COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 42 (2018) 2150–2180

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12568

Spatializing Emotion: No Evidence for a Domain-General Magnitude System

Benjamin Pitt,[a,b] Daniel Casasanto[a,b]

[a]*Department of Psychology, University of Chicago*
[b]*Department of Human Development, Cornell University*

Received 19 October 2016; received in revised form 14 September 2017; accepted 22 September 2017

Abstract

People implicitly associate different emotions with different locations in left-right space. Which aspects of emotion do they spatialize, and why? Across many studies people spatialize emotional valence, mapping positive emotions onto their dominant side of space and negative emotions onto their non-dominant side, consistent with theories of metaphorical mental representation. Yet other results suggest a conflicting mapping of emotional intensity (a.k.a., emotional magnitude), according to which people associate more intense emotions with the right and less intense emotions with the left — regardless of their valence; this pattern has been interpreted as support for a domain-general system for representing magnitudes. To resolve the apparent contradiction between these mappings, we first tested whether people implicitly map either valence or intensity onto left-right space, depending on which dimension of emotion they attend to (Experiments 1a, b). When asked to judge emotional valence, participants showed the predicted valence mapping. However, when asked to judge emotional intensity, participants showed no systematic intensity mapping. We then tested an alternative explanation of findings previously interpreted as evidence for an intensity mapping (Experiments 2a, b). These results suggest that previous findings may reflect a left-right mapping of spatial magnitude (i.e., the size of a salient feature of the stimuli) rather than emotion. People implicitly spatialize emotional valence, but, at present, there is no clear evidence for an implicit lateral mapping of emotional intensity. These findings support metaphor theory and challenge the proposal that mental magnitudes are represented by a domain-general metric that extends to the domain of emotion.

Keywords: ATOM; Conceptual metaphor theory; Emotion; Magnitude; Mental metaphor; Space; Valence

---

[Correction added on December 19, 2017, after first online publication: Corresponding author's email address was changed from casasanto@cornell.edu to casasanto@alum.mit.edu.]

Correspondence should be sent to Daniel Casasanto, Department of Human Development, Cornell University, Martha Van Rensselaer Hall, 116 Reservoir Ave, Ithaca, NY 14853. E-mail: casasanto@alum.mit.edu

1. Introduction

People implicitly associate different emotions with different locations in the left-right space, mapping points along a continuum of emotions onto an imaginary lateral spatial continuum. Which dimensions of emotions do people spatialize, and why? According to one proposal, people spatialize emotions according to their intensity (a.k.a. “emotional magnitude”), associating less intense emotions with the left side of space and more intense emotions with the right (Holmes & Lourenco, 2011; henceforth H&L). This proposed *Intensity Mapping* has been interpreted as evidence for a novel extension of a generalized magnitude system that is hypothesized to underlie the mental representation of quantities (Walsh, 2003). However, as we detail below, such an Intensity Mapping conflicts with another, well-established spatialization of emotion — the *Valence Mapping* — according to which people associate positive emotions with their dominant side of space and negative emotions with their non-dominant side (Casasanto, 2014, for review). Here, we test both the Valence and Intensity Mappings and suggest a resolution to the apparent contradiction between them.

1.1. Evidence for a lateral mapping of emotional valence

According to many studies, people implicitly spatialize emotional valence, mapping positive and negative emotions onto their dominant and non-dominant sides, respectively. Right-handers, therefore, tend to associate positive emotions (e.g., happiness) with the right side of space, and negative emotions (e.g., anger) with the left (Casasanto, 2011, 2014). This Valence Mapping has been observed across a variety of tasks and populations. For example, right- and left-handers express different preferences for items located on their left and right: choosing which of two products to buy, which of two job applicants to hire, which person to date or to befriend, determining which of two alien creatures looks more honest, and even deciding which of two political candidates to vote for (Casasanto, 2009; Kim, Krosnick, & Casasanto, 2015; Zhao et al., 2016). In all of these cases, right-handers tended to prefer the product, person, or creature presented on their right side, but left-handers tended to prefer the one on their left. This pattern persisted even when people made judgments orally, without using their hands to respond (Casasanto, 2009; Casasanto & Chrysikou, 2011; de la Fuente, Casasanto, Martínez-Cascales, & Santiago, 2017; de la Fuente, Casasanto, & Santiago, 2015). Children as young as 5 years old already make evaluations according to handedness and spatial location, judging animals shown on their dominant side to be nicer and smarter than animals on their non-dominant side (Casasanto & Henetz, 2012).

The Valence Mapping appears to be unaffected by variations in cultural practices and attitudes. The same Valence Mapping found in U.S. and European participants has also been found in Moroccan participants, who read Arabic (which is written from right to left), and who have strong cultural taboos associated with the left (de la Fuente, Casasanto, Román, & Santiago, 2015). A body-based Valence Mapping is found both in neurotypicals and in patients with impaired motor systems. Unilateral stroke patients who have lost the fluent use of one side of the body tend to associate "good" with their functionally dominant side of space (Casasanto & Chrysikou, 2011).

The implicit Valence Mapping influences people's memory and their motor actions, as well as their judgments. In one experiment, participants were shown the locations of fictitious positive and negative events on a map, and they were asked to recall the locations later. Memory errors were predicted by the valence of the event and the handedness of the participant: Right-handers were biased to locate positive events too far to the right and negative events too far to the left on the map, whereas left-handers showed the opposite biases, and the amount of error was predicted by the strength of participants' handedness (Brunyé, Gardony, Mahoney, & Taylor, 2012). Beyond the laboratory, the association of "good" with the dominant side can be seen in left- and right-handers' spontaneous speech and gestures. In the final debates of the 2004 and 2008 U.S. presidential elections, positive speech was more strongly associated with right-hand gestures and negative speech with left-hand gestures in the two right-handed candidates (Bush, Kerry), but the opposite association was found in the two left-handed candidates (McCain, Obama; Casasanto & Jasmin, 2010).

Reaction time (RT) tasks have also shown lateral space-valence congruity effects, and they are of particular relevance to this study. Across multiple experiments, right- and left-handers were faster to classify centrally presented words as positive when responding with their dominant hand, and faster to classify words as negative when responding with their non-dominant hand (Kong, 2013; Song, Chen, & Proctor, 2017; de la Vega, De Filippis, Lachmair, Dudschig, & Kaup, 2012; de la Vega, Dudschig, De Filippis, Lachmair, & Kaup, 2013). A similar pattern was found when people judged positive and negative emotional faces with the dominant and non-dominant hands (Kong, 2013). Together, results of more than two-dozen experiments conducted across 10 laboratories converge to provide strong and generalizable evidence for an implicit left-right Valence Mapping, which has been shown in children and adults, across six languages and cultures, using a variety of methods (e.g., questionnaires, diagram tasks, memory tasks, spontaneous gestures, RT tasks) and a variety of materials (e.g., words, maps, drawings, faces).

The Valence Mapping is an example of a "mental metaphor" (see Casasanto, 2009, 2016; Lakoff & Johnson, 1980): an implicit association in long-term memory between two analog continuums, the first serving as a concrete *source domain* (i.e., space) and the second as a relatively abstract *target domain* (i.e., emotional valence). By importing the inferential structure of source domains like space into the target domain, mental metaphors allow people to represent and reason about abstract domains like time (e.g., Bonato, Zorzi, & Umilta, 2012, for review) and number (e.g., Wood, Willmes, Nuerk, & Fischer, 2008; for review), as well as emotion. According to theories of metaphorical mental representation, these mappings are constructed on the basis of experience with the physical and social environment. In the case of the Valence Mapping, associations between lateral space and emotional valence appear to arise from the physical experiences we have with our hands: People come to associate positive emotions with the side of space on which they can usually act more fluently with their dominant hand, and negative emotions with the side on which they act more clumsily, with the non-dominant hand (Casasanto & Chrysikou, 2011;

see also Ping, Dhillon, & Beilock, 2009; Oppenheimer, 2008; Reber, Schwarz, & Winkielman, 2004). In right-handers, this lopsided experience gives rise to a mental metaphor linking positive emotions with the right and negative emotions with the left.

1.2. Evidence for a lateral mapping of emotional intensity

According to one study, people spatialize emotional intensity (which the authors refer to as "emotional magnitude"), associating less intense emotions with the left and more intense emotions with the right (Holmes & Lourenco, 2011). Participants responded to photographs of emotional faces that varied in both valence (e.g., happy vs. angry) and intensity (e.g., happy vs. extremely happy). As indexed by their RTs, participants appeared to associate less intense emotions with the left and more intense emotions with the right — regardless of whether the emotional valence was positive or negative (H&L, Experiment 2). These data were interpreted as support for a generalized magnitude system for representing time, space, quantity, and other mental magnitudes (Walsh, 2003; for critiques, see Casasanto, Fotakopoulou, & Boroditsky, 2010; Merritt, Casasanto, & Brannon, 2010; Bottini & Casasanto, 2013). Specifically, these findings were interpreted as evidence for an extension of a generalized magnitude system to a "hyper-general system of magnitude representation" that is "so abstract as to encompass even socio-emotional cues" (Holmes & Lourenco, 2011, p. 316).

H&L present data consistent with the proposal that people implicitly associate emotional intensity with left-right space: an Intensity Mapping. However, these data conflict with the Valence Mapping. However, an Intensity Mapping predicts that very negative emotions (e.g., extremely angry) should be associated with the *right* (because they are more intense), the Valence Mapping predicts that these negative emotions should be associated with the *left*, in right-handers (because they have negative valence). More broadly, according to an Intensity Mapping, very positive and very negative emotions should be on the same side of space, but according to the Valence Mapping they should be on opposite sides of space in people's minds.

