

Spatial congruity effects reveal metaphors, not markedness

Sarah Dolscheid^{1,2} (sarah.dolscheid@mpi.nl) Cleve Graver⁴ (gravc243@newschool.edu) Daniel Casasanto^{3,4} (casasand@newschool.edu)

¹Max Planck Institute for Psycholinguistics, Nijmegen, NL

²International Max Planck Research School for Language Sciences, Nijmegen, NL

³Donders Center for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, NL

⁴Department of Psychology, The New School for Social Research, New York, USA

Abstract

Spatial congruity effects have often been interpreted as evidence for metaphorical thinking, but an alternative markedness-based account challenges this view. In two experiments, we directly compared metaphor and markedness explanations for spatial congruity effects, using musical pitch as a testbed. English speakers who talk about pitch in terms of spatial height were tested in speeded space-pitch compatibility tasks. To determine whether space-pitch congruency effects could be elicited by any marked spatial continuum, participants were asked to classify high- and low-frequency pitches as 'high' and 'low' or as 'front' and 'back' (both pairs of terms constitute cases of marked continuums). We found congruency effects in high/low conditions but not in front/back conditions, indicating that markedness is not sufficient to account for congruency effects (Experiment 1). A second experiment showed that congruency effects were specific to spatial words that cued a vertical schema (tall/short), and that congruency effects were not an artifact of polysemy (e.g., 'high' referring both to space and pitch). Together, these results suggest that congruency effects reveal metaphorical uses of spatial schemas, not markedness effects.

Keywords: metaphor, polarity correspondence, markedness, musical pitch, space

Introduction

Are *high* hopes somewhere in the air? Or what about *rising* prices? And where exactly are you when you are feeling *down*? Spatial metaphors like these are very common in language. Moreover, according to conceptual metaphor theory, people not only talk in terms of space but they also *think* metaphorically (i.e. spatially) (Lakoff & Johnson, 1980). Whereas arguments in favor of this claim were initially based on linguistic data (and thus circular in nature), psychological experiments have now shown that spatial representations importantly contribute to people's understanding of domains like time (Casasanto & Boroditsky, 2008), social dominance (Schubert, 2005), or valence (Meier & Robinson, 2004).

Many of these psychological studies base their findings on binary compatibility tasks. In one experiment, for instance, participants were asked to classify dimensions in a metaphoric target domain (i.e., valence: judge the positive or negative valence of a word), while, at the same time aspects of the spatial source domain (i.e, location; up and down) were varied. In line with "GOOD is UP" metaphors, people were faster to evaluate positive words when they

appeared in a high spatial location compared to a low location (and vice versa for negatively valenced words) (Meier & Robinson, 2004). Similarly, participants made faster judgments about social power when words for powerful people are at the top of a display and powerless people at the bottom (e.g., 'king' above 'slave', rather than vice versa; Schubert, 2005). These "metaphoric congruency effects" (Lakens, 2012), with faster performance for congruent compared to incongruent trials, have been taken as evidence that metaphoric target domains automatically activate congruent spatial information, supporting claims of conceptual metaphor theory (e.g. Meier & Robinson, 2004; Schubert, 2005).

On an alternative account, however, it has been argued that congruency effects may be better explained as polarity alignment effects, also called markedness effects¹ (Lakens, 2012). Like many other continuums in language and mind, metaphoric source and target domains (e.g. *height* or *happiness*) are considered to be bipolar. That is, they consist of an unmarked or +polar endpoint (e.g. *high*, *happy*), and an opposing marked or -polar endpoint (*low*, *sad*). Unmarked endpoints (+polar) are commonly defined as the default, evaluatively positive or broader dimension as opposed to the marked (-polar) ones (see e.g., Lehrer, 1985; Proctor & Cho, 2006; for a critical approach see Haspelmath, 2006). Moreover, there is evidence that polarity differences affect cognitive processing. Participants show faster reaction times for unmarked (+polar) dimensions as compared to marked (-polar) ones (Clark, 1969; Seymour, 1974). Reaction time benefits for congruent metaphoric dimensions (like *happy* and *up*) could thus alternatively be explained by an additive processing advantage for +polar endpoints (e.g. happy +polar, up +polar): Across many studies, perceptual and linguistic judgments are faster when the poles of marked continuums are aligned (e.g., 'good' matched with 'up') than when they are misaligned (e.g., 'good' matched with 'down'; Clark, 1969; Lakens, 2012; Proctor & Cho, 2006). The existence of markedness effects in binary response compatibility tasks raises a question: Does polarity alignment offer an alternative, non-metaphorical explanation for "metaphor congruency effects" like those reported by Meier & Robinson (2004) and Schubert (2005), which rely on

