Original Articles

Metaphors we learn by: Directed motor action improves word learning

Daniel Casasanto[a,b,c,*], Angela de Bruin[d]

[a] *Department of Psychology, University of Chicago, USA*
[b] *Department of Human Development, Cornell University, USA*
[c] *Department of Psychology, Cornell University, USA*
[d] *Basque Center on Cognition, Brain and Language, Donostia, Spain*

ABSTRACT

Can performing simple motor actions help people learn the meanings of words? Here we show that placing vocabulary flashcards in particular locations after studying them helps students learn the definitions of novel words with positive or negative emotional valence. After studying each card, participants placed it on one of two shelves (top or bottom), according to its valence. Participants who were instructed to place positive cards on the top shelf and negative cards on the bottom shelf, consistent with metaphors that link "good" with "up," remembered the words' definitions better than participants who followed the opposite spatial mapping, and better than control participants who placed all of the cards on the desktop. Saying "up" and "down" after studying the cards was ineffective, suggesting a privileged role for motor action in activating space-valence associations that partly constitute the meanings of emotionally charged words. These results provide a first demonstration that mental metaphors can be activated strategically to improve (or impair) word learning: We call this the *strategic use of mental metaphor* (SUMM) effect. Even when multiple factors known to enhance encoding of verbal materials into long-term memory were matched across conditions (e.g., study time, repetition, distinctiveness, depth of processing), metaphor-congruent motor actions led to better elaborated, more durable memories.

1. Introduction

When people talk about ideas with positive and negative emotional valence, they often use spatial metaphors. "Good" is associated with the top and "bad" with the bottom of an imaginary spatial continuum (e.g., her spirits *soared*; her hopes *sank*; Lakoff & Johnson, 1980, 1999). Beyond language, it appears that people conceptualize positive and negative ideas, in part, using mental metaphors: mappings between nonlinguistic representations in the concrete domains of space and motion, which can be experienced through perceptuo-motor interactions with the physical world, and in the relatively abstract domain of emotional valence which can only be experienced through introspection or interoception.

Links between vertical motion and valence can be seen in spontaneous expressions of pride and shame found across cultures (e.g., raising the arms above the head; slumping the shoulders; Tracy & Matsumoto, 2008). Inducing an upright or slumped posture in the laboratory can influence the amount of pride people express (Stepper & Strack, 1993) and the efficiency with which they retrieve positive and negative memories (Riskind, 1983), demonstrating a causal relationship between vertical motion and emotion.

Just as socially meaningful postures and gestures can cue positive or negative feelings and memories, so can meaningless upward and downward motor actions. Casasanto and Dijkstra (2010, Expt. 2) randomly assigned participants to move marbles either upward or downward, from one cardboard box to another, while retrieving and retelling autobiographical memories in response to neutral-valence prompts (e.g., *Tell me about something that happened yesterday*). Participants recounted positive memories more often during upward movements and negative memories more often during downward movements, demonstrating that repetitive motor actions with no semiotic value can activate the spatio-motor representations that, by hypothesis, partly constitute emotional memories.

If metaphor-congruent motor actions can influence memory retrieval, can they also be harnessed to improve learning? Here we investigated whether activating mental metaphors from vertical space to valence can help people learn the definitions of positive and negative words, and if so, whether it is more effective to activate these associations via verbal or nonverbal means.

---

2. Experiment 1: Can metaphor-congruent motor actions improve word learning?

Experiment 1 tested whether placing vocabulary flashcards in metaphor-congruent locations after studying them can help participants learn the definitions of words with positive and negative emotional valence.

2.1. Methods

Participants Right-handed Dutch speakers (N = 72) from the Radboud University participated for payment.

Materials Sixteen pronounceable Dutch pseudowords were created, which carried no strong positive or negative connotations, according to a pretest. Sixteen Dutch nouns were selected, half with positive and the other half with negative valence according to the ratings of their English translation equivalents in the ANEW corpus (Bradley & Lang, 1999). Although matching was not required by the fully counter-balanced design, the positive Dutch words were matched with the negative words for syllables, letters, concreteness, and lexical frequency. The assignment of each pseudoword to a positive or negative Dutch word was counterbalanced across participants (Appendix A). We made 2 sets of 16 laminated flashcards. Each card had a pseudoword on one side and either its positive or negative Dutch counterpart on the other side. Participants studied the cards while seated at a table. In front of them was a laptop computer, used only during the test phase. To their right were two cardboard shelves, one approximately 20 cm above the tabletop and the other 20 cm below the tabletop. (These shelves were removed in the baseline condition.)