1.3. Is there a hyper-general magnitude system that extends to emotions?

Here, we conducted four experiments to resolve this apparent contradiction, and to evaluate the strength of the evidence for an Intensity Mapping — and by inference, for a hyper-general system of magnitude representation that extends to the domain of emotion. In Experiments 1a, b, we explored the possibility that people have *both* a Valence Mapping and an Intensity Mapping in long-term memory, but that at any moment only one of these contradictory mappings is activated strongly enough to produce the predicted pattern of behavior (see Casasanto & Bottini, 2014). We tested whether people spatialize emotion according to whichever emotional dimension is more salient, depending on the context (broadly consistent with suggestions of H&L's). To preview our results, we found evidence for the Valence Mapping but no evidence for an Intensity Mapping. Therefore, in Experiments 2a, b, we tested an alternative explanation for H&L's findings based on the *spatial magnitude* (as opposed to the *emotional magnitude*) of a salient feature of their stimuli, in order to determine whether previous findings should be interpreted as evidence for an Intensity Mapping.

2. Experiment 1a: Do people spatialize whichever dimension of emotion they attend to?

Currently, the available evidence seems to support two lateral spatializations of emotion, one based on valence and the other based on intensity. Although Valence Mappings and Intensity Mappings conflict, it is possible that people spatialize emotion according to either valence or intensity flexibly, depending on which dimension is more salient, due to characteristics of the stimuli or the task. In the stimuli that H&L used, emotional intensity may have been more salient than valence. They used photographs that depicted happy and angry facial expressions at four levels of intensity, ranging from "neutral" to "extreme" (Tottenham et al., 2009). Thus, within the same category of valence (e.g., within "happy"), there were gradations in intensity. Perhaps the construction of these stimuli directed participants' attention to the dimension of emotional intensity more than to valence, causing them to spatialize emotion according to intensity rather than valence.

In Experiment 1a, we sought a simple solution to the apparent conflict between the Valence Mapping and an Intensity Mapping. We tested whether orienting participants toward either valence or intensity would cause them to show the corresponding spatial mapping. Participants responded to emotional words and judged them according to either their valence or their intensity, by pressing buttons on the left or right of a keyboard. Although H&L used faces as stimuli, we reasoned that words would allow us to orient participants' attention to different dimensions of emotion more easily than faces would. However, the emotional connotations of faces have been argued to be automatic (Tracy & Robins, 2008) and largely invariant (Dolan, 2002; Smith, Cottrell, Gosselin, & Schyns, 2005; cf. Barrett, Mesquita, & Gendron, 2011), the connotations of words are notoriously variable (Casasanto & Lupyan, 2015; Wittgenstein, 1953). Thus, different aspects of words' meanings can be highlighted flexibly, depending on the task context. Additionally, using words as stimuli rather than faces allowed us to avoid the confound between spatial magnitude and emotional intensity found in H&L's stimuli, which we describe in section 4.

2.1. Method

2.1.1. Participants

Thirty-two right-handed adults from the University of Chicago community who spoke fluent English participated in the main experiment for payment or course credit, after giving informed written consent. Half were randomly assigned to make speeded valence judgments ($n = 16$) and the other half to make speeded intensity judgments ($n = 16$). Twenty-four other participants from the same community completed a questionnaire used to measure normative judgments of the stimulus words' meanings.

2.1.2. Materials

Four emotional words were selected, on the basis of a previous experiment in French speakers (Carbè & Gevers, 2013): “horrible” (very negative), “bad” (negative), “good” (positive), and “perfect” (very positive). These words yielded a crossing of Emotional Valence (positive, negative) with Emotional Intensity (more intense, less intense). This four-item fully crossed design is analogous to the many space-number association experiments that test for left-right mappings of two large numbers (e.g., 8, 9) and two small numbers (e.g., 1, 2), crossed with the numbers’ parity (odd, even; e.g., Fischer, Castel, Dodd, & Pratt, 2003; Gevers, Lammertyn, Notebaert, Verguts, & Fias, 2006; Lindemann, Abolafia, Pratt, & Bekkering, 2008).

In order to quantify differences among the stimulus words, we asked a group of participants ($N = 24$) who did not participate in the main experiment to rate the four words on intensity and valence. For this norming task, the words appeared in a vertical column along the midline of a printed page. Each participant judged the emotional valence of the words on a scale from -5 (most negative) to 5 (most positive) and, on another page, the emotional intensity of the words on a scale from 0 (least intense) to 10 (most intense). They wrote their ratings on a horizontal line to the right of each word, all using their right hand. The position of the words on the page and the order of valence and intensity judgments was counterbalanced across participants.

2.1.3. Procedure

In both the valence and intensity RT tasks, the four stimulus words appeared one at a time in black text in the center of a computer screen. Participants in the valence task judged whether each word expressed a “positive emotion” or “negative emotion.” Participants in the intensity task judged whether each word expressed a “more intense emotion” or “less intense emotion.” Participants were instructed to respond as quickly and accurately as possible by pressing one of two keys (the “*z*” and the “/” keys on the English-US QWERTY keyboard) using the index finger of each hand. If no response was given, trials automatically ended after 2-s. In one block of trials, participants pressed the key on the left to indicate a negative/less intense emotion and the key on the right to indicate a positive/more intense emotion. In a second block, this mapping was reversed, and the order of blocks was counterbalanced across participants. Participants performed eight practice trials at the start of each block, after which the four stimulus words were presented in random order 24 times, composing a total of 192 critical trials per participant over two blocks.

After testing, participants were debriefed to determine whether they were aware of the purpose of the experiment, and they then completed a language history questionnaire and the Edinburgh Handedness Inventory (EHI; Oldfield, 1971).

2.2. Results

2.2.1. Valence and intensity ratings

To allow direct comparison across tasks, ratings were standardized for each rater and each task. The four stimulus words varied significantly in their valence ratings (F(3,

69) = 112.30, $p < .00001$) and their intensity ratings ($F(3, 69) = 31.92$, $p < .00001$). Mean standardized valence ratings ranged from "horrible" (−1.08 ± 0.05) to "bad" (−0.40 ± 0.11), "good" (0.56 ± 0.07), and "perfect" (0.92 ± 0.10). Mean standardized intensity ratings ranged from "good" (−0.77 ± 0.12), the least intense, to "bad" (−0.54 ± 0.11), "perfect" (0.57 ± 0.15), and "horrible" (0.73 ± 0.15). There was no significant difference between the range of the valence ratings ($M = 2.14 \pm 0.03$) and the range of the intensity ratings ($M = 2.07 \pm 0.05$; $t(23) = 1.13$, $p = .27$).

2.2.2. Accuracy

Four subjects failed to follow instructions and were replaced. The error rate was numerically higher in the valence task (5.6%) than in the intensity task (4.3%; difference = 1.3%), but this difference was not significant ($\chi^2(1) = 1.72$, $p = .19$).[1] Inaccurate trials were excluded from the RT analyses.

2.2.3. RT analyses

To evaluate and compare the Valence Mapping and the Intensity Mapping, we used the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R (R Development Core Team, 2017). RTs were first log transformed to approximate a normal distribution of residuals and then entered into a mixed effects model with response hand and standardized valence or intensity ratings as predictors, and with random slopes and intercepts for participants. The Valence Mapping was indexed by the interaction between response hand and standardized valence ratings in the valence task, and the Intensity Mapping was indexed by the interaction between response hand and standardized intensity ratings in the intensity task.

For comparison with H&L's findings, we also report the Valence and Intensity Mappings as subject-specific regression slopes, regressing dRT values (dRT = right-hand – left-hand RT) for each word over its standardized valence or intensity rating.[2]

These regression slopes also allowed us to address a secondary question. Although our primary aim was to contrast spatial mappings of valence and intensity, these data can also address the question: Are space-emotion mappings categorical or continuous? The answer bears on the status of these spatial mappings as mental metaphors, which are hypothesized to be continuous (see General Discussion; see also Casasanto & Gijssels, 2015; Freddi, Brouillet, Cretenet, Heurley, & Dru, 2016). We used nonparametric tests on the subject-specific regression slopes to determine whether any observed Valence Mapping and Intensity Mapping were categorical or continuous.

2.2.4. RTs: Valence judgments

Following H&L, RTs greater than 2.5 standard deviations from subject means were removed (2.69% of accurate responses). The RTs of remaining trials had a mean of 577 ms and a standard deviation of 159 ms.

Of primary interest, the mixed-effects model of RTs revealed a significant Valence Mapping ($\chi^2(1) = 5.76$, $p = .02$; Fig. 1, left); participants associated negatively valenced words with left space and positively valenced words with right space.

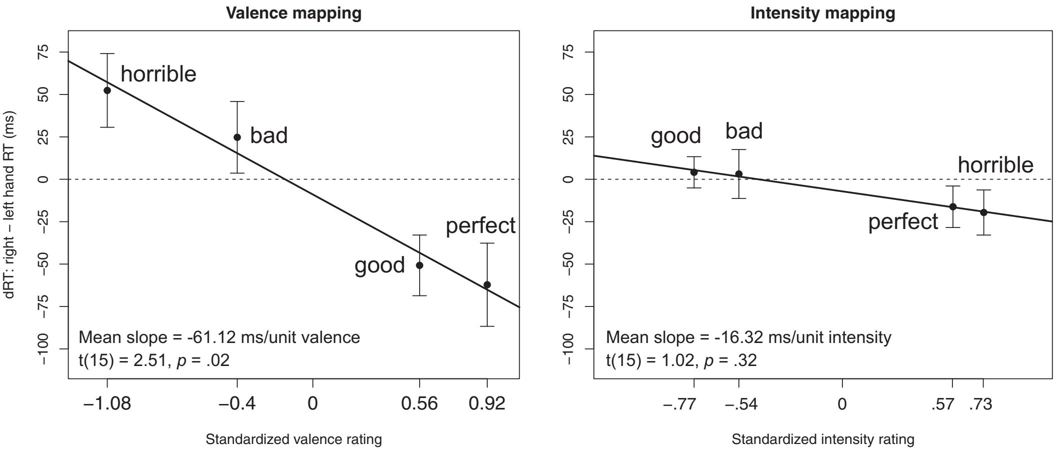

Fig. 1. Results of Experiment 1a. Left: Significant Valence Mapping in the valence task. Right: Non-significant Intensity Mapping in the intensity task. Error bars show the standard error of the mean.

