.04)	[.47, .61]	1	.57 (.04)	[.48, .65]	2	.63 (.04)	[.55, .70]	3
*a*_pos_cong	.57 (.03)	[.52, .62]	1	.60 (.04)	[.53, .67]	2	.60 (.03)	[.54, .66]	3
*b*_neg	.46 (.03)	[.40, .53]	1	.46 (.05)	[.37, .55]	2	.47 (.04)	[.40, .54]	3
*b*_pos	.55 (.05)	[.46, .65]	1	.56 (.04)	[.48, .65]	2	.57 (.05)	[.47, .67]	3

Figure 12

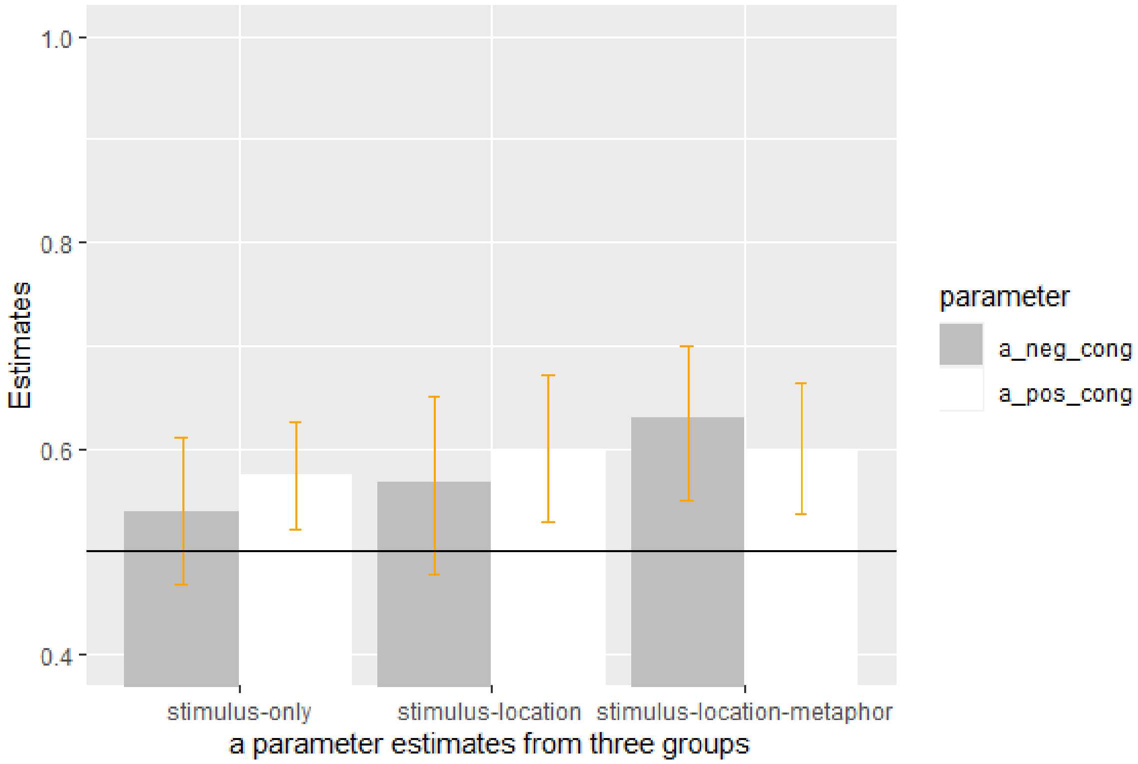
Estimates for Parameter D (Group-Level) from Three Groups in Experiment 3



Note. Error bars indicate 95%BCI. *D* parameters represent the probabilities of correctly recognising a positive/negative emoji that had been presented at the up/down location during learning as an old item, with the subscripts indicate different valence-verticality conditions.

Figure 13

Estimates for Parameter a (Group-Level) from Three Groups in Experiment 3



Note. Error bars indicate 95%BCI. *a* parameters represent the probabilities of guessing the metaphor-congruent vertical location as the source of a present positive/negative emoji.

H1 regarding the metaphor-incongruency effect on item memory was not supported in any of the three groups, Group 1: $pp_{\text{D\_up\_neg<D\_down\_neg}}$ = .70, $pp_{\text{D\_down\_pos<D\_up\_pos}}$ = .62; Group 2, $pp_{\text{D\_up\_neg<D\_down\_neg}}$ = .27, $pp_{\text{D\_down\_pos<D\_up\_pos}}$ = .47; Group 3, $pp_{\text{D\_up\_neg<D\_down\_neg}}$ = .16, $pp_{\text{D\_down\_pos<D\_up\_pos}}$ = .93.

H2 regarding negativity advantage on item memory was also not supported in any of the three groups, Group 1: $pp_{\text{D\_up\_neg<D\_up\_pos}}$ = .98, $pp_{\text{D\_down\_neg<D\_down\_pos}}$ = .80; Group 2: $pp_{\text{D\_up\_neg<D\_up\_pos}}$ = .93, $pp_{\text{D\_down\_neg<D\_down\_pos}}$ = 1.00; Group 3: $pp_{\text{D\_up\_neg<D\_up\_pos}}$ = .83, $pp_{\text{D\_down\_neg<D\_down\_pos}}$ = .63. From the posterior probability and the visual illustration of *D* parameter estimates in Figure 12, a trend can be identified that positive emojis were memorised relatively better overall compared to negative ones, which contradicted the hypothesis.

H3 regarding metaphor-congruent source guessing biases, on the contrary, was substantially supported across three groups. In Group 1, guessing biases of emojis presented at metaphor-congruent locations were strongly supported with positive emojis ($pp_{a\_pos\_cong<0.5}$ = .00) but not with negative ones ($pp_{a\_neg\_cong<0.5}$ = .13). Group 2 showed strong evidence from positive emojis again ($pp_{a\_pos\_cong<0.5}$ = .00), and marginal from negative ones ($pp_{a\_neg\_cong<0.5}$ = .07). Whereas Group 3 provided strong support from both valence contexts ($pp_{a\_pos\_cong<0.5}$ = .00; $pp_{a\_neg\_cong<0.5}$ = .00).

H4 regarding the influence of awareness on metaphoric effects was only partially supported, mainly by the effect on guessing biases. Figure 13 clearly showed an increasing trend of metaphor-congruent source guessing biases along with the rise of valence-verticality association awareness. On the contrary, since the metaphor-incongruency effect on item memory was not demonstrated at all across the three groups, the expected enlarging effect from higher awareness was obviously not identified.

In general, metaphoric effects influenced source guessing biases quite substantially, but did not influence item memory as predicted. To take Experiment 1 results into account altogether, the metaphoric effects seem to show a consensual pattern, meaning “true” for source guessing but “false” for item memory. The increasing trend of guessing biases across three groups showed that higher awareness of the metaphor can strengthen the metaphor-congruent expectation which participants may have used to rely on to generate metaphor-congruent source guessing responses.

Moreover, the negativity advantage on item memory was not supported with emojis as materials. One potential reason is that even the emojis with supposedly negative expressions still have some comical feature, which may interfere with the impression of negativity and as such may hinder the deeper cognitive processing because of negativity. Consequently, the negativity advantage was absent. Though this still cannot explain why positive emojis were memorised slightly better compared to negative ones. Further evidence is in need to discuss this finding.

As an issue in the current experiment, the overall item memory estimates were lower compared to Experiment 1, indicating a higher difficulty level of the task. This suggests that memorising emojis is generally more difficult than memorising words. This is potentially due to two reasons. First, the similarity between emojis may be higher than between words. For example, smiley faces with or without blush are both positive emojis but can be difficult to discriminate. But there is no such problem when memorising words. Second, the less descriptive nature of emojis makes it harder for internal verbal rehearsal during encoding and consequently may have affected the recognition performance.

Further studies are in need to investigate the absence of substantial metaphoric effects on item memory. It needs to be considered whether it is in general harder for metaphoric effects to be manifested on item memory, compared to source guessing biases, regardless of memory materials.

Experiment 4

As introduced in Chapter 1, embodied cognition theory suggests that, in cognitive processes abstract concepts are associated with concrete concepts automatically, and that metaphors provide a scaffold for people to use concrete concepts to process abstract concepts cognitively (Lakoff & Johnson, 1980, 1999). Neuroscience research also demonstrates similar neural activity when participants were in a task to discriminate positive or negative words as when to discriminate words with up or down connotation (Quadflieg et al., 2011).

According to this argument, if we expect the valence-verticality metaphoric association to affect item memory and source guessing, the related concrete concepts with high or low vertical connotations should trigger the same effects. The current experiment was expected to investigate the metaphoric effects with an entirely physically ground context.

Methods

Design

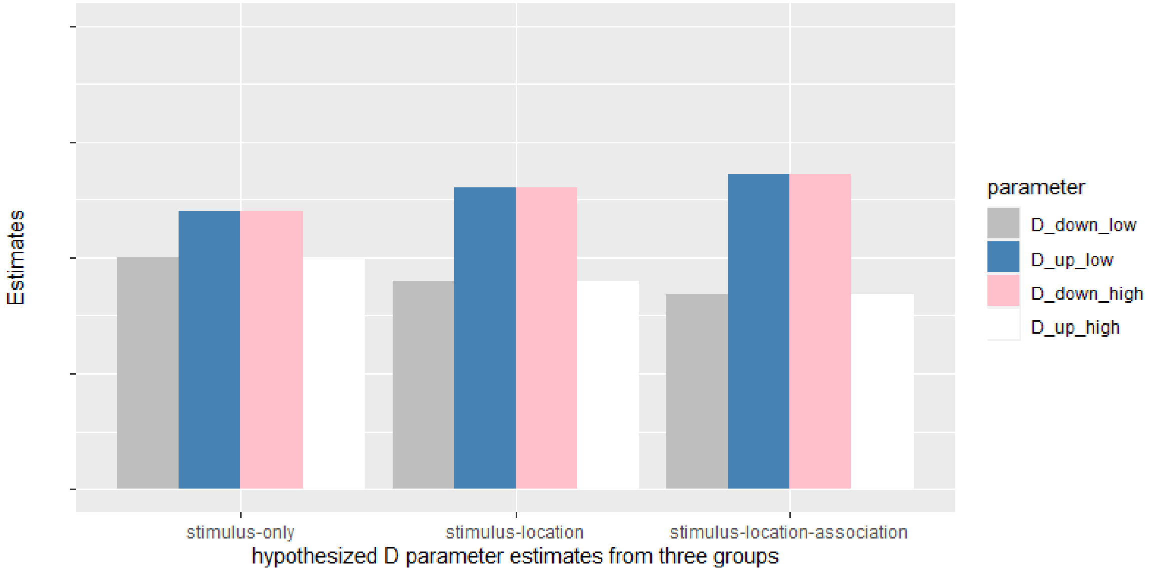
The design of this experiment was almost identical to Experiment 1, with the only difference of replacing the valence variable with a connotation variable, yielding the following design: 2 connotations (high vs. low, within-subjects) × 2 verticality (up vs. down, within-subjects) × 3 awareness conditions (stimulus-only vs. stimulus-location vs. stimulus-location-association, between-subjects).

A similar adapted 2HTSM model as in Experiment 1 was used, except that positive/negative valence was substituted by high/low vertical connotation.

The hypotheses were: Hypothesis 1 (H1): Connotation-incongruency effect on item memory, namely, participants should remember stimuli better when presented at connotation-incongruent locations compared to at connotation-congruent locations: *D*_up_low > *D*_down_low, *D*_down_high > *D*_up_high; Hypothesis 2 (H2): Connotation-congruent guessing biases, namely, when unable to recall the present location of an item, participants should guess it was presented at the location congruent to its connotation rather than the incongruent location: *a*_high_cong (= *g*_high_cong) > 0.5, *a*_low_cong (= *g*_low_cong) > 0.5; Hypothesis 3 (H3): Higher awareness of the spatial information and the connotation-verticality association was expected to increase the connotation-incongruency effect on item memory and the connotation-congruent source guessing biases. Due to the absence of valence, there was no negativity advantage expected in the current experiment. Consequently, the hypothesised pattern of *D* parameters has changed, see Figure 14 for a visual illustration.

Figure 14

Hypothesised Estimates for Parameter D in Experiment 4



Note. D parameters represent the probabilities of correctly recognising a high/low word that had been presented at the up/down location during learning as an old item, with the subscripts indicate different connotation-verticality conditions.

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. 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 9. See Appendix A.4 for a full list of the selected materials.

Table 9

Comparison of High and Low Connotation Words Stimuli (M±SD) in Experiment 4

	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

Note. "Length" stands for how many letters in each word, "Syllable" stands for how many syllables in each word; "***" indicates $p < .001$.

Participants

N = 105 English native speakers (N = 32, 28, 42, respectively in each group after data exclusion) were recruited from Cardiff University to take part in this study, of which 16 were males, mean age = 20.22 years (*SD* = 3.04). Participants got 1 course credit for their participation.

Procedure

This experiment was conducted online due to the COVID-19 pandemic when lab-based tests were not allowed. Informed consent was obtained from each participant before the experiment started. Moreover, due to the limitations of online data collection, the screen could not be set vertically as was done in the previous, lab-based experiments. Also, the size of screen depended on participants' own end devices. All other procedures were identical to Experiment 1.

Results and Discussion

Due to concerns about data quality in online data-collection, the data were trimmed based on the average completing duration of testing phase using the Tukey criterion. That is, the duration of each participant needed to be within the range of [Q1 - 3*IQR, Q3 + 3*IQR], where Q1 represents the lower quantile and Q3 represents the upper quantile, and IQR represents the interquartile range (see Tukey, 1977). Three participants were excluded because of too long duration, indicating no constant concentration during the test.

Similar model selection and estimation procedures as in Experiment 1 were conducted. 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.

Group-level mean estimates of the posterior distribution for the parameters and corresponding 95%BCI are reported in Table 10. Figure 15 and 16 illustrate item memory and guessing biases from the three groups.

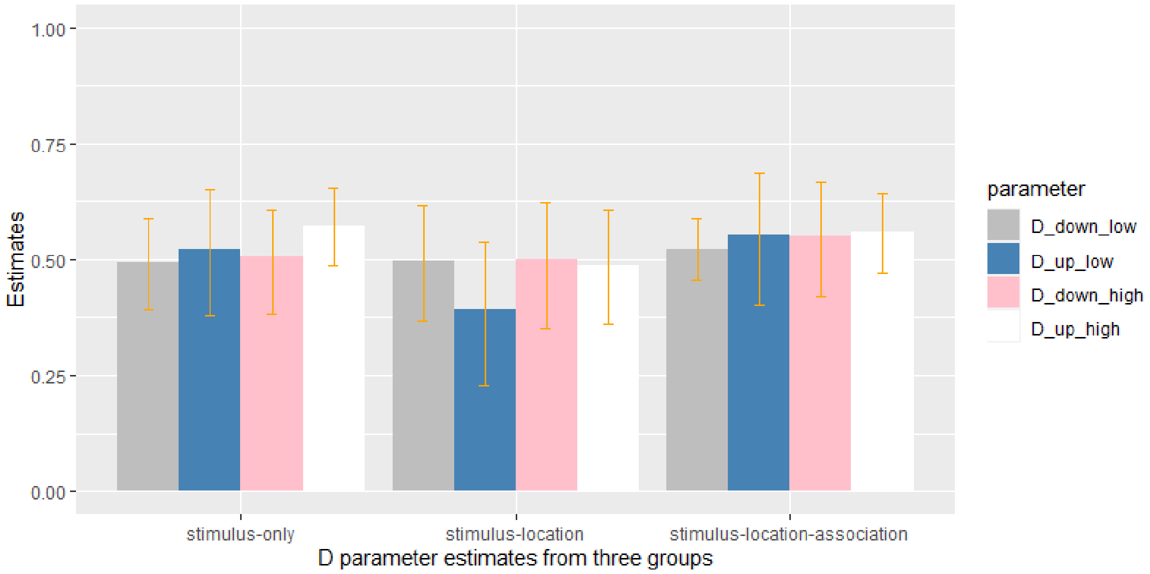
Table 10

Estimates for Parameters (Group-Level) from the Selected Models in Experiment 4

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

Figure 15

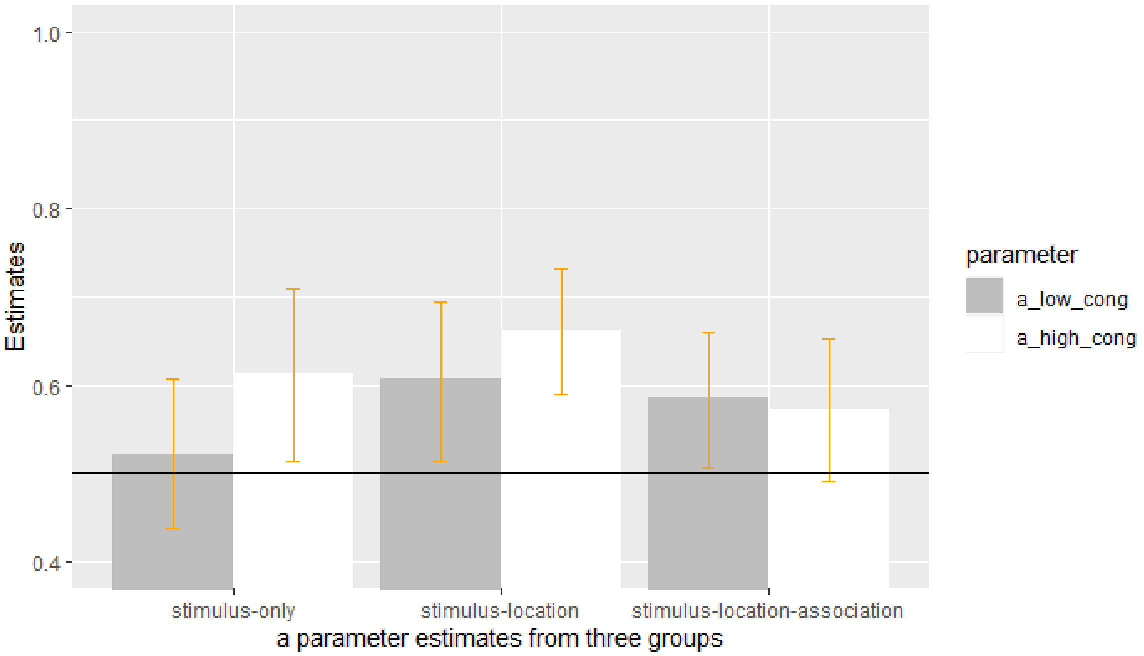
Estimates for Parameter D (Group-Level) from Three Groups in Experiment 4



Note. Error bars indicate 95%BCI. *D* parameters represent the probabilities of correctly recognising a high/low word that had been presented at the up/down location during learning as an old item, with the subscripts indicate different connotation-verticality conditions.