Procedure Each participant performed 2 study phases and 2 test phases. During the first study phase, participants studied the definitions of 8 pseudowords and then performed a recognition memory test. This study-test procedure was then repeated for 8 new pseudowords with items counterbalanced across blocks. Within each block, half of the pseudowords had positive Dutch definitions and the other half had negative definitions. Participants received written instructions before each phase, and the experimenter was present throughout the experiment to ensure the instructions were followed.

Study phase All participants were told that they would be learning the definitions of "words in an alien language," which had either strongly positive or negative meanings. Their job was to study each flashcard for 6 s (timed by a metronome) and to memorize its definition for a subsequent test. They were told that, while studying each card, they should think about whether the word had a positive or negative definition, and then place the card in a particular spot, which varied according to the condition they were assigned to.

Each participant was randomly assigned to study the flashcards under one of three conditions: making good-is-up movements (i.e., metaphor-congruent condition; N = 24), good-is-down movements (i.e., metaphor-incongruent condition; N = 24), or baseline movements (N = 24). Participants were told that recent scientific findings had shown that thinking about valence and making corresponding movements improves word learning. In the good-is-up and good-is-down conditions, participants received examples of Dutch idioms in which positive or negative words were connected to a high or low location. For the good-is-up condition, these examples were: 'in de zevende hemel zijn' ('to be in seventh heaven') and 'je down voelen' ('feeling down'). In the good-is-down condition, participants received the following examples: 'met beide benen op de grond staan' ('to have both feet on the ground') and 'het hoog in de bol hebben' (literally 'having it high in the head', meaning 'being arrogant').

Participants in both the good-is-up and good-is-down conditions were told to place each card on either the upper or the lower shelf, according to the word's valence. Good-is-up participants placed positive cards on the upper shelf and negative cards on the lower shelf; good-is-down participants did the opposite (positive cards down, negative cards

up). Participants in the baseline condition performed the same task, except that after studying each word and thinking about its valence they placed the card on the desktop, in one pile, on the spot where the shelves were attached in the other conditions. Thus, the encoding instructions for each 6-second study period were the same across all conditions (i.e., memorize the alien word's definition while thinking about its valence); what differed across conditions was the motor action that participants performed after studying each card.

Participants studied the cards one at a time, in random order. When the participant finished studying all of the cards, the experimenter shuffled them and returned them to the participant for a second and third round of studying (i.e., each card was studied 3 times). The experimenter confirmed that all participants placed the cards in the correct locations after studying them.

Test phase After each study phase, participants performed a 2-alternative forced choice recognition test. Each pseudoword appeared in the center of the laptop screen for 6 s, after which 2 of the studied Dutch words (one target and one foil) appeared on the left and right, with their positions counterbalanced across participants. Each Dutch word appeared once as the target word and once as the foil. Half of the target-foil pairs had the same valence while the other half differed in valence. Participants pressed the Q or the P key to indicate whether the correct definition of the pseudoword appeared on the left or the right. Responses were self-paced.

2.2. Results and discussion

Overall, the motor actions that participants performed during encoding had a highly significant effect on their recognition memory, according to an omnibus mixed-effects binary logistic regression with Subjects and Items as repeated random factors and Encoding Condition as a fixed factor (Baayen, Davidson, & Bates, 2008). Accuracy was highest in the metaphor-congruent good-is-up condition, lowest in the metaphor-incongruent good-is-down condition, and intermediate in the metaphor-neutral baseline condition (Wald $\chi^2 = 14.70$, df = 2, p = .001, $\eta_p^2 = 0.013$; Fig. 1).[1] The same model was used to test pairwise differences between the conditions (good-is-up vs. good-is-down: Difference = 0.08, Wald $\chi^2 = 14.34$, df = 1, p = .0001, $\eta_p^2 = 0.020$; good-is-up vs. baseline: Difference = 0.04, Wald $\chi = 4.17$, df = 1, p = .04, $\eta_p^2 = 0.006$; baseline vs. good-is-down: Difference = 0.04, Wald $\chi^2 = 3.63$, df = 1, p = .057, $\eta_p^2 = 0.005$; all tests 2-tailed; data are available as supplementary materials).