To determine whether this Valence Mapping was categorical or continuous, we used Spearman's Rho to obtain a rank-order correlation coefficient for each participant, which indexes the degree to which their dRTs varied monotonically with valence (inferences about monotonicity required a rank-order test, not an lmer). Rho values were standardized using a Fisher's Z transformation, and Z-values were tested against zero by one-sample t-test ($t(15) = 2.66$, $p = .02$), confirming the results of the parametric analysis reported above and providing evidence that the mapping between valence and space was continuous.

2.2.5. RTs: Intensity judgments

RTs > 2.5 standard deviations from subject means were removed (2.52% of accurate responses). The RTs of the remaining trials had a mean of 547 ms and a standard deviation of 146 ms, and they did not differ significantly from RTs in the valence task ($\chi^2(1) = 1.41$, $p = .23$).

Of primary interest, in the mixed-effects model of RTs, the interaction between response hand and intensity ratings did not approach significance ($\chi^2(1) = 0.80$, $p = .37$; Fig. 1, right), indicating that participants in the intensity task did not show an Intensity Mapping.

In the nonparametric test, the mean Z-value did not differ from zero ($t(15) = 1.33$, $p = .20$), confirming the results of the parametric analysis, and showing no evidence of an Intensity Mapping.

2.2.6. RTs: Comparison of valence and intensity effects

To test the difference between the significant Valence Mapping observed in the valence task and the non-significant Intensity Mapping observed in the intensity task, we used a linear mixed-effects model on log-transformed RTs with response hand, standardized valence and intensity ratings, and task (valence or intensity) as predictors, and with random slopes and intercepts for participants (Barr, Levy, Scheepers, & Tily, 2013). The three-way interaction between response hand, ratings, and task was not significant ($\chi^2(1) = 2.06$, $p = .15$).

2.3. Discussion

When judging the emotional valence of English words, participants associated negative valence with the left and positive valence with the right, showing the Valence Mapping typical of right-handers (Casasanto, 2014). This Valence Mapping appears to be continuous: The most negative word (horrible) was associated most strongly with left-hand responses, the most positive word (perfect) was associated most strongly with right-hand responses, and dRTs were intermediate for words with intermediate valences (bad, good; see also Freddi et al., 2016). By contrast, when judging the emotional intensity of the same words, the same participants did not systematically associate different levels of emotional intensity with different spatial locations. The relationship between response hand and intensity did not approach significance, providing no support for an implicit spatial mapping of intensity (cf., Holmes & Lourenco, 2011).

Before accepting the conclusion that people did not spatialize emotional intensity, even when they were making explicit intensity judgments, first it is important to consider whether participants were, in fact, judging the intensity of the words' referents reliably. For instance, if the intensity task were more difficult than the valence task, this difference could cause intensity judgments to be more variable and less likely to yield a reliable effect. However, an examination of the accuracy and RT data suggests that the intensity task was not, in fact, more difficult than the valence task. Participants had high accuracy in both tasks (above 95%), and their accuracy was numerically *higher* in the intensity task than in the valence task. Likewise, on average, participants were 30 ms faster in the intensity task than in the valence task. These differences were not statistically significant (*ps* > .19), but if these speed and accuracy data tell us anything about relative task difficulty, they suggest that the valence task was the harder task.

Next, it is important to evaluate whether the task instructions affected how people spatialized emotion implicitly; that is, whether our task manipulation was effective in changing the way people spatialized emotion. An alternative possibility is that participants in the intensity task could have spatialized the words according to their *valence*, implicitly, even when the instructions required them to make intensity judgments, explicitly. A further analysis rules out this possibility. The Valence Mapping was significantly stronger for participants who made explicit valence judgments (in the valence task) than for participants who made explicit intensity judgments (in the intensity task; $\chi^2(1) = 6.25$, $p = .01$). There was no evidence of an implicit mapping of valence during the intensity task ($\chi^2(1) = 0.02$, $p = .88$). Thus, the task instructions were successful in causing participants to spatialize emotion differently in the valence versus intensity tasks, but orienting their attention to emotional intensity did not cause them to spatialize this dimension of emotion.

Finally, in principle, our ability to detect an Intensity Mapping could be affected by the range of intensity ratings: If the range of ratings was more restricted for intensity than for valence, it could be more difficult to detect an Intensity Mapping than a Valence Mapping. As reported above, however, the difference between the valence and intensity ranges did not approach statistical significance. We address these concerns further in Experiment 1b.

To summarize the results of Experiment 1a, the accuracy and RT data confirm that, during the valence task, participants were accurately judging valence explicitly and continuously spatializing valence implicitly. By contrast, during the intensity task, although participants were accurately judging intensity explicitly they were not spatializing intensity implicitly.

3. Experiment 1b: A second test of valence and intensity mappings

In Experiment 1a, we predicted that if people implicitly associate both valence and intensity with left-right space, then participants would (i) spatialize valence when making speeded valence judgments (consistent with de la Vega et al., 2012, 2013; Kong, 2013), and (ii) spatialize emotional intensity when making speeded intensity judgments (consistent with the conclusions of Holmes & Lourenco, 2011). Only the first prediction was upheld. The goal of Experiment 1b was to test the same predictions with greater power, generalizability, and sensitivity.

To increase power, we doubled the number of participants and octupled the number of items; increasing the sample of items from 4 to 32 also increased the generalizability of the results. To increase sensitivity, we collected word ratings from each participant, after the RT task, and predicted dRTs on the basis of these subject-specific word ratings.

3.1. Method

3.1.1. Participants

Sixty-four right-handed adults from the University of Chicago community participated for payment or course credit after giving informed written consent. Half were randomly assigned to the valence task (n = 32) and the other half to the intensity task (n = 32).

3.1.2. Materials

The set of four emotional words from Experiment 1a was expanded to include 32 words for which valence (positive, negative) was crossed with emotional intensity (high intensity, low intensity): *accepted, aloof, attractive, bad, beloved, brilliant, capable, courteous, determined, devastated, disliked, drowsy, ecstatic, energized, exhausted, good, gorgeous, gregarious, hated, hesitant, hideous, horrible, idiotic, ignorant, insulting, obstinate, perfect, pleased, prepared, rested, unattractive, unhappy.*

3.1.3. Procedure

The procedure was identical to that of Experiment 1a except for the following changes. Participants performed 32 practice trials at the start of each block (one for each word), after which these words were presented in random order five times, composing 320 critical trials over two blocks. After this speeded RT task, participants rated the valence and intensity of each word. Words appeared on the computer screen one at a time in randomized order and participants spoke aloud their numerical ratings, on the same scales used in Experiment 1a. The order of the valence rating and intensity rating tasks was counterbalanced across participants.

3.2. Results

3.2.1. Valence and intensity ratings

To allow direct comparison across tasks, ratings were standardized for each subject and each task. The 32 stimulus words varied significantly in their valence ratings ($F(31, 9797) = 1738$, $p < .00001$) and their intensity ratings ($F(31, 9753) = 438.7$, $p < .00001$). The range of the raters' intensity ratings ($M = 3.38 \pm 0.06$) was significantly greater than the range of their valence ratings ($M = 3.06 \pm 0.05$ *SD*; $t(63) = 3.90$, $p = .0002$).

3.2.2. Accuracy

In Experiment 1a, valence and intensity ratings from an independent group of participants determined the valence and intensity of the four stimulus words, allowing easy identification of error trials. In Experiment 1b, by contrast, the ratings of the valences and intensities of the 32 stimulus words were specific to the individual participants in the RT tasks. This allowed us to accommodate idiosyncratic differences in word meaning across individuals but precluded any straightforward identification of "errors." However, we can report the rates at which participants gave no response in the RT task (i.e., missed trials). The miss rate was very small (less than 1%) overall, and it did not differ between the valence task (0.55%) and the intensity task (1.30%; Difference = 0.75%, $\chi^2(1) = 2.60$, $p = .11$).

3.2.3. RT analyses

To evaluate and compare the Valence Mapping and the Intensity Mapping, we used the lme4 package (Bates et al., 2014) in R (R Core Team, 2014), as in Experiment 1a. RTs were first log transformed to approximate a normal distribution of residuals and then entered into a mixed effects model with response hand and standardized word ratings (i.e., subject-specific valence or intensity ratings) as predictors, and with random slopes and intercepts for participants and words. The Valence Mapping was indexed by the interaction between response hand and standardized valence ratings and the Intensity Mapping was indexed by the interaction between response hand and standardized intensity ratings.

For comparison with H&L's findings, we also report the Valence and Intensity Mappings as subject-specific regression slopes, regressing dRT values (dRT = right-hand – left-hand RT) for each word over its standardized valence or intensity rating. In a secondary analysis, these regression slopes also allowed us to conduct nonparametric tests to determine whether any Valence Mapping or Intensity Mapping observed was categorical or continuous.

3.2.4. RTs: Valence judgments

Missed trials (0.55%) and trials with RTs greater than 2.5 standard deviations from subject means (3.2%) were excluded from RT analyses. The RTs of the remaining trials had a mean of 702 ms and a standard deviation of 180 ms.

Of primary interest, the mixed-effects model of RTs revealed a highly-significant Valence Mapping ($\chi^2(1) = 13.95$, $p = .0002$; Fig. 2, left); participants associated negatively valenced words with left space and positively valenced words with right space.

In a nonparametric test, we used Spearman's Rho to obtain a rank-order correlation coefficient for each participant. Rho values were standardized using a Fisher's *Z* transformation, and *Z*-values were tested against zero by one-sample *t*-test ($t(31) = 2.47$, $p = .02$), confirming that the mapping between valence and space was continuous.

3.2.5. RTs: Intensity judgments

Missed trials (1.30%) and trials with RTs greater than 2.5 standard deviations from subject means (2.9%) were excluded from RT analyses. The RTs of remaining trials had a mean of 767 ms and a standard deviation of 240 ms, and they were significantly slower than Valence judgments (Difference = 65 ms; $\chi^2(1) = 4.44$, $p = .04$; for discussion, see section 3.3.5).

Of primary interest, the mixed-effects model of RTs, the interaction between response hand and intensity ratings did not approach significance ($\chi^2(1) = 0.36$, $p = .55$; Fig. 2, right), providing no evidence of an Intensity Mapping.