Figure 16

Estimates for Parameter a (Group-Level) from Three Groups in Experiment 4



Note. Error bars indicate 95%BCI. *a* parameters represent the probabilities of guessing the connotation-congruent vertical location as the source of a present high/low word.

H1 regarding connotation-incongruency effect on item memory was not supported in any of the three groups, Group 1: $pp_{\text{D\_up\_low<D\_down\_low}}$ = .35, $pp_{\text{D\_down\_high<D\_up\_high}}$ = .83; Group 2: $pp_{\text{D\_up\_low<D\_down\_low}}$ = .89, $pp_{\text{D\_down\_high<D\_up\_high}}$ = .43; Group 3: $pp_{\text{D\_up\_low<D\_down\_low}}$ = .34, $pp_{\text{D\_down\_high<D\_up\_high}}$ = .54.

H2 regarding connotation-congruent source guessing biases was substantially supported across three groups. In Group 1, guessing biases of high-connotation words presented at connotation-congruent location was strong ($pp_{\text{a\_high\_cong<0.5}}$ = .01) but not of low-connotation words ($pp_{\text{a\_low\_cong<0.5}}$ = .31). Whereas in both Group 2 and 3, the connotation-congruent source guessing biases were supported in both connotation contexts, Group 2: $pp_{\text{a\_high\_cong<0.5}}$ = .00; $pp_{\text{a\_low\_cong<0.5}}$ = .01; Group 3: $pp_{\text{a\_high\_cong<0.5}}$ = .04; $pp_{\text{a\_low\_cong<0.5}}$ = .02.

H3 regarding the influence of awareness on item memory and source guessing was only partially supported. Again, since the absence of predicted connotation-incongruency effect on item memory in all three groups, the expected enlarging effect from higher awareness was subsequently not identified. Regarding source guessing biases, there was an increasing tendency of the connotation-congruent source guessing biases from Group 1 to Group 2, but not from Group 2 to Group 3, see Figure 16.

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 biases demonstrated in Experiment 1 and 3. In the current, entirely physically grounded, context, the congruency between present location and stimulus' vertical connotation provides an expectation for participants to rely on to generate source guessing responses. And the similarity of metaphor-congruent and connotation-congruent source guessing tendencies support the argument that the valence processing is mapped onto the physical dimension of verticality, because valence, as an abstract concept per se, is not tied to any spatial dimension and should not create any source expectations by itself.

Similar to in Experiment 3, the subtle incongruency effect on item memory demonstrated in Experiment 1 Group 1 was not found in the current experiment. The connotation-incongruency did not show any influence on item memory. In Experiment 3, it was discussed that more evidence was in need to consider the possibility of a null metaphoric effect on item memory. Here, the null effect in the current experiment provides some support for this argument. According to the simulation theory (Barsalou, 1999, 2008), if the “good is up” metaphor plays a role in effects on item memory of valenced stimuli, it is supposed to function via the automatic activation of the related physical concept, verticality. In this vein, directly using vertical concepts was supposed to trigger a similar or even stronger effect. On the contrary, the expected effect was not indicated, which suggest an unsuccessful pathway for the metaphor to be effective on item memory.

Experiment 5

Further investigation was considered to explore the updated hypothesis on substantial metaphoric effects on source guessing biases but null on item memory. But because of the

COVID-19 pandemic, face-to-face experiments were banned in the U.K. Under this circumstance, a series of lab-based replication studies in China was conducted instead in order to provide better control of the experimental environment. Specifically, Experiment 5a, 5b and 5c replicated Experiment 1, 3 and 4, respectively, in a Mandarin speaking context.

To provide background support for this replication, previous literature suggests that the "good is up" metaphor is not limited to English culture or language context only, but has also been demonstrated in Mandarin (Wu et al., 2019), German (Dudschig et al., 2015), Russian and French (Luodonpää-Manni & Viimaranta, 2010) contexts. Specifically, Chinese people also associate positive valence with up location and negative valence with down location (Wu et al., 2019), which provides opportunity to conduct this series of replications in China.

On the basis of the previous four experiments, the current experiment expected that metaphor-incongruency or connotation-incongruency would have no effect on item memory, whereas source guessing biases with respect to metaphor-congruent or connotation-congruent locations were still expected.

Another modification in this experiment was that only the Group 1 "stimulus-only" awareness condition was kept. The reason for this was that in Experiment 1 to 4, the metaphoric effects on item memory were manifested only in the stimulus-only awareness condition, and not even consistently across studies. On the other hand, the influence of awareness was not a central research question in the present project. Consequently, the replications focused on one awareness condition to see if consistent results could be found across three sub-studies.

Experiment 5a

Methods

Design

Experiment 5a used a within-subjects design of 2 valence (positive vs. negative) × 2 verticality (up vs. down). Hypotheses were: Hypothesis 1 (H1): No metaphor-incongruency effect on item memory: D_up_neg = D_down_neg, D_down_pos = D_up_pos; Hypothesis 2 (H2): Negativity advantage on item memory: D_up_neg > D_up_pos; D_down_neg > D_down_pos; Hypothesis 3 (H3): Guessing biases towards metaphor-congruent locations:

*a*_pos_cong (= *g*_pos_cong) > 0.5, *a*_neg_cong (= *g*_neg_cong) > 0.5.

Materials

132 words were selected from Affective Norms for English Words (ANEW, Bradley & Lang, 1999). They were then translated to two-character Mandarin words for pre-test (N = 38) to obtain subjective valence ratings from Chinese participants, in order to assure validity. 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.5 for a full list of the selected materials.

Participants

N = 47 Mandarin native speakers were recruited from Wuhan University to take part in this study, of which 25 were males, 22 were females, mean age = 19.74 years (SD = 1.65). Participants got paid by 1 course credit or ¥10 cash for their participation.

Procedure

Participants were tested individually in laboratory rooms at Wuhan University. The procedure was identical to Experiment 1 except that in the instructions at the beginning of the experiment, participants were all only told to memorise the stimuli.

Results and Discussion

Identical model selection and estimation procedure as in Experiment 1 was conducted. The final selected model fitted well: Model 5b: $p_{T1} = .535$ and $p_{T2} = .502$. Group-level mean estimates of the posterior distribution for the parameters and corresponding 95%BCI are reported in Table 11. And same as previous experiments, 95%BCI and Bayesian posterior probability (*pp*) were used to report hypotheses test.

Table 11

Estimates for Parameters (Group-Level) from the Selected Model in Experiment 5a

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*_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]
*b*_neg	.52 (.05)	[.42, .63]
*b*_pos	.35 (.05)	[.25, .45]

H1 regarding no metaphor-incongruency effect on item memory was supported: *D*_up_neg - *D*_down_neg = -.06 [-.19, .05]; *D*_down_pos - *D*_up_pos = -.03 [-.15, .08]. As a comparison, there was a marginal metaphor-incongruency effect demonstrated in Experiment 1 Group 1 on negative words (*D*_up_neg > *D*_down_neg), but not for positive words.

H2 regarding negativity advantage on item memory was partially supported by stimuli from down location, $pp_{\text{D\_down\_neg<D\_down\_pos}}$ = .00, but not from up location, $pp_{\text{D\_up\_neg<D\_up\_pos}}$ = .25. On the contrary, the effect was demonstrated in Experiment 1 Group 1 with words presented at up location (*D*_up_neg > *D*_up_pos), but not from down location.

H3 regarding metaphor-congruent source guessing biases were marginally supported with positive words, $pp_{\text{a\_pos\_cong<0.5}}$ = .09, and not supported with negative words, $pp_{\text{a\_neg\_cong<0.5}}$ = .87. As a comparison, metaphor-congruent source guessing biases in Experiment 1 Group 1 were more substantially detected, with strong evidence with negative words ($pp_{\text{a\_neg\_cong<0.5}}$ = .00) and marginal evidence with positive words ($pp_{\text{a\_pos\_cong<0.5}}$ = .06).

Experiment 5b

Methods

Design

Experiment 5b used the same design as Experiment 5a, which was a within-subjects design of 2 valence (positive vs. negative) × 2 verticality (up vs. down) but with emojis as materials instead. The three hypotheses were identical to Experiment 5a as well.

Materials

The same 80 emojis from Lisbon Emoji and Emoticon Database (LEED, see Rodrigues et al., 2018) as used in Experiment 3 pre-test were also used here and has been pre-tested (*N* = 38) again to obtain subjective valence ratings from Chinese participants. Considering the relatively low estimates of item memory in Experiment 3, Experiment 5b 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.6 for a full list of the selected materials.

Participants

N = 43 Mandarin native speakers were recruited from Wuhan University to take part in this study, of which 18 were males, 25 were females, mean age = 19.35 years (*SD* = 1.12). Participants got paid by 1 course credit or ¥10 cash for their participation.

Procedure

Participants were tested individually in laboratory rooms at Wuhan University. The procedure was identical to Experiment 3 except that in the instructions at the beginning of the experiment, participants were all only told to memorise the stimuli.

Results and Discussion

Identical model selection and estimation procedure as in Experiment 3 was conducted. The final selected model fitted well: Model 5b: $p_{T1} = .540$ and $p_{T2} = .577$. Group-level mean estimates of the posterior distribution for the parameters and corresponding 95%BCI are reported in Table 12.

Table 12

Estimates for Parameters (Group-Level) from the Selected Model in Experiment 5b

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*_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]
*b*_neg	.49 (.04)	[.41, .57]
*b*_pos	.63 (.03)	[.57, .70]

H1 regarding no metaphor-incongruency effect on item memory was supported: *D*_up_neg - *D*_down_neg = .03 [-.11, .17]; *D*_down_pos - *D*_up_pos = -.08 [-.24, .08]. Similarly, metaphor-incongruency effect on item memory was not detected in Experiment 3 Group 1 either.

H2 regarding negativity advantage on item memory was marginally supported by stimuli from down location, $pp_{\mathrm{D\_down\_neg<D\_down\_pos}}$ = .09, but not from up location, $pp_{\mathrm{D\_up\_neg<D\_up\_pos}}$ = .19. On the contrary, negativity advantage on item memory was not detected in Experiment 3 Group 1 at all, with emojis from either up or down location.

H3 regarding metaphor-congruent source guessing biases was marginally supported with both valence contexts, $pp_{\mathrm{a\_pos\_cong<0.5}}$ = .07, $pp_{\mathrm{a\_neg\_cong<0.5}}$ = .06. As a comparison, the expected source guessing biases were detected in Experiment 3 Group 1 with positive emojis ($pp_{\mathrm{a\_pos\_cong<0.5}}$ = .00), but not with negative emojis ($pp_{\mathrm{a\_neg\_cong<0.5}}$ = .13).

Experiment 5c

Methods

Design

Experiment 5c used a within-subjects design of 2 connotation (high vs. low) × 2 verticality (up vs. down). Hypotheses were: Hypothesis 1 (H1): No connotation-incongruency effect on item memory: *D*_up_low = *D*_down_low, *D*_down_high =

*D*_up_high; Hypothesis 2 (H2): Guessing biases towards connotation-congruent locations: *a*_high_cong (= *g*_high_cong) > 0.5, *a*_low_cong (= *g*_low_cong) > 0.5.

Materials

96 words from Lebois and colleagues' (2015) research materials were translated to two-character Mandarin words, and synonyms due to translation were deleted before pre-test ($N = 38$). Eighty words, half with high verticality connotations, the other half with low verticality connotations, were selected with mean rating 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, $t(78) = -21.94$, $p < .001$. See Appendix A.7 for a full list of the selected materials.

Participants

$N = 42$ Mandarin native speakers were recruited from Wuhan University to take part in this study, of which 24 were males, 18 were females, mean age = 19.86 years ($SD = 1.03$). Participants got paid by 1 course credit or ¥10 cash for their participation.

Procedure

Participants were tested individually in laboratory rooms at Wuhan University. The procedure was identical to Experiment 4 except that 5c was a lab-based study and the screen was set vertically as in Experiment 1-3 in order to increase the salience of the vertical dimension in terms of the difference between the up and down locations. Similar to in Experiment 5a and 5b, participants were all only told to memorise the stimuli without any additional awareness of the spatial information or connotation-verticality association.

Results and Discussion

Identical model selection and estimation procedure as in Experiment 4 was conducted. The final selected model fitted well: Model 5b: $p_{T1} = .442$ and $p_{T2} = .325$. Group-level mean estimates of the posterior distribution for the parameters and corresponding 95%BCI are reported in Table 13.

Table 13

Estimates for Parameters (Group-Level) from the Selected Model in Experiment 5c

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*_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]
*b*_low	.31 (.05)	[.21, .42]
*b*_high	.33 (.06)	[.22, .44]

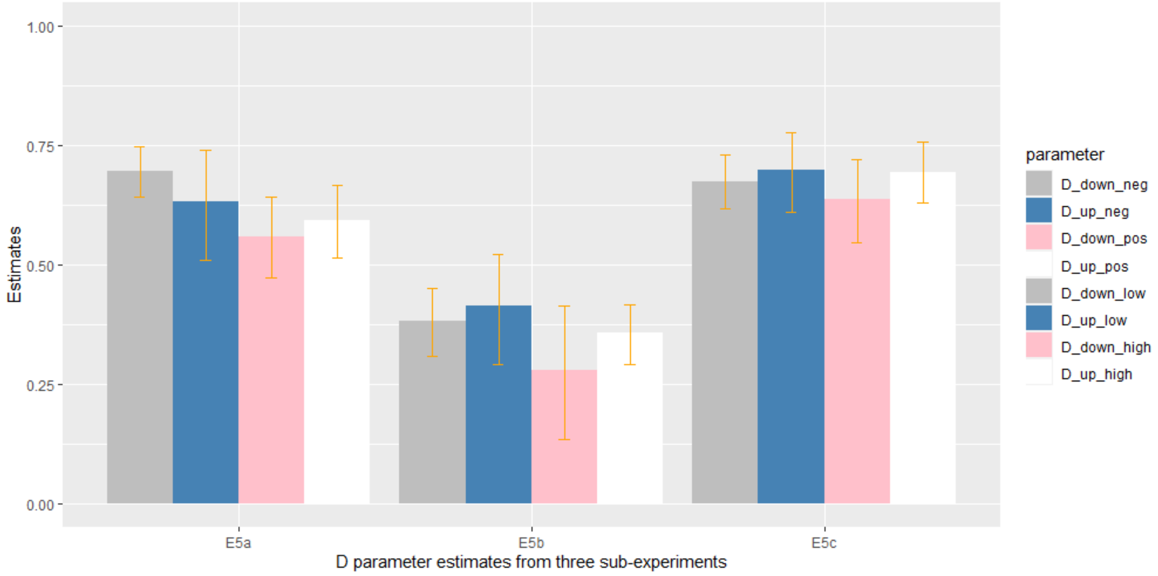
H1 regarding no connotation-incongruency effect on item memory was supported: *D*_up_low - *D*_down_low = .02 [-.08, .12]; *D*_down_high - *D*_up_high = -.06 [-.17, .05]. Similarly, connotation-incongruency effect on item memory was not detected in Experiment 4 either.

H2 regarding connotation-congruent source guessing biases was substantially supported with both connotation contexts, $pp_{\text{a\_high\_cong}<0.5}$ = .00, $pp_{\text{a\_low\_cong}<0.5}$ = .00. As a comparison, the expected guessing biases were detected in Experiment 4 Group 1 with high-connotation words ($pp_{\text{a\_high\_cong}<0.5}$ = .01), but not with low-connotation words ($pp_{\text{a\_low\_cong}<0.5}$ = .31).

Figure 17 and 18 illustrate *D* and *a* parameter estimates across Experiment 5a to 5c.