¹ Here, the terms "markedness" and "polarity" will be used interchangeably.

dimensional compatibility in binary speeded response tasks? And if so, what would this mean for theories of metaphorical mental representation?

Crucially, not all of the evidence for metaphoric thinking comes from (binary) congruency effects. Rather, it has been shown that people's metaphoric representations of domains like time or musical pitch map onto space in a continuous analog fashion (Casasanto, 2010; Dolscheid, Shayan, Majid, & Casasanto, 2013). English speakers, for instance, who talk about musical pitch in terms of spatial height (high vs. low pitch; see e.g. Stumpf, 2006) also associate higher pitches with higher positions in space in nonlinguistic psychophysical tasks. In one study, participants were asked to reproduce musical pitches while watching lines varying in spatial height. Since lines were presented at multiple positions (i.e., 9 levels of height) in a random order, effects of space on pitch could not be attributed to (binary) polarity. Rather, participants' pitch reproductions were affected by the spatial information in a continuous way; tones accompanied by higher lines were reproduced at a higher frequency on average than the same tones accompanied by lower lines, resulting in a linear influence of height on pitch (Dolscheid et al., 2013). In this study, responses were not speeded, and the metaphor-congruity effects did not rely on the kind of binary stimulus-response compatibility that is believed to give rise to polarity alignment effects (Proctor & Cho, 2006).

Furthermore, some mappings between space and musical pitch go against markedness. Whereas speakers of many languages (including English) refer to pitch in terms of spatial height, other languages like Farsi or Turkish encode pitch in terms of spatial thickness (Shayan, Ozturk, & Sicoli, 2011). These thickness-pitch metaphors follow a reversed polarity alignment. Thick (+polar) refers to a low frequency pitch (−polar), whereas thin (−polar) refers to a high frequency pitch (+polar). Since Farsi speakers implicitly represent pitch in terms of thickness (Dolscheid et al., 2013), spatial schemas appear to be more important than polarity alignment.

Although experiments like Dolscheid et al.'s (2013) provide evidence for metaphorical mental representation that cannot be explained by markedness, the role of markedness in binary compatibility tasks remains controversial. Do source-target congruity effects merely show polarity alignment? Or do they reveal metaphoric associations? While metaphors and polarity are often indistinguishable in compatibility tasks (see also Lakens, 2012), we predict that when markedness and metaphor are juxtaposed, congruity effects will support metaphoric thinking, not markedness. What should matter is whether the words that participants have to classify in binary compatibility tasks activate the appropriate spatial schema (e.g., in the case of space-pitch mappings for English speakers, it should be a vertical spatial schema). That is, schema-appropriateness should be necessary, and markedness may not be sufficient to produce congruity effects.

In Experiment 1, we tested compatibility in height-pitch metaphors for 2 pairs of spatial terms, both paradigm cases of marked continuums (Clark, 1973). One pair corresponds to the poles of the correct spatial continuum (high-low), the other to the poles of an incorrect spatial continuum (front-back). High and front both constitute the unmarked or +polar endpoint, whereas low and back represent the marked or −polar endpoint (see e.g., Clark, 1973; Landsberg, 1995). Participants were asked to make binary speeded judgments on high-frequency and low-frequency pitches, classifying pitches either in a polarity-congruent way (e.g. high pitches as *high* or *front*), or in a polarity-incongruent way (e.g. high pitches as *low* or *back*). If polarity alignment drives space-pitch congruity effects, then similar effects should be found when pitch is mapped to any marked linear spatial continuum, regardless of its orientation: High/low and front/back should both produce pitch-congruity effects. Alternatively, if activating a particular spatial schema for pitch is critical (i.e., the schema that is encoded in the participants' language), then high/low should result in a congruency effect, but front/back should not.