In the nonparametric test, the mean *Z*-value did not differ from zero ($t(31) = 0.71$, $p = .48$), confirming the null effect found in the parametric analysis, and showing no evidence of an Intensity Mapping.

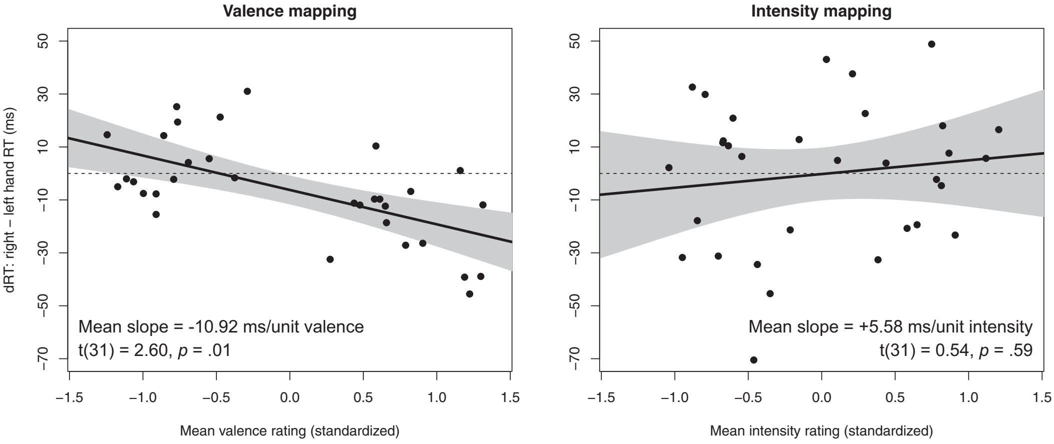

Fig. 2. Results of Experiment 1b. Left: Significant Valence Mapping in the valence task. Right: Non-significant Intensity Mapping in the intensity task. Dots show mean rating and dRT values for each word. Shaded areas indicate 95% confidence intervals around the lines of best fit.

3.2.6. RTs: Comparison of valence and intensity effects

To test the difference between the significant Valence Mapping observed in the valence task and the non-significant Intensity Mapping observed in the intensity task, we used a linear mixed-effects model on log-transformed RTs with response hand, standardized word ratings, and task as predictors, and with random slopes and intercepts for participants and words. The three-way interaction between response hand, word rating, and task was significant, indicating that the Valence Mapping was reliably stronger than the Intensity Mapping ($\chi^2(1) = 8.03$, $p = .005$).

3.3. Discussion

The results of Experiment 1b replicated the results of Experiment 1a, extending them to a larger sample of participants and a larger set of words: We find a significant Valence Mapping but no significant Intensity Mapping. Before accepting the conclusion that people implicitly spatialize emotional valence on the lateral axis but not emotional intensity, first it is important to consider potential alternative explanations for the pattern of data found in Experiments 1a, b.[3]

3.3.1. Power

Could a lack of statistical power explain the null effects of intensity in Experiments 1a, b? In Experiment 1a the effect of emotional intensity trended slightly in the predicted direction, raising the possibility that a more powerful test would reveal a significant Intensity Mapping. Yet this possibility was not borne out in Experiment 1b, which included twice as many subjects and eight times as many items as Experiment 1a. Rather than showing a stronger effect in the predicted direction, Experiment 1b showed a non-significant trend in the opposite direction from the trend we found in Experiment 1a, and from what H&L reported.

3.3.2. Generalizability

The small number of stimuli used in Experiment 1a left open the possibility that, for unknown reasons, those particular words were not appropriate for eliciting an Intensity Mapping. The larger stimulus set in Experiment 1b presumably constitutes a more representative sample of emotional words in English.

3.3.3. Specificity

In Experiment 1a, we analyzed RTs on the basis of averaged valence and intensity norms collected in an independent sample of participants. It is possible that the normative ratings of words' intensities did not match the subjective ratings of the individual participants in the RT experiment, which could cause the normative ratings to be a poor predictor of subject-specific RTs. Using subject-specific word ratings in Experiment 1b alleviated this concern.

3.3.4. Ranges of ratings

Our ability to detect an Intensity Mapping could be affected by the range of intensity ratings. In Experiment 1a, although there was no statistical difference between the ranges of the valence and intensity ratings, the mean range of participants' intensity ratings was slightly smaller than the mean range of their valence ratings (about 3% smaller on average). In principle, a greater range in ratings could have made it easier to detect a Valence Mapping than an Intensity Mapping in Experiment 1a. Yet in Experiment 1b the difference in ranges was reversed. As reported above, the range of the intensity ratings was significantly *greater* than the range of the valence ratings. No matter whether the range of the intensity ratings was greater or lesser than the range of the valence ratings, the Valence Mapping was significant and the Intensity Mapping was non-significant.

3.3.5. *Reliability of intensity judgments*

In principle, our ability to detect an Intensity Mapping could be affected by the reliability with which participants judged the relative intensities and relative valences of the words. Specifically, if participants found it more difficult to judge the relative intensities of the stimulus words than to judge their relative valences, their intensity judgments could be more variable and therefore less likely to show reliable RT effects. However, upon examination of the accuracy and RT data from Experiments 1a, b, overall there is no evidence that intensity judgments were more difficult than valence judgments (sections 3.3.5.1–3.3.5.2). A further analysis demonstrates that participants were judging intensity reliably (section 3.3.5.3).

3.3.5.1. Accuracy: If accuracy were significantly lower in the intensity task than in the valence task, this would support the suggestion that intensity judgments were more difficult than valence judgments. This was not the case; accuracy in Experiment 1a was over 95% overall, and it was numerically *higher* in the intensity task (96%) than in the valence task (94%; see section 2.2.2). The miss rate in Experiment 1b was very low overall (about 1%), and it did not differ significantly across tasks (see section 3.2.2). When the data from both experiments were combined, there was no significant difference in the miss rate across tasks, whether considering all stimulus words ($\chi^2(1) = 2.01$, $p = .16$), or considering only the four words that were used in both Experiments 1a, b ($\chi^2(1) = 1.64$, $p = .20$).

3.3.5.2. RTs: If RTs were significantly slower in the intensity task than in the valence task, this would support the suggestion that intensity judgments were more difficult than valence judgments. Yet there was no consistent difference in RTs across tasks. In Experiment 1a, RTs were numerically *faster* in the intensity task than in the valence task (see sections 2.2.4–2.2.5), whereas in Experiment 1b, this pattern reversed (see sections 3.2.4–3.2.5). When the data from both experiments were combined, there was no significant difference in RTs across tasks, whether considering all words ($\chi^2(1) = 0.31$, $p = .58$) or the original four words only ($\chi^2(1) = 0.51$, $p = .48$).

3.3.5.3. Symbolic distance effect: We tested the extent to which participants reliably judged the relevant dimensions of the stimulus words using the *symbolic distance effect* (Moyer & Landauer, 1967). In this effect, people are faster to judge larger differences (e.g., 1 vs. 5) than smaller differences (e.g., 4 vs. 5). This effect is orthogonal to our effect of interest, and although the word "distance" appears in its name, the symbolic distance effect does not imply any spatial mapping (Herrera, Macizo, & Semenza, 2008; de Hevia, Girelli, & Vallar, 2006). Although the symbolic distance effect has been tested most often on concurrent pairs of stimuli, it can be also be evaluated for serially presented stimuli like the emotional words in Experiment 1b (see Ginsburg, van Dijck, Previtali, Fias, & Gevers, 2014; Experiment 1b included enough items to calculate the symbolic distance effect meaningfully, but Experiment 1a did not).

If participants in the intensity task were not reliably judging the relative intensities of the stimulus words, then there should be no symbolic distance effect in the intensity task in Experiment 1b. We evaluated the symbolic distance effect in the valence and intensity tasks using a mixed-effects model in which log-transformed RTs were predicted by response hand and the absolute value of participants' standardized valence or intensity ratings, with random slopes and intercepts for subjects and words. Participants showed a significant distance effect in both the valence task ($\chi^2(1) = 6.49$, $p = .01$) and the intensity task ($\chi^2(1) = 14.04$, $p = .0002$); they were faster to respond to words with extreme valence or intensity ratings than to words with intermediate ratings. This finding indicates that participants reliably judged both the relative valences and the relative intensities of the stimulus words, and it shows that relative intensity was systematically influencing RTs. The distance effect was numerically stronger in the intensity task, suggesting that the relative intensities of words were judged at least as reliably as their relative valences. Therefore, the absence of an Intensity Mapping in Experiment 1b cannot be attributed to participants failing to judge the words' intensities reliably.

To summarize these points, there is no clear evidence that the intensity task was more difficult than the valence task, from either the accuracy or the RT data. Moreover, the presence of a highly significant symbolic distance effect in the intensity task (that was, if anything, stronger than in the valence task) confirms that participants in the intensity task were reliably judging the relative emotional intensities of the stimulus words, even though they were not reliably spatializing them.

3.3.6. Summary of findings from Experiment 1

The results of Experiments 1a, b showed a highly significant Valence Mapping, consistent with previous results (Casasanto, 2014, for review), but they contrast with H&L's results, which were interpreted as showing an Intensity Mapping. In Experiment 2, we tested a possible explanation for this apparent contradiction.

4. Experiment 2: Emotional or spatial magnitude?

Why did H&L find an Intensity Mapping whereas we did not? The answer to this question bears on whether there is, in fact, any evidence for a "hyper-general system of magnitude representation" that extends to the domain of emotion. A potential answer comes from an examination of H&L's stimuli. For the two experiments in which H&L

reported an Intensity Mapping (their Experiments 1 and 2), the stimuli were photographs of six actors making facial expressions that varied in their emotional intensity (what H&L called emotional magnitude): neutral, happy, very happy, extremely happy, etc. When classifying the gender of these faces using left- and right-hand response keys, participants showed a pattern of RTs that was consistent with an Intensity Mapping.

