Vertical motions in the learning of spatial, emotion, and social words

Sho Akamine[1] Tsuyoshi Kohatsu[2] Ayaka Oshiro[3] Manami Sato[3]
[1]Max Planck Institute for Psycholinguistics[2] University of the Basque Country [3]Okinawa International University
sho.akamine@mpi.nl

Abstract

People express and understand abstract concepts in terms of concrete concepts. For example, the abstract concept of "good–bad" is structured via the concrete source domain of "up–down." Recent studies demonstrate that activating a concrete source domain by performing or perceiving vertical motions can enhance the learning of valence words (e.g., joy, misery). This study attempts to replicate these findings and extend them to spatial and social words.

1 Introduction

Theories of embodied cognition propose that individuals' perceptual experiences shape the structure of concepts in their minds [1-7] For instance, according to conceptual metaphor theory [8], abstract concepts (e.g., time) are expressed and constructed in terms of concrete concepts (e.g., money) that individuals have perceived and stored via their interactions with the physical world [2,4,8]. One type of conceptual metaphor is *orientational metaphors*, in which abstract target domains (e.g., positive or negative valence) are structured in terms of the more concrete source domains of spatial orientations (e.g., up and down) [4,8].

Recent findings empirically support cognitive mappings between the source domains of space and motion and the target domain of emotional valence. For example, socially meaningful postures and gestures as well as meaningless upward and downward motor actions can generate positive or negative feelings and memories [9].

Further, a recent study showed that metaphor-congruent motor actions can improve word learning [10]. In the study, 72 L1-Dutch speakers were asked to learn 16 words in an "alien" language (i.e., pseudowords), which were paired with eight positive emotion and eight negative emotion Dutch words, using paper flashcards. After studying a flashcard for six seconds, the participants placed it on either an upper or a lower shelf based on the condition to which they were assigned. For instance, in the metaphor-congruent condition, participants were instructed to place flashcards with positive definitions on the upper shelf and flashcards with negative definitions on the lower shelf; thus, the directions metaphorically associated with the meaning matched the direction of the motor action (i.e., positive meaning–upward motion; negative meaning–downward motion). In contrast, participants assigned to the metaphor-incongruent condition were told to place flashcards with positive meanings on the lower shelf and flashcards with negative meanings on the upper shelf; thus, the directionality was mismatched. Participants in the neutral condition pilled all the flashcards together on their desktop. Each participant performed these actions with each flashcard three times. They then took a two-alternative forced-choice recognition test. The results showed the highest proportion of correct responses in the metaphor-congruent condition, followed by the neutral and then the metaphor-incongruent condition. Based on these results, the researchers suggested that metaphor-congruent motor actions can improve the learning of emotionally charged words, presumably because the metaphor-congruent motor actions activate the source domain of the emotion words, facilitating the process of connecting existing concrete concepts (upward/downward direction) and novel words, whereas metaphor-incongruent motor actions impede such connections [10].

One question raised by Casasanto and de Bruin's study [10] is which parts of a motor action (e.g., motor planning, kinesthetic feedback, and visual feedback) contribute to the learning enhancement effect. Another question is whether the learning enhancement effect they found would apply to the acquisition of words that activate vertical spatial schemas either directly (spatial words) or metaphorically (emotion or social words).

Akamine et al. [11] conducted an experiment to investigate whether the visual perception of vertical motions, either alone or along with the performance of motor action, would improve the learning of valence words. The study found that: (i) accuracy was highest in the metaphor-congruent condition regardless of modality, which is consistent with the previous findings [10]; and (ii) reaction times in the visual- perception-without-motor-actions condition were significantly faster than in the visual-perception-with- motor-actions condition. This unexpected finding might suggest that the motor actions impeded learning, presumably because the actions occupied too much of the working memory capacity that the participants required for their word learning.

To extend the previous findings, the present study tests two hypotheses: (i) visual perception of vertical motions improves the learning of words that activate up/down spatial concepts either metaphorically (emotion and social words) or directly (spatial words), and (ii) the degree of learning enhancement effect does not change for spatial, emotion, and social words.