Experiment 1

Methods

Participants Twenty-four English speakers with no reported hearing problems participated for payment (5\$ per 30 minutes). Four participants were excluded from analyses for not following instructions (i.e. they responded according to the wrong response mapping throughout at least one condition). They were replaced by a new sample of 4 participants who had not previously participated in the task.

Materials and Procedure Participants were asked to classify tones (one high and one low pitch) as quickly and accurately as possible by pressing buttons on the QWERTY keyboard (Q and P-keys). Stimuli were presented on an Apple iMac using Vision Egg 2.6 (Straw, 2008). Sounds were generated by Audacity software (<http://audacity.sourceforge.net/>) and comprised two pure tones (frequency: 262 and 440 hertz). Each tone lasted 400 ms. Participants listened to one tone at a time, via sealed headphones. Immediately following the offset of each tone, two response options (e.g., high, low) appeared, one on the bottom left and the other on the bottom right of the screen. Participants were instructed to classify the sound by pressing the button located under the corresponding word (e.g., high or low) as fast and accurately as possible. The left-right locations of the spatial terms varied randomly from trial to trial so that participants could not predict the location of the correct word in advance.

Spatial terms (high-low vs. front-back) were presented in 2 blocks, a high-low block and a front-back block. Within each block, spatial terms were crossed with 2 mappings (congruent, incongruent). The order of blocks was counterbalanced across participants. The order of congruity was counterbalanced within each block. Across blocks,

incongruent and congruent conditions were always presented in alternation. Before each condition, participants received 6 practice trials with feedback. Participants were also given an example illustrating the respective mapping before the practice trials.

Each condition consisted of 24 trials, 96 trials in total. In half of the trials a high pitch was presented, in the other half a low pitch. In the high-low congruent condition, the high pitch had to be classified as *high* and the low pitch as *low*. In the high-low incongruent condition, the high pitch had to be categorized as *low* and the low pitch as *high*. In the front-back congruent condition the high pitch had to be categorized as *front* and the low pitch as *back* (according to patterns of polarity/markedness). In the front-back incongruent condition the low pitch had to be categorized as *front* and the high pitch as *back*.

Results

All data were analyzed using R (version 2.14.2; <http://www.r-project.org/>) and the R packages *lme4* (Bates & Maechler, 2009) and *languageR* (Baayen, 2009; cf. Baayen, 2008). We carried out linear mixed-effects regression models of Space (high-low versus front-back) and Congruency (congruent, incongruent) on accuracy and RTs. Using the principle of backward selection, we started out with a full (conservative) model which took into consideration not only the random intercept but also the random slopes of subject whenever it was appropriate (i.e., when the factor was a within-subject factor). Random intercepts and slopes of items were not included in the analysis due to the small number of items (4 words: high/low, front/back). To interpret the significance, we adopted the criterion that a given cosine was significant if the absolute value of the *t*-statistic (or *z*-statistic) exceeded 2 (Baayen, 2008).

Accuracy The mean accuracy for all target trials was 92.4% ($SD = 8.1$). For high/low conditions, accuracy was 92.4% ($SD = 9.7$), and for front/back conditions, accuracy was 92.5% ($SD = 11.3$). For congruent conditions, accuracy was 95.9% ($SD = 4.5$) and for incongruent conditions it was 88.9% ($SD = 15.5$). Analyzing accuracy by using a logistic mixed effects model on binary accuracy data yielded no main effects or interaction of Space (high/low, front/back) and Congruency (congruent, incongruent), (Space: $z = |1.3|$; Congruency: $z = |0.2|$; Space by Congruency: $z = |1.2|$).

Reaction times Reaction times of the button presses were analyzed by linear mixed effects models. Only correct trials were considered which resulted in the exclusion of 7% of the data. Responses greater or less than ± 2 SDs away from each participant's average RTs were also excluded, which resulted in the removal of 6% of the accurate trials.

There was no significant main effect of Space ($t = |0.3|$). The model yielded a significant main effect of Congruency ($t = |3.5|$) and a significant interaction of Congruency by Space ($t = |3.3|$). A linear mixed effect model of Congruency

on reaction times restricted to the level of high/low, yielded a significant effect of Congruency ($t = |4.5|$), demonstrating a congruity effect of high/low conditions. Restricting the model to the level of front-back yielded no significant effect of Congruency ($t = |0.2|$) (see Figure 1).