However, as acknowledged by H&L, these findings might also be “due to specific facial features that varied across the range of stimuli, rather than the magnitude of emotional expression per se” (p. 318). One such feature was the “size of the mouth; more happy faces had a larger mouth opening than less happy faces” (p. 318).

Although H&L dismissed this confound between emotional intensity and spatial magnitude, it is of concern for two reasons. First, across cultures, people tend to focus on the mouths of emotional faces (Jack, Blais, Scheepers, Schyns, & Caldara, 2009; Koda & Ruttkay, 2014; Smith et al., 2005; cf., Blais, Jack, Scheepers, Fiset, & Caldara, 2008), and this tendency is particularly strong in U.S. participants (Jack, Caldara, & Schyns, 2012; Yuki, Maddux, & Masuda, 2007). Second, and crucially, previous data show that people implicitly spatialize *area* along a left-to-right continuum. In one experiment, participants were faster to make speeded judgments on smaller circles with the left hand and larger circles with the right hand (Ren, Nicholls, Ma, & Chen, 2011), producing a dRT effect much like H&L’s. In another experiment, people showed an implicit size mapping even when size was irrelevant to the task (Sellaro, Treccani, Job, & Cubelli, 2014).

Given that (a) people implicitly associate smaller spatial areas with the left and larger areas with the right, and (b) H&L’s more intense faces have larger mouths, it is possible that the emotional magnitude effect H&L reported was, instead, a *spatial magnitude* effect. To test this alternative explanation, here we measured the mouth sizes of the faces in H&L’s stimuli (gratefully obtained from H&L) and used them in place of emotional intensity to predict dRTs in H&L’s Experiments 1 and 2 (Experiments 2a, b here).

4.1. Experiment 2a: An alternative account of H&L’s Experiment 1 results

4.1.1. Method

H&L’s Experiment 1 used 24 faces from the NimStim database (Tottenham et al., 2009). The photographs depicted six actors, each making four emotional expressions, which H&L labeled as neutral, happy, very happy, and extremely happy. Here a coder blind to the experimental hypothesis used Adobe Photoshop to select and measure the area of the mouth (including the lips) of each face used in H&L’s Experiment 1.

4.1.2. Results and discussion

4.1.2.1. Analysis of mouth sizes: After the mouth sizes were measured, measurements were averaged for each emotional expression (reported in units of 1000 pixels): Neutral ($M = 3.27$; ±0.27); Happy ($M = 3.64$; ±0.40); Very happy ($M = 7.58$; ±0.48); Extremely happy ($M = 10.53$; ±0.75). A first analysis confirmed that mouth size varied significantly across emotional expressions (with random intercepts for actors; $\chi^2(1) = 51.79$, $p = .0000000003$; Fig. 3).

Fig. 3. Top: The face stimuli used in H&L Experiment 1, with measured mouth size highlighted in black. Bottom: Relative mouth size (black circles) and standard errors (dotted lines) by emotional expression.

A second analysis tested whether mouth size covaried with emotional intensity. For each image, the mouth size (from our measurements) was used to predict the intensity ratings (i.e., emotional magnitude ratings) obtained from H&L, with random intercepts for actors and expressions. Mouth size reliably predicted emotional intensity ratings (r = .89; $\chi^2(1) = 15.95$, p = .0007).

4.1.2.2. Predicting dRT from mouth size: The finding of a correlation between mouth size and emotional intensity confirms that spatial magnitude could, in principle, give rise to H&L's effects. To test whether mouth size can, in fact, explain the dRT effects that H&L found we conducted a further analysis that was analogous to H&L's analysis, using mouth size instead of emotional intensity ratings to predict dRTs. For each participant and image, dRTs (obtained from H&L) were regressed over standardized mouth size (from our measurements) to produce a spatial magnitude slope for each participant. The mean slope differed significantly from zero (M = −11.84 ms/unit area; $t(17) = 2.75$, p = .01; Fig. 4), indicating that faces with smaller mouths were associated with the left and faces with larger mouths were associated with the right, consistent with an implicit mapping of spatial magnitude onto left-right space (a Spatial Magnitude Mapping).

4.1.2.3. Independent effects of intensity and mouth size: In a final analysis, we tested the independent contributions of mouth size and emotional intensity to the RT effects. Although mouth size and emotional intensity ratings were highly correlated (r = .89), in principle, emotional intensity could predict dRTs above and beyond mouth size (or vice versa). To evaluate the contribution of emotional intensity controlling for mouth size, we residualized emotional intensity with respect to mouth size and regressed dRT over these

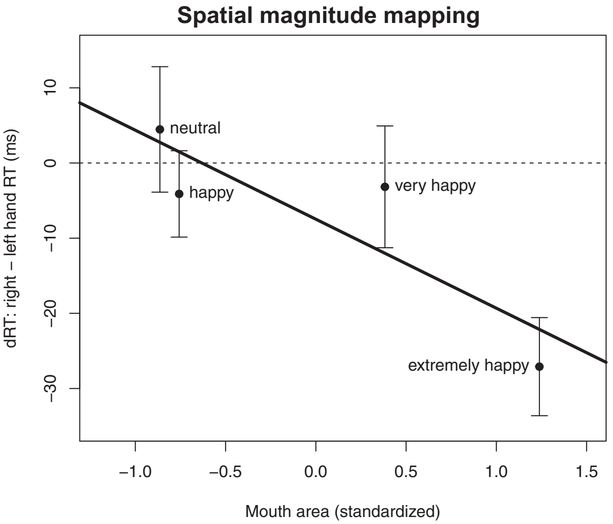

Fig. 4. Results of Experiments 2a. Mouth area reliably predicted dRTs from H&L's Experiment 1. Error bars show the standard error of the mean.

residuals. When the effect of mouth size was controlled, the mean slope of the effect of emotional intensity on dRT no longer differed from zero ($M = -6.54$ ms/unit intensity; $t(17) = 0.46$, $p = .65$). Next, to evaluate the contribution of mouth size controlling for emotional intensity, we residualized mouth size with respect to emotional intensity and regressed dRT over these residuals. When the effect of intensity was controlled, the mean slope of the effect of mouth size on dRT no longer differed from zero ($M = -5.70$ ms/unit area; $t(17) = 0.45$, $p = .66$).

These analyses show that there were no significant contributions of mouth size independent of emotional intensity, or of emotional intensity independent of mouth size, to H&L's RT effects. As such, it is not possible to interpret H&L's results as evidence for an Intensity Mapping; the results are equivocal between an Intensity Mapping and a Spatial Magnitude Mapping.

4.2. Experiment 2b: An alternative account of H&L's Experiment 2 results

Using the same methods as in Experiment 2a, here we tested for a confounding role of mouth size in the RT data of H&L's Experiment 2.

4.2.1. Method

H&L's Experiment 2 used 30 faces from the NimStim database (Tottenham et al., 2009). The photographs depicted six actors making each of five emotional expressions, which H&L labeled as neutral, happy, angry, extremely happy, and extremely angry. A coder blind to the experimental hypothesis used Adobe Photoshop to select and measure the area of the mouth (including the lips) of each face (see Fig. 5).

4.2.2. Results

4.2.2.1. Analysis of mouth sizes: After the mouth sizes were measured, measurements were averaged for each emotional expression (reported in units of 1,000 pixels): Neutral ($M = 3.27$; ±0.27); Happy ($M = 3.64$; ±0.40); Angry ($M = 2.26 \pm 0.20$); Extremely happy ($M = 10.53$; ±0.75); Extremely angry ($M = 6.75 \pm 0.85$). A first analysis confirmed that the mean mouth size varied significantly across the emotional expressions (with random intercepts for actors; $\chi^2(1) = 60.58$, $p = .00000000002$; Fig. 5).

A second analysis tested whether mouth size covaried with emotional intensity. For each image, the mouth size (from our measurements) was used to predict the intensity rating, with random intercepts for actors and expressions. Mouth size reliably predicted emotional intensity ratings ($r = .62$; $\chi^2(1) = 6.22$, $p = .01$).

4.2.2.2. Predicting dRT from mouth size: To test whether mouth size can explain H&L's effects in their second experiment, we conducted a further analysis, analogous to H&L's, using mouth size instead of emotional magnitude ratings to predict dRTs. For each participant and image, dRTs (obtained from H&L) were regressed over standardized mouth size (from our measurements) to produce a spatial magnitude slope for each participant. The mean slope differed significantly from zero ($M = -10.14$ ms/unit area; $t(45) = 2.72$,