2 Methodology

2.1 Participants

Fifty-nine (43 female, 16 male) right-handed native Japanese speakers living in Okinawa participated in the experiment. Their age range was 18–27 years (M_{age} = 21.7, SD_{age} = 1.62). The participants received gift cards.

2.2 Word Materials

Pseudowords We created 54 English-based pseudowords by replacing the second or later consonantal syllable onset of basic English words with another consonant (e.g., "orber" based on "order"). Half of the pseudowords consisted of two syllables, and the other half consisted of three.

Associated Meaning Each pseudoword was paired with a real Japanese word as the meaning of the pseudoword. Of the 54 pseudowords, 18 were paired with Japanese translations of emotionally charged English words from the ANEW corpus [12], 18 with social words, and 18 with spatial words. Each word category had two subcategories, one associated with the source domain of UP and the other associated with the source domain of DOWN. Thus, half of the emotionally charged words carried positive meanings (e.g., joy), and the other half carried negative meanings (e.g., war). Likewise, half of the social words had positive meanings, implied by higher social status (e.g., CEO), and the other half had negative meanings, implied by lower social status (e.g., poor). Of the spatial words, half were what we call "up" words, meaning they refer to higher spatial positions (e.g., cloud), and the other half were "down" words, indicating lower spatial positions (e.g., floor). These categories are summarized in Table 1.

Table 1. Summary of associated meanings

Category	Meaning	Source domain	Example
Emotion	Positive	UP	喜び (joy)
	Negative	DOWN	不幸 (misery)
Social	Positive	UP	社長 (CEO)
	Negative	DOWN	無職 (unemployed)
Spatial	Up	UP	雲 (cloud)
	Down	DOWN	床 (floor)

2.3 Experimental Design

Nine conditions were created by manipulating the two independent variables of directional congruency (congruent, incongruent, control) and word category (emotion, social, spatial).

In the experiment, participants saw a series of digital flashcards, each of which, after six seconds, automatically moved until it left the screen and was replaced by the next flashcard. In the congruent condition, the flashcard moved in the same direction as the source domain of the word on the flashcard: upward for positive (emotion/social words) and up (spatial words) meanings, and downward for negative (emotion/social words) and down (spatial words) meanings. In the incongruent condition, the flashcards moved in the opposite direction from the associated meaning: flashcards with positive/up meanings moved downward and ones with negative/down meanings moved upward. In the control condition, the flashcards moved rightward or leftward randomly. The three conditions are represented in Figure 1.

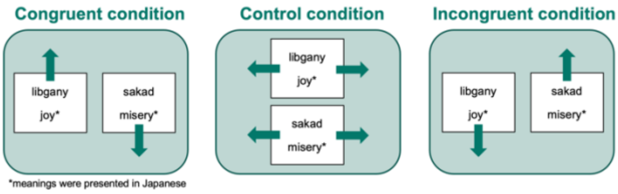

Figure 1. Directional congruency conditions

2.4 Procedure

Learning Phase Participants first completed a learning phase of three blocks, each with six words for each word category, for a total of 18 words per block (i.e., three positive emotion, three negative emotion, three positive social, three negative social, three up spatial, and three down spatial), and a total of 54 words for the entire learning phase. The congruency conditions were assigned to each word such that they were equally distributed within blocks.

Each block of the learning phase presented the 18 words three times. The participants saw a pseudoword with its Japanese translation on a digital flashcard and were asked to study it for six seconds while thinking about whether the word carried a positive/up or negative/down meaning. After six seconds, the flashcards automatically moved upward, downward, or right/leftward depending on the condition. Once the participant had seen all 18 words, the next round began, with the same 18 words presented in a different order. The same procedure was repeated for each of the three sets of 18 words, so that the participants studied all 54 words, three times each, during the learning phase.