Only the metaphor-congruent good-is-up condition is likely to be strategically useful for students or educators, who are unlikely to want to use metaphor-incongruent motor actions to *impair* learning. Still, the metaphor-incongruent good-is-down condition is important inferentially, in at least two ways. First, the difference between the baseline and good-is-down conditions provides an independent conceptual replication and extension of the good-is-up metaphor-congruity effect. (This difference is reported as marginally significant (p = .057) in a two-tailed test; we note however, that good-is-up metaphors make an *a priori* directional prediction. Therefore, a 1-tailed test is licensed, according to which this effect is clearly significant at p = .029.) Second, and crucially, the addition of the good-is-down condition allows us to evaluate the independent contribution of two variables: metaphor congruity and *distinctiveness at encoding*, a factor known to improve memory (Schacter, Israel, & Racine, 1999).

---

[1] Half of target-foil pairs differed in valence. In these instances, accurate answers could potentially have been due to participants remembering the valence rather than the exact meaning of the target word. We therefore conducted an additional analysis including valence match (same/different). Valence match did not affect accuracy (Wald $\chi^2 = 0.163$, df = 2, p = .686) and the effect of instructions remained significant when Match was controlled (Wald $\chi^2 = 14.72$, df = 2, p = .001, $\eta_p^2 = 0.013$).
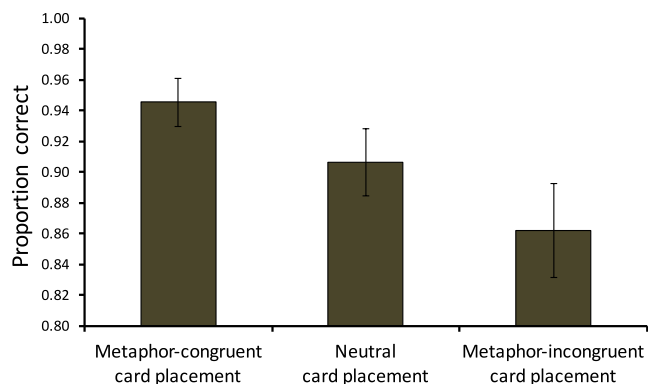
Fig. 1. Results of Experiment 1. Participants who placed flashcards in metaphor-congruent locations after studying them (left column) remembered the definitions of positive and negative words better than participants who placed the cards in metaphor-incongruent locations (right column), and better than participants in the baseline condition, who placed all cards on the tabletop (middle column). Error bars indicate s.e.m.

Whereas cards were placed in a single pile during the baseline condition, they were sorted into two piles during the metaphor-congruent and metaphor-incongruent movement conditions. Separating the cards into two piles presumably made the encoding episodes more distinctive in the up/down movement conditions than in the baseline condition. If we had only included the baseline and metaphor-congruent conditions in our design, the results would have been hard to interpret: In principle, the observed difference between the baseline and good-is-up conditions could be driven either by metaphor congruity, distinctiveness, or some combination of these variables. The inclusion of the good-is-down condition allowed us to resolve this ambiguity.

In the full experimental design, metaphor congruity and distinctiveness were orthogonal (i.e., statistically independent). Thus, we could examine the effects of these variables independently. A comparison of the good-is-up vs. good-is-down conditions allowed us to test for an effect of metaphor congruity independent of distinctiveness (since distinctiveness was identical across these conditions). As detailed above, this comparison showed a highly significant 8-percentage-point effect of metaphor congruity on accuracy, controlling for distinctiveness. To determine whether there was any effect of distinctiveness independent of metaphor congruity, we averaged the data from the good-is-up and good-is-down conditions (i.e., combining the 2-pile conditions), and compared the mean of the (higher-distinctiveness) 2-pile conditions to the mean of the (lower-distinctiveness) 1-pile baseline condition. This analysis showed that the mean accuracy in the combined 2-pile conditions was 91%; the mean accuracy in the 1-pile baseline condition was also 91% (difference between higher-distinctiveness and lower-distinctiveness conditions = 0%).