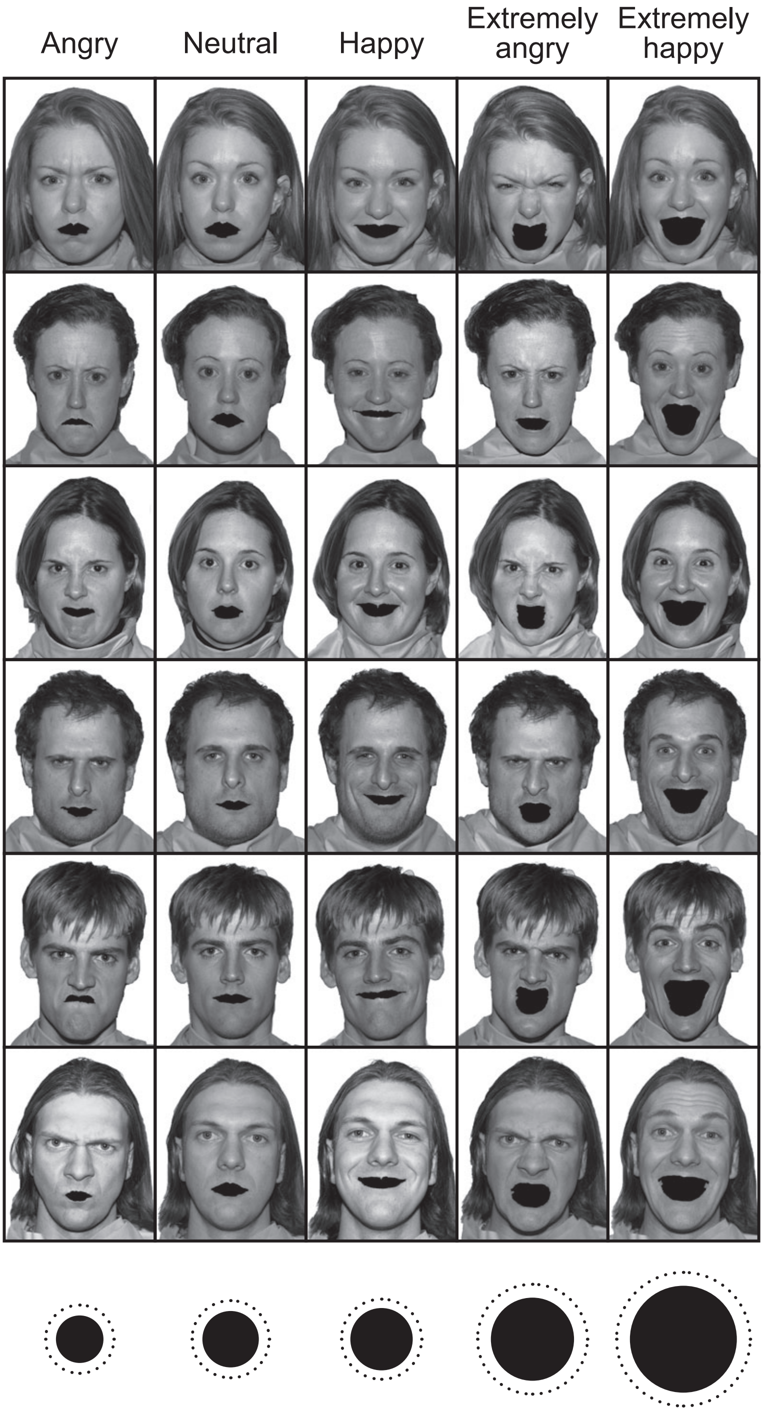

Fig. 5. Top: The face stimuli used in H&L Experiment 2, with measured mouth size highlighted in black. Bottom: Relative mouth size (black circles) and standard errors (dotted lines) by emotional expression.

p = .009; Fig. 6), suggesting an implicit mapping of spatial magnitude (i.e., mouth size) onto left-right space like that observed in Experiment 2a.

4.2.2.3. Independent effects of intensity and mouth size: In a final analysis, we tested the independent contributions of mouth size and emotional intensity to the RT effects. To evaluate the contribution of emotional intensity controlling for mouth size, we residualized emotional intensity with respect to mouth size and regressed dRT over these residuals. When the effect of mouth size was controlled, the mean slope of the effect of emotional intensity on dRT no longer differed from zero ($M = -10.48$ ms/unit intensity; $t(45) = 1.40$, $p = .17$); the Intensity Mapping disappeared. Next, to evaluate the contribution of mouth size controlling for emotional intensity, we residualized mouth size with respect to emotional intensity and regressed dRT over these residuals. When the effect of intensity was controlled, the mean slope of the effect of mouth size on dRT no longer differed from zero ($M = -3.31$ ms/unit intensity; $t(45) = 0.73$, $p = .47$). As in Experiment 2a, these analyses show that there were no significant contributions of mouth size independent of emotional intensity or of emotional intensity independent of mouth size to H&L's RT effects.

4.3. Discussion

Experiment 2 investigated an alternative explanation for the RT effects reported in H&L's first two experiments, which the authors interpreted as evidence for an Intensity Mapping. We showed that the intensity (i.e., emotional magnitude) of H&L's face stimuli was strongly correlated with the area (i.e., spatial magnitude) of the mouths, which is a salient facial feature that people use spontaneously to process emotional expressions (Jack

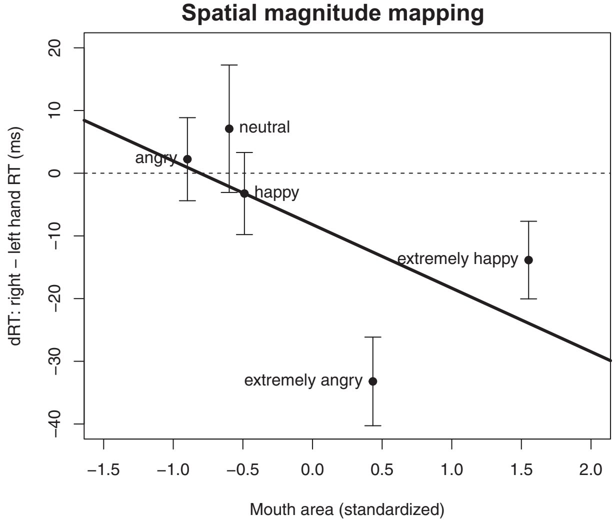

Fig. 6. Results of Experiments 2b. Mouth area reliably predicted dRTs from H&L's Experiment 2. Error bars show the standard error of the mean.

et al., 2009; Koda & Ruttkay, 2014; Smith et al., 2005). Faces that expressed more intense emotions also had larger mouths. Moreover, mouth size reliably predicted the RT effects in H&L's data, and controlling for mouth size eliminated the reported effect of emotional intensity.

Together, the results of Experiments 2a, b suggest that the dRT effects in H&L's Experiments 1 and 2 cannot be attributed to the emotional intensity of the stimulus faces, which was confounded with the size of their mouths. Was mouth size *solely* responsible for H&L's reported effects? Our analyses do not support this conclusion, either, since controlling for intensity also eliminated the effect of mouth size. Rather, H&L's effects may be due to variance shared between the emotional intensity and the mouth size of the faces.

Why were we unable to distinguish between the effect of mouth size and the effect of emotional intensity in H&L's data? One possibility is that that we simply did not have enough power to distinguish the independent effects of these two variables, but that with more data, one might find an effect of emotional intensity above and beyond the effect of mouth size, or vice versa. Another possibility is that mouth size cannot, in principle, be separated from the emotional intensity of faces because the size of a mouth is "perceptually integral" with the intensity of the emotion it expresses (Garner, 1974). Unlike words, faces express emotion largely by virtue of their form. To the extent that the emotional intensity of a face *is determined by* the size of the mouth, mouth size and emotional intensity may not be separable in our perception of emotional faces (and therefore not separable in the data). To the extent that mouth size is perceptually integral to the emotional intensity of faces, participants in H&L's Experiments 1 and 2 may have "categorize[d] stimuli not on the basis of experimenter-intended stimulus dimensions, but on the basis of some other property that is "emerging" from the integral stimuli" (Lewandowsky, Roberts, & Yang, 2006, p. 1678).

Mouth size and emotional intensity are clearly correlated not only in H&L's stimuli, but also in people's experience of faces more generally. Does that mean that H&L's data can be interpreted as evidence for an Intensity Mapping? No. Regardless of whether mouth size and emotional intensity are perceptually integral, H&L's results do not license the conclusion that either one of these variables is responsible for the observed RT effects. Our analyses indicate that H&L's RT effects in their Experiments 1 and 2 *could* be generated by emotional intensity, but they could just as well be generated by spatial magnitude (or by a combination of these variables). Is one of these possibilities more likely? Arguably, a Spatial Magnitude interpretation of H&L's effects is best supported by the available data, given that (a) a left-right Spatial Magnitude Mapping has already been shown repeatedly (Ren et al., 2011; Sellaro et al., 2014), whereas (b) H&L's study was the first purported demonstration of an Intensity Mapping, and (c) we failed to find any evidence for an intensity mapping when we removed the spatial magnitude confound by using emotional words.

5. General discussion

In four experiments, we investigated how people spatialize different aspects of emotion, and whether emotion-space mappings provide evidence for a hyper-general system for representing mental magnitudes. We sought to reconcile an apparent contradiction between previous studies showing an implicit left-right mapping of emotional valence (Casasanto, 2014; for review) with the results of a study positing an implicit left-right mapping of emotional intensity (Holmes & Lourenco, 2011). In Experiments 1a, b, we tested whether people implicitly spatialize emotion according to either valence or intensity, depending on the context. When judging the emotional valence of words, participants reliably showed the Valence Mapping typical of right-handers. By contrast, when judging the emotional intensity of the same words, participants showed no evidence of an Intensity Mapping, inconsistent with the findings of H&L. Experiments 2a, b provided a possible explanation for this discrepancy. In H&L's stimuli, emotional intensity was confounded with spatial magnitude: Faces expressing greater emotional intensity also had larger mouths. Mouth size reliably predicted the pattern of RTs reported by H&L, and controlling for mouth size eliminated the effect of emotional intensity. Together, these experiments (a) extend the evidence for an implicit lateral mapping of emotional valence, and for theories of metaphorical mental representation more broadly, (b) provide no evidence for the implicit lateral mapping of emotional intensity posited by H&L, (c) provide an alternative explanation for the RT effects that H&L interpreted as showing an Intensity Mapping, and (d) call into question the existence of a "hyper-general system of magnitude representation" that extends to the domain of emotion.

5.1. Is there any evidence for an intensity mapping?

In light of these findings, what is the evidence that people implicitly spatialize emotional intensity on a left-right continuum? Currently, the evidence for this claim rests on a paper by H&L comprising three experiments. To evaluate the status of this claim, we examine each of these experiments in detail.

5.1.1. H&L's Experiment 1

In H&L's Experiment 1, participants saw photographs of faces that varied from neutral to extremely happy and classified their gender using left- and right-hand response keys. Although the pattern of RTs was consistent with an Intensity Mapping, there are at least two other accounts of these results.

First, these results can easily be explained by a Valence Mapping. Faces showing greater emotional intensity were also more positive in valence than faces showing less emotional intensity. Therefore, as H&L acknowledged, "left-to-right orientation might reflect mappings of valence, rather than [emotional] magnitude" (p. 318). Given that an implicit left-right Valence Mapping has been shown by more than two dozen experiments (see Casasanto, 2014), the confound between valence and intensity, alone, suggests that the results of H&L's Experiment 1 cannot be interpreted as support for an Intensity Mapping; that is, if data are *equally consistent* with a well-validated hypothesis (Valence Mapping) and with a new hypothesis (Intensity Mapping), then there is no basis for interpreting them as support for the new hypothesis.