Figure 17

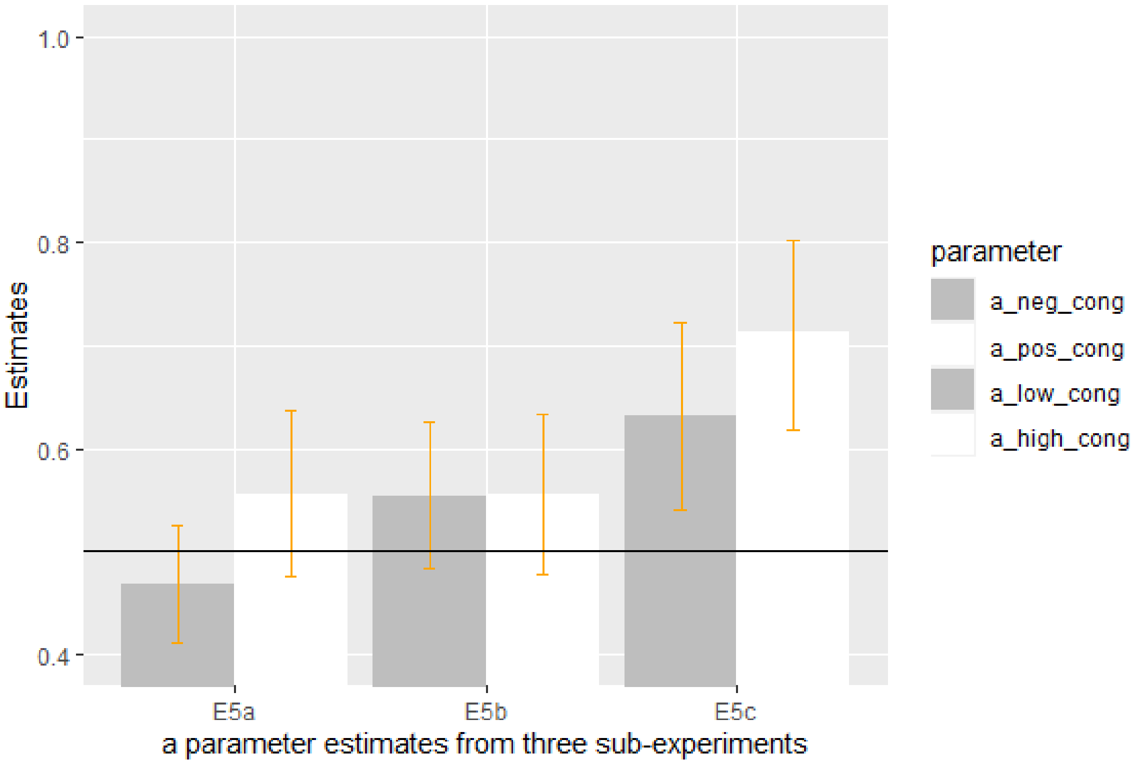
Estimates for Parameter D (Group-Level) from Three Sub-experiments in Exp 5



Note. Error bars indicate 95%BCI. Same colour was used to represent negative valence from Experiment 5a & 5b and low connotation from Experiment 5c, same for positive valence and high connotation, for easiness of comparison. *D* parameters represent the probabilities of correctly recognising a positive/negative/high/low stimulus that had been presented at the up/down location during learning as an old item, with the subscripts indicate different valence-verticality or connotation-verticality conditions.

Figure 18

Estimates for Parameter a (Group-Level) from Three Sub-experiments in Exp 5



Note. Error bars indicate 95%BCI. Same colour was used to represent negative valence from Experiment 5a & 5b and low connotation from Experiment 5c, same for positive valence and high connotation, for easiness of comparison. *a* parameters represent the probabilities of guessing the metaphor-congruent or connotation-congruent vertical location as the source of a present positive/negative/high/low stimulus.

Discussion

Across three sub-experiments, the updated hypothesis on no metaphor-incongruency or connotation-incongruency effects on item memory has been substantially supported. To take results from Experiment 1, 3 and 4 into account as well, except for the marginal effect detected in Experiment 1, there was no other evidence to suggest effects of the “good is up” metaphor or spatial connotation on item memory. Altogether, this research suggests that presentation at metaphor-congruent location or metaphor-incongruent location in the encoding phase does not change item memory performance of valenced stimuli. And this remains the same for words with physically vertical location connotations.

Source guessing biases were relatively consistent across three sub-experiments. Although metaphor-congruent source guessing biases were only partially supported in Experiment 5a and 5b, Experiment 5c provided strong evidence for connotation-congruent source guessing biases. Potentially, this is because that the link attained between physical stimuli and spatial location is closer compared to the link between valence and spatial location conceptually, with the former creating stronger expectations of the congruency that then guides the source guessing. Moreover, according to the previous findings on the awareness factor, there was an increasing trend for source guessing biases along with the rising awareness of spatial information and the association between valence or connotation and verticality. As a result, the stimulus-only awareness condition chosen in Experiment 5 was expected to have a relatively lower level of guessing biases. Nevertheless, Experiment 5 still provides some evidence to suggest the predicted source guessing biases. This finding under the circumstance of no explicit information of the metaphor suggests a spontaneous use of the metaphoric association in generating source guessing responses.

A negativity advantage on item memory, as expected in Experiment 5a and 5b was only partially supported in 5a and marginally demonstrated in 5b. The evidence, in general, was not substantial, especially if we also take into account the fact that the related evidence obtained in Experiment 1 and 3 was also only weak.

Discussion of this chapter

In this chapter, five experiments were conducted to investigate the metaphoric effects of “good is up” in the context of recognition memory. The metaphoric effects on recognition processes are conceptualised to be possibly effective in two ways, that is, via source guessing strategies and/or via item discrimination. More specifically, metaphor-congruent source guessing biases and better discrimination performance for items presented at metaphor-incongruent locations were hypothesised, but substantial evidence was only obtained for source guessing biases but not for item discrimination.

Metaphor creates expectation of valence-verticality congruency

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 in situations when participants were unable to recall the presented location of a valenced stimulus. This is in line with previous research on schema-congruent guessing biases which suggest 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, this research indicates that metaphors can function in the source guessing process similar to schemas, to create the expectation and to guide guessing responses toward a direction congruent with the expectations.

Some evidence for metaphor-congruent source guessing biases was obtained even when the metaphor was not mentioned explicitly in the stimulus-only and stimulus-location awareness groups, which means the association between valence and verticality was implicit but, to some extent, solid in people’s minds.

Combined with the connotation-congruent source guessing biases demonstrated in Experiment 4 and 5c, the results support the argument that valence, as an abstract concept, is processed in a similar way as the concept of physical height underlying vertical connotations in the processing of material objects and situations that have meanings tied to emphasised verticality (e.g., “SKY” and up vs. “PIT” and down). This furthermore suggests 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). And connotation-congruent guessing biases are generally stronger than metaphor-congruent guessing, as manifested in Experiment 5, which may suggest that 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 according spatial locations.

The “good is up” metaphor does not influence item memory

On the other hand, regarding a hypothesized metaphoric effect on item memory performance, our results were in favour of a null effect. Previous studies which reported no effects of the “good is up” metaphor on recognition memory all used different paradigms compared to 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 encoding phase (Experiment 3 from Crawford et al., 2014). The present research started with a conceptual replication of Crawford and colleagues' (2014) Experiment 2 which suggests that the vertical location of valenced words in the encoding phase can influence recognition memory performance, but did not conclusively support their argument.

The differences of 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 the recognition memory process: item discrimination per se, versus guessing components of various kinds that actually form an integral part of performance in recognition paradigms. The current analysis allows for a more accurate estimate of the pure item memory performance in a recognition situation as compared to the conventional ANOVA measures used in Crawford et al (2014).

Negativity advantage on item memory is on and off

The assumed negativity effect on item memory was not consistently manifested in experiments using valenced stimuli, including valenced words (Experiment 1 & 5a) and emojis (Experiment 3 & 5b). The hypothesis of negativity advantage was partially supported in Experiment 1 Group 1 and Experiment 5a. But in Experiment 3, positive emojis were generally memorised better compared to negative ones which is on the contrary to the hypothesis. But the contradictory results from Experiment 3 were not replicated in Experiment 5b which also used emojis but with Chinese participants, in which a marginal support for the negativity advantage was found.

In general, it seems to be easier to demonstrate a negativity advantage on item memory with word stimuli compared to with emojis. As discussed in Experiment 3, this might due to the somewhat comical feature of emojis, which may interfere with the impression of negativity and as such may hinder the deeper cognitive processing attributing to negativity. Another potential reason is the generally poorer item memory performance of emojis compared to of word materials. It may be harder to identify any differences between valence contexts if the general estimates were lower. Combined the two potential explanations, negativity advantage was more difficult to show with emojis as memory materials.

But even the evidence obtained with words stimuli cannot be considered substantial, which contradicted the hypothesized negativity bias proposed by previous research (Baumeister et al., 2001; Rozin & Royzman, 2001). A large body of research has provided empirical evidence to suggest that people direct more attention to negative information and consequently recognise and recall negative information better than positive information (see Unkelbach et al., 2020 for a review). But some researchers suggest that the negativity advantage on memory is attributed more to arousal rather than the negative meaning per se (Mather & Sutherland, 2009, 2011). According to this argument, it potentially explains why the negativity advantage on item memory was not substantially detected in Experiment 1, due to the control of equalised mean arousal scores for positive and negative words. At any rate, this explanation is only speculative and inconclusive due to the lack of an arousal measure for emoji stimuli (in Experiment 3 and 5b) and for the Mandarin word stimuli (in Experiment 5a). The present research is, to my knowledge, the first time to apply source monitoring paradigm to investigate item memory of stimuli with different valence, it is good to identify new issues for further investigation. Probably future research can take a closer look at this phenomenon and provide better insight.

Awareness condition influence metaphoric effects

The results across three awareness conditions demonstrate that the awareness of spatial location (Group 2) and of the valence- or connotation-verticality association (Group 3) facilitates the occurrence of congruent source guessing biases. As mentioned above when discussing metaphor-congruent source guessing biases, the metaphor builds up the expectation of valence-verticality congruency which guides the source guessing responses.

Along with the rise of awareness of the association, it is naturally expected that the congruent source guessing biases might be enlarged. Results suggest the enlarging effect not only for metaphor-congruent guessing biases, but also for the connotation-congruent guessing biases. There were indeed some exceptions to the observation of congruent guessing biases being enlarged, as a trend; nevertheless, the effects from awareness were observed in general.

It was also expected that the awareness of metaphoric association would enhance metaphoric effects on item memory, but due to the absence of support for the basic effect itself, this argument was not supported.

Limitations

Apparently, this series of experiments has a few limitations, the most obvious one is that many different kinds of materials were involved such as words, emojis and unfamiliar language characters, which is why a consistent difficulty level was hard to achieve across experiments. An example is the relatively high difficulty level of emoji memorising as mentioned above. All predicted effects were quantified by parameter estimates or their differences, and the estimates were dependent on the difficulty level of the task. As a result, the true effect sizes of investigated effects in different experiments using different stimuli may vary.

Another potential issue is that the overall lower estimates of item memory parameters *D* in Experiment 3 and 5b indicate that participants found it harder to memorise emojis compared to words. Potential reasons were discussed in Experiment 3 including the lower distinctiveness and the less descriptive nature of emojis. The differences between emojis with same valence can be undiagnostic to participants, and consequently may have caused difficulty for participants when trying to distinguish emojis from each other based on the minor differences in the testing phase. Moreover, 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, would not share this feature.

Another limitation was due to the unavoidable online data collection in Experiment 4 due to the COVID-19. Because of this, the optimal control over the experimental environment could not be achieved, which may be problematic. Participants had to use their own devices at home to take part, where the vertical screen sizes were limited compared to previous lab-based experiments, which may have hindered the manifestation of connotation-incongruency effects, if any. Fortunately, Experiment 5b conducted in China, which was a replication of Experiment 4, seemed to yield similar results in terms of the influence of vertical connotations in the non-valence, physical domain.

Conclusion of this chapter

To conclude, the “good is up” metaphoric effects on item memory were not demonstrated substantially in the series of five experiments, but source guessing biases towards metaphor-congruent locations were found substantial in general.

Chapter 3 Replication of Meier & Robinson (2004)

In the previous chapter, a series of experiments are reported that were conducted to investigate the “good is up” metaphoric effects on recognition memory processes. The results show that valence can lead to metaphor-congruent source guessing substantially, but does not differentiate item memory performance of stimuli from metaphor-congruent and metaphor-incongruent sources.

When discussing the unexpected null metaphoric effect on item memory, some previous research which also showed null effects of the “good is up” metaphor on memory was presented (e.g., McMullan, 2016). A potential explanation is considered that the metaphoric effects on cognition may not be as strong as first reported in the original research that identified the effects. In line with this concern, some doubts of Meier & Robinson’s (2004) studies are worth noticing. Meier & Robinson’s (2004) research was the first one to empirically propose that the “good is up” metaphor influences cognition processes such as affective judgements and attention shifting. Although there is plenty of research in support of metaphoric effects on cognition, as reported in the introduction chapter, there are also some failed replications of Meier & Robinson’s (2004) studies. McMullan (2016) conducted a close replication of Meier & Robinson’s (2004) Study 1 (having participants judge the valence of word stimuli presented at metaphor-congruent or -incongruent locations) but found no significant metaphoric effects (see Experiment 3 in McMullan, 2016). Huang & Tse (2015) replicated metaphoric effects from Meier & Robinson’s (2004) Study 1 in their first experiment, but only on positive words, not on negative ones. Then, in the subsequent 5 experiments when replicating Meier & Robinson’s (2004) Study 2 (having participants discriminate letters presented at metaphor-congruent or -incongruent locations after priming with valenced words) with an adapted paradigm, they consistently failed to find the metaphoric effects, namely, the valence prime did not direct participants’ attention to metaphor-congruent locations automatically, no matter when participants were required to verbally rehearse 6-digits simultaneously as additional working memory load or not, no matter valence judgement of the prime was by key-pressing or vocally, no matter the valence prime was a word or a picture.

Following this concern, the current chapter is going to present Experiment 6 and 7 as replications of Study 1 and 2 from Meier & Robinson (2004), respectively, to investigate metaphor-congruency effects on affective judgement as well as on attention directing. Besides replicating the original studies on metaphoric effects using positive and negative words as materials (Experiment 6a & 7a), affective emojis are used in two parallel studies with the same design (Experiment 6b & 7b). As a result, it turned out that affective emojis function in a similar pattern as affective words, but the expected metaphoric effects were not replicated in the current research. Potential reasons and implications are discussed at the end of this chapter.

Experiment 6a & 6b

Experiment 6a and 6b investigated whether affective judgements are facilitated when the valenced stimuli's present vertical locations are congruent with the "good is up" metaphor. Based on results from Meier & Robinson (2004), it was expected that positive stimuli are evaluated faster if presented at the top of the computer screen, and so are negative ones when presented at the bottom of the screen, compared to when presented at metaphor-incongruent locations. At the same time, the effect was expected to be demonstrated with both valenced words and emojis as materials. Namely, the inherent valence feature was expected to be the key point of triggering metaphoric effects, irrespective of valence being expressed by stimuli with more or less semantic information.

Methods

Design

Experiment 6 used a 2 valence (positive vs. negative) × 2 verticality (up vs. down) within-subjects design, which was identical to Study 1 from Meier & Robinson (2004). Participants were required to evaluate each word or emoji as having either a positive or negative meaning in each trial, and the vertical locations of the stimuli on the screen (up or down) were varied randomly so that half stimuli of each valence were presented at metaphor-congruent locations and the other half were presented at metaphor-incongruent locations.

Hypothesis: Metaphor-congruency facilitates affective evaluation. Namely, participants were expected to respond faster to positive stimuli when presented at the up location, and to negative stimuli when presented at the down location.

Materials

In Experiment 6a, same materials as in Experiment 1 in Chapter 2 were used. In total, 40 positive words and 40 negative words from the Affective Norms for English Words (ANEW, Bradley & Lang, 1999) were used, with the mean valence scores of positive words and negative words equal to 7.20 (*SD* = 0.37) and 2.86 (*SD* = 0.63) respectively on a 9-point rating scale. Arousal scores, length, syllable length and frequencies of positive and negative words did not differ significantly, details see Table 4 in Chapter 2.

In Experiment 6b, the same materials as in Experiment 3 in Chapter 2 were used. 72 pre-tested emojis from the Lisbon Emoji and Emoticon Database (LEED, Rodrigues et al., 2018) were used, half positive and the other half negative, with mean valence ratings equal to 3.73 (*SD* = 0.73) and -3.72 (*SD* = 0.72), respectively, on a scale ranging from -7 to +7 indicating extremely negative to extremely positive. The selected emojis were all made into pictures of 72×68 pixels in size, to achieve a consistent presentation.

Participants

As a replication, more participants were recruited in each sub-experiment compared to the original study (*N* = 34 in Meier & Robinson, 2004) and the previous replication (*N* = 57 in McMullan, 2016).

In Experiment 6a, *N* = 91 English native speakers were recruited from Cardiff University via an online university-wide Experimental Management System to take part in this study, of which 77 were females, 7 were males and 7 did not report their gender, mean age *M* = 19.30 years (*SD* = 1.53).

In Experiment 6b, *N* = 79 fluent English speakers were recruited from the same source as Experiment 6a, of which 63 were females, 11 were males and 5 did not report their gender, mean age *M* = 19.78 years (*SD* = 3.65).

In both experiments, participants took part in the study in exchange of 1 course credit or £2 cash.

Procedure

Experiment 6a and 6b were both conducted online via Pavlovia (https://pavlovia.org/), using an identical procedure as in Meier & Robinson's (2004) Study 1. Experimental programs were written in PsychoPy (Peirce, 2009; Peirce et al., 2019). In each trial, a fixation cue (+++) was presented at the centre of the screen for 300 ms at the beginning. Following the central cue, a subsequent fixation cue (+++) was presented for 300 ms at the location either above or below (determined randomly) the screen centre with a distance of 15% vertical height of the screen (exact distance was dependent on participants' devices). Then, a third fixation cue (+++) was presented for 300 ms at the location either above or below (in the same vertical direction as the last cue) the screen centre with a distance of 30% vertical height of the screen. The series of fixation cues was intended to direct participants' attention to where the stimulus was going to appear, thereby to reduce variances due to participants' random gaze exploration on screen. The testing stimulus, either word or emoji, was then presented at the location either above or below the screen centre (in the same vertical direction as the two previous cues) with a distance of 40% vertical height of the screen. Words were presented in white font colour, and emojis were presented as pictures, all on a black background and horizontally centred. Participants were required to classify each stimulus as positive or negative as fast and as accurately as possible. Responses were made by pressing "Q" or "P" keys, representing "positive" or "negative" judgement responses, counterbalanced across participants ("Q" = "positive" & "P" = "negative" vs. "P" = "positive" & "Q" = "negative"). If a response was inaccurate, the word "INCORRECT" appeared in a red font for 1500 ms before the next trial began. Trials with accurate responses were separated by a blank screen for 500 ms as an inter-trial interval (ITI).