On the basis of these analyses we can conclude that metaphor congruity, our variable of interest, had a highly significant effect, controlling for distinctiveness. By contrast, distinctiveness (a variable not of interest) had no measurable effect, controlling for metaphor congruity. The null effect of distinctiveness in this study should not be surprising. In order to elicit distinctiveness effects, researchers typically induce participants to associate different to-be-remembered items with *dozens* of different retrieval cues (e.g., see Arndt & Reder, 2003) – not just two different retrieval cues (i.e., two piles), as in the present experiment.

To conclude this point, although there was a difference in distinctiveness between our baseline and up/down movement conditions, distinctiveness and metaphor congruity were orthogonal in the experimental design; this feature allowed us to determine that there was no independent effect of distinctiveness on accuracy and, more importantly, to rule out the possibility that distinctiveness could provide

an alternative, non-metaphorical explanation for our observed metaphor-congruity effects. Other factors known to influence memory outcomes, including item novelty, repetition, study time, study spacing, and depth of processing (i.e., use of a semantic-level orienting task) were also controlled across conditions.

To summarize the results of Experiment 1, participants who placed the flashcards on the shelves according to the good-is-up/bad-is-down mapping scored better than those who placed them according to the opposite space-valence mapping, and also better than participants in the baseline condition. Placing flashcards in metaphor-congruent locations improved learning. We call this a *strategic use of mental metaphors* (SUMM) effect.[2]

3. Experiment 2: extending the SUMM effect to neutral words

In Experiment 1, the "alien words" that participants learned had strongly valenced definitions. The goal of Experiment 2 was to extend strategic use of mental metaphors to a broader population of words. Here the alien words had neutral definitions, but participants were asked to generate either a positive association or a negative association for each word, and to place the flashcards in the appropriate locations according to this positive or negative association.

3.1. Methods

Participants Twenty-four right-handed Dutch speakers from Radboud University participated for payment.

Materials and procedure The materials and procedure were identical to those used in Experiment 1, with the following two exceptions. First, the 16 pseudowords from Experiment 1 were paired with 16 new Dutch words with neutral valences, according to the ANEW ratings of their English translation equivalents (Bradley & Lang, 1999; Appendix B). Second, before each study phase, the experimenter instructed participants to form novel positive associations for half of neutral words (e.g., if the word were *grass*, you could think about how grass reminds you of playing soccer), and negative associations with the other half of the words (e.g., if the word were *grass*, you could think about how grass gives you hay fever), and recorded the associations that participants generated. Essentially, participants turned the neutral words into valenced words for the purposes of the experiment. Positive-association and negative-association words were distributed evenly across the 2 study phases, with their order pseudorandomized, and with the assignment of valence to the words counterbalanced across participants. Half of the participants were assigned to the good-is-up condition, and the other half to the good-is-down condition. The experimenter confirmed that participants placed the flashcards in the appropriate locations after studying them (i.e., top or bottom shelf), according to the positive or negative valence of the associations they had formed.

---

[2] A reviewer suggested running a second baseline condition in which participants received no instructions about how to study the words (i.e., no instruction to think about word valence), before placing the card in the "neutral" position on the desktop. This new baseline condition (N = 12) produced exactly the same proportion of correct responses (Mean = 0.91) as the baseline condition reported in the main text, suggesting that explicitly considering the valence of the items did not affect recognition memory performance independent of the action manipulation. Ordinarily, orienting participants to "deeper," meaning based encoding of words improves recollection (Craik & Tulving, 1975). Our "deep encoding" (valence instructions) and "free encoding" (no valence instructions) baseline conditions may have produced identical results because the stimuli were so strongly positive and negative (e.g., miracle, joy vs. rape, torture) that their valences were processed automatically, with or without orienting instructions to think about this aspect of meaning.