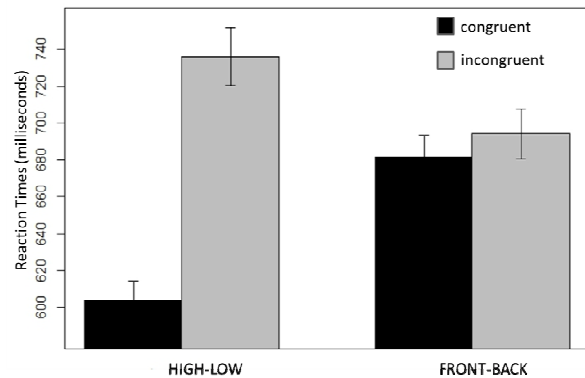


Figure 1: The influence of Space (high-low; front-back) and Congruency (congruent; incongruent) on pitch categorization (plotted in milliseconds).

Discussion

In Experiment 1 we find congruency effects for high/low but not for front/back conditions, suggesting that activating the appropriate spatial schema (i.e., spatial height) is what is relevant in such binary response compatibility tasks. Words that activate a different (irrelevant) spatial schema (front-back), however, do not result in a congruity effect. This finding indicates that congruity effects cannot be attributed to markedness (polarity alignment), since 'front' and 'back' also name the unmarked and marked ends of a (sagittally oriented) linear spatial continuum.

Experiment 2

Whereas high/low terminology is conventional for pitch in English, front/back is not. Maybe we only find a congruency effect in the case that is lexicalized, but not in the other case (front/back)? This skeptical interpretation would not change the fact that markedness is not sufficient to elicit congruency effects, but it would call into question our claim about activating the right spatial schema. Do we find congruency effects only because participants were using the polysemous words high/low, which can refer to "height" in both space and pitch?

To rule out this alternative explanation, in Experiment 2, we compared congruity effects in two pairs of spatial terms: tall/short and big/small. Neither pair of spatial expressions can be used in conventional English to describe the height of musical pitches (i.e., their frequency). If high/low congruency effects were driven by polysemy, then neither of these pairs of spatial terms should produce a congruity effect. However, if space-pitch congruity effects result from using words that activate a vertical spatial schema, then "tall" and "short" should produce a congruity effect because they are *schematically appropriate* (even though they are

lexically inappropriate). By contrast, “big” and “small” should not produce any space-pitch congruity effect, because these terms refer to 3-dimensional size, and should not activate the appropriate 1-dimensional vertical spatial schema (Dirven & Taylor, 1988; Taylor, 2002).

In addition to testing whether the height-pitch congruity effect in Experiment 1 depended on the polysemy of “high” and “low,” Experiment 2 also provides a second test of the sufficiency of markedness to produce space-pitch congruity effects. “Big” is the unmarked (positive) end and “small” the marked (negative) end of the big-small continuum. Therefore, markedness predicts that judgments should be faster when “big” is matched with “high” than when “small” is matched with “high.”

Methods

Participants Twenty-four English speakers with no reported hearing problems participated for payment (\$5 per 30 minutes). One participant was excluded from analyses for not following instructions (i.e. the participant responded according to the wrong response mapping throughout one condition). He was replaced by a new participant who had not previously participated in the task.

Materials and Procedure The same procedure as in Experiment 1 was used, with the following exceptions. Rather than classifying pitches as high-low or front-back, participants classified them as tall-short for one block and big-small for the other.

In the tall-short congruent condition the high pitch had to be categorized as *tall* and the low pitch as *short*. In the tall-short incongruent condition the low pitch had to be categorized as *tall* and the high pitch as *short*. In the big-small congruent condition the high pitch had to be categorized as *big* and the low pitch as *small* (according to patterns of markedness). In the big-small incongruent condition the low pitch had to be categorized as *big* and the high pitch as *small*.

