When Is Language a Window into the Mind? Looking Beyond Words to Infer Conceptual Categories

Kevin J. Holmes (kjholmes@berkeley.edu)
Department of Linguistics, University of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720 USA

Phillip Wolff (pwolff@emory.edu)
Department of Psychology, Emory University, 36 Eagle Row, Atlanta, GA 30322 USA

Abstract

Language is often regarded as a rich source of evidence about the mind. However, a number of findings challenge this position, at least at the level of words: Where languages differ in their lexical distinctions, conceptual differences are not always observed. We ask here how language might serve as a window into the mind despite an apparently loose connection between words and concepts. We propose that prominent conceptual distinctions, though not necessarily captured by individual words, may be revealed by elements of meaning shared by multiple words. Testing this hypothesis in the domain of space, we show that clusters of spatial terms, identified through dimensionality reduction analyses of semantic similarity data, align with conceptual categories spontaneously accessed during the perceptual discrimination of spatial relations. These findings suggest that aspects of semantic structure beyond the level of words may provide considerable insight into the conceptual system. Implications for research on linguistic relativity are discussed.

Keywords: language and thought; word meaning; concepts; semantic structure; space; categorical perception.

Introduction

Many cognitive scientists regard language as a window into the mind (Chomsky, 1975; Lakoff, 1987; Pinker, 2007). Complicating this view, however, is the observation that languages differ dramatically in how they partition the world by name (Malt & Wolff, 2010). Critically, this semantic diversity is not necessarily mirrored by corresponding conceptual diversity: Where languages differ in their lexical distinctions, conceptual differences are not always observed (e.g., Malt et al., 1999; Munnich, Landau, & Dosher, 2001; Papafragou, Hulbert, & Trueswell, 2008). Such findings suggest that, at least at the level of words, language may not be a particularly good window into the mind. In this research, we look beyond individual words to identify other aspects of semantic structure that might prove more tightly connected to the conceptual system. In particular, we propose that prominent conceptual distinctions may be revealed by elements of meaning shared by multiple words. Investigating this hypothesis in the domain of space, we identify clusters of spatial prepositions with similar meanings and assess the extent to which those meanings are spontaneously accessed during the nonlinguistic processing of spatial relations. Our ultimate conclusion will be that language can provide an illuminating window into the mind—if you know where to look.

Dissociations between words and concepts

A large literature documents the pervasiveness of semantic diversity, with cross-linguistic variation in word meaning observed in such disparate domains as artifacts (Malt et al., 1999), spatial relations (Levinson et al., 2003), and number (Frank et al., 2008), among many others. If language is a window into the mind at the level of words, such diversity should also be observed at the conceptual level. That is, speakers of different languages should perform differently on relevant nonlinguistic tasks, in a manner that aligns with the lexical distinctions of their respective languages.

Although this prediction has been supported by a number of studies investigating the Whorfian hypothesis (see Wolff & Holmes, 2011), other studies have shown striking asymmetries in performance on linguistic and nonlinguistic tasks. Malt et al. (1999) found that speakers of English, Spanish, and Chinese differed markedly in how they named a set of common household containers (e.g., bottles, jars, etc.), yet showed remarkable agreement when sorting the objects based on overall similarity. Munnich et al. (2001) observed that English, Japanese, and Korean speakers differed in their naming, but not their memory, of spatial locations. Papafragou et al. (2008) found that English and Greek speakers described motion events differently despite showing similar attentional patterns when viewing the events. Together, these findings suggest that the distinctions picked out by words are not invariably salient at the conceptual level, implying some degree of dissociation between words and conceptual representations.