3.2. Results and discussion

As in Experiment 1, participants remembered the meanings of words better when they were instructed to place flashcards in metaphor-congruent locations after studying them (good-is-up Proportion Correct = 0.92 ± 0.02) than when they were instructed to place them in metaphor-incongruent locations (good-is-down Proportion Correct = 0.85 ± 0.03; Difference = 0.07, Wald χ^2 = 4.30, df = 1, p = .04, η_p^2 = 0.012). The interaction of Experiment (Valenced Words, Neutral Words) and Condition (good-is-up, good-is-down) did not approach significance (Wald χ^2 = 0.54, df = 1, p = .46, η_p^2 = 0.00001), indicating that the strength of the metaphor-congruity effect did not differ between conventionally-valenced words and words for which positively- or negatively-valenced associations were generated *ad hoc*.

4. Experiment 3: Are metaphor-congruent words as effective as motor actions?

In Experiments 1 and 2, the mental metaphor good-is-up / bad-is-down was activated using upward and downward motor actions. Experiment 3 tested whether using the words "up" and "down" has a similar effect on learning.

4.1. Methods

Participants Forty-eight right-handed Dutch-speakers from Radboud University participated for payment.

Materials and procedure Half of the participants were randomly assigned to the good-is-up condition and the other half to the good-is-down condition. The materials and procedure were identical to those used in Experiment 1, with the following exception. After studying each card and considering the word's valence, rather than placing the cards up or down the participants said the words "omhoog" ("up") or "omlaag" ("down") aloud, with the correct response depending upon the valence of the word and the condition to which the participant was assigned. The experimenter confirmed that all participants produced the correct spatial words. After studying each item and saying "up" or "down" according to its valence, participants placed the card on the tabletop, in the same spot as during the baseline condition of Experiment 1.

4.2. Results and discussion

Saying "up" and "down" after studying the novel words' definitions had no effect on recognition memory performance. Accuracy did not differ between the Verbal good-is-up condition (Proportion Correct = 0.88 ± 0.02) and the Verbal good-is-down condition (Proportion Correct = 0.89 ± 0.02; Wald χ^2 = 0.46, df = 1, p = .50, η_p^2 = 0.001), nor did these condition means differ from that of the baseline condition from Experiment 1 (Wald χ^2 = 1.64, df = 2, p = .44, η_p^2 = 0.001). When upward and downward motor actions were replaced with the words "up" and "down," the metaphor congruity effect disappeared.

5. General discussion

Activating the mental metaphor good-is-up / bad-is-down via directed motor actions can help people learn the definitions of novel words with positive or negative emotional valence. Cognitive scientists and educators have emphasized the importance of linguistic metaphors for acquiring new concepts (Gentner & Wolff, 2000) and learning foreign languages (Danesi, 1995). This study provides a first demonstration that mental metaphors – independent of linguistic metaphors – can be mobilized to promote learning, establishing a *strategic use of mental metaphors* (SUMM) effect. (For other relationships between valence, space, and memory see Brunyé, Gardony, Mahoney, & Taylor, 2010;

Casasanto & Dijkstra, 2010; Crawford, Margolies, Drake, & Murphy, 2006).

In Experiment 1, participants who placed vocabulary flashcards in metaphor-congruent locations after studying them remembered the words' definitions 8 percentage points better than participants who had placed the cards on the shelves according to the opposite spatial mapping. Importantly, participants in the metaphor-congruent (good-is-up) condition also remembered the definitions significantly better than participants in the baseline condition who placed all of the cards on the tabletop. Participants in the metaphor-incongruent (good-is-down) condition tended to perform below baseline. Placing flashcards in metaphor-congruent locations improved learning, and placing them in metaphor-incongruent locations impaired learning.

In Experiment 2, the SUMM effect was extended to novel words whose conventional meanings are neutral in valence. These results suggest that this method could be effective for learning the meaning of any content word (e.g., nouns, verbs, adjectives, adverbs), so long as a positive or negative association can be generated.