Results

Before data analysis, outliers with low accuracy rate were detected and excluded in each experiment using the Tukey criterion (Tukey, 1977). One participant was excluded in Experiment 6a, and three participants were excluded in Experiment 6b.

Trials with inaccurate responses were excluded from the analysis in both experiments. 6613 out of the original 7200 observations (91.85%) remained in Experiment 6a, and 5198 out of the original 5472 observations (94.99%) remained in Experiment 6b. Subsequently, RT data 2.5*SD*s above or below the mean latency were replaced with the *M*±2.5*SD* value respectively (McMullan, 2016; Meier & Robinson, 2004). Latency data were then log-transformed to normalise the distribution (Ratcliff, 1993).

Instead of a two-way repeated measure analysis of variance (ANOVA) which was previously used by Meier & Robinson (2004) and McMullan (2016), a linear mixed model (LMM) approach was applied following the appeal to transit from ANOVA to mixed model analysis (Boisgontier & Cheval, 2016). ANOVA holds the assumption of the independence of observations which is not fulfilled in repeated measures where multiple observations are collected from each participant. Linear mixed models have been developed to take this dependence into consideration (Baayen et al., 2008), therefore, providing more accurate estimates of the effects and non-inflated Type I errors (Singmann & Kellen, 2019). Moreover, LMM approach shows the advantages including allowing unbalanced data to be used, avoiding information loss due to prior averaging over stimuli or participants, and the widespread availability and ease of use in different contexts with categorical or continuous predictors (Judd et al., 2012). Therefore, LMM can be especially useful in the current research to analyse the unbalanced dataset with different numbers of observations from each participant due to the exclusion of incorrect responses. An R package “afex” (Singmann et al., 2015) was used to construct LMMs in the analysis of this chapter, and a Satterthwaite method was used to fit the models.

Model construction procedure followed the theory proposed by Barr and colleagues (2013) and the online step-by-step instructions provided by Singmann & Kellen (2019) as supplement materials for their paper[7]. The analysis started from a maximal model with by-participant random slopes for valence, verticality, and their interactions. Because the items

---

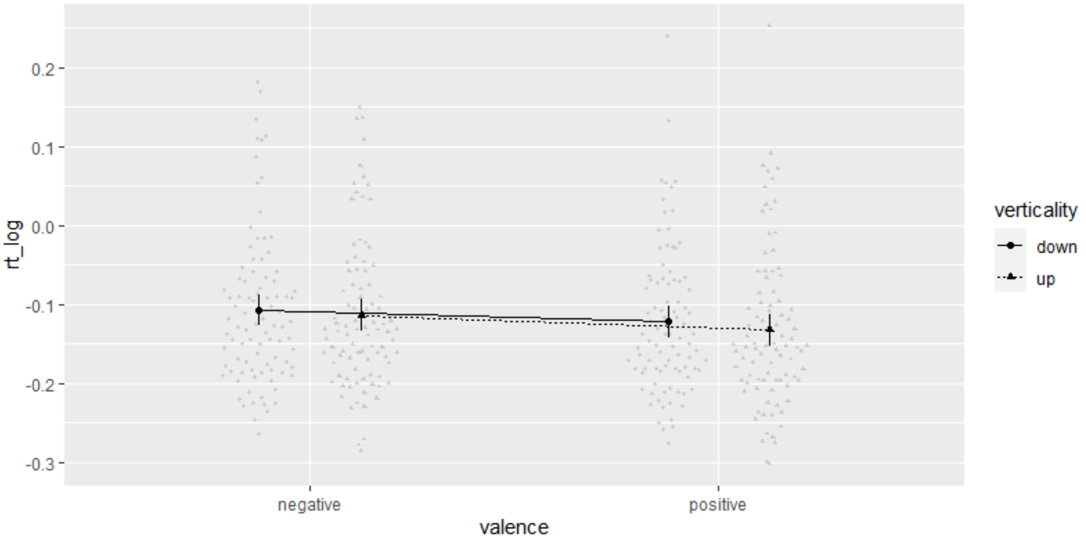
[7] https://cran.r-project.org/web/packages/afex/vignettes/afex_mixed_example.html

were randomly allocated to being presented at either up or down location for each participant, so there are no by-item random slopes in the maximal model. Then based on model-fit results, the model was modified step by step until optimal. Details of model construction procedure and according R codes are described in Appendix C.1 and C.2, along with information about the particular random-effect structure adopted for each model in each experiment of this chapter. Results are going to be reported based on the final constructed model. Effect sizes (Cohen's *dz*) are reported across the experiments for those effects that are interpreted as relevant to the main hypothesis.

In Experiment 6a, a main effect of valence was demonstrated, *F*(1, 89.67) = 19.44, *p* < .001, *dz* = .46, participants responded faster to categorise positive words (*M* = 803 ms, *SD* = 216 ms) than negative words (*M* = 831 ms, *SD* = 220 ms). A main effect of verticality was also demonstrated, *F*(1, 90.12) = 5.41, *p* = .022, *dz* = .24, participants responded faster to words presented at up location (*M* = 810 ms, *SD* = 218 ms) than to those presented at down location (*M* = 824 ms, *SD* = 219 ms). The expected interaction of valence and verticality was not detected, *F*(1, 6366.35) = 0.41, *p* = .524. Figure 19 illustrates the null interaction of valence by verticality using the log-transformed reaction time data.

Figure 19

Mean Log-transformed RT in Each Valence × Verticality Condition in Experiment 6a

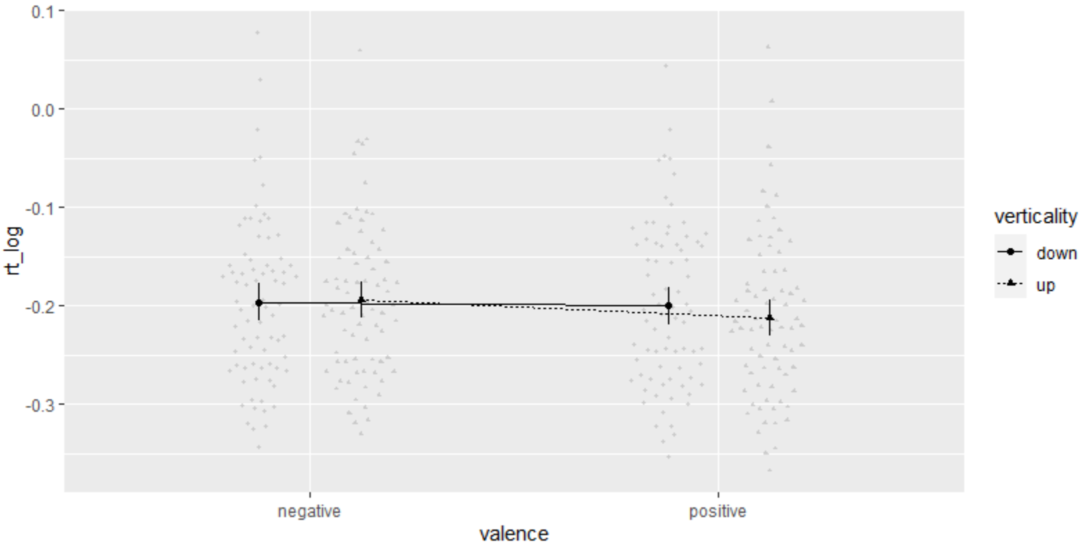


In Experiment 6b, a main effect of valence was also demonstrated, *F*(1, 75.11) = 9.45,

p = .003, dz = .34. Participants again responded faster to categorising positive emojis (M = 659 ms, SD = 153 ms) than negative emojis (M = 673 ms, SD = 150 ms). But non-significant main effect of verticality was demonstrated, $F(1, 74.27) = 1.96$, p = .165, dz = .16. The interaction of valence and verticality was marginally significant, $F(1, 72.03) = 3.83$, p = .054. Post hoc shows that participants responded faster to positive emojis when presented at up location compared to when at down location ($t(147) = 2.37$, p = .038, dz = .28; descriptive statistics of RT: M_{up_pos} = 652 ms, SD_{up_pos} = 153 ms, M_{down_pos} = 666 ms, SD_{down_pos} = 153 ms), whereas the contrast was not significant for negative emojis when presented at down location compared to when at up location ($t(146) = -0.358$, p = .721, dz = .04; descriptive statistics of RT: M_{down_neg} = 673 ms, SD_{down_neg} = 159 ms, M_{up_neg} = 673 ms, SD_{up_neg} = 141 ms). Figure 20 illustrates the marginal interaction of valence by verticality using the log-transformed reaction time data.

Figure 20

Mean Log-transformed RT in Each Valence × Verticality Condition in Experiment 6b



Discussion

Of the most interest, the interaction of valence and verticality was only marginally identified in Experiment 6b with affective emojis as materials, and not demonstrated in Experiment 6a with valenced words as materials for the affective judgement task. More specifically, the contrast of reaction time to metaphor-congruent and -incongruent present stimuli was only manifested in Experiment 6b with positive emojis, but not in other contexts. In general, the evidence found in the current experiment is relatively weak in terms of supporting the argument of metaphor-congruency's facilitation on affective judgement. The results are inconsistent with Meier & Robinson's (2004) Study 1, but more in line with McMullan's (2016) Experiment 3 which was a close replication of the former. McMullan (2016) discussed several possible explanations of the failed replication of the original study and appealed for more replications. As a step forward from a close replication, the current research used both affective words and emojis in two individual experiments. More influence factors affecting the materials were considered and controlled, such as word arousal norm, word length, syllable length and word frequency. Emoji materials were selected from pre-test results to assure a general consensus rating on valence of each emoji. Moreover, a LMM analysis approach was applied in the current experiment in order to get more accurate estimates of the effects. In general, an improvement of the methods was achieved, but consequently, the expected metaphor-congruency effect did not substantially show up.

A main effect of valence was demonstrated in both experiments (6a and 6b) where participants showed shorter reaction time to correctly evaluate positive stimuli compared to negative stimuli, which was in line with results from Meier & Robinson (2004) and McMullan (2016). Previous research suggests that people process positive information faster than negative information (Bargh et al., 1992, see Unkelbach et al., 2020 for a review). Positive valence facilitates judgements even when the task is irrelevant to valence, for example, when asked to respond whether a stimulus is a word or a non-word, participants identified positive words more frequently and faster (Unkelbach et al., 2010). Generally, current results supported previous findings showing that positive valence can facilitate processing speed of the information.

A main effect of verticality was demonstrated in Experiment 6a where participants responded faster to words presented at up location than to those at down location, regardless of valence. But the effect was not significant in Experiment 6b. This effect was also demonstrated in McMullan's (2016) Experiment 3 but not in Meier & Robinson's (2004) Study 1. Altogether, it seems that when there is a significant interaction of valence by verticality, the main effect of verticality is not significant, on the contrary, the main effect of verticality only appears when the interaction was not significant. Speculatively, this phenomenon makes sense because if metaphor-congruency effect is strong, participants are supposed to respond faster to negative stimuli when presented at down location compared to when at up location, thereby consequently hindering the manifestation of a main effect of verticality. On the contrary, the main effect of verticality can only show up when the expected metaphoric effect is not strong enough, if any.

Another interesting observation is that participants generally responded faster to emojis than words. This seems to support the expectation from Chapter 2 that emojis may be more direct in terms of their affective meanings compared to word stimuli, with less semantic information and less complexity, such that emojis may be processed faster than words within the same paradigm.

Experiment 7a & 7b

Experiment 7a and 7b were intended to investigate whether the mere act of evaluating emojis without a concurrent manipulation of verticality can direct spatial attention to metaphor-congruent locations or not, as a replication of Meier & Robinson's (2004) Study 2, using affective words and emojis as materials, respectively.

Methods

Design

Experiment 7 used a 2 valence (positive vs. negative) × 2 verticality (up vs. down) within-subjects design, which was identical to Study 2 from Meier & Robinson (2004). Participants were required to evaluate an affective prime presented in the centre of the screen at the beginning of each trial, then to indicate whether the following letter was P or Q, randomly presented at the top or bottom location.

Hypothesis: Metaphor-congruency facilitates letter discrimination. Affective primes were expected to direct spatial attention to metaphor-congruent locations and consequently to facilitate responses to letters presented at metaphor-congruent locations. Namely, participants were expected to faster discriminate letters presented at the top of the screen after seeing positive primes, and faster to letters presented at the bottom after seeing negative primes.

Materials

Same materials as in Experiment 6a & 6b were used in the current two experiments.

Participants

As a replication, more participants were recruited in each sub-experiment compared to the original study (*N* = 28 in Meier & Robinson, 2004).

In Experiment 7a, *N* = 83 English native speakers were recruited from the same online source as in Experiment 6, of which 76 were females, 4 were males and 3 did not report their gender, mean age *M* = 19.48 years (*SD* = 2.28).

In Experiment 7b, *N* = 76 fluent English speakers were recruited from the same source, of which 68 were females, 7 were males and 1 did not report their gender, mean age *M* = 19.96 years (*SD* = 2.17).

In both experiments, participants got paid by 1 course credit or £2 cash for their participation.

Procedure

Experiment 7a and 7b were both conducted online via the same platform as in Experiment 6. In each trial, a fixation cue (+++) was presented at the centre of the screen for 1000 ms at the beginning. Following the central cue, a randomly determined positive or negative priming stimulus, which was a word in Experiment 7a or an emoji in Experiment 7b, was presented at the centre of the screen for 1000 ms. Words were presented in white, and emojis were presented as pictures, all on a black background and horizontally centred. During prime presentation, participants were required to keep looking at the prime and to distinguish whether it was positive or negative, but no responses were required. Immediately afterwards, a target letter (P or Q) was presented randomly at the top or bottom of the screen and participants were asked to discriminate the letter as fast and as accurately as possible by pressing key "P" or "Q". If a response was inaccurate, the word "INCORRECT" appeared in a red font for 1500 ms before the next trial began. Trials with accurate responses were separated by a blank screen for 500 ms as an inter-trial interval (ITI).

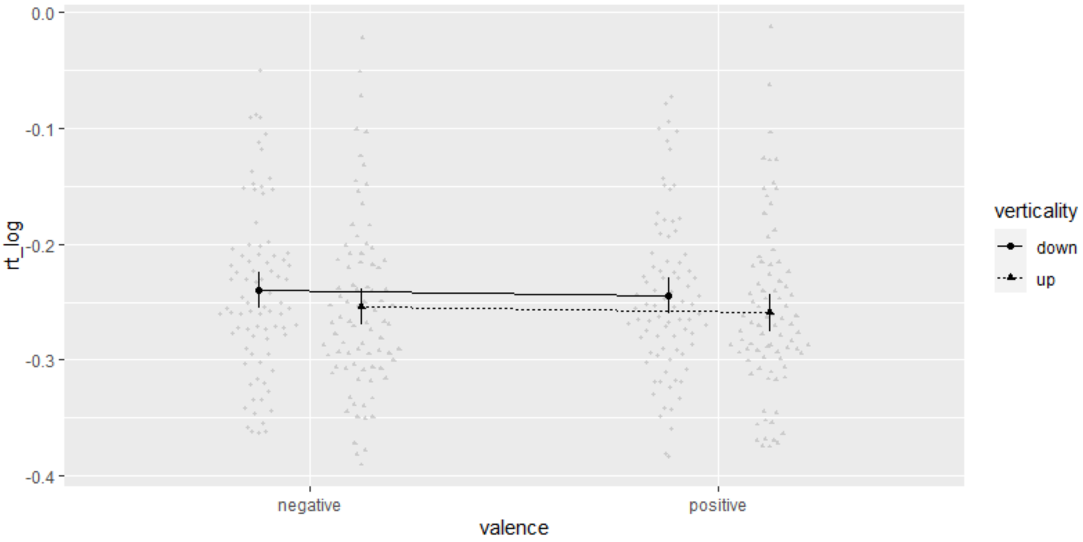
Results

The Tukey criterion [Q1 - 3*IQR, Q3 + 3*IQR] on accuracy rate as in Experiment 6a & 6b was used again in the current analysis to detect outlier participants. No outliers were found. Then, trials with inaccurate responses were excluded from the analysis in both experiments. 6268 out of the original 6640 observations (94.40%) remained in Experiment 7a, and 5202 out of the original 5472 observations (95.07%) remained in Experiment 7b. Subsequently, RT data 2.5*SD*s above or below the mean latency were replaced with the *M*±2.5*SD* value respectively. Latencies were then log-transformed to normalise the distribution (Ratcliff, 1993). LMM approach was applied for analysis, detailed model construction procedure see Appendix C.3 and C.4.

In Experiment 7a, a main effect of verticality was demonstrated, $F(1, 81.12) = 18.43$, $p < .001$, $dz = .46$, participants responded faster to distinguish letters presented at the top of the screen ($M = 580$ ms, $SD = 114$ ms) than to those presented at the bottom ($M = 598$ ms, $SD = 108$ ms), regardless of prime's valence. No main effect of valence was demonstrated, $F(1, 6105.13) = 2.89$, $p = .089$, $dz = .19$, which means participants showed similar speed of letter discrimination after positive primes ($M = 586$ ms, $SD = 109$ ms) and after negative primes ($M = 592$ ms, $SD = 113$ ms). Of the most interest, the expected interaction of valence and verticality was not detected, $F(1, 6106.02) = 0.01$, $p = .905$. Figure 21 illustrates the null interaction of valence by verticality on log-transformed reaction time.