Second, here we show that the findings of H&L's Experiment 1 are also consistent with a Spatial Magnitude Mapping, in which smaller mouth size is associated with the left and larger mouth size with the right (see our Experiment 2a). In sum, although the findings of H&L's Experiment 1 could have resulted from an Intensity Mapping (thus providing the first evidence for such a mapping), they can also be explained by two previously established mappings: the left-right mapping of emotional valence and the left-right mapping of spatial magnitude (Ren et al., 2011; Sellaro et al., 2014).

5.1.2. *H&L's Experiment 2*

H&L's Experiment 2 is resistant to the first of these alternative explanations, but it is still subject to the second. Unlike the results of H&L's first experiment, the results of their second experiment cannot be explained by a Valence Mapping. However, here we show that these results can be explained by a Spatial Magnitude Mapping (see our Experiment 2b). Such an effect would not be surprising given people's tendencies to attend to the mouth when assessing emotional faces (Jack et al., 2009; Koda & Ruttkay, 2014) and to spatialize physical size from left to right even when size is irrelevant to the task (Ren et al., 2011; Sellaro et al., 2014). To the extent that participants were spatializing mouth size, the results of H&L's Experiments 1 and 2 do not bear on the spatialization of any dimension of emotion.

5.1.3. *H&L's Experiment 3*

H&L's Experiment 3 provides no evidence of an Intensity Mapping, even bracketing any consideration of confounds between intensity and valence, or between intensity and spatial magnitude. Participants in H&L's Experiment 3 classified faces according to either happiness (happy/not happy) or anger (angry/not angry). The pattern of RTs depended on the type of judgment participants made: Those who judged happiness associated the angriest faces with the left and the happiest faces with the right; those who judged anger showed the opposite mapping, associating the happiest faces with the left and the angriest faces with the right.

Although H&L interpreted these two patterns as evidence of the flexibility of an "emotional magnitude line" (i.e., an Intensity Mapping), both patterns are inconsistent with any linear mapping of emotional intensity onto left-right space. However, an Intensity Mapping predicts that people should associate the angriest and happiest faces with the same side of space, participants in H&L's Experiment 3 mapped them onto opposite sides of space. These results are, therefore, incompatible with the proposed emotional magnitude line. (Likewise, these results cannot be explained by a Spatial Magnitude Mapping, given that mouth size is confounded with emotional intensity.) Although participants responded differentially to faces with different emotions, they did not spatialize faces according to emotional intensity (or spatial magnitude).

Instead, these findings may be readily explained by *polarity alignment* (Proctor & Cho, 2006), also called *markedness* (Clark, 1969; Dolscheid & Casasanto, 2015; Lakens, 2012; Lynott & Coventry, 2014). Like many other analog continuums in language and mind, *happiness* and *anger* are both bipolar (or marked) continuums. That is, they consist of a

+polar (or unmarked) endpoint (*happy*, *angry*) and an opposing –polar (or marked) endpoint (*unhappy*, *not angry*). Lateral space is also a bipolar continuum, where *right* has +polarity and *left* has –polarity, according to linguistic and cultural conventions that link "right" with "positive" (e.g., phrases like "the *right* answer; customs like raising the right hand to swear an oath; cf., Huber et al., 2014).

In polarity alignment effects, participants show faster RTs when the poles of two continuums align than when they misalign (see Clark, 1969; Proctor & Cho, 2006). This is exactly the effect that H&L observed in their Experiment 3: Participants responded faster when the poles of the designated emotional continuum aligned with the poles of the lateralized response (*happy/angry: right; unhappy/not angry: left*) than when they misaligned (*happy/angry: left; unhappy/not angry: right*). Therefore, overall, the results of H&L's Experiment 3 show what appears to be a classic polarity alignment effect (Clark, 1969; Proctor & Cho, 2006).[4]

5.1.4. Summary of evidence for an intensity mapping

None of the three experiments presented by H&L provide any clear evidence for an Intensity Mapping. Although the results of H&L's Experiments 1 and 2 were *consistent* with an Intensity Mapping, they could also be explained by the Valence Mapping (for Experiment 1) or by a Spatial Magnitude Mapping (for Experiments 1–2), which have both been shown repeatedly. The results of H&L's Experiment 3 were not consistent with an Intensity Mapping, at all; rather, they are consistent with an effect of polarity alignment.

In our Experiments 1a, b, participants showed no evidence of an Intensity Mapping, even when they were making intensity judgments explicitly (and even though they spatialized another aspect of emotion and showed a clear symbolic distance effect in the intensity task). In principle, future studies could reveal an Intensity Mapping in certain contexts. However, given the current evidence, if people spatialize emotional intensity at all, it is unlikely that they activate a left-right mapping of intensity with the same strength or automaticity with which they activate left-right mappings of valence (Casasanto, 2014; for review), spatial magnitude (Ren et al., 2011; Sellaro et al., 2014), number (e.g., Wood et al., 2008; for review), time (Bonato, Zorzi & Umlita, 2012, for review), or politics (e.g., Oppenheimer & Trail, 2010).

In sum, neither H&L's study nor ours shows any clear evidence for an Intensity Mapping.

5.2. Is there any other evidence for a "hyper-general system of magnitude representation?"

Aside from H&L's study, one other study has sought to extend the putative generalized magnitude system to the domain of emotion, testing a well-established mapping between vertical space and emotional valence (Santiago, Ouellet, Román, & Valenzuela, 2012). Although the authors suggested that emotional valence may be represented by a generalized magnitude system, this argument is problematic. As specified in the original proposal

(Walsh, 2003), a generalized magnitude system could only support the representation of *prothetic* domains; that is, domains in which people experience quantitative variation (e.g., more/less size, duration, brightness, loudness, etc.) Emotional valence, however, is not a prothetic domain. Rather, valence has more in common with *metathetic* domains like hue, pitch, and timbre, in which distinctions are qualitative, not quantitative (Stevens, 1957; also see Bottini & Casasanto, 2013). An "hour" has *more duration* than a "minute," but "happiness" does not have *more valence* than "sadness." Even if two emotions like happiness and sadness were experienced with equal intensity, their valence would still be qualitatively different. Therefore, valence cannot be represented by the proposed generalized magnitude system, and spatial mappings of valence do not support the proposed hyper-general system of magnitude representations.

5.3. Is the Valence Mapping a mental metaphor?

Although our primary aim was to investigate whether people can spatialize emotions according to valence or intensity, these data address an additional question: Is the left-right mapping of valence categorical or continuous? Either possibility is plausible a priori: Since the spatial mapping of valence is determined by which hand is dominant — and hands are discrete — the mapping could be categorical (i.e., all levels of "positive" could be equally associated with the dominant side). Alternatively, associations between emotional valence and lateral space could reflect a mental metaphor, which is defined as an association in long-term memory between two analog continuums (Casasanto, 2016). When people activate a mental metaphor, they construct a point-to-point mapping between the continuums, representing differences in the target domain (e.g., emotional valence) via differences in the source domain (e.g., lateral space). Therefore, like other mental metaphors, space-valence metaphors should entail a continuous mapping: For right handers, more positive emotions should be more strongly associated with the right than less positive emotions and more negative emotions should be more strongly associated with the left than less negative emotions.

A previous experiment using an explicit spatialization task has suggested such a continuous mapping of valence. When asked to spatialize facial expressions of different emotions along a lateral line, participants placed the more strongly valenced expressions farther out to the sides (according to their handedness) and placed the more moderately valenced expressions progressively closer to the center (Freddi et al., 2016). Likewise here, Experiments 1a, b show evidence of a continuous valence mapping in a task that measures space-valence associations implicitly: The more positive the emotional valence, the faster participants were to respond with their right hand (compared to their left), and the more negative the emotional valence the faster they were to respond with their left hand (compared to their right). These data present the strongest evidence to date that the *implicit* lateral mapping of emotional valence is continuous, and they support the claim that lateral space-valence associations reflect a mental metaphor in which people represent the abstract continuum of emotional valence, in part, via associations with the more concrete continuum of lateral space.

6. Conclusion

Here, we showed that people implicitly map emotional valence onto left-right space (a Valence Mapping), but we found no spatial mapping of emotional intensity (no Intensity Mapping), even when participants were judging the emotional intensity of the stimuli explicitly. We provided an alternative explanation for previous data interpreted as support for an Intensity Mapping. These previous findings can be explained in terms of (a) the Valence Mapping, (b) a left-right mapping of spatial magnitude, or (c) polarity alignment (i.e., markedness effects), or a combination of these factors.

The idea of an Intensity Mapping was motivated by the search for a hyper-general magnitude system for representing multiple prothetic domains, including non-physical "magnitudes" such as emotional intensity (Holmes & Lourenco, 2011). Yet, across multiple experiments from two laboratories, no clear evidence for an Intensity Mapping has been found. In the absence of any such evidence, there is no empirical support for a hyper-general magnitude system that extends to the domain of emotion. Rather, the present findings extend the evidence for the Valence Mapping, which is one of many mental metaphors through which people spatialize abstract domains — whether or not they are prothetic — according to the specifics of their physical and social experiences.

Acknowledgments

Research was supported by a James S. McDonnell Foundation Scholar Award (#220020236) and NSF grant (BCS #125710) to DC. A preliminary report on Experiments 1a, b and 2b was presented at the 38th Annual Conference of the Cognitive Science Society (Philadelphia). Thanks to Wim Gevers and Katia Carbé for helpful discussions; to Katia Carbé and Nikolai Maximay for help with data collection; to Geoffrey Brookshire, Laura Staum Casasanto, Tom Gijssels, and Ché Lucero for discussion of statistical analyses; and to Kevin Holmes and Stella Lourenco for their collegiality and for sharing their data.