Several factors might account for this word-concept mismatch. Word meanings are shaped, to a much greater degree than conceptual knowledge, by historical forces such as language contact and past speakers' concerns (Malt, Gennari, & Imai, 2010), and by communicative pressures, such as the need to maximize informativeness and minimize cognitive load (Kemp & Regier, 2012). As a consequence, the words of a language will tend to reflect the language's history and support efficient communication, but may often fail to capture salient conceptual distinctions – despite the long-standing intuition that they should (cf. Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

The semantic clusters hypothesis

Although the factors outlined above render individual words an unreliable guide to conceptual representations, there may be other ways in which language can provide insight into the conceptual system. Several recent approaches combine semantic data from multiple languages, motivated by the idea that cross-linguistically frequent semantic distinctions may be linked to prominent, perhaps even universal, conceptual ones (Malt et al., 2011; Regier, Khetarpal, & Majid, in press). However, a similar idea can be applied to a single language: Elements of meaning that are shared by many words – and hence apply across a wide range of communicative contexts – may be particularly likely to capture key conceptual distinctions. Words that share the same element of meaning can be likened to snapshots of the same underlying concept: No single word will capture the concept on its own, but by examining multiple words with closely related meanings, the concept may emerge (cf. Regier et al., in press). Accordingly, conceptually salient distinctions may be revealed by clusters of related words. We call this proposal the *semantic clusters hypothesis*.

Testing this hypothesis requires (1) identifying clusters of words within a given domain, and (2) assessing their conceptual salience. The first step may be achieved by obtaining a measure of the similarities among all of the words in a domain. A common method for collecting semantic similarity data is to have people divide words into groups based on their meanings (e.g., Wolff & Song, 2003). Words with similar meanings will tend to be grouped together often, while words with dissimilar meanings will rarely be grouped together. These co-occurrences may be combined across participants to construct a similarity matrix, which in turn may be analyzed using dimensionality reduction techniques, such as multidimensional scaling (MDS). Any clusters of words, or *latent categories*, within the semantic similarity space for the domain are likely to be revealed by such techniques.

The second step requires examining the extent to which the latent categories factor into cognitive processes unrelated to language. One way to establish the role of categories in nonlinguistic processing is to show that the category membership of a set of items influences how the items are perceived. Items from different categories are often easier to tell apart than items from the same category, even after controlling for the physical distance between the items – a phenomenon known as *categorical perception* (CP; Goldstone & Hendrickson, 2010). In the case of latent categories, CP could be tested by having people discriminate among items from the domain of interest, with the items coming from either different latent categories or the same latent category. CP would be indicated by superior discrimination on between- compared to within-category trials. Such an effect, if found, would indicate that the latent categories are spontaneously accessed in a nonlinguistic context, providing evidence for their conceptual salience.

We adopted the approach outlined above to test the semantic clusters hypothesis in the domain of space, a perennial battleground in research on the language-thought interface (Li & Gleitman, 2002; Majid et al., 2004). In Experiment 1, participants sorted a large inventory of spatial prepositions into groups, and MDS was used to identify latent categories. Experiment 2 examined the conceptual salience of these categories, using CP as a diagnostic. Recent evidence suggests that CP is stronger in the left hemisphere than the right (Gilbert et al., 2006), even for unnamed categories (Holmes & Wolff, 2012) – consistent with specialization of the left hemisphere for categorical processing independent of language (Kosslyn et al., 1989). Thus, even though the items within a latent category might share no common name, we expected CP for such categories to be left-lateralized.

Experiment 1

The goal of Experiment 1 was to obtain a measure of the semantic structure of the spatial domain, from which clusters of prepositions could be identified.

Method

Participants Sixty-three Emory University undergraduates, all native English speakers, participated for course credit or payment. One participant was excluded for not following instructions.

Materials An inventory of English spatial prepositions was assembled by adapting a comprehensive list from Landau and Jackendoff (1993). Forty-two prepositions were selected from the original list, omitting archaic (e.g., *betwixt*, *without*), intransitive (e.g., *apart*, *downstairs*), non-spatial, (e.g., *ago*, *despite*) and predominantly metaphorical (e.g., *in line with*) prepositions, and those requiring a phrasal verb construction (e.g., *through*, as in "pierce through"). The resulting inventory is shown in Table 1.