Figure 21

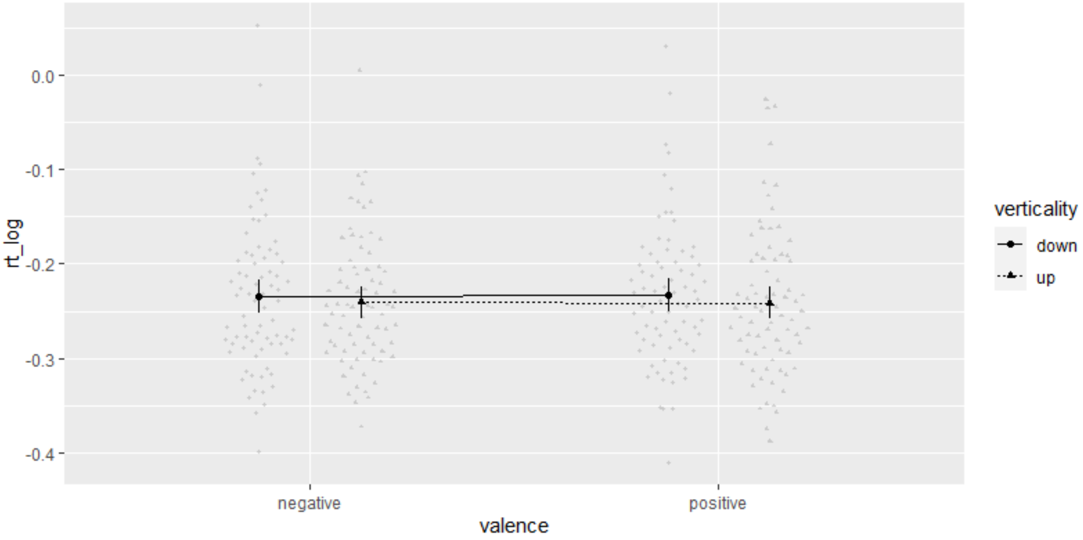
Mean Log-transformed RT in Each Valence × Verticality Condition in Experiment 7a



In Experiment 7b, a similar result pattern as in Experiment 7a appeared. A main effect of verticality was demonstrated, *F*(1, 73.62) = 4.62, *p* = .035, *dz* = .23, participants responded faster to distinguish letters presented at the top of the screen (*M* = 602 ms, *SD* = 117 ms) than to those presented at the bottom (*M* = 611 ms, *SD* = 124 ms). No main effect of valence was demonstrated, *F*(1, 5053.84) = 0.00, *p* = .977, *dz* = .00, participants showed similar reaction time to discriminate letters after positive primes (*M* = 607 ms, *SD* = 124 ms) and after negative primes (*M* = 606 ms, *SD* = 117 ms). The expected interaction of valence and verticality was again not detected, *F*(1, 5050.70) = 0.08, *p* = .777. Figure 22 illustrates the null interaction of valence by verticality on log-transformed reaction time.

Figure 22

Mean Log-transformed RT in Each Valence × Verticality Condition in Experiment 7b



Discussion

The expected interaction of valence and verticality was not demonstrated in either Experiment 7a or 7b, which means that irrespective of using affective words or emojis as primes, participants' attention did not shift to the metaphor-congruent spatial locations afterwards. The results were against Meier & Robinson's (2004) Study 2 which suggests that making an evaluative judgement can direct attention to the metaphor-congruent visual space.

One possible reason is the modification in the current experiment. Originally, Meier & Robinson (2004) required a vocal response from participants to judge the valence of the prime in order to trigger the presentation of target letter, only the trials with correct affective judgement responses to the primes were included in the analysis. But due to the technical limitation that came with online data collection, I did not manage to replicate this procedure. Instead, participants were instructed to discriminate positive or negative primes in their mind during a fixed duration of 1000 ms. For some trials, participants may have incorrectly judged the valence of the prime which could consequently have influenced the target letter discrimination afterwards. Moreover, the reaction time of judging the valence of some primes may have been shorter or longer than of others, or, some participants may generally have been faster or slower in completing the affective judgement task, than others. Consequently, the fixed duration of prime presentation in each trial for all participants may have hindered the attention shifting effect to be manifested, if any. Although these post hoc speculative reasons might have been prevented the metaphor-congruency effect from showing up, this would also indicate that the expected effect was subtle and easily affected or diminished.

Different to Meier & Robinson's (2004) Study 2, there was a consistent main effect of verticality demonstrated in the two experiments which showed that participants responded faster toward letters presented at up location than at down location. This effect was consistent with previous research which showed that participants tend to look up faster compared to look down in a visual search task (Previc & Blume, 1993), and attention is distributed with an upward bias unconsciously (Feng, 2011). Presumably, in the original Meier & Robinson (2004) study, the metaphor-congruency effect was strong enough to be detectable on the background of the general upward bias of attention, in which case the main effect of verticality was not detected. But in the current experiment, with the absence of metaphoric effect, the verticality effect was demonstrated.

Discussion of this chapter

Across two experiments each with two sub-experiments using affective words and emojis as materials, respectively, there was no substantial evidence to support the “good is up” metaphoric effects on affective judgement nor on attention directing.

In Experiment 6a & 6b, participants did not judge the valence faster when stimuli presented at metaphor-congruent locations as expected, with the only exception that when using positive emojis as materials, the up location facilitates the reaction speed. But to combine with the other non-significant results from the same experiment, the evidence cannot be considered substantial enough. Especially considering that the original Study 1 from Meier & Robinson (2004) only used valenced word stimuli as materials and found the metaphor-congruency facilitating effect on affective judgement, whereas Experiment 6a in the current research did not manage to replicate their finding. In Experiment 7a & 7b, after primed with affective stimuli, participants did not discriminate letters faster that were presented at metaphor-congruent locations than those at metaphor-incongruent locations. In general, I did not find substantial evidence to support the metaphoric effects reported by Meier & Robinson (2004). In contrast, the results were in line with the null effect reported by McMullan (2016) and Huang & Tse (2015).

Open Science Collaboration (2012) listed several possible reasons of a failed replication which can be considered. The first possible explanation is that the original effect is false, namely perhaps the original metaphoric effects identified by Meier & Robinson (2004) was attributed to a Type I error. However, the opposite could alternatively be true instead, namely that the failed replication in the current research was attributed to a Type II error. Importantly, however, no arbitrary conclusions can be made merely based on either one failed or one supportive replication.

The second putative explanation is that the actual size of the effect is smaller than originally reported, which makes it more difficult to detect in replications. A meta-analysis of intelligence research showed that effects declining over time systematically outnumber effects increasing at a ratio of about 2:1 (Pietschnig et al., 2019) and the researchers proposed that the phenomenon may not be confined to intelligence research only, but expected to generalize to psychological science and perhaps empirical sciences in general. Moreover, Pietschnig and colleagues (2019) also found that effect misestimations were more substantial when initial studies had smaller sample sizes and reported larger effects, which was the case of Meier & Robinson's (2004) two studies. The two studies were both with large effect sizes (Study 1: $\eta_p^2 = .16$; Study 2: $\eta_p^2 = .23$; which were not reported in the original paper but calculated from the reported ANOVA results) and relatively small sample sizes, $N = 34$ and $N = 28$ respectively. Whereas the effect size was smaller in McMullan's (2016) replication of Study 1, $\eta_p^2 = .01$, with a larger sample size, $N = 57$.

The third possibility is that the method or analysis of either the original or replication study is flawed. McMullan (2016) mentioned a detail in his Discussion that word frequency was not controlled in Meier & Robinson's (2004) research, which can be problematic.

Potential influence factors such as word frequency, word length and syllable length were all considered when selecting materials in the current research. For emoji materials, the subjective valence ratings were pre-tested to assure a good control. Moreover, LMM as a new analysis approach was applied to provide more accurate estimates of effects compared to ANOVA. And a post hoc test was conducted after the marginal interaction identified in Experiment 6b to show that the expected effect was on positive stimuli only, which was missing in Meier & Robinson’s (2004) paper after reporting the interaction in both studies. In general, a “purer” effect was expected to be demonstrated in the current results, with less noise.

The last reason to consider is the differences of methodology between the original research and the replication. As mentioned above, due to the feature of online data collection, the environment of experiments could not have been as much controlled as is deemed desirable for lab-based experiments. Besides, the procedure of Experiment 7 did not to 100% replicate Meier & Robinson's (2004) Study 2, out of the same reason. The flaws of online data collection may contribute critically to this failed replication.

To sum up, the experiments from this chapter failed to replicate original findings on metaphor-congruency effects from Meier & Robinson (2004), but they echoed non-significant findings from previous replications, together provided more evidence to doubt, at least the effect robustness of the original studies. But at the same time, the limitations of the current research cannot be ignored which may contribute to the non-significant results significantly. Better controlled and implemented replications are appealed in the future to further investigate the metaphoric effects.

Chapter 4 General Discussion

In this chapter, I am going to integrate and discuss the findings obtained in all 7 empirical experiments of this research. Starting by outlining the main aims and hypotheses of the current research, results from each experiment are going to be summarised and discussed subsequently. Following that, I will sequentially focus on the discussion of the metaphor-congruent source guessing biases, the predicted, but not materialised, metaphoric effects on item memory, potential reasons and implications of the failed replications in the current research, a comparison between word and emoji materials between various experiments, and a comparison between positive and negative valence. Then the limitations of the current research are going to be discussed along with the appeal for future investigations with a better control. Finally, the conclusion section summarises this research and suggests that the "good is up" metaphor can affect cognition but in a more subtle way than expected.

Current project

The main research question I investigated in this thesis was the metaphoric association of valence and verticality on recognition processes. The source monitoring paradigm was introduced which means that in the testing phase, participants were required to recognise whether a presented item was learned or not from an earlier learning phase, and if an item was classified by the participants as learned, they were subsequently required to indicate whether the item was presented at up or down location during learning. The source monitoring paradigm was introduced for two reasons. One is to provide more accurate estimates of item memory performance that excludes the probability that participants give correct answers to items as old or new just by guessing, the other is to analyse metaphoric effects on source guessing biases at the same time along with item memory performance. According to the Attention Elaboration Hypothesis (AEH) and previous findings from Crawford and colleagues (2014), it was expected that when presented at metaphor-incongruent locations, a valenced stimulus may violate the expectation of metaphor-congruency (positive-up & negative-down) and consequently receives more attention and more conceptual elaboration. Such a stimulus would therefore be processed more deeply, and better memorised than when presented at metaphor-congruent locations. Meanwhile, metaphor-congruent source guessing biases were expected based on previous research which used the source monitoring paradigm and found expectation-congruent source guessing biases with other materials (Bayen et al., 2000; Ehrenberg & Klauer, 2005; Kuhlmann et al., 2016; Wulff & Kuhlmann, 2020).

A series of 5 experiments were conducted starting from a conceptual replication of the Crawford and colleagues' (2014) study which manipulated valence and vertical present location in the encoding phase. Apart from the hypothesized metaphoric effects, a negativity advantage on item memory was expected according to previous research (Baumeister et al., 2001; Rozin & Royzman, 2001), which means negative stimuli were expected to be memorised better than positive stimuli. Additionally, awareness of spatial location and the valence-verticality metaphoric association was manipulated by providing different instructions for the learning phase. This manipulation required participants to memorise stimuli only (Group 1: stimulus-only), or to memorise stimuli with the presented location (Group 2: stimulus-location), or to memorise the stimuli with the presented location after being made aware of the valence-verticality metaphoric association (Group 3: stimulus-location-metaphor), respectively. It was expected that the raised awareness can increase the expectation of metaphoric congruency and consequently increase the memory of items that violate the expectation (stimuli presented at metaphor-incongruent locations), as well as increase the metaphor-congruent source guessing biases. Experiment 1 used valenced words as materials, same as the Crawford and colleagues' (2014) study. As a result, regarding the metaphoric effects on item memory, only a marginal metaphor-incongruency effect on item memory was demonstrated in Group 1 with negative words and not demonstrated in any other conditions. The expected negativity advantage on item memory was only demonstrated in Group 1 with words presented at up location and not demonstrated in any other conditions. However, across the three groups, a generally consistent demonstration on metaphor-congruent source guessing biases was found. The biases showed a trend to increase for positive stimuli along with the rise of awareness, however, to decrease for negative stimuli. In other words, the predicted effects of awareness on source guessing biases were partially demonstrated.

Experiment 2 used characters from an unfamiliar language (subjectively rated as valence-neutral) as materials for the memory task (instead of valenced words in Experiment 1) as a comparison to demonstrate the essential role of valence to elicit biased source guessing and the (expected) memory performance differences. As predicted, when asked to recognise neutral stimuli learned from either up or down location in the source monitoring paradigm, participants showed neither different item memory performance nor biased source guessing due to the verticality variable.

Experiment 3 was again a parallel study to Experiment 1 which used positive and negative emojis to replace valenced words as materials to investigate whether the hypothesized metaphoric effects can be demonstrated on item memory and source guessing biases with materials with less semantic information but more direct on affective valence. As a result, no evidence was obtained to support metaphoric effects on item memory. But similar to Experiment 1, metaphor-congruent source guessing biases were substantially manifested, and an increasing trend of the bias was identified along with the rise of awareness of the valence-verticality association. The expected negativity advantage on item memory was not found in any conditions.

Experiment 4 was designed based on the embodied cognition theory which suggests that metaphors function as a scaffold to let people use concrete concepts to process abstract concepts cognitively. In line with this argument, Experiment 4 created an entirely physically grounded context which used words with high or low connotations as materials. Connotation-verticality incongruency effects on item memory and congruent source guessing biases were expected. As a result, when memorising words with high or low vertical connotation, participants did not show any differences on item memory due to the congruency or incongruency of connotation and vertical present location. However, connotation-congruent source guessing biases were generally substantial, and the biases were stronger in Group 2 than in Group 1 along with the rise of awareness of spatial information, but participants in Group 3 did not show the strongest biases when being made aware of the vertical connotations of the word stimuli.

On the grounds of the consistent failure of identifying metaphoric effects on item memory, it was hypothesized to be null in Experiment 5. Experiment 5 conducted three sub-experiments as replications of Experiment 1, 3 and 4 in a Mandarin speaking context. Due to the previous evidence obtained in different language contexts other than English (Dudschig et al., 2015; Luodonpää-Manni & Viimaranta, 2010), especially in Mandarin context (Wu et al., 2019), the “good is up” metaphor was expected to show similar effects on cognitive processes in Mandarin context as in English context. Based on results from the previous 4 experiments, a null metaphoric effect on item memory as well as metaphor-congruent source guessing biases were expected in Experiment 5. Without manipulating awareness of metaphoric information anymore, Experiment 5 only used a 2 (valence/connotation) × 2 verticality within-subjects design, with valenced words, emojis, and vertical connotation words as materials, respectively. Altogether, a null effect on item memory was demonstrated, as well as substantial congruent source guessing biases.

Summarising the 5 experiments in Chapter 2, the hypothesized metaphoric effects on recognition processes were only demonstrated on source guessing biases, but not on item memory performance. Metaphor-congruent source guessing biases mean that when participants cannot recall where a stimulus was presented, they tend to guess in a way that is congruent to the “good is up” metaphor, which is to guess that positive stimuli were learned more from the up location and negative ones from the down location, even in situations where the metaphor was not explicitly mentioned, and irrespective of valenced words or emojis being used as materials. On the contrary, item memory of metaphor-congruent or incongruent presented valenced stimuli did not substantially differ from each other, with the only evidence marginally manifested in Group 1 from Experiment 1. This is against the hypothesis which predicts that the stimuli violating the metaphor-congruency would attract more attention and generate more elaboration in the encoding phase, and consequently be memorised better. Potentially, this null effect indicates that the valence-verticality association based on the “good is up” metaphor is not strong enough, in which case the metaphor-incongruently presented stimuli cannot elicit as much violation as predicted, and consequently cannot significantly attract as much surplus attention to cause better memory performance as compared to congruently presented stimuli. This pattern of results was not only identified with valence-verticality context, but also in a non-metaphor but pure physically grounded context with concrete words with high or low vertical connotation. This indicates that not only the metaphoric association is not strong enough to cause significantly more elaboration when violated, a more obvious and direct physical association is not strong enough either. But the null effect on item memory alone cannot deny the existence of the association between valence and its spatial implications as metaphor. The consistent results of metaphor-congruent and connotation-congruent source guessing biases have provided evidence for the “good is up” metaphor, as well as for embodied cognition as such. People automatically map positive and negative valence onto up and down locations in a similar way as people also map high and low connotations onto the vertical locations. Altogether, this chapter indicates that the metaphor can influence cognition, but maybe in a more subtle way than expected.

Following the concern of the strength of metaphoric effects on cognition, Chapter 3 conducted two experiments to replicate Meier & Robinson’s (2004) research which was the first to empirically demonstrate that the “good is up” metaphor significantly influences cognitive processes such that metaphor-congruency facilitates affective judgements and attention directing. Meier & Robinson’s (2004) research is seen as a cornerstone of the research on the “good is up” metaphor, however, doubts have been raised in recent years due to failed replications (Huang & Tse, 2015; McMullan, 2016). Experiment 6 and 7 in the current project replicated Study 1 and 2 from Meier & Robinson (2004), respectively. Apart from using valenced words as materials in the replication experiments (Experiment 6a & 7a), positive and negative emojis were used in two parallel experiments (Experiment 6b & 7b) and were expected to show the same metaphoric effects. As a result, when required to judge the valence of a stimulus presented at either up or down location (Experiment 6a & 6b), participants did not substantially respond faster when it was presented at metaphor-congruent locations as predicted, with the only exception that the up location facilitated the reaction time to positive emojis. Overall, the evidence cannot be considered as substantial. In Experiment 7a & 7b, when required to discriminate letter P or Q presented at either up or down location after being primed with a positive or negative stimulus, participants did not respond faster to letters presented at metaphor-congruent locations than those at metaphor-incongruent locations. In general, the current research does not support the metaphoric effects on cognitive processes reported by Meier & Robinson (2004), but is in line with the failed replications reported by McMullan (2016) and Huang & Tse (2015). The parallel experiments which used affective words and emojis respectively have demonstrated a similar result pattern and suggests a consistency of the findings.