Results

All data were analyzed using R (version 2.14.2; <http://www.r-project.org/>) and the R packages *lme4* (Bates & Maechler, 2009) and *languageR* (Baayen, 2009; cf. Baayen, 2008). We carried out linear mixed-effects regression models of Space (tall-short versus big-small) and Congruency (congruent, incongruent) on accuracy and RTs. Using the principle of backward selection, we again started out with a full (conservative) model which took into consideration not only the random intercept but also the random slopes of subject whenever it was appropriate (i.e., when the factor was a within-subject factor). Random intercepts and slopes of items were not included in the analysis due to the small number of items (4 words: tall/short, big/small). To interpret the significance, we adopted the criterion that a given cosine was significant if the absolute value of the *t*-statistic (or *z*-statistic) exceeded 2 (Baayen, 2008).

Accuracy The mean accuracy for all target trials was 94.8% ($SD = 11.4$). For tall/short conditions, accuracy was 94.6% ($SD = 13.1$), and for big/small conditions, accuracy was 94.9% ($SD = 10.1$). For congruent conditions, accuracy was 96.2% ($SD = 6.7$) and for incongruent conditions it was 93.3% ($SD = 17.1$). Analyzing accuracy by using a logistic mixed effects model on binary accuracy data yielded no main effects or interaction of Space (tall/short, big/small) and Congruency (congruent, incongruent), (Space: $z=|1.0|$; Congruency: $z=|1.0|$; Space by Congruency: $z=|1.0|$).

Reaction times Reaction times of the button presses were analyzed by linear mixed effects models. Only correct trials were considered which resulted in the exclusion of 4% of the data. Responses greater or less than ± 2 SDs away from each participant’s average RTs were also excluded, which resulted in the removal of 4% of the accurate trials.

There was no significant main effect of Congruency ($t=|0.9|$). The model yielded a significant main effect of Space ($t=|2.3|$) and a significant interaction of Congruency by Space ($t=|3.2|$). A linear mixed effect model of Congruency on reaction times restricted to the level of tall/short, yielded a significant effect of Congruency ($t=|3.0|$), demonstrating a congruity effect of tall/short conditions. Restricting the model to the level of big-small yielded no significant effect of Congruency ($t=|1.5|$) (see Figure 2).

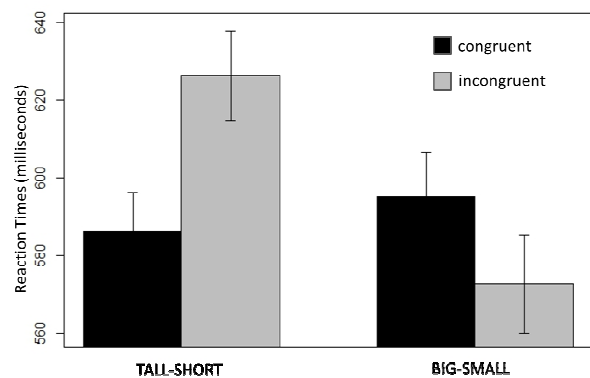


Figure 2: The influence of Space (tall-short; big-small) and Congruency (congruent; incongruent) on pitch categorization (plotted in milliseconds).

Discussion

In Experiment 2 we find a congruity effect in tall/short but not in big/small conditions. Therefore, congruity effects cannot be attributed to polysemy or markedness. Rather, space-pitch compatibility is based on activating the appropriate spatial schema, which serves as the source domain for English speakers' mental representations of musical pitch.

General Discussion

In two experiments, we show binary response-time congruity effects attributable to metaphorical thinking, but not to markedness. Classifying pitches with vertical spatial terms elicited space-pitch congruity effects, but no comparable effects are found when people were asked to classify pitches with terms that name the poles of other marked spatial continuums (front vs. back; big vs. small). Polarity alignment (a.k.a. markedness), therefore, is not sufficient to produce space-pitch congruity effects. Rather, schema-appropriateness is necessary, supporting theories of metaphorical mental representation.

Moreover, congruity effects are not restricted to polysemous words like “high” and “low,” which can be used for both space and pitch. Rather, congruity effects can also be found for words like “tall” and “short,” which have no musical senses, but which activate a vertical spatial schema: the “active ingredient” in the observed space-pitch congruity effects.