Test Phase Fifteen minutes after the learning phase, participants completed the two-alternative forced choice recognition test. In each trial, a fixation cross with a gray background appeared for 500 ms, and then a pseudoword was presented for 3000 ms. Next, they saw another fixation cross with a white background, followed by two side-by-side options for the meaning of the pseudoword; participants selected between the two meanings by pressing F for the left option and J for the right (see Figure 2). The two options were the meaning of the target pseudoword (i.e., correct response) and the meaning of another of the pseudowords that had been in the same block in the learning phase (i.e., incorrect response). The entire experiment was programmed in PsyToolkit [13-14].

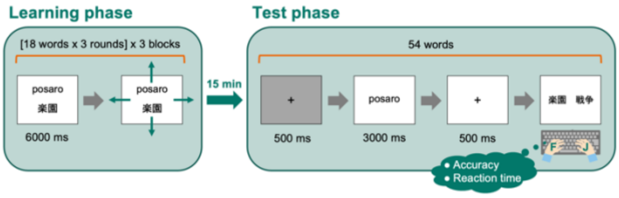

Figure 1. Flow of the experiment

2.5 Analysis

Data Exclusion Data from four participants with a mean accuracy below 68.5% were removed (based on a binomial test with an alpha of 0.01); this procedure resulted in the removal of 216 responses (54 items x 4 participants).

For the reaction time analyses, two timed-out responses (i.e., 10000 ms) and 50 exceedingly fast responses (i.e., below 400 ms) were removed from the data. In addition, 118 responses with reaction times of more than 2.5 *SD* above or below the participant's mean were removed.

As a result, we performed statistical analyses on 2970 responses from 55 participants for accuracy and on 2800 responses for reaction time.

Coding We adopted a treatment coding scheme for coding two categorical independent variables: congruency and word category. For congruency, the control condition was set as the reference level (0). We created two treatment levels that contrasted the reference level with the congruent condition (*ctl_cg*) and with the incongruent condition (*ctl_incg*). Likewise, the spatial word category was set as the reference level for word category, and two treatment levels contrasting the reference level with the emotion condition (*spt_emt*) and with the social condition (*spt_soc*) were created.

Models A mixed-effects regression model and a linear mixed-effects model were performed for accuracy and reaction time, respectively, using *lme4* [15] and the *lmerTest* package [16] in R [17]. Both models included aforementioned four treatment levels as fixed effects, and participants and items as random intercepts. The model structure is summarized below:

Accuracy/RT ~ *ctl_cg* * *spt_emt* + *ctl_cg* * *spt_soc* + *ctl_incg* * *spt_emt* + *ctl_incg* * *spt_soc* + (1|*participant*) + (1|*item*)

3. Results

Accuracy The mean accuracy for each congruency condition across the three word categories is visualized in Figure 3. Figures 3 and 4 use these abbreviations: InCg = incongruent condition; Ctl = control condition; Cg = congruent condition; error bars represent standard errors.

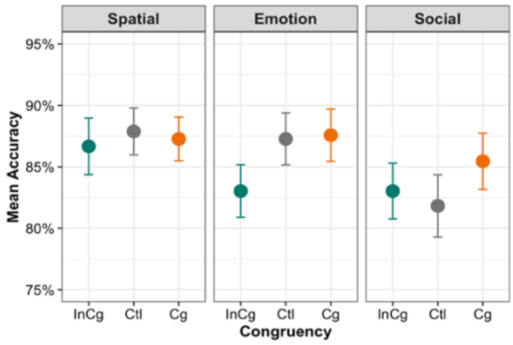

Figure 2. Mean accuracy

A mixed-effects regression model revealed significantly lower accuracy in the control condition for social words (M = 81.8%, SD = 18.8%) than for spatial words (M = 87.9%, SD = 14.1%) [β = -0.50, SE = 0.24, z = -2.07, p = .04]. A power analysis using the simr package [18] with 500 simulations indicated a power of 52.4% (47.9% – 56.9%). No other significant effects were found.

Reaction Time The distribution of reaction times for each congruency condition across the three word categories is plotted in Figure 4. The x in the boxplots represents mean reaction time. A linear mixed-effects model revealed no significant effects.