To sum up, the seven experiments considered together provide new insights and a more differentiated assessment of the involvement of the “good is up” metaphor in cognitive processing. Consistently manifested source guessing biases suggest the existence of the “good is up” metaphor and its effects on the heuristic part of recognition judgments. But the lack of evidence, in the present set of studies, in support of metaphoric effects on affective judgement, attention directing and item memory altogether suggest that the metaphoric effects may be limited.

“Good is up” is like a schema, but with weak strength

The current research shows that in a source monitoring task, when unable to recall where a valenced stimulus was presented in the learning phase, participants tend to guess in a direction congruent to the “good is up” metaphor, which means to guess more “up” as the source location when presented with a positive stimulus and more “down” location when presented with a negative stimulus. This result is consistent with previous source monitoring research on schema-relevant topics which suggests that the prior knowledge of a schema that implies the congruency between a particular item and one particular source can bias source guessing toward a schema-congruent direction, with various contexts such as profession schemas (Bayen et al., 2000), social schemas (Ehrenberg & Klauer, 2005), age stereotypes (Kuhlmann et al., 2016; Wulff & Kuhlmann, 2020), and objects in rooms schemas (i.e., typical objects from a bedroom or bathroom, see Bayen et al., 2000), etc.

The evidence obtained in the current research suggests that the “good is up” metaphor exists in our prior knowledge, and people rely on this piece of information to generate a source guessing response when unable to recall where a valenced stimulus had been presented, vertically.

It seems however that the metaphor does not have an effect as strong as other types of schematic knowledge have on cognition. In the abovementioned previous research on schema-congruent source guessing, no explicit instructions on the schemas were in need to trigger the effect on source guessing biases. For example, when investigating the source guessing biases with a profession schema context, Bayen and colleagues (2000) instructed participants to memorise some statements from two virtual characters, “Tom” and “Jim”, without mentioning that some statements were lawyer-related or doctor-related (mixed with other schema-irrelevant statements in a random sequence). It was only briefly mentioned that “Tom” was a doctor and “Jim” was a lawyer before the testing phase (see Experiment 2 in Bayen et al., 2000). Under such circumstances, a significant schema-congruent source guessing bias was demonstrated. As a comparison, in the stimulus-only condition (Group 1) in Experiment 1 and 3 where the metaphoric association was not explicitly mentioned, the metaphor-congruent source guessing was sometimes marginal (e.g., with positive words in Experiment 1) or even absent (e.g., with negative emojis in Experiment 3). Although along with the rising awareness of spatial location and the metaphoric association (a comparison across the three between-subjects groups), there is an increasing trend of metaphor-congruent guessing biases, results still show that the metaphoric effects on source guessing biases are weaker compared to the guessing biases based on schematic information in the abovementioned studies on schemas, and an explicit instruction to direct the attention to spatial information or even the metaphoric association helps to enlarge the effect on source guessing biases.

Another comparison was between the metaphor-congruent and connotation-congruent guessing biases when using valenced stimuli and words with vertical connotations as materials, respectively. In the same paradigm and same language context (Experiment 1 & 3 vs. 4; Experiment 5a & 5b vs. 5c), connotation-congruent source guessing biases were more substantially demonstrated than metaphor-congruent ones. This means that the expectation of congruency between stimuli's semantic vertical connotation and the present location is stronger than the expectation of congruency between positive/negative valence and the present up/down location, which is understandable as the physical connotation is more concrete and direct. From an embodied cognition perspective, when thinking about the word "sky", people 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 more mental processes to be mapped onto vertical locations. In other words, the distance in association between valence and verticality is further than the distance between high/low connotation and verticality conceptually.

To sum up, the valence-verticality metaphoric association plays an effective role in cognition, more specifically, it provides information for people to rely on to generate source guessing responses based on an expectation of metaphor-congruency. The effect is relatively weak compared to schematic information or connotation-verticality association.

Heuristics based on the "good is up" metaphor

In a wider framework, the metaphor-congruent source guessing biases may be discussed under the aspect of heuristics.

Heuristics refer to "strategies for executing the processing of operations involved in complex problem solving" (Mumford & Leritz, 2005, p. 203), which ignore part of the information with the aim to make decisions more quickly and frugally with reduced efforts than more complex methods (Gigerenzer & Gaissmaier, 2011; Shah & Oppenheimer, 2008). In the experiments reported in Chapter 2, when required to respond the source of a particular item and unable to recall the piece of spatial information, participant may turn to heuristics for help to generate a guessing response. As a result, the heuristics may have led participants to guessing more of a metaphor-congruent source rather than a metaphor-incongruent one. The biases suggest that the "good is up" metaphor can be seen as effective for the deployment of heuristics.

Fischer (2014) suggests that *metaphor heuristics* build upon linguistically realised conceptual metaphors which bridges two domains. Decisions on one domain can be inferred by the known facts about the other domain through employing mappings constitutive of the metaphors, both consciously or unconsciously. When relying on the metaphor heuristics based on the “good is up” metaphor, the mappings between positive valence and up vertical location were employed, same as between negative valence and down location, in which case participants tended to guess more of metaphor-congruent source locations than metaphor-incongruent ones. As suggested by the classical accuracy-effort trade-off view, the decisions based on heuristics may imply greater errors than those based on more “rational” considerations with logic or statistical models incorporating more information (Payne et al., 1993). In the current source monitoring task context, although the probabilities of a valenced stimulus that has been presented at up or down locations are both 50%, participants guessed the source in a metaphor-congruent way with higher probability. Tversky & Kahneman (1974) suggest that biases in decision making reveal heuristics of thinking under uncertainty. In this vein, the demonstrated metaphor-congruent source guessing biases stem from the reliance on the metaphor heuristics reflect the valence-verticality association in mind.

Furthermore, Tversky & Kahneman (1974) also suggest that when the input information is highly redundant or correlated, highly consistent patterns are more often observed based on the heuristics. In line with this argument, the rising awareness of the spatial information and metaphoric association in the current research may have potentially emphasized the metaphor heuristics and resulted in an increase of the metaphor-congruent source guessing biases.

Heuristics are undeniably useful as they enable people to make decisions according to simple rules of thumb instead of engaging in more complex information processing at the cost of a large amount of efforts (Shah & Oppenheimer, 2008), especially because uncertainty cannot be entirely avoided in daily life when making decisions. In this vein, the human mind resembles an adaptive toolbox with various heuristics tailored for specific kinds of decision-making problems, similar to how a handyman’s toolbox contains multiple hammers and screwdrivers (Gigerenzer, 2008). And the “good is up” metaphor helps to form one of the tools/heuristics.

“Somewhat expected” vs. “highly expected”

As discussed above, it seems that the metaphoric association of valence and verticality has an impact on cognition but to a relatively subtle degree, which provides another possible influence factor to explain the null effect of metaphor on item memory, which is the extent of expectation violation.

It has long been noticed that information violating expectations can be memorised better than the information congruent to expectations (Hastie & Kumar, 1979; Mäntylä & Bäckman, 1992; Rojahn & Pettigrew, 1992; Sakamoto & Love, 2004; Vakil et al., 2003), potentially due to that when exposed to unexpected information, people generally pay more attention due to the processing dysfluency that results from expectation disconfirmation (Roese & Sherman, 2007). In line with previous research and the AEH theory, a metaphor-incongruent effect on item memory was expected such that valenced stimuli presented at metaphor-incongruent vertical locations in the encoding phase would be memorised better than those presented at metaphor-congruent locations, and some empirical evidence was found to support this hypothesis (Crawford et al., 2014).

However, it is relatively less noticed and empirically tested that the strength of expectation plays a potential moderating role in influencing the memory for expected or unexpected information (Stangor & McMillan, 1992). Previous research shows that when the expectations were enlarged in a more overt way (e.g., by explicit schematic descriptions of personal traits and behaviours, such as a friendly person acts in a generally good-hearted manner), better recall performance of expectation-incongruent items was demonstrated, whereas better memory of expectation-congruent items was demonstrated when the expectations were induced in a less overt way (i.e., without the explicit emphasis, see Heider et al., 2007). In research investigating source memory, it was also suggested that the inconsistency effect predicted by the AEH only occurs when the expectation strength is high (Küppers & Bayen, 2014). This argument potentially explains the null metaphoric effects on item memory as observed here. 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 performance between metaphor-congruent and incongruent stimuli. This argument immediately elicits the following question of why the strength of expectation is insufficient. Two possible explanations are that, the materials do not sufficiently express valence, or the verticality dimension is not sufficiently activated when valenced stimuli presented in the encoding phase, or even both. The current research cannot provide confirmative evidence for either possible reason, this question might be an interesting topic for further investigations.

Furthermore, this speculation of weak strength of the metaphoric association may also apply to the null metaphoric effect on affective judgement and attention directing demonstrated in the experiments reported in Chapter 3 where I failed to replicate the findings from Meier & Robinson (2004). Long since the Stroop effect (Stroop, 1935) was identified, shorter reaction time has been seen as a manifestation of congruency. In a typical Stroop task, when participants were required to report the font colour of the presented words, the words presented in the colours incongruent to its semantic meaning were responded with longer latencies compared to those congruent ones. A number of studies on metaphors used reaction time differences to indicate metaphoric associations (e.g., “power is up”, Schubert, 2005; “divine is up”, Meier et al., 2007), including the research conducted by Meier & Robinson (2004) on the “good is up” metaphor. Similar to the Stroop pattern, even when the location of presentation is irrelevant to the task of judging the valence of stimuli or detecting letters presented at either vertical location, it is still expected to influence the response time via metaphor-congruency or incongruency, implicitly. However, this hypothesis was not supported in the current research, which may be attributed to a relatively weak strength of the expectation of valence-verticality congruency. As the expectation strength can be influenced or regulated by contexts, this discussion is not suggesting that the metaphoric association cannot prompt the expected effects at all. More potential reasons of the failed replications are going to be discussed in the next section.

Obviously, the explanation on the grounds of insufficiently activated expectation is purely speculative without empirical support. The reaction time method can only identify whether the congruency is effective in the current experiment or not, but we are not able to measure the extent of expectation strength. Empirically, it is hard to define or measure what is a strong enough expectation to trigger the incongruency effect on memory performance or to induce a congruency effect on affective judgement and attention directing. When investigating the incongruent effect on memory, Küppers & Bayen (2014) just distinguished the different levels of expectation strength by a vague and arbitrary distinction between "very expected" versus "somewhat expected". The distinction lacks a clear criterion, which is especially needed in more specific contexts. For example, what degree of expectation strength would one need in order to demonstrate effects on item memory, or on affective judgments, or on the direction of attention. Between all these cases, the critical levels of expectation strength might be different.

In general, from the results of this research, it is proposed that probably the "good is up" metaphoric association only creates a "somewhat expected" congruency expectation rather than a highly expected one between positive valence and up location, as well as between negative valence and down location.

Failure of replications is still meaningful

Across the seven experiments, both the metaphor-incongruency effect on item memory and the metaphor-congruency effect on affective judgement and attention directing were not successfully replicated. Quite a few possible factors may have contributed to the results and some of them have been discussed in Chapter 3, including the probably smaller true effect sizes than originally reported and the differences of analysis and methodology between the replications and the original studies.

Replications with null effects are still meaningful. Replication plays a fundamental role in science to verify empirical findings (Francis, 2012; Schmidt, 2016). Replications aim to demonstrate whether or not the results can be reproduced when separated from the specific circumstances, such as time, place or people, under which the findings were originally identified (Schmidt, 2016). Null effects in empirical research are unavoidable because statistics in experimental psychology indicates that the findings come with probabilities, rather than a definite yes or no for a hypothesized effect. Replications with confirming or disconfirming evidence help to carefully evaluate the credibility of the original empirical findings, and this procedure is essential for a self-correction of not only psychological science, but all science fields in a broader sense (Ioannidis, 2012).

A publication bias favouring statistically significant results has long been noticed (Begg, 1994), especially in the social sciences (Franco et al., 2014). Fanelli (2010b) analysed 2434 papers published across various disciplines and found that Psychology and Psychiatry showed the highest odds, 91.5%, of reporting significant results, whereas the average odds across all disciplines is 84% and the lowest is 70.2% in Space Science. A growing competitive environment and the ‘‘publish or perish’’ culture in academia may have increased this bias among scientists (Fanelli, 2010a), to make them not only to focus more on research yielding significant results, but also to emphasize the significant part of the findings when the hypothesis is partially supported. For example, Crawford and collegues (2014), who originally reported the metaphor-incongruency effect on memory, claimed that “in both free recall and recognition tasks, we find a memory advantage for words that had been studied in metaphor incongruent locations (positive down, negative up)” (p. 1) in the abstract, but their statistics showed only a marginally significant contrast of memory of negative words when presented at up or down location in the two experiments using free recall and recognition task respectively (*p* = .056 and *p* = .061), whereas a non-significant contrast was found for positive words in either experiment. In this situation, their abstract might be a bit misleading to present a conclusion of a metaphor-incongruent advantage on memory. It is understandable that when summarising the findings, some nuance might be omitted, but as a result, it gives a more arbitrary statement which only emphasizes the significant results. I am not focusing on criticising any specific researchers or assuming that some researchers purposely exaggerated their findings. What I suggest is that the publication bias may encourage researchers to produce “a good story” with more confirmative evidence for the hypotheses, rather than “a true story” that contains more detailed descriptions of the non-significant results.

Notably, there is a growing trend to accept and support more non-significant results and failed replications. Along with the rising awareness of the crisis of confidence in psychological science (Pashler & Wagenmakers, 2012), a bunch of researchers have started appealing for more well-structured replications, no matter with successful or failed results, to go public online (see Carpenter, 2012; Open Science Collaboration, 2012). A systematically better research practice was appealed by the *open science movement*. Munafò and colleagues (2017) define open science as “the process of making the content and process of producing evidence and claims transparent and accessible to others” (p. 5). In line with this appeal, a growing interest in transparency, accessibility, reproducibility of scientific research has led the academia to become more interested in the *pre-registration* of research. In pre-registration, researchers describe the hypotheses, methods, and analyses before the planned research is conducted, so that it can be externally verified (van ’t Veer & Giner-Sorolla, 2016). Analytically, instead of the traditional null-hypothesis significance testing, various approaches have been recommended, such as estimation based on effect sizes, confidence intervals and meta-analysis (Cumming, 2014) as well as Bayesian methods of data analysis (Francis, 2012). It is expected that the wide adoption of these measures is able to optimise the scientific process and to reduce the publication bias, in order to yield more reliable scientific findings.

In general, although the current research failed to replicate the original findings of metaphoric effects on item memory, affective judgement and attention directing, it is still valuable to the development of research on the metaphor topic. The current research has adopted a relatively better research practice compared to the original ones, for example, it provided better control of the materials such that considering the potential influence of word frequency and syllable length. Furthermore, more recent developed analyses approaches were applied in the current research and the results were reported in a more recommended way. Specifically, a Bayesian hierarchical MPT modelling approach was used in Chapter 2 and the parameter estimates were reported with a 95% BCI range; then in Chapter 3, a novel approach, linear mixed model, was applied, and effect sizes were reported as well.

Together, I believe this research has provided its implications for further research on the “good is up” metaphor, especially concerning the strength of the metaphoric effects on cognitive processes. Of course, this argument is not suggesting this research is flawless. There are some caveats which may have potentially contributed to the failure of the replications and will be discussed later in the Limitations section.

Words vs. emojis

Previous research widely used positive and negative words as valenced stimuli to investigate the “good is up” metaphor and its effects (Meier & Robinson, 2004; see also Huang & Tse, 2015; Schaverien, 2013). Verbal words are the most common way to express affective meanings, but along with the rise of social media and virtual communication online, emojis have recently become another popular approach to communication. Adobe’s 2021 Global Emoji Trend Report (Hunt, 2021) suggests that emojis help people express their emotions in digital communication in a way of better integrating tone of voice, gestures and emotional reactions with imagery, as compared to with words alone. Emojis show emotional expressions in a more direct way which help people from different cultures overcome language barriers when communicating (Hunt, 2021) and a semantic evaluation research shows a general consistency of emoji meanings across different language contexts (Barbieri et al., 2016). In the current research, emojis were introduced as affective materials with positive and negative valence, as a comparison to the traditional word stimuli. It was of interest whether the hypothesized metaphoric effects would appear in the same way with the two kinds of different stimuli.

Across these experiments, the results pattern of metaphoric effects when using valenced word stimuli or emojis did not differ significantly from each other, which indicates that it is the inherent valence feature which matters for the metaphoric effects rather than through which way it was manifested. At least within the range of visual stimuli, it does not matter whether verbal stimuli like words or imagery stimuli like emojis are used to represent valence in terms of the “good is up” metaphor. However, it is possible that this consistent pattern may have also appeared due to the fact that the predicted effects were reliably demonstrated only on source guessing biases and not substantially on item memory or affective judgement and attention directing. In other words, since there were no significant metaphoric effects identified in most of the tested cognitive processes, it might be hard for the differences between word and emoji stimuli on metaphoric effects to appear, if any should exist.

Word and emoji stimuli still showed some interesting contrasts in other aspects, such as item memory performance and affective judgement speed. Poorer item memory performance of emojis compared to words was observed consistently in Chapter 2 with both British and Chinese participants, which was potentially due to some shared features across emojis with the same valence. For example, many positive emojis shared the same features of a smiling mouth and smiling eyes, with the subtle differences between each other to be whether they had a blush or not, whether the mouth or eyes were open or not, etc. However, the positive words do not have this high similarity between each other. As a result, the similarity differences may have contributed to the item memory performance differences of the two kinds of stimuli. Regarding the differences of reaction time of affective judgement to words and to emojis in Experiment 6, a potential explanation has been discussed which is that emojis may contain less complexity while being more direct on affective meanings than words. Frequent daily use of emojis has made participants more familiar with the emoji stimuli (Jones et al., 2020), which may have contributed to the faster reaction time to emojis as well.