Table 1: Spatial prepositions used in Experiment 1.

about	atop	in	past
above	before	in back of	to the left of
across	behind	in front of	to the right of
after	below	inside	to the side of
against*	beneath	near	toward
along	beside	off	under
alongside	between	on	underneath
amid	beyond	on top of	up
among	by	opposite	within
around	down	outside	
at	far from	over	

*excluded from analyses

Each of the prepositions was printed in bold at the top of a 4" × 6" index card. Below each term were two example sentences reflecting prototypical spatial usages of the term.

Procedure The experiment consisted of two phases. In the first phase, participants were presented with the randomly ordered stack of index cards and were asked to write a definition for each preposition based on the two example sentences. The purpose of this task was to encourage participants to think relatively deeply about the meanings of the prepositions. For a subset of participants, the term *against* was inadvertently omitted from the stack of cards; as a result, this term was excluded from analyses.

In the second phase, participants were asked to divide the index cards into as many groups as they felt were appropriate. They were told that the prepositions in each group should have "essentially the same meaning." Participants were given as much time as they needed to complete both phases of the experiment.

Results and discussion

The number of groups of prepositions ranged from 5 to 29 (M = 14.1, SD = 5.8). The raw sorting data were converted into a pairwise similarity matrix, with the similarity between each pair of prepositions taken to be the proportion of participants who grouped them together. For example, if all 62 participants grouped *above* and *below* together, the similarity between them would be 62 ÷ 62 = 1; if 31 participants grouped *above* and *below* together, the similarity between them would be 31 ÷ 62 = .5, and so on.

The similarity matrix was submitted as input to a MDS algorithm, ALSCAL (ordinal model), and solutions of increasing dimensionality were generated. Because the largest decline in stress (a measure of the degree of fit between the actual and estimated inter-item distances) occurred between 1 and 2 dimensions, the 2-dimensional solution (stress = .26) is shown in Figure 1.[1]

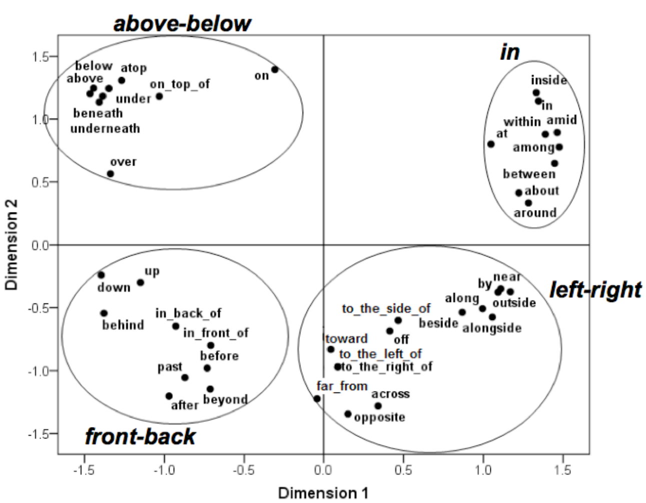

Figure 1: Multidimensional scaling solution of sorting data from Experiment 1. K-means clusters are marked on the solution and labeled for descriptive purposes.

To help identify clusters within the solution, the estimated inter-item distances were combined into a new pairwise similarity matrix. This matrix was then submitted as input to a series of K-means clustering analyses, using increasing values of K (i.e., number of clusters). In these analyses, substantial reduction in within-cluster variance occurred up to K = 4, with only minimal further reduction thereafter. These results suggest that the MDS similarity space is most optimally partitioned into four clusters. These clusters –labeled *above-below*, *front-back*, *left-right*, and *in* – are marked on the solution in Figure 1.[2] Notably, three of the clusters contain words that are essentially opposite in meaning. This suggests that the clusters cannot be reduced to individual word meanings, but instead may be viewed as latent categories.[3] The next experiment investigated the conceptual salience of these categories; that is, the extent to which they play a role in the nonlinguistic processing of spatial relations.