In Experiment 3, when the instructions to place the cards on the upper or lower shelves were replaced with instructions to say the words "up" or "down," the metaphor-congruity effect disappeared. On the basis of results from Experiments 1 and 2 alone, it was not clear whether motor actions, *per se*, had any effect on memory performance, or whether merely thinking "up" and "down" was driving the observed effects. The results of Experiment 3, however, suggest that directed motor actions played a critical role in improving or impairing memory. Further experiments are needed to distinguish the contributions of the various constituents of these directed actions (e.g., motor planning, kinesthetic feedback, visual feedback; cf. Crawford et al., 2006).

Together, these experiments show that mental metaphors can be used strategically to improve word learning, and that the way in which mental metaphors are activated (verbally vs. nonverbally) may determine the effectiveness of this strategy. Although the directed actions people made here were not communicative, these results converge with those of studies showing effects of co-speech gesture on learning (Goldin-Meadow & Beilock, 2010; Goldin-Meadow, 2010; Kelly, McDevitt, & Esch, 2009). The magnitude of the metaphor-congruity effect for valenced words was not large (4 percentage points above baseline), but even this difference would be sufficient to qualitatively change a grade on a vocabulary test (i.e., from an "A-minus" to an "A" in the US system). Furthermore, the goal of this study was to provide proof of concept. Two simple changes would be likely to boost the observed effects. First, recognition memory accuracy on both of our baseline tasks was 91%: already near ceiling. It is remarkable, therefore, that significant improvement above baseline was observed. In classroom settings, students would likely need to learn a larger number of items which could reduce accuracy overall and allow larger benefits of metaphor congruity to emerge. Second, adjusting the timing of study and test phases could allow greater differentiation between the baseline and metaphor-congruent conditions. More practice studying with this method could yield stronger effects (here words were studied for less than 5 min), as could a longer retention interval between study and test. Indeed, benefits of communicative hand gestures on learning have been shown to increase over time (Goldin-Meadow, 2015): The same could be true for benefits of non-communicative motor actions.

Mechanism of the SUMM effect. Why did metaphor-congruent motor actions improve word learning? Our study was designed to rule out explanations based on several factors known to enhance long-term memory for verbal materials. First, study time, item spacing, and item repetition were held constant across all conditions. Second, Depth of Processing (Craik & Lockhardt, 1972) was also held constant across conditions. "Deep" processing was operationalized by Craik and Tulving (1975) as memory encoding that results from semantic (meaning-based) judgments about studied words. Our participants made the same "deep" semantic evaluation of words' valences across all conditions (with the exception of the added baseline condition, see

footnote). Third, moving a flashcard to one location or another provided a source-memory cue (Mitchell & Johnson, 2009) that was not available in the baseline condition. Relatedly, separating the cards into two piles made the encoding episodes slightly more distinctive in the 2-pile up/down movement conditions than in the 1-pile baseline condition, and could have made the resulting memory traces less susceptible to inter-item interference (Arndt & Reder, 2003). Yet, none of these facts can predict or explain the observed results. In principle, linking items to more distinctive actions or locations could have boosted memory performance in both the metaphor-congruent *and* metaphor-incongruent movement conditions, independent of metaphor congruity; but this is not the pattern we found. Distinctiveness cannot explain why performance in the metaphor-congruent condition was better than in the metaphor-incongruent condition (even though distinctiveness was identical across these conditions). Likewise, distinctiveness cannot explain why performance in the more distinctive 2-pile metaphor-incongruent condition was *worse* than in the less distinctive 1-pile baseline condition: Distinctiveness predicts the opposite relationship between these conditions. Finally, the mean of the higher-distinctiveness 2-pile conditions (i.e., the metaphor-congruent and metaphor-incongruent conditions combined) did not differ from the mean of the lower-distinctiveness 1-pile condition: Both means were 91%. Thus, there was no measurable independent effect of distinctiveness. In sum, neither study time, study spacing, repetition, depth of processing, nor distinctiveness can be the principle underlying these metaphor congruity effects, since all of these factors were matched across the metaphor-congruent and metaphor-incongruent conditions.