Altogether, although there are differences between the effects showed with words and emojis, they have generally yielded a consistent pattern of results in the investigation on metaphoric effects, which in total provide more reliable evidence for the findings of the current research.

Positive vs. negative

The asymmetry of positive and negative valence has been widely noticed that people pay more attention to negative information, memorise it better and weigh it more heavily compared to positive information, whereas positive information is processed faster and shows stronger congruency effects in the evaluation priming paradigm (see Unkelbach et al., 2020 for a review). The current research partially supports the argument, because positive stimuli were found to be processed faster than negative ones in Experiment 6, but the expected negativity advantage in item memory was not substantially demonstrated in Chapter 2, with the only confirmative evidence obtained from Experiment 1 and 5a.

The shorter processing time for positive stimuli has received wide support from previous research. Bargh and colleagues (1992) found that the more positive a word was subjectively evaluated, the faster it was responded to in a “good or bad” valence classification task. Evidence was also obtained in a lexical decision task where other factors such as word length, frequency, first phoneme, and arousal controlled, positive words were still responded to faster than negative words (Estes & Adelman, 2008). In line with this “positivity advantage”, the current research also found that participants responded faster to positive stimuli in the valence classification task compared to negative ones, and consistent between both words and emojis as materials. Similar response time differences between positive and negative stimuli were also demonstrated in the original research on the “good is up” metaphor conducted by Meier & Robinson (2004). In general, this finding has been consistently supported.

There is more controversy with regards to the negativity effect on memory performance. Although there was evidence to suggest a negativity advantage on memory performance, there was also research suggesting a null effect. Ortony and colleagues (1983) might have been among the first to report a negativity advantage on recognition memory. In an unexpected recognition test after reading 80 sentences half of which with a positive tone and the other half with a negative tone, participants discriminated old or new sentences with a negative tone better compared to positive ones. This advantage has been replicated in a number of later experiments (Inaba et al., 2005; Ohira et al., 1998; Robinson-Riegler & Winton, 1996). This effect was also demonstrated in the research conducted by Crawford and colleagues (2014) where they reported the metaphor-incongruency effect on memory performance. However, as discussed in Chapter 2, some researchers suggest that the arousal extent matters greater than the negative valence of the stimuli regarding the memory performance differences (Mather & Sutherland, 2009, 2011). But the lack of arousal measure for emoji stimuli (in Experiment 3 and 5b) and for the Mandarin word stimuli (in Experiment 5a) makes it hard to make a conclusive speculation in the current research.

Another interesting discussion is about people's ambivalent evaluation of some valenced stimuli. For some stimuli, the affective evaluation can be quite straightforward and consistent across most people, such as "love", "happiness", "murder" and "death". For others, however, the evaluation requires more efforts, where multiple factors are considered and appropriately compared (Cunningham et al., 2004). For example, when asked to classify the word "cake", some people might think more about its sweet taste and its relevance in celebrating positive events, however, some may subsequently consider its unhealthy ingredients and the contained amounts of sugar and fat, as well as the relevant feelings of body shame. Research shows that a consensus of valence rating on a particular item can still be demonstrated even when participants hold different extents of ambivalence (Kaplan, 1972). The ambivalence shows the complexity of the valence concept, and the process of labelling a stimulus as positive or negative might differ across participants and across stimuli. Potentially in future research on the valence concept, the positivity and negativity can be measured independently as an exploration of their separate effects on cognitive processes.

Limitations

There are a few limitations of the current research, including the gender representativeness of participants, the control over emoji materials and the lack of experimental control in online data collection. Potential implications will be discussed, and hopefully, future research on this topic can address these potential issues.

Participants

A large proportion of the participants in the current research are females, which might have implications for interpreting the results. Gender differences have been frequently noticed in research on emotions (Bradley et al., 2001; Hamann & Canli, 2004; McManis et al., 2001; Thomsen et al., 2005). Males' and females' brain activity differs when processing positive and negative emotional images (Kemp et al., 2004), such that males show a stronger brain activity for positive visual stimuli than females whereas females show a stronger brain activity for negative stimuli (Wrase et al., 2003). But in both abovementioned studies, gender did not significantly influence the self-reported valence ratings on the emotional stimuli, neither positive nor negative ones (Kemp et al., 2004; Wrase et al., 2003). However, other research found that females perceive negative stimuli to be more negative than males do (Stevens & Hamann, 2012). Specifically for emojis, females rated the negative emojis as more negative than did the males, whereas no gender differences were detected for the valence ratings on positive emojis (Jones et al., 2020). There was also evidence on gender differences on memory of affective stimuli. Highly emotional pictures were generally memorised better in participants of both genders, compared to more neutral stimuli, but the difference was more pronounced for females than males (Canli et al., 2002).

Due to the lack of consideration of the gender differences and also the recruitment convenience, participants in the current research were mainly recruited from the university campus, which caused a larger proportion of female participants, with the only exception of in Experiment 5 with Chinese context. It is suggested that future research using affective stimuli should consider the gender differences as a potential influence factor for participants recruitment.

Materials

The limitation of materials primarily concerns the emojis stimuli. The potential influence factors inherent to the materials were not as well controlled as they were for word materials. For word stimuli, the materials were selected from a widely used database ANEW (Bradley & Lang, 1999) where norms of various aspects had been measured and provided, such as arousal and frequency. But as emojis became of interest as affective stimuli only in recent years, the LEED (Rodrigues et al., 2018) is the only database that I am aware of for norms of emojis, and the norms were obtained from a Portuguese context only. Considering the potential discrepancies of valence ratings between Portuguese participants and

British/Chinese participants in the current research, a pre-test was conducted to ensure a consensus of valence ratings. However, variables such as arousal and frequency were not measured and controlled which can be a critical disadvantage. Apart from the lack of considering these variables, it might also be hard to practically control the usage frequency variable of emoji materials, on one hand due to individual differences, such as gender and age (Prada et al., 2018), for example, females reported greater overall use of emojis than males (Jones et al., 2020); on the other hand, the positive emojis were in general more frequently used than negative ones (Li & Yang, 2018).

Generally speaking, the control over the emoji stimuli was not as good as over words; this is something can be improved in future studies. Despite this disadvantage, the experiments using the two kinds of materials have yielded a similar pattern of results, which emphasizes the key role of inherent positive or negative valence of these materials with respect to metaphoric effects.

Online data collection

As briefly mentioned in Chapter 3, the limitations of online data collection can be worrying for the reliability of the experiment results. Experiment 4, 6 and 7 in the current research were all conducted online and participants were guided entirely by a written program on a website during the whole procedure, without the ability to communicate to the experimenter.

Online data collection for research is not a new thing. Due to the advantages of time saving, lower cost, ease of data entry, format flexibility and the ability to capture additional response-set information, web-based data collection approach has been fast developing and widely applied in all fields in the recent years (Granello & Wheaton, 2004). But more often, web-based data collection was used for questionnaires rather than experiments. Experiments normally require a higher standard for environment, such as to provide a consistent and quiet lab space for participants to be able to focus on the tasks, and to guarantee that participants can have the chance to ask for experimenters whenever they have questions about the procedure. Specifically in the current research, in order to enlarge the spatial difference between up and down location, the computer screen in the lab was set vertically instead of the usual horizontal set-up in Experiment 1, 2, 3 and 5. The flexibility to change the hardware in experiments cannot be achieved in the context of online experiments. However, due to the global pandemic of COVID-19, lab-based experiments which involving face-to-face interaction had been banned for a long time, and online experiment has become a not perfect but helpful alternative.

Apart from the lack of control over the experimental environment, the limitations of online experiments also include participants' lack of motivation(Lefever et al., 2007). For example, some participants took part for course credits and participated without strong motivations or interest in the research per se, in which case they may be less focused during the experiments and may have added error variance to the dataset, especially in Experiment 6 & 7 where response time mattered directly to the hypothesized effects. Although for lab-based experiments this cannot be fully avoided either, the web-based research might augment the unwillingness component, due to the lack of supervision.

Conclusion

The "good is up" metaphor does not only play a role in our daily language use, but also plays a role in our cognitive processes. 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 reveals the heuristics based on the valence-verticality association. However, previous findings of the metaphoric effects on item memory, affective judgement and attention directing were not replicated in the current research, indicating that the metaphor may affect cognition in a more subtle way than expected. Future investigations on this topic with better control is appealed to altogether provide a more reliable and conclusive suggestion.

References

Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin & Review*, *20*(2), 326–333. https://doi.org/10.3758/s13423-012-0342-7

Arnold, N. R., Bayen, U. J., & Smith, R. E. (2015). Hierarchical multinomial modeling approaches: An application to prospective memory and working memory. *Experimental Psychology*, *62*(3), 143–152. https://doi.org/10.1027/1618-3169/a000287

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barbieri, F., Kruszewski, G., Ronzano, F., & Saggion, H. (2016). How cosmopolitan are emojis? Exploring emojis usage and meaning over different languages with distributional semantics. *Proceedings of the 24th ACM International Conference on Multimedia*, 531–535. https://doi.org/10.1145/2964284.2967278

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, *62*(6), 893–912. https://doi.org/10.1037/0022-3514.62.6.893

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*(4), 548. https://doi.org/10.1037/0033-295X.97.4.548

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

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), 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*(1), 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

Begg, C. B. (1994). Publication bias. In C. Harris & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). Russell Sage Foundation.

Bobrow, D. G., & Norman, D. A. (1975). Some principles of memory schemata. In A. Collins & D. G. Bobrow (Eds.), *Representation and understanding* (pp. 131–149). Academic Press. https://doi.org/10.1016/B978-0-12-108550-6.50010-0

Boisgontier, M. P., & Cheval, B. (2016). The anova to mixed model transition. *Neuroscience & Biobehavioral Reviews*, *68*, 1004–1005. https://doi.org/10.1016/j.neubiorev.2016.05.034

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *13*(2), 185–189. https://doi.org/10.1111/1467-9280.00434

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

Bowdle, B. F., & Gentner, D. (2005). The Career of Metaphor. *Psychological Review*, *112*(1), 193–216. https://doi.org/10.1037/0033-295X.112.1.193

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion*, *1*(3), 300–319. https://doi.org/10.1037/1528-3542.1.3.300

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.

Brookshire, G., Ivry, R., & Casasanto, D. (2010). Modulation of motor-meaning congruity effects for valenced words. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 1940–1945). Cognitive Science Society.

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

Canli, T., Desmond, J. E., Zhao, Z., & Gabrieli, J. D. E. (2002). Sex differences in the neural basis of emotional memories. *Proceedings of the National Academy of Sciences*, *99*(16), 10789–10794. https://doi.org/10.1073/pnas.162356599

Carpenter, S. (2012). Psychology’s bold initiative. *Science*, *335*(6076), 1558–1561. https://doi.org/10.1126/science.335.6076.1558

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

Crawford, E. L. (2009). Conceptual metaphors of affect. *Emotion Review*, *1*(2), 129–139. https://doi.org/10.1177/1754073908100438

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

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of Aattitudes. *Journal of Cognitive Neuroscience*, *16*(10), 1717–1729. https://doi.org/10.1162/0898929042947919

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

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

Elandt-Johnson, R. C. (1971). *Probability models and statistical methods in genetics.* John Wiley.

Erdfelder, E., & Bredenkamp, J. (1998). Recognition of script-typical versus script-atypical information: Effects of cognitive elaboration. *Memory & Cognition*, *26*(5), 922–938. https://doi.org/10.3758/BF03201173

Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion*, *8*(4), 441–444. https://doi.org/10.1037/1528-3542.8.4.441

Fanelli, D. (2010a). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PLoS ONE*, *5*(4), e10271. https://doi.org/10.1371/journal.pone.0010271

Fanelli, D. (2010b). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*, *5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Fay, A. J., & Maner, J. K. (2012). Warmth, spatial proximity, and social attachment: The embodied perception of a social metaphor. *Journal of Experimental Social Psychology*, *48*(6), 1369–1372. https://doi.org/10.1016/j.jesp.2012.05.017

Feng, J. (2011). *Top-down influences on selective attention across the extended visual field*. [Doctoral dissertation, University of Toronto].

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

Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review*, *19*(6), 975–991. https://doi.org/10.3758/s13423-012-0322-y

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

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/1177011136

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45–56. https://doi.org/10.1037/0003-066X.52.1.45

Gibbs, R. W., Bogdanovich, J. M., Sykes, J. R., & Barr, D. J. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language*, *37*(2), 141–154. https://doi.org/10.1006/jmla.1996.2506

Gibbs, R. W., Costa Lima, P. L., & Francozo, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics*, *36*(7), 1189–1210. https://doi.org/10.1016/j.pragma.2003.10.009

Giessner, S. R., & Schubert, T. W. (2007). High in the hierarchy: How vertical location and judgments of leaders' power are interrelated. *Organizational Behavior and Human Decision Processes*, *104*(1), 30–44. https://doi.org/10.1016/j.obhdp.2006.10.001

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*(1), 20–29. https://doi.org/10.1111/j.1745-6916.2008.00058.x

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Globig, 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

Gottwald, J. M., Elsner, B., & Pollatos, O. (2015). Good is up—spatial metaphors in action observation. *Frontiers in Psychology*, *6*, 1605. https://doi.org/10.3389/fpsyg.2015.01605

Granello, D. H., & Wheaton, J. E. (2004). Online data collection: Strategies for research. *Journal of Counseling & Development*, *82*(4), 387–393. https://doi.org/10.1002/j.1556-6678.2004.tb00325.x

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

Hamann, S., & Canli, T. (2004). Individual differences in emotion processing. *Current Opinion in Neurobiology*, *14*(2), 233–238. https://doi.org/10.1016/j.conb.2004.03.010

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, *37*(1), 25–38. https://doi.org/10.1037/0022-3514.37.1.25

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

Heider, J. D., Scherer, C. R., Skowronski, J. J., Wood, S. E., Edlund, J. E., & Hartnett, J. L. (2007). Trait expectancies and stereotype expectancies have the same effect on person memory. *Journal of Experimental Social Psychology*, *43*(2), 265–272. https://doi.org/10.1016/j.jesp.2006.01.004

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*(1), 21–47. https://doi.org/10.1007/BF02294263

Huang, Y., & Tse, C. S. (2015). Re-examining the automaticity and directionality of the activation of the spatial-valence “good is up” metaphoric association. *PLoS ONE*, *10*(4), e0123371. https://doi.org/10.1371/journal.pone.0123371

Hunt, P. D. (2021). *World Emoji Day 2021: How emoji can help create a more empathetic world, for all of us*. https://blog.adobe.com/en/publish/2021/07/15/global-emoji-trend-report-2021.html#gs.8kmgev

Inaba, M., Nomura, M., & Ohira, H. (2005). Neural evidence of effects of emotional valence on word recognition. *International Journal of Psychophysiology*, *57*(3), 165–173. https://doi.org/10.1016/j.ijpsycho.2005.01.002

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, *7*(6), 645–654. https://doi.org/10.1177/1745691612464056

Jamrozik, A., McQuire, M., Cardillo, E. R., & Chatterjee, A. (2016). Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin and Review*, *23*(4), 1080–1089. https://doi.org/10.3758/s13423-015-0861-0

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

Jones, L. L., Wurm, L. H., Norville, G. A., & Mullins, K. L. (2020). Sex differences in emoji use, familiarity, and valence. *Computers in Human Behavior*, *108*, 106305. https://doi.org/10.1016/j.chb.2020.106305

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin*, *77*(5), 361–372. https://doi.org/10.1037/h0032590

Kemp, A. H., Silberstein, R. B., Armstrong, S. M., & Nathan, P. J. (2004). Gender differences in the cortical electrophysiological processing of visual emotional stimuli. *NeuroImage*, *21*(2), 632–646. https://doi.org/10.1016/j.neuroimage.2003.09.055

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

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., & 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., 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.

Langston, W. (2002). Violating orientational metaphors slows reading. *Discourse Processes*, *34*(3), 281–310. https://doi.org/10.1207/S15326950DP3403_3

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

Lefever, S., Dal, M., & Matthíasdóttir, Á. (2007). Online data collection in academic research: Advantages and limitations. *British Journal of Educational Technology*, *38*(4), 574–582. https://doi.org/10.1111/j.1467-8535.2006.00638.x

Li, L., & Yang, Y. (2018). Pragmatic functions of emoji in internet-based communication---a corpus-based study. *Asian-Pacific Journal of Second and Foreign Language Education*, *3*(1), 16. https://doi.org/10.1186/s40862-018-0057-z

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

Mäntylä, T., & Bäckman, L. (1992). Aging and memory for expected and unexpected objects in real-world settings. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1298–1309. https://doi.org/10.1037/0278-7393.18.6.1298

Mather, M., & Sutherland, M. (2009). Disentangling the effects of arousal and valence on memory for intrinsic details. *Emotion Review*, *1*(2), 118–119. https://doi.org/10.1177/1754073908100435

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science*, *6*(2), 114–133. https://doi.org/10.1177/1745691611400234

Matzke, D., Dolan, C. V, Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205–235. https://doi.org/10.1007/s11336-013-9374-9

McDermott, K. B., & Roediger, H. L. (2018). Memory (encoding, storage, retrieval). In A. Butler (Ed.), *General Psychology (Fall 2018). Psychology Curricular Materials. 2.* (pp. 118–141). Valparaiso University.