Experiment 2

In Experiment 2, participants were presented with multiple pictures showing spatial relations from the *above-below*, *left-right*, and *front-back* categories.[4] Their task was to decide whether the pictures were perceptually identical or one of the pictures (the target) was different from the others (the distractors). On "different" trials, the target was from either the same category as the distractors (within-category; e.g., *above* vs. *below*) or a different category (between-category; e.g., *above* vs. *left*). CP would be revealed by faster or more accurate performance on between- than within-category trials. Note that because the target and distractors had different names on both within- and between-category trials, CP – if observed – would reflect the influence of categorical rather than linguistic representations. Given evidence that CP for both named and unnamed categories is left-lateralized (Holmes & Wolff, 2012), we expected that latent categories would likewise yield left-lateralized CP. To examine this possibility, the location of the target was varied, with left-lateralized CP indicated by stronger CP effects when the target is presented in the right visual field (RVF; i.e., left hemisphere) than the left visual field (LVF).

[1] The 3-dimensional solution (stress = .17) provided little additional information. The third dimension could be interpreted as reflecting a distinction between metric (e.g., *far from*, *near*) and nonmetric (e.g., *above*, *to the left of*) prepositions, but this dimension also distinguished the four clusters in Figure 1 reasonably well.

[2] Other dimensionality reduction techniques (hierarchical clustering and principal components analysis) yielded similar results, suggesting that the clusters are not an artifact of MDS (Holmes, 2012).

[3] In addition to identifying clusters within the 2-dimensional similarity space, the dimensions themselves may also be interpreted. These dimensions seem to capture broad distinctions among spatial relations in the world. The y-axis reflects a distinction between topological and projective relations (Levinson et al., 2003); most of the prepositions in the *above-below* and *in* clusters refer to relations between contiguous objects, whereas those in the *front-back* and *left-right* clusters specify a frame of reference. The x-axis is less easily interpreted. Several researchers have noted that the *above-below* and *front-back* axes are perceptually asymmetric with respect to canonical body position, whereas the *left-right* axis is perceptually symmetric (e.g., Clark, 1973). However, the *in* cluster is not well captured by this distinction; relations of containment and proximity are not readily characterized in terms of symmetry.

[4] The *in* category was not included because it was the only category that did not contain terms with opposite meanings, making CP more difficult to assess than for the other categories.

Method

Participants Twenty-two Emory University undergraduates, all right-handed native English speakers, participated for course credit or payment. Four participants were excluded, 3 for low accuracy (< 65% correct on "different" test trials) and 1 for a mean reaction time (RT) greater than 2.5 standard deviations above the mean for all participants.

Materials The materials were 12 pictures of a bird and an airplane (see Figure 2). Each picture displayed the objects from one of 3 perspectives (front, side, or top view). There were 4 pictures from each perspective, each showing the bird in a different location. The distance from the bird to the airplane, as determined by their closest edges, was the same across locations.

Figure 2: Stimuli used in Experiment 2. (a) front view: *above*, *below*, *left*, and *right*. (b) side view: *above*, *below*, *front*, and *back*. (c) top view: *left*, *right*, *front*, and *back*.

In the discrimination task, each display consisted of a fixation marker surrounded by 4 pictures, all from the same perspective (see Figure 3). In each picture, the center of the airplane subtended 11.5° (h) × 12.8° (v) visual angle.

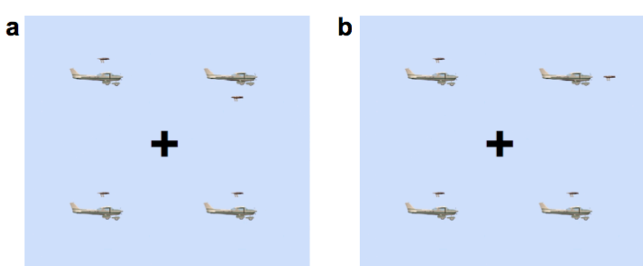

Figure 3: Examples of displays used in Experiment 2. (a) within-category trial (*below* target, *above* distractors). (b) between-category trial (*front* target, *above* distractors).