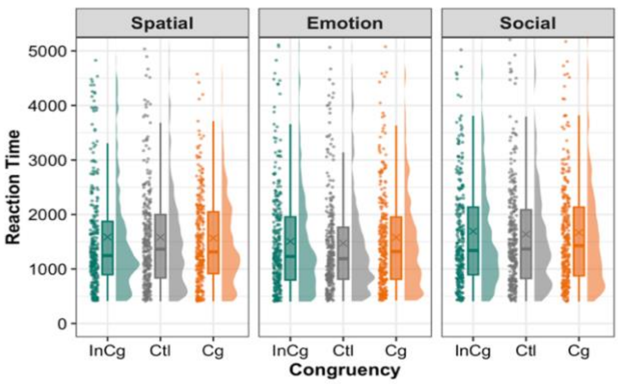

Figure 3. Distribution of reaction time

4. Discussion

This study tested two hypotheses: (i) visual perception of vertical motions improves the learning of words that activate up/down spatial concepts either metaphorically (emotion and social words) or directly (spatial words), and (ii) the degree of learning enhancement effect does not change for spatial, emotion, and social words.

The results did not support the first hypothesis, as there was no significant effect of perceiving congruent vertical motions on either accuracy or reaction times. We speculate two possible explanations.

First, unlike previous studies in which participants completed a task with only one congruency condition (between-subject design) [10] or in three blocks, each associated with one condition (within-subject block design) [11], we used a within-subject randomized block design in which each block included all three conditions. Thus, for instance, participants learned one word in the congruent condition and the next word in the incongruent condition, which may have led to effects that interfered with each other, making the effects less powerful. Supporting this explanation, the mean accuracy for the congruent condition was higher than that for the incongruent condition regardless of word category.

Second, reanalyzing the data from the previous studies using a similar model as in the present study, we did not find significant effects of congruent motions (β = 0.63, SE = 0.44, z = 1.42, p = 0.16 for [10]; β = 0.50, SE = 0.27, z = 1.90, p = 0.06 for [11]). Also, the statistical power (32.0% for [10]; 48.2% for [11]) was inadequate [18–19]. These findings suggest that a between-subject design might be more suitable for investigating effects of visually perceiving congruent or incongruent vertical motions.

As for the second hypothesis, we found mixed results. Although the difference was not statistically significant, the mean accuracy was higher for the congruent condition than for the incongruent condition, suggesting a learning enhancement (interference) effect of perceiving meaning congruent (incongruent) vertical motions. However, this pattern was less prominent for spatial words. This implies that perceiving meaning-congruent/incongruent motions improves/interferes with word learning only when the mapping between the target meaning and the domain of spatial orientations is metaphorical, but more research is needed to further our understanding of this possibility. Lastly, significantly lower accuracy in the control condition for social words than spatial words might indicate that social words are more difficult to learn.

Acknowledgement

This study is supported by KAKENHI #19H01263 (PI: Manami Sato).

References

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

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

[3] Bergen, B. K. (2015). Embodiment. In E. Dabrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 10–30). Walter de Gruyter GmbH & Co KG.

[4] Lakoff, G., & Johnson, M. L. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.

[5] Shapiro, L. A. (2019). *Embodied cognition* (2nd ed.). Routledge/Taylor & Francis Group.

[6] Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156. https://doi.org/10.1111/1467-9280.00326

[7] Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, *4*, 58. https://doi.org/10.3389/fpsyg.2013.00058

[8] Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. University of Chicago Press.

[9] 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

[10] Casasanto, D., & de Bruin, A. (2019). Metaphors we learn by: Directed motor action improves word learning. *Cognition*, *182*, 177–183. https://doi.org/10.1016/j.cognition.2018.09.015

[11] Akamine, S., Omine, A., Kohatsu, T., Niikuni, K., & Sato, M. (2021). Visual perception of vertical motions improves valence word learning. *Conference handbook for the 22nd Annual International Conference of Japanese Society for Language Sciences (JSLS22)*, 79–82.

[12] 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.

[13] Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

[14] Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

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

[16] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

[17] R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

[18] Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

[19] Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00863