McGlone, M. S., & Harding, J. L. (1998). Back (or forward?) to the future: The role of perspective in temporal language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1211–1223. https://doi.org/10.1037/0278-7393.24.5.1211

McManis, M. H., Bradley, M. M., Berg, W. K., Cuthbert, B. N., & Lang, P. J. (2001). Emotional reactions in children: Verbal, physiological, and behavioral responses to affective pictures. *Psychophysiology*, *38*(2), 222–231. https://doi.org/10.1111/1469-8986.3820222

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., Hauser, D. J., Robinson, M. D., Friesen, C. K., & Schjeldahl, K. (2007). What’s" up" with God? Vertical space as a representation of the divine. *Journal of*

Personality and Social Psychology, *93*(5), 699–710. https://doi.org/10.1037/0022-3514.93.5.699

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

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *2*(1), 1–21. https://doi.org/10.1016/S0022-5371(63)80063-8

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

Mumford, M. D., & Leritz, L. E. (2005). Heuristics. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 203–208). Elsevier. https://doi.org/10.1016/B0-12-369398-5/00168-7

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 1–9. https://doi.org/10.1038/s41562-016-0021

Murphy, G. L. (1996). On metaphoric representation. *Cognition*, *60*(2), 173–204. https://doi.org/10.1016/0010-0277(96)00711-1

Núñez, R. E., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, *30*(3), 401–450. https://doi.org/10.1207/s15516709cog0000_62

Ohira, H., Winton, W. M., & Oyama, M. (1998). Effects of stimulus valence on recognition memory and endogenous eyeblinks: Further evidence for positive-negative asymmetry. *Personality and Social Psychology Bulletin*, *24*(9), 986–993. https://doi.org/10.1177/0146167298249006

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*(6), 657–660. https://doi.org/10.1177/1745691612462588

Ortony, A., Turner, T. J., & Antos, S. J. (1983). A puzzle about affect and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 725–729. https://doi.org/10.1037/0278-7393.9.4.725

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

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253

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

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology*, *1*(1), 33–60. https://doi.org/10.1080/14792779108401856

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

Pietschnig, J., Siegel, M., Eder, J. S. N., & Gittler, G. (2019). Effect declines are systematic, strong, and ubiquitous: A meta-meta-analysis of the decline effect in intelligence research. *Frontiers in Psychology*, *10*, 2874. https://doi.org/10.3389/fpsyg.2019.02874

Prada, M., Rodrigues, D. L., Garrido, M. V, Lopes, D., Cavalheiro, B., & Gaspar, R. (2018). Motives, frequency and attitudes toward emoji and emoticon use. *Telematics and Informatics*, *35*(7), 1925–1934. https://doi.org/10.1016/j.tele.2018.06.005

Previc, F. H., & Blume, J. L. (1993). Visual search asymmetries in three-dimensional space. *Vision Research*, *33*(18), 2697–2704. https://doi.org/10.1016/0042-6989(93)90229-P

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

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

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339. https://doi.org/10.1037/0033-295X.95.3.318

Robinson-Riegler, G. L., & Winton, W. M. (1996). The role of conscious recollection in recognition of affective material: Evidence for positive-negative asymmetry. *The Journal of General Psychology*, *123*(2), 93–104. https://doi.org/10.1080/00221309.1996.9921263

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

Roese, N. J., & Sherman, J. W. (2007). Expectancy. In A. W. Kruglanski & E. T. Higgins

(Eds.), *Social psychology: Handbook of basic principles* (pp. 91–115). The Guilford Press.

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology*, *31*(2), 81–109. https://doi.org/10.1111/j.2044-8309.1992.tb00958.x

Rouder, J. N., Morey, R. D., Speckman, P. L., & Pratte, M. S. (2007). Detecting chance: A solution to the null sensitivity problem in subliminal priming. *Psychonomic Bulletin & Review*, *14*(4), 597–605. https://doi.org/10.3758/BF03196808

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, *133*(4), 534–553. https://doi.org/10.1037/0096-3445.133.4.534

Santana, E., & de Vega, M. (2011). Metaphors are embodied, and so are their literal counterparts. *Frontiers in Psychology*, *2*, 90. https://doi.org/10.3389/fpsyg.2011.00090

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

Schaverien, P. M. (2013). *The effect of valence and arousal on spatial attention*. [Master's thesis, Victoria University of Wellington].

Schmidt, S. (2016). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 581–596). American Psychological Association. https://doi.org/10.1037/14805-036

Schubert, T. W. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology*, *89*(1), 1–21. https://doi.org/10.1037/0022-3514.89.1.1

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

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

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2015). *afex: Analysis of factorial experiments*. R package version 0.13-145.

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In D. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology* (pp. 4–31). Routledge. https://doi.org/10.4324/9780429318405-2

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.

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*(1), 167–183. https://doi.org/10.1016/j.jmp.2009.06.007

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

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information : A review of the social and social developmental literatures. *Psychological Bulletin*, *111*(1), 42–61. https://doi.org/10.1037/0033-2909.111.1.42

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

Stevens, J. S., & Hamann, S. (2012). Sex differences in brain activation to emotional stimuli: A meta-analysis of neuroimaging studies. *Neuropsychologia*, *50*(7), 1578–1593. https://doi.org/10.1016/j.neuropsychologia.2012.03.011

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Thomsen, D. K., Mehlsen, M. Y., Viidik, A., Sommerlund, B., & Zachariae, R. (2005). Age and gender differences in negative affect—Is there a role for emotion regulation? *Personality and Individual Differences*, *38*(8), 1935–1946. https://doi.org/10.1016/j.paid.2004.12.001

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences*, *105*(33), 11655–11660. https://doi.org/10.1073/PNAS.0802686105

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Unkelbach, C., Alves, H., & Koch, A. (2020). Negativity bias, positivity bias, and valence asymmetries: Explaining the differential processing of positive and negative information. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 62, pp. 115–187). Academic Press. https://doi.org/10.1016/bs.aesp.2020.04.005

Unkelbach, C., von Hippel, W., Forgas, J. P., Robinson, M. D., Shakarchi, R. J., & Hawkins, C. (2010). Good things come easy: Subjective exposure frequency and the faster processing of positive information. *Social Cognition*, *28*(4), 538–555. https://doi.org/10.1521/soco.2010.28.4.538

Vakil, E., Sharot, T., Markowitz, M., Aberbuch, S., & Groswasser, Z. (2003). Script memory for typical and atypical actions: Controls versus patients with severe closed-head injury. *Brain Injury*, *17*(10), 825–833. https://doi.org/10.1080/0269905021013196

van ’t Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004

von Hecker, U., Klauer, K. C., & Sankaran, S. (2013). Embodiment of social status: Verticality effects in multilevel rank-orders. *Social Cognition*, *31*(3), 374–389. https://doi.org/10.1521/soco.2013.31.3.374

von Hecker, U., Klauer, K. C., Wolf, L., & Fazilat-Pour, M. (2016). Spatial processes in linear ordering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(7), 1003–1033. https://doi.org/10.1037/xlm0000220

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

Wrase, J., Klein, S., Gruesser, S. M., Hermann, D., Flor, H., Mann, K., Braus, D. F., & Heinz, A. (2003). Gender differences in the processing of standardized emotional visual stimuli in humans: A functional magnetic resonance imaging study. *Neuroscience Letters*, *348*(1), 41–45. https://doi.org/10.1016/S0304-3940(03)00565-2

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

Yasuda, S. (2010). Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *Tesol Quarterly*, *44*(2), 250–273. https://doi.org/10.5054/tq.2010.219945

Zanolie, K., van Dantzig, S., Boot, I., Wijnen, J., Schubert, T. W., Giessner, S. R., & Pecher, D. (2012). Mighty metaphors: Behavioral and ERP evidence that power shifts attention on a vertical dimension. *Brain and Cognition*, *78*(1), 50–58. https://doi.org/10.1016/j.bandc.2011.10.006

Zarzeczna, N., von Hecker, U., Proulx, T., & Haddock, G. (2020). Powerful men on top: Stereotypes interact with metaphors in social categorizations. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(1), 36–65. https://doi.org/10.1037/xhp0000699

Appendices

Appendix A. Materials of all the experiments

A.1 Positive and negative words in Experiment 1

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.

A.2 Neutral character stimuli in Experiment 2

ዮ ኳ ብ ት ን ታ

ህ ሏ ዪ ይ ቃ ለ

ኩ ሯ ኑ ም ቄ ሙ

ኮ ኗ ሯ ኴ ፑ ዖ

ሴ ሚ ዲ ኢ ጄ ቷ

ጎ ሱ ኧ ሩ ዕ አ

ዌ ቹ ው ኪ ዝ ቯ

ኡ ቅ

A.3 Positive and negative emojis in Experiment 3

<u>Positive emojis:</u>

<u>Negative emojis:</u>

A.4 High and low connotation words in Experiment 4

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.

A.5 Positive and negative Mandarin words in Experiment 5a

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.

A.6 Positive and negative emojis in Experiment 5b

Positive emojis:

Negative emojis:

A.7 High and low connotation Mandarin words in Experiment 5c

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. DICs of alternative models in Experiment 1-5

As described in Chapter 2, 4 alternative models were considered for each dataset which were Model 5b, 5c, 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 support (Millar, 2009; Spiegelhalter et al., 2002). Regarding comparison between Model 5 and 6 across all datasets, DICs either showed preference of Model 5b/5c ($DIC_{6c/6d} - DIC_{5b/5c} > 6$), or did not show any preferences (all ΔDICs ≤ 6) in which case Model 5b/5c was selected due to consideration of parsimony. Then regarding the comparison between Model 5b and 5c, as explained in Chapter 2, if Model 5c was preferred by the smaller DIC ($DIC_{5b} - DIC_{5c} > 6$), Model 5c was selected, otherwise Model 5b was selected. The DICs of the 4 alternative models for each dataset are provided in the tables below.

Table 14

DICs of Alternative Models in Experiment 1

	Group 1	Group 2	Group 3
Model 5b	1375	1294*	1418
Model 5c	1368*	1295	1410*
Model 6c	1372	1309	1420
Model 6d	1378	1306	1414

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

Table 15

DICs of Alternative Models in Experiment 2

	Group 1	Group 2
Model 5b	809.4*	795*
Model 5c	815.6	796.8
Model 6c	809.6	803.5
Model 6d	815.8	803

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

Table 16

DICs of Alternative Models in Experiment 3

	Group 1	Group 2	Group 3
Model 5b	1463*	1421	1437*
Model 5c	1466	1411*	1439
Model 6c	1463	1426	1436
Model 6d	1461	1420	1443

Note. "*" represents the selected model in each Group

Table 17

DICs of Alternative Models in Experiment 4

	Group 1	Group 2	Group 3
Model 5b	1331*	1142*	1682*
Model 5c	1331	1139	1683
Model 6c	1334	1149	1690
Model 6d	1334	1151	1686

Note. "*" represents the selected model in each Group

Table 18

DICs of Alternative Models in Experiment 5a-5c

	E5a	E5b	E5c
Model 5b	1873*	1772*	1610*
Model 5c	1878	1772	1614
Model 6c	1875	1776	1623
Model 6d	1878	1771	1621

Note. "*" represents the selected model in each Group

Appendix C. LMM construction procedure in Experiment 6-7

As described in Chapter 3, the model construction procedure followed the steps proposed by Barr and colleagues (2013) with the practical guidance provided by Singmann & Kellen (2019), which started from a maximal model with by-participant random slopes for valence, verticality and their interactions then modified until optimal. Model construction was achieved using an R package "afex" (Singmann et al., 2015). Final models differed from each other due to different results of each modification step, details see below.

C.1 LMM of Experiment 6a

The maximal model with by-participant random slopes for valence, verticality and their interaction, with Satterthwaite approximation method used (method = “S”):

```
m1s <- mixed(rt_log ~ valence*verticality +
            (valence*verticality|participant),
          eliminated, method = "S",
          control = lmerControl(optCtrl = list(maxfun = 1e6)))
```

The maximal model did not converge successfully, indicated by a “singular fit” warning. In this case, the usual first step is removing the correlations among the random terms, which mean excluding correlation among by-participant random effect terms (i.e., no correlations among random slopes themselves and between the random slopes and the random intercept):

```
m2s <- mixed(rt_log ~ valence*verticality +
            (valence*verticality||participant),
          eliminated, method = "S",
          control = lmerControl(optCtrl = list(maxfun = 1e6)),
          expand_re = TRUE)
```

Again, a “singular fit” warning showed up indicating that the model was over-parameterized and did not converge successfully. Then the estimated random effect terms were checked by the following command:

```
summary(m2s)$varcor
```

The output showed that the estimated SD of the random slopes of the two-way interactions was zero, so the interaction was excluded from the model:

```
m3s <- mixed(rt_log ~ valence*verticality +
              (valence+verticality||participant),
            eliminated, method = "S",
            control = lmerControl(optCtrl = list(maxfun = 1e6)),
                expand_re = TRUE)
```

The current model successfully converged. But before deciding the final model, it was in need to check whether the correlation terms can be added again without running into any problems:

```
m4s <- mixed(rt_log ~ valence*verticality +
            (valence+verticality|participant),
          eliminated, method = "S",
          control = lmerControl(optCtrl = list(maxfun = 1e6)),
          expand_re = TRUE)
```

The model failed to converge. As a result, m3s was the final selected model in Experiment 6a.

C.2 LMM of Experiment 6b

The model construction procedure also started from a similar maximal model as in Experiment 6a:

```
m1s <- mixed(rt_log ~ valence*verticality +
          (valence*verticality|participant),
        eliminated, method = "S",
            control = lmerControl(optCtrl = list(maxfun = 1e6)))
```

The maximal model did not converge successfully. Excluded correlation among by-participant random effect terms in the next step:

```
m2s <- mixed(rt_log ~ valence*verticality +
          (valence*verticality||participant),
        eliminated, method = "S",
        control = lmerControl(optCtrl = list(maxfun = 1e6)),
            expand_re = TRUE)
```

This Model fitted well, then the estimated random effect terms were checked by the following command:

```
summary(m2s)$varcor
```

The output showed no estimated SD of the random slopes was equal to zero, so m2s became the final selected model in Experiment 6b.

C.3 LMM of Experiment 7a

The model construction procedure also started from a similar maximal model as in the previous two experiments:

```
m1s <- mixed(rt_log ~ valence*verticality +
          (valence*verticality|participant),
        eliminated, method = "S",
            control = lmerControl(optCtrl = list(maxfun = 1e6)))
```

The maximal model did not converge successfully. Excluded correlation among by-participant random effect terms:

```
m2s <- mixed(rt_log ~ valence*verticality +
          (valence*verticality||participant),
        eliminated, method = "S",
        control = lmerControl(optCtrl = list(maxfun = 1e6)),
```

```
                expand_re = TRUE)
```

A “singular fit” warning showed up indicating that the model was over-parameterized and did not converge successfully. Then the estimated random effect terms were checked by the following command:

```
summary(m2s)$varcor
```

The output showed that the random effect SD for the valence term was estimated to be zero. However, the random slope of valence cannot be removed while retaining the two-way interaction of valence and verticality, the next model started to remove the interaction first:

```
m3s <- mixed(rt_log ~ valence*verticality +
               (valence+verticality||participant),
             eliminated, method = "S",
             control = lmerControl(optCtrl = list(maxfun = 1e6)),
                 expand_re = TRUE)
```

Not surprisingly, a “singular fit” warning showed up again and the inspection of the estimates showed that the random effect SD for the valence term was still estimated to be zero:

```
summary(m3s)$varcor
```

Remove random slope of valence in the next step:

```
m4s <- mixed(rt_log ~ valence*verticality +
             (verticality||participant),
           eliminated, method = "S",
           control = lmerControl(optCtrl = list(maxfun = 1e6)),
               expand_re = TRUE)
```

The Model fitted well. Again, before deciding the final model, it was in need to check whether the correlation terms can be added again without running into any problems:

```
m5s <- mixed(rt_log ~ valence*verticality +
             (verticality|participant),
           eliminated, method = "S",
           control = lmerControl(optCtrl = list(maxfun = 1e6)),
           expand_re = TRUE)
```

The Model fitted well. As a result, m5s became the final selected model in Experiment 7a.

C.4 LMM of Experiment 7b

The model construction procedure also started from a similar maximal model as in the previous experiments:

```
m1s <- mixed(rt_log ~ valence*verticality +
          (valence*verticality|participant),
        eliminated, method = "S",
            control = lmerControl(optCtrl = list(maxfun = 1e6)))
```

The maximal model did not converge successfully indicated by a “singular fit” warning. Excluded correlation among by-participant random effect terms:

```
m2s <- mixed(rt_log ~ valence*verticality +
          (valence*verticality||participant),
        eliminated, method = "S",
        control = lmerControl(optCtrl = list(maxfun = 1e6)),
            expand_re = TRUE)
```

A “singular fit” warning showed up indicating that the model was over-parameterized and did not converge successfully. Then the estimated random effect terms were checked by the following command:

```
summary(m2s)$varcor
```

The output showed that the random effect SDs for the both the valence term and the interaction of valence and verticality were estimated to be zero. The next model started to remove the interaction first:

```
m3s <- mixed(rt_log ~ valence*verticality +
            (valence+verticality||participant),
          eliminated, method = "S",
          control = lmerControl(optCtrl = list(maxfun = 1e6)),
              expand_re = TRUE)
```

Not surprisingly, a “singular fit” warning showed up again and the inspection of the estimates showed that the random effect SD for the valence term was still estimated to be zero:

```
summary(m3s)$varcor
```

Remove random slope of valence in the next step:

```
m4s <- mixed(rt_log ~ valence*verticality +
          (verticality||participant),
        eliminated, method = "S",
        control = lmerControl(optCtrl = list(maxfun = 1e6)),
            expand_re = TRUE)
```

The Model fitted well. Again, the correlation terms were tried to be added:

```
m5s <- mixed(rt_log ~ valence*verticality +
          (verticality|participant),
```

```
            eliminated, method = "S",
            control = lmerControl(optCtrl = list(maxfun = 1e6)),
            expand_re = TRUE)
```

A “singular fit” warning showed up. As a result, m4s became the final selected model in Experiment 7b.