Design and procedure There were 3 blocks of trials, each consisting of 16 practice trials and 192 test trials. All displays in each block were from a single perspective (front, side, or top). The order of the blocks was counterbalanced.

On half of the test trials in each block, the 4 pictures in the display were identical ("same" trials). This resulted in 4 unique "same" displays in each block, with each display presented 24 times. On the other half, 3 pictures (distractors) were identical and the fourth (target) was different ("different" trials). There were 2 kinds of "different" trials: (a) within-category, in which target and distractors were from the same category (*above-below*, *left-right*, or *front-back*); and (b) between-category, in which target and distractors were from different categories (see Figure 3). Across "different" trials, each picture served as the target at all 4 positions in the display (2 LVF, 2 RVF), resulting in 48 unique "different" displays per block (16 within-category, 32 between-category), each presented twice. Trials were presented in random order.

On each trial, a fixation marker appeared centrally for 500 ms, followed by one of the displays for 200 ms (to discourage eye movements). Participants indicated whether there was an "odd one out" (i.e., target) by pressing the "S" key for same (i.e., no odd one out) or the "D" key for different, using their left and right index fingers, respectively. The next trial began after participants logged a response. Feedback was provided after practice, but not test, trials.

Following the discrimination task, participants wrote a brief description of the relative locations of the bird and airplane in each of the 12 pictures, presented on index cards. This served as a manipulation check to verify that participants interpreted the pictures as showing the spatial relations they were intended to depict.

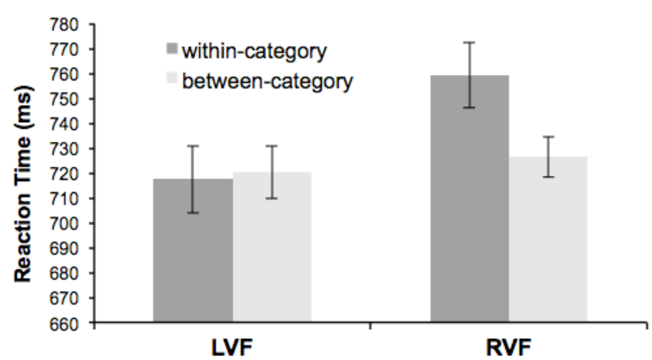

Figure 4: Results of Experiment 2.
Error bars are 95% within-subjects confidence intervals.
LVF = left visual field; RVF = right visual field.

Results and discussion

As shown in Figure 4, the categories *above-below*, *left-right*, and *front-back* elicited CP: Participants were faster to discriminate spatial relations from different categories than from the same category. Notably, this effect was found only in the RVF, indicating that CP was left-lateralized, consistent with a left hemisphere specialization for categorical processing (Kosslyn et al., 1989) and previous left-lateralized CP studies (e.g., Holmes & Wolff, 2012).

On the discrimination task, mean accuracy was 89.1% (*SD* = 7.5), with no difference between "same" and "different" trials ($p > .1$). Subsequent analyses focused on the "different" trials, for which CP could be assessed. Trials in which participants responded incorrectly (12.2%) or RT was greater than 2.5 SD from individual means (2.8%) were excluded. A 2 (visual field: LVF vs. RVF) × 2 (category relation: within- vs. between-category) repeated-measures analysis of variance (ANOVA) on RT for the remaining trials yielded main effects of visual field, $F(1, 17) = 8.83$, $p = .009$, and category relation, $F(1, 17) = 8.53$, $p = .01$, and an interaction, $F(1, 17) = 11.60$, $p = .003$. Participants responded faster on between- than within-category trials with RVF targets, $t(17) = 5.17$, $p < .0001$, but not LVF targets, $t(17) = .34$, $p > .7$ (see Figure 4), indicating left-lateralized CP. An analogous ANOVA on the accuracy data yielded no significant effects (all ps > .05), suggesting that there was no speed-accuracy tradeoff.