Notes

1. Incorrect trials included trials in which the participant gave the wrong response as well as trials in which the participant failed to respond: misses. We report the miss rate here to facilitate comparison with Experiment 1b, in which the miss rate is the only measure of accuracy available. (We thank a reviewer for suggesting this addition.) Of 6,656 critical RT trials, there were a total of 20 missed trials. The miss rate did not differ across the two tasks (Valence: 9 misses = 0.27%; Intensity: 11 misses = 0.33%; $\chi^2(1) = 0.25$, $p = .62$).
2. Although these slopes could be used for inferential statistics (see Fias, Brysbaert, Geypens, & Géry, 1996), using them here would be inappropriate for several

reasons. Baayen, Davidson, and Bates (2008) show that this by-participant regression approach inflates Type 1 error rates. Furthermore, because Fias et al.'s method collapses over large amounts of data (here, a 96:1 compression), it is not suitable for testing the higher-order interactions on which our experimental questions depend.

3. We thank a reviewer for pointing out that emotional valence is highly correlated with emotional motivation in our stimuli (as it is in the world): Words with positive valence (e.g., *determined*) are associated with approach-motivated behavior, whereas words with negative valence (e.g., *hesitant*) are associated with avoidance-motivated behavior. This confound makes it difficult to empirically distinguish a mapping of emotional valence from a mapping of emotional motivation in our data. However, whereas previous RT tasks have shown a left-right mapping of valence controlling for motivation (e.g., Root, Wong, & Kinsbourne, 2006), the converse has not been shown.
4. Holmes and Lourenco (2011) were aware of the possibility of polarity alignment effects, and they provided evidence that their results could not be explained by alignment between the polarities of the *spatial response* (left-right) and the *gender* of the faces (male-female). But alignment between space and gender is not the only potential source of polarity alignment effects. As we illustrate here, the poles of the spatial response can be mis/aligned with the poles of the emotion continuum on which participants judged the faces in Experiment 3: poles that were made salient by the explicit positive versus negative contrast in the wording used (happy vs. unhappy; angry vs. not angry).

References

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.

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.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Available at: http://CRAN.R-project.org/package=lme4. Accessed March 1, 2015.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE*, *3*(8), e3022.

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2257–2273.

Bottini, R., & Casasanto, D. (2013). Space and time in the child's mind: Metaphoric or ATOMic? *Frontiers in Psychology*, *4*, 803.

Brunyé, T. T., Gardony, A., Mahoney, C. R., & Taylor, H. A. (2012). Body-specific representations of spatial location. *Cognition*, *123*(2), 229–239.

Carbè, K., & Gevers, W. (2013). *Investigating the spatial representation of magnitude information and emotional valence concepts*. Paper presented at The 32nd European Workshop on Cognitive Neuropsychology, Bressanone.

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351–367.

Casasanto, D. (2011). Different bodies, different minds: The body-specificity of language and thought. *Current Directions in Psychological Science*, *20*(6), 378–383.

Casasanto, D. (2014). Bodily relativity. In L. Shapiro (Ed.), *Routledge handbook of embodied cognition* (pp. 108–117). New York: Routledge.

Casasanto, D. (2016). A shared mechanism of linguistic, cultural, and bodily relativity. *Language Learning*, *66*(3), 714–730.

Casasanto, D., & Bottini, R. (2014). Mirror reading can reverse the flow of time. *Journal of Experimental Psychology: General*, *143*(2), 473.

Casasanto, D., & Chrysikou, E. G. (2011). When left is "right": Motor fluency shapes abstract concepts. *Psychological Science*, *22*(4), 419–422.

Casasanto, D., Fotakopoulou, O., & Boroditsky, L. (2010). Space and time in the child's mind: Evidence for a cross-dimensional asymmetry. *Cognitive Science*, *34*, 387–405.

Casasanto, D., & Gijssels, T. (2015). What makes a metaphor an embodied metaphor? *Linguistics Vanguard*, *1*(1), 327–337.

Casasanto, D., & Henetz, T. (2012). Handedness shapes children's abstract concepts. *Cognitive Science*, *36*, 359–372.

Casasanto, D., & Jasmin, K. (2010). Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *PLoS ONE*, *5*(7), e11805.

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543–566). Cambridge, MA: MIT Press.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, *76*(4), 387.

de la Fuente, J., Casasanto, D., Martínez-Cascales, J. I., & Santiago, J. (2017). Motor imagery shapes abstract concepts. *Cognitive Science*, *41*(5), 1350–1360. https://doi.org/10.1111/cogs.12406.

de la Fuente, J., Casasanto, D., Román, A., & Santiago, J. (2015). Can culture influence body-specific associations between space and valence? *Cognitive Science*, *39*, 821–832.

de la Fuente, J., Casasanto, D., & Santiago, J. (2015). Observed actions affect body-specific associations between space and valence. *Acta Psychologica*, *156*, 32–36.

de la Vega, I., De Filippis, M., Lachmair, M., Dudschig, C., & Kaup, B. (2012). Emotional valence and physical space: Limits of interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 375.

de la Vega, I., Dudschig, C., De Filippis, M., Lachmair, M., & Kaup, B. (2013). Keep your hands crossed: The valence-by-left/right interaction is related to hand, not side, in an incongruent hand–response key assignment. *Acta Psychologica*, *142*(2), 273–277.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191–1194.

Dolscheid, S., & Casasanto, D. (2015). Spatial congruity effects reveal metaphorical thinking, not polarity correspondence. *Frontiers in Psychology*, *6*(1836), 1–11.

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition*, *2*, 95–110.

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, *6*(6), 555.

Freddi, S., Brouillet, T., Cretenet, J., Heurley, L. P., & Dru, V. (2016). A continuous mapping between space and valence with left-and right-handers. *Psychonomic Bulletin & Review*, *23*(3), 865–870.

Garner, W. R. (1974). *The processing of information and structure*. Oxford, UK: Psychology Press.

Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., & Fias, W. (2006). Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychologica*, *122*(3), 221–233.

Ginsburg, V., van Dijck, J. P., Previtali, P., Fias, W., & Gevers, W. (2014). The impact of verbal working memory on number–space associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 976.

Herrera, A., Macizo, P., & Semenza, C. (2008). The role of working memory in the association between number magnitude and space. *Acta Psychologica*, *128*(2), 225–237.

de Hevia, M. D., Girelli, L., & Vallar, G. (2006). Numbers and space: A cognitive illusion? *Experimental Brain Research*, *168*(1–2), 254–264.

Holmes, K. J., & Lourenco, S. F. (2011). Common spatial organization of number and emotional expression: A mental magnitude line. *Brain and Cognition*, *77*(2), 315–323.

Huber, S., Klein, E., Graf, M., Nuerk, H. C., Moeller, K., & Willmes, K. (2014). Embodied markedness of parity? Examining handedness effects on parity judgments. *Psychological Research*, *79*(6), 963–977.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, *19*(18), 1543–1548.

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, *141*(1), 19.

Kim, N., Krosnick, J., & Casasanto, D. (2015). Moderators of candidate name-order effects in elections: An experiment. *Political Psychology*, *36*(5), 525–542.

Koda, T., & Ruttkay, Z. (2014). Eloquence of eyes and mouth of virtual agents: Cultural study of facial expression perception. *AI & Society*, *32*(1), 17–24.

Kong, F. (2013). Space–valence associations depend on handedness: Evidence from a bimanual output task. *Psychological Research*, *77*(6), 773–779.

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 726.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*(2), 195–208.

Lewandowsky, S., Roberts, L., & Yang, L. X. (2006). Knowledge partitioning in categorization: Boundary conditions. *Memory & Cognition*, *34*(8), 1676–1688.

Lindemann, O., Abolafia, J. M., Pratt, J., & Bekkering, H. (2008). Coding strategies in number space: Memory requirements influence spatial–numerical associations. *The Quarterly Journal of Experimental Psychology*, *61*(4), 515–524.

Lynott, D., & Coventry, K. (2014). On the ups and downs of emotion: Testing between conceptual-metaphor and polarity accounts of emotional valence–spatial location interactions. *Psychonomic Bulletin & Review*, *21*(1), 218–226.

Merritt, D. J., Casasanto, D., & Brannon, E. M. (2010). Do monkeys think in metaphors? Representations of space and time in monkeys and humans. *Cognition*, *117*, 191–202.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215* (5109), 1519–1520.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241.

Oppenheimer, D. M., & Trail, T. E. (2010). Why leaning to the left makes you lean to the left: Effect of spatial orientation on political attitudes. *Social Cognition*, *28*(5), 651.

Ping, R. M., Dhillon, S., & Beilock, S. L. (2009). Reach for what you like: The body's role in shaping preferences. *Emotion Review*, *1*(2), 140–150.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/. Accessed March 1, 2015.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*(4), 364–382.

Ren, P., Nicholls, M. E., Ma, Y. Y., & Chen, L. (2011). Size matters: Non-numerical magnitude affects the spatial coding of response. *PLoS ONE*, *6*(8), e23553.

Root, J. C., Wong, P. S., & Kinsbourne, M. (2006). Left hemisphere specialization for response to positive emotional expressions: A divided output methodology. *Emotion*, *6*(3), 473.

Santiago, J., Ouellet, M., Román, A., & Valenzuela, J. (2012). Attentional factors in conceptual congruency. *Cognitive Science*, *36*(6), 1051–1077.

Sellaro, R., Treccani, B., Job, R., & Cubelli, R. (2014). Spatial coding of object typical size: Evidence for a SNARC-like effect. *Psychological Research*, *79*(6), 950–962.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*(3), 184–189.

Song, X., Chen, J., & Proctor, R. W. (2017). Role of hand dominance in mapping preferences for emotional-valence words to keypress responses. *Acta Psychologica*, *180*(August), 33–39.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249.

Tracy, J. L., & Robins, R. W. (2008). The automaticity of emotion recognition. *Emotion*, *8*(1), 81.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford, UK: Blackwell.

Wood, G., Willmes, K., Nuerk, H. C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly*, *50*(4), 489.

Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, *43*(2), 303–311.

Zhao, X., He, X., Zhang, W., Chen, G., Chen, Q., & Huang, L. (2016). Interpersonal choice: The advantage on the left or on the right? *International Journal of Psychology*, https://doi.org/10.1002/ijop.12388