Another principle of successful memory, however, may underlie the observed metaphor congruity effects: Elaboration. Although Depth of Processing is operationalized in terms of the orienting tasks used, since the first tests of the Levels of Processing framework researchers have posited that the reason deep encoding enhances memory is that it encourages the "spread and elaboration" of activation through a network of related information in long-term memory (Craik & Tulving, 1975, pg. 291). Elaboration of a memory trace is encouraged by congruity between the meaning of a word and the context in which it is encoded (Schulman, 1974). The amount of elaboration can be described in terms of the number of associated "semantic features" activated at encoding: the more the better (Craik & Tulving, 1975, pg. 291.)

Hypothetical semantic features associated with our stimulus words might include their denotations, other related words and concepts, relevant mental images, and crucially for the present discussion, relevant spatial schemas. Reaction-time experiments suggest that people activate vertical spatial schemas for positive and negative words with a high degree of automaticity, even when space and valence are irrelevant to the task and participants are unaware of the space-valence manipulation (Casasanto, Brookshire, & Ivry, 2015), suggesting that these schemas are typically part of the semantic representation of valenced words. When participants were performing metaphor-congruent actions, activation of the correct spatial schemas was encouraged by the encoding context, promoting the activation of a network of information that was well elaborated, since the upward schema associated with a target item like "champion" would also be associated with other semantic features of this word (e.g., with ideas like "winning," and images like raising the arms in victory).

Metaphor-incongruent flashcard placement had the opposite effect, for a combination of reasons: First, when participants were instructed to place the positive cards on the bottom shelf, presumably the appropriate (upward) spatial schema had to be inhibited to avoid

programming the wrong motor action. Second, the inappropriate (downward) spatial schema had to be activated in order to perform the correct motor action, and this inappropriate spatial schema decreased the elaboration of the memory trace. Whereas a metaphor-congruent spatial schema would increase activation of the target word's semantic associates, a metaphor-incongruent spatial schema would decrease their activation (e.g., a downward schema during the encoding of "champion" is incompatible with ideas like "winning" and images like raising the arms in victory, and may inhibit these features of the target word's semantic network.) In summary, metaphor-congruent motor actions may improve memory performance because they increase elaboration of the semantic network activated while participants encode the target words, whereas metaphor-incongruent motor actions decrease semantic elaboration.

Metaphorical actions vs. words Why did metaphor-congruent motor actions improve word learning, whereas using a closely matched verbal strategy did not? Learning the definitions of novel words is a type of verbal paired-associates task. Therefore, on one possible explanation, saying the words "up" and "down" could have interfered with learning more than executing upward and downward motor actions did. Yet, the data are not consistent with this explanation, for two reasons. First, if producing words had interfered with encoding or consolidation processes, this would have resulted in an overall decrement in recognition memory performance compared to baseline. But in fact, memory performance in the verbal good-is-up and good-is-down conditions (Experiment 3) was indistinguishable from performance in the baseline condition (Experiment 1): There was no decrement. Second, although verbal interference could plausibly impair performance in both conditions, it cannot account for the lack of difference between the metaphor-congruent and metaphor-incongruent conditions in Experiment 3. The details of the data necessitate another explanation.

On an alternative account, directed motor actions may have influenced how well participants learned emotionally charged words, whereas spatial words did not influence learning, simply because performing directed motor actions *entails* activating a spatio-motor representation, whereas saying a word like "up" or "down" does not. In some contexts these words may be processed shallowly, without activating any spatio-motor schemas or cuing any neural activity in motor circuits that control directed actions; this may be especially likely when the words "up" and "down" are repeated 24 times each in a span of 5 minutes, as they were in Experiment 3. By contrast, even highly practiced reaching and placing actions require a detailed motor plan. Motor actions may be more effective than spatial words at activating spatio-motor representations, and therefore at influencing the learning of valenced words, because directed motor actions cannot be produced without spatio-motor representations, whereas spatial words can be (e.g., under conditions of semantic satiation). Motor actions, therefore, may play a privileged role in activating the mental metaphors we learn by.

Acknowledgements

Appendix A

Stimuli used in Experiments 1 and 3. All pseudowords were paired with a positive Dutch word for half of the participants (column 2) and a negative Dutch word for the other half of the participants (column 3). The English translations have their ANEW valence ratings indicated in parentheses.