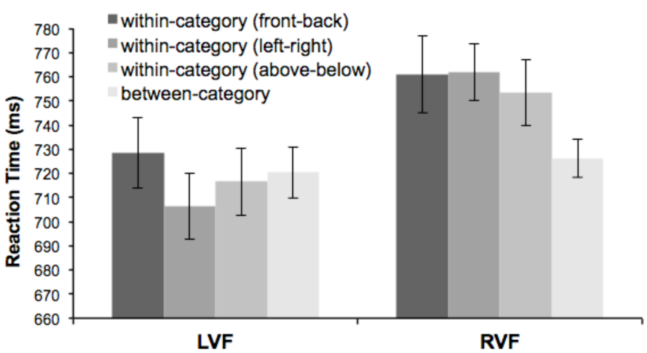

Figure 5: Results of Experiment 2 by category.

Planned comparisons revealed left-lateralized CP for each of the 3 categories assessed. Within-category trials were divided according to the category membership of target and distractors: *above-below* (i.e., *above* target and *below* distractors, or *below* target and *above* distractors), *left-right*, and *front-back*. For RVF targets, discrimination was faster on between-category trials than on each of the 3 kinds of within-category trials [*above-below*: $t(17) = 2.17$, $p = .04$; *left-right*: $t(17) = 2.66$, $p = .02$; *front-back*: $t(17) = 2.13$, $p = .05$; see Figure 5]. For LVF targets, none of these differences reached significance (ps > .2).

The results of the picture description task were as expected. Across the 12 pictures, 87.5% of the descriptions included prepositions from the intended category (e.g., "the bird is above the plane" or "the plane is below the bird" for *above* pictures); for 6 of the pictures, there was 100% agreement. The descriptions of the 4 *front-back* pictures were the most variable. Six participants consistently described these pictures using horizontal or vertical terms (e.g., "to the right of" for the *front* picture in Figure 2b), implying that they viewed them as 2-dimensional. Importantly, however, any ambiguity in the stimuli could not itself account for left-lateralized CP, as the existence of multiple interpretations for a given picture would presumably lead to slower responses on both within- and between-category trials (if not more so for the latter, given that individual participants occasionally used the same preposition to describe pictures from different categories, but never the same category). In addition, none of the 216 descriptions included superordinate terms (e.g., "horizontal" for *left-right* pictures), suggesting that left-lateralized CP was not driven by linguistic representations.

In sum, the results of Experiment 2 provide evidence for the conceptual salience of the *above-below*, *left-right*, and *front-back* categories identified in Experiment 1. The findings support the semantic clusters hypothesis in showing that clusters of related words align with conceptual categories that are spontaneously accessed during nonlinguistic processing, with consequences for simple perceptual decisions.

General Discussion

Recent research on the language-thought interface has led to a paradox. Although language has long been viewed as a window into the mind, a number of studies have suggested that word meanings may often be dissociated from conceptual representations. The present research offers a potential resolution to this discrepancy: Language may be a better reflection of the conceptual system at the level of clusters of words than at the level of individual words. According to the semantic clusters hypothesis, clusters of words capture salient conceptual distinctions. Consistent with this hypothesis, groups of spatial prepositions that clustered together in a semantic similarity space in Experiment 1 aligned with conceptual categories that yielded CP in Experiment 2. These findings suggest that language and the conceptual system may share a common underlying structure that is obscured when focusing solely on individual word meanings and their conceptual analogues. In related work, we have shown that clusters of words elicit *stronger* CP effects than individual words (Holmes, 2012), further supporting our conclusions.