In most cases, the polarities of metaphorical source and target domains are aligned (e.g. Lakens, 2012; Lakoff & Johnson, 1980). For instance, happy, powerful, good, and high in pitch are all UP (the positive end of this spatial continuum), whereas their antonyms are DOWN (the negative end of the continuum). This relationship between metaphor and markedness makes it hard to determine the cause of many response compatibility effects. However, the polarities of metaphorical source and target domains are not always aligned. Musical pitch provides one domain, in which the marked end of the source domain (space) can be matched to the unmarked end of the target domain (pitch). Farsi speakers, for instance, represent pitch in terms of thickness. In Farsi speakers’ language and thought, the unmarked pole of the spatial continuum (thick) is aligned with the marked pole of the pitch continuum (low frequency). Thus, metaphors and markedness can dissociate in Farsi – at least to the extent that markedness can be established in a principled way.

Making psychological predictions on the basis of markedness is problematic because researchers may disagree on how markedness is defined, and even on which end of a given continuum is marked. Whereas Schubert (2005) describes “powerful” as the marked and “powerless” as the unmarked endpoint of the “power” continuum, others have suggested the reverse (e.g. Lakens, 2012). In addition to these inconsistencies, it is not always clear what markedness actually means. By definition, quite a number of attributes like frequency, familiarity, or fluency, seem to be subsumed under the umbrella term *markedness* (see Haspelmath, 2006). In one experiment, for instance, Lakens (2012) manipulated polarity by adjusting the frequency of the ‘marked’ endpoint. While usually marked attributes like bad or down (–polar) occur less frequently, this was no longer the case for a group of Laken’s participants. Critically, these participants also no longer showed a congruity effect, which was taken as evidence for a polarity account. However, in line with Haspelmath (2006),

it is questionable why one should talk about polarity when actually frequency is driving the effects. Unlike markedness, which is a notoriously ambiguous construct (e.g., Haspelmath, 2006, enumerates 12 distinct usages of this term in cognitive science), metaphors in language are more widely agreed upon. Expressions like “a high soprano” and “a low bass” make clear predictions about the spatial mappings that people should be activating for pitch, and therefore what congruity effects should be found: Linguistic metaphors tell us which end is “up.”

Here we find an impact of spatial schemas on source-target congruity as predicted by metaphors in language. Our results suggest an automatic, Stroop-like interference effect of metaphorical associations, converging with other findings of height-pitch congruity effects. In one task, for instance, participants made judgments about musical timbre while spatial height information was varied on a computer screen. Although pitch was irrelevant to the task, people’s judgments were affected by the alignment of tonal and spatial height (Evans & Treisman, 2010), suggesting a highly automatic source-target mapping (see also Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006; for limits of automaticity see Brookshire, Ivry, & Casasanto, 2010).

Unlike previous experiments, here the spatial source domain was not manipulated physically but rather via linguistic stimuli (i.e., we presented words like high/low; tall/short etc.). This allowed us to directly assess effects of polysemy. In Experiment 2, height-pitch congruity effects could not simply be attributed to lexical overlap (high/low for space and pitch). Rather, we found that words activating a similar vertical schema (tall/short) were sufficient to trigger space-pitch congruity effects even if the words were lexically inappropriate. One could argue, however, that congruity effects in tall/short conditions were still indirectly driven by polysemy. Participants may have activated high/low terminology when classifying pitches, which then in turn led to semantic priming from high to tall, and low to short. However, although we cannot entirely rule out such priming effects, this explanation is unlikely to account for our results, for several reasons. According to Latent Semantic Analysis (LSA; <http://lsa.colorado.edu/>), ‘tall’ is more strongly related to ‘short’ (LSA cosine: .48) than to ‘high’ (LSA cosine: .31). Moreover, ‘short’ is about equally strongly related to ‘high’ (LSA cosine: .30) as to ‘low’ (LSA cosine: .31). Since activation is expected to spread between the most strongly related items (Collins & Loftus, 1975), simple spreading activation would have wiped out a tall-short congruity effect rather than producing it. Moreover, although big is more closely related to high than to low (LSA cosine: .18 versus .12) congruity effects remain absent in big/small conditions. The non-significant big/small effect even points into the *opposite direction* (see Figure 2), suggesting that semantic priming is unlikely to drive the observed patterns of results.² Thus, while spatial

² The trend toward a big-low congruity effect could be driven by underlying associations between size and pitch (e.g., see Evans & Treisman, 2009) – but not by markedness or semantic priming.

terms like high/tall/big may be semantically related and overlap in markedness, we find that activating the appropriate vertical spatial schema is critical for producing space-pitch congruity effects.