Pseudoword	Positive Dutch equivalent	Negative Dutch equivalent
bantig	vakantie (holiday; 8.2)	depressie (depression; 1.9)
dranlig	overwinning (victory; 8.3)	armoede (poverty; 1.7)
geng	wonder (miracle; 8.6)	dood (death; 1.6)
grem	paradijs (paradise; 8.7)	kanker (cancer; 1.5)
greumig	vakantie (holiday; 8.2)	depressie (depression; 1.9)
heep	succes (success; 8.3)	ellende (misery; 1.9)
keeg	kampioen (champion; 8.4)	tragedie (tragedy; 1.8)
kradig	vreugde (joy; 8.6)	marteling (torture; 1.6)
krucht	liefde (love; 8.7)	verkrachting (rape; 1.3)
noeg	paradijs (paradise; 8.7)	kanker (cancer; 1.5)
reustig	liefde (love; 8.7)	verkrachting (rape; 1.3)
rittig	vreugde (joy; 8.6)	marteling (torture; 1.6)
rupt	kampioen (champion; 8.4)	tragedie (tragedy; 1.8)
scheum	overwinning (victory; 8.3)	armoede (poverty; 1.7)
sliemend	wonder (miracle; 8.6)	dood (death; 1.6)
zelvig	succes (success; 8.3)	ellende (misery; 1.9)

Appendix B

Stimuli used in Experiment 2. The English translations have their ANEW valence ratings indicated in parentheses.

Pseudoword	Dutch equivalent
bantig	badkamer (bathroom; 4.8)
dranlig	gewoonte (habit; 5.3)
geng	stoel (chair; 5.4)
grem	familie (family; 5.5)
greumig	standbeeld (statue; 4.9)
heep	verf (paint; 5.3)
keeg	lente (spring; 5.3)
kradig	salade (salad; 4.5)
krucht	poster (poster; 4.9)
noeg	viool (violin; 4.9)
reustig	onderwijs (education; 5)
rittig	humor (humor; 5.1)
rupt	reis (voyage; 5.1)
scheum	haardroger (hairdryer; 4.5)
sliemend	radio (radio; 5.4)
zelvig	brief (letter; 4.8)

Appendix C. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2018.09.015.

References

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language, 48*(1), 1–15.

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,* 390–412.

Bradley, M. M., & Lang, P.J. (1999). Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report C-1, Gainesville, FL. The Center for Research in Psychophysiology, University of Florida.

Brunyé, T. T., Gardony, A., Mahoney, C. R., & Taylor, H. A. (2010). Body-specific representations of spatial location. *Cognition, 123,* 229–239.

Casasanto, D., Brookshire, G., & Ivry, R. (2015). Meaning is not a reflex: Context dependence of spatial congruity effects. *Cognitive Science, 39*(8), 1979–1986.

Casasanto, D., & Dijkstra, K. (2010). Motor action and emotional memory. *Cognition, 115*(1), 179–185.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268.

Crawford, L. E., 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,* 1153–1169.

Danesi, M. (1995). Learning and teaching languages: The role of 'conceptual fluency'. *International Journal of Applied Linguistics, 5,* 3–20.

Gentner, D., & Wolff, P. (2000). Metaphor and knowledge change. In E. Districh, & A. Marbnau (Eds.). *Cognitive dynamics: Conceptual change in humans and machines* (pp. 295–342). Mahwah, NJ: Lawrence Erlbaum Associates.

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition, 2*, 1–19.

Goldin-Meadow, S. (2015). From action to abstraction: Gesture as a mechanism of change. *Developmental Review, 38*, 167–184.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science, 5*(6), 664–674.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes, 24*, 313–334.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Chicago: University of Chicago Press.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135*(4), 638.

Riskind, J. H. (1983). Nonverbal expressions and the accessibility of life experience memories: A congruence hypothesis. *Social Cognition, 2*, 62–86.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1–24.

Schulman, A. I. (1974). Memory for words recently classified. *Memory & Cognition, 2*(1), 47–52.

Stepper, S., & Strack, F. (1993). Proprioceptive determinants of emotional and none-motional feelings. *Journal of Personality and Social Psychology, 64*, 211–220.

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of National Academy of Sciences of the United States of America, 105*(33), 11655–11660.