We used CP as a tool for assessing conceptual salience, but our findings also inform the nature of CP itself. In particular, our findings show that CP can occur in the perception of relations, not just in the perception of objects or object properties – the focus of the vast majority of previous CP research (Goldstone & Hendrickson, 2010). Our findings also lend support to the generality of left-lateralized CP effects, recently contested in the domain of color (e.g., Witzel & Gegenfurtner, 2011), though they challenge the dominant linguistic interpretation of such effects (cf. Gilbert et al., 2006). Given that left-lateralized CP occurred for categories whose members share no common label, the phenomenon appears to be driven by categories rather than their names (Holmes & Wolff, 2012), despite the propensity to link left hemisphere processing to language.

Although our findings demonstrate a clear connection between language and the conceptual system, they do not address the origins of this connection, including the possibility that language is the causal agent. The spatial clusters identified here, though shown to be conceptually salient in the minds of English speakers, are not necessarily universal. In principle, the clusters might vary cross-linguistically, and those differences could lead to differences in nonlinguistic spatial processing. On the one hand, such differences seem unlikely because the elements of meaning associated with the clusters presumably reflect, and are constrained by, structure in the world to a much greater extent than are individual word meanings (see Malt et al., 2010). On the other hand, striking cross-linguistic differences in the meanings of spatial terms have been documented (e.g., Levinson et al., 2003), suggesting the possibility of further differences at deeper levels of semantic structure.

If such differences exist, an investigation of analogous conceptual differences would provide a strong test of the semantic clusters hypothesis. Though agnostic with respect to the universality of semantic structure, the hypothesis would predict that clusters of words within a particular language should be conceptually salient for speakers of that language. Thus, speakers of languages with different sets of clusters should show correspondingly different patterns of CP for those clusters. Notably, this kind of Whorfian effect would be unlike any previously reported, in that it would be driven by categories not explicitly encoded in the semantic system – and of which many language users likely have no conscious awareness. Probing the existence of such effects may represent the next frontier in research on linguistic relativity.

Acknowledgments

We are grateful to Larry Barsalou, Harold Gouzoules, Stella Lourenco, Laura Namy, Marjorie Pak, and Terry Regier for helpful feedback, Ted Guio for graphic design, and Tyler Brown and Lee Kugelmann for help with transcribing and coding descriptions. This research was supported by an APA Dissertation Research Award to KJH.

References

Chomsky, N. (1975). *Reflections on language*. New York: Pantheon.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*, 819-824.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *PNAS*, *103*, 489-494.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *WIREs Cognitive Science*, *1*, 69-78.

Holmes, K. J. (2012). *Language as a window into the mind: The case of space*. Doctoral dissertation, Department of Psychology, Emory University.

Holmes, K. J., & Wolff, P. (2012). Does categorical perception in the left hemisphere depend on language? *Journal of Experimental Psychology: General*, *141*, 439-443.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*, 1049-1054.

Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J., & Gabrieli, J. D. E. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 723-735.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*, 217-265.

Levinson, S. C., Meira, S., & The Language and Cognition Group (2003). 'Natural concepts' in the spatial topological domain—Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*, *79*, 485-516.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*, 265-294.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*, 108-114.

Malt, B. C., Ameel, E., Gennari, S., Imai, M., Saji, N., & Majid, A. (2011). Do words reveal concepts? *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 519-524). Austin, TX: Cognitive Science Society.

Malt, B. C., Gennari, S., & Imai, M. (2010). Lexicalization patterns and the world-to-words mapping. In B. C. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience*. New York: Oxford University Press.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*, 230-262.

Malt, B. C., & Wolff, P. (Eds.) (2010). *Words and the mind: How words capture human experience*. New York: Oxford University Press.

Munnich, E., Landau, B., & Dosher, B. A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition*, *81*, 171-207.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*, 155-184.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York: Penguin Books.

Regier, T., Khetarpal, N., & Majid, A. (in press). Inferring semantic maps. *Linguistic Typology*.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.

Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, *11*.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *WIREs Cognitive Science*, *2*, 253-256.

Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology*, *47*, 276-332.