Conclusions

Metaphor congruency effects have been challenged by a polarity account, claiming that binary response compatibility effects may be better explained by markedness than by metaphorical thinking (Lakens, 2012). Indeed, metaphor and polarity are often hard to distinguish. However, here we show that when polarity and metaphor are juxtaposed, congruity effects support metaphorical thinking, not polarity.

Furthermore, these results show that it is not necessary to use polysemous words to produce source-target congruity effects (i.e., words that can refer to both the metaphorical source and target domains). Words that activate a vertical schema (e.g., tall/short) produce a space-pitch congruity effect despite being lexically inappropriate. Words that activate a different spatial schema (e.g., front/back, big/small) do not produce any space-pitch congruity effect, despite naming the poles of other marked spatial continuums.

Together, these results indicate that activating the appropriate spatial schema is the “active ingredient” in space-pitch congruity effects – not polysemy or markedness – supporting theories of metaphorical mental representation.

Acknowledgments

We thank Tom Gijssels for assistance in programming and for help with the experimental set-up, and Geoff Brookshire for help with R-analyses. This research was supported by a Travel Grant of Radboud University Nijmegen and an IMPRS fellowship to SD and by a James S. McDonnell Foundation Scholar Award (#220020236) to DC.

References

Baayen, H. R. (2009). LanguageR. *R package*.
 Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
 Bates, D., & Maechler, M. (2009). Package “lme4”(Version 0.999375-32): linear mixed-effects models using S4 classes. Available (April 2011) at <http://cran.r-project.org/web/packages/lme4/lme4.pdf>.
 Brookshire, G., Ivry, R., & Casasanto, D. (2010). Modulation of motor-meaning congruity effects for valenced words. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1940–1945). Austin, TX: Cognitive Science Society.
 Casasanto, D. (2010). Space for thinking. In V. Evans & P. Chilton (Eds.), *Language, cognition and space: The state of the art and new directions* (pp. 453–478). London: Equinox Publishing.
 Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, *76*(4), 387–404.
 Clark, H. H., Carpenter, P. A., & Just, M. A. (1973). On the meeting of semantics and perception. In W. G. Chase (Ed.), *Visual information processing* (pp. 311–382). Chicago: Academic Press.
 Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407–428.
 Dirven, R., & Taylor, J. (1988). The Conceptualization of Vertical Space in English: The Case of Tall. In B. Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics* (pp. 207–229). Amsterdam: Benjamins.
 Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The Thickness of Musical Pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, Advance online publication. doi:10.1177/0956797612457374.
 Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of vision*, *10*(1), 6–12.
 Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*(1), 25–70.
 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), 726–736.
 Lakoff, G., & Johnson, M. (1980). *Metaphors we live by* (Vol. 111). Chicago London: The University of Chicago Press.
 Landsberg, M. E. (1995). *Syntactic iconicity and linguistic freezes: the human dimension* (Vol. 9). De Gruyter Mouton.
 Lehrer, A. (1985). Markedness and antonymy. *Journal of Linguistics*, *21*(2), 397–429.
 Meier, B. P., & Robinson, M. D. (2004). Why the Sunny Side Is Up. Associations Between Affect and Vertical Position. *Psychological Science*, *15*(4), 243–247.
 Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416–442.
 Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*(2), 113–129.
 Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. *Journal of personality and social psychology*, *89*(1), 1–21.
 Seymour, P. H. (1974). Asymmetries in judgments of verticality. *Journal of Experimental Psychology*, *102*(3), 447–455.
 Shayan, S., Ozturk, O., & Sicoli, M. A. (2011). The thickness of pitch: Crossmodal metaphors in Farsi, Turkish, and Zapotec. *The Senses and Society*, *6*(1), 96–105.
 Stumpf, C. (2006). *Tonpsychologie* (Vol. 1). Leipzig: Hirzel.
 Taylor, J. R. (2002). Near synonyms as co-extensive categories: “high” and “tall” revisited. *Language Sciences*, *25*(3), 263–284.