Distribution Agreement

Signature:

Kevin J. Holmes	Date

Language as a Window into the Mind:
The Case of Space

By

Kevin J. Holmes
Doctor of Philosophy

Psychology

Phillip Wolff
Advisor

Lawrence W. Barsalou
Committee Member

Harold Gouzoules
Committee Member

Stella F. Lourenco
Committee Member

Laura L. Namy
Committee Member

Marjorie Pak
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Language as a Window into the Mind:
The Case of Space

By

Kevin J. Holmes
B.A., M.A., Stanford University, 2005

Advisor: Phillip Wolff, Ph.D.

An abstract of
a dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2012

Abstract

Language as a Window into the Mind:
The Case of Space

By Kevin J. Holmes

Many cognitive scientists regard language as a rich source of evidence about the human mind. Much research over the past forty years has been driven by the assumption that words reveal underlying concepts. At the same time, cross-linguistic work has shown that languages differ dramatically in how they partition the world by name. To maintain the premise that words align with concepts, this linguistic diversity would have to be mirrored by corresponding conceptual diversity, consistent with the Whorfian hypothesis that language shapes thought. However, a number of recent findings are incompatible with this position: Where languages differ in their word meanings, conceptual differences are often lacking. Such evidence calls into question the notion that words are a direct route to concepts. In this dissertation, I examine how language might serve as a window into the mind despite the lack of alignment between words and concepts. In particular, I propose that similarities in meaning across multiple words, identifiable as cohesive clusters within the semantic structure of a domain, map onto prominent conceptual distinctions. I call this proposal the *semantic clusters hypothesis*. According to this hypothesis, language is a better reflection of the conceptual system at the level of clusters of words than at the level of words themselves.

A series of five experiments used space as a test bed for investigating the semantic clusters hypothesis. In these experiments, clusters of spatial terms identified through dimensionality reduction analyses of semantic similarity data (Experiment 1) aligned with conceptual distinctions influential in the nonlinguistic processing of spatial relations (Experiments 2-3). Further, clusters that were more differentiated at the semantic level were also more salient at the conceptual level (Experiments 4-5). These findings suggest that despite the failures of individual words to reveal concepts, aspects of semantic structure beyond the level of words may provide an illuminating window into the mind. The contributions of *macrosemantics*, the approach to meaning exemplified by the present research, are discussed with respect to ongoing debates on the relationship between language and thought.

Language as a Window into the Mind:
The Case of Space

By

Kevin J. Holmes
B.A., M.A., Stanford University, 2005

Advisor: Phillip Wolff, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2012

Acknowledgments

This dissertation was made possible by the incredible opportunities afforded to me as a member of the psychology community at Emory, and by the tremendous support that I have received from so many people over the course of my intellectual journey thus far. Above all, I am grateful to my advisor, Phillip Wolff, for helping me shape my ideas, for pushing me to think broadly and boldly, for showing me the art in science, and for encouraging me in all aspects of my work. Having prime access to Phil's remarkable mind these past five-plus years is not something that will be easy to give up. I am also indebted to Stella Lourenco for being a wonderful mentor and role model. Stella's intellect, ingenuity, and passion for science are inspiring, and I secretly want to be her when I grow up. I owe a very special debt to Laura Namy for making me feel at home from my first day at Emory, for helping me navigate the bumps in my road, and for always looking out for me both professionally and personally.

Before my time at Emory, several people were instrumental in guiding me along the path to graduate school. I thank Anne Fernald, Natasha Kirkham, and especially Michael Ramscar at Stanford for introducing me to cognitive science, for cultivating my burgeoning interest in the study of language and mind, and for teaching me the tools of the trade. In the years between my undergraduate and graduate studies, my friendships with Valorie Salimpoor and Lauren Clepper enriched my thinking and broadened my personal horizons, blurring the line between work and play in the best possible way. A heartfelt thanks to these academic siblings and dear friends, who have had a more profound impact on me than they will ever know.

At Emory, a number of faculty members have been incredibly generous with their time and have supported me in countless ways. I thank Larry Barsalou, Harold Gouzoules, Rob Hampton, Lynne Nygaard, Marjorie Pak, and Philippe Rochat, each of whom has helped me grow as a scholar. I am also grateful to all the members of the Cognitive and Linguistic Systems Lab over the years ("our CLS family") for their dedication, curiosity, and good humor. I especially thank Jason Shepard for many stimulating conversations, and several outstanding undergraduates for their contributions to this dissertation research. Nirvi Shah helped me get the project off the ground, Bianca Copello took it in new and exciting directions, Ted Guio created the initial perceptual stimuli, and Tyler Brown and Lee Kugelmann transcribed and coded picture descriptions. In completing this dissertation, I have benefited from the support of an APA Dissertation Award, funding from the Laney Graduate School, and the invaluable assistance and generosity of Paula Mitchell and the rest of the Emory Psychology staff.

I have been immensely lucky to share my graduate school experience with Jeni Pathman, Jenni Stevens, Umay Suanda, and Malini Suchak, all wonderful friends and colleagues from the very beginning, and Sabrina Sidaras, Christy Wilson-Mendenhall, Kile Ortigo, and Christina Tzeng, whose camaraderie, advice, and support were much appreciated along the way. I also want to thank my Atlanta tennis buddies for all the fun and excitement we've had on and off the court, and for keeping me sane and grounded in the world outside academia.

I don't know where I would be without the amazing Brett Kinney, who hopped aboard this journey with me and gave me a new, deeper purpose for everything I do.

Every day with Brett is full of learning, wonder, and sheer delight. I am so very lucky to have him in my life, and I can’t wait to see where our next adventures take us.

Finally, I thank my family, both living and no longer with us. My mother, Louise Holmes, worked tirelessly to give me the best opportunities in education and in life. I couldn’t have made it this far without her unwavering love, support, and belief in me. I dedicate this dissertation to her and to the memory of my father, John C. Holmes, and my grandmother, Siu Dip Soong, who would have been so proud to see me reach this point.

Table of Contents

Chapter

1 Introduction **1**
- 1.1 Language as a window into the mind 1
- 1.2 Words and the world 2
- 1.3 The problem of linguistic diversity 4
- 1.4 Dissociations between words and concepts 6
- 1.5 Why are words a bad window on concepts? 8
- 1.6 The semantic clusters hypothesis 10
  - 1.6.1 The differentiation principle 12
  - 1.6.2 Testing the hypothesis 14
- 1.7 Space as a test bed 16
  - 1.7.1 Overview of the dissertation and empirical predictions 16

2 Inferring semantic structure **18**
- 2.1 The semantics of space 18
- 2.2 Dimensionality reduction methods 20
- 2.3 Experiment 1: Sorting prepositions 23
  - 2.3.1 Method 24
    - 2.3.1.1 Participants 24
    - 2.3.1.2 Materials 24
    - 2.3.1.3 Procedure 25
  - 2.3.2 Results and discussion 26
    - 2.3.2.1 Multidimensional scaling 27
    - 2.3.2.2 *K*-means clustering 32
    - 2.3.2.3 Hierarchical clustering 34
    - 2.3.2.4 Principal components analysis 36
    - 2.3.2.5 General conclusions 40
    - 2.3.2.6 Correspondence with individual data 41
    - 2.3.2.7 Relative differentiation of the clusters 44
- 2.4 Summary 46

3 Assessing conceptual salience 47

3.1 Interpreting semantic structure 47

3.2 Categorical perception as a diagnostic of conceptual salience 49

3.2.1 Words and CP 50

3.3 Using CP to test the semantic clusters hypothesis 52

3.4 Experiment 2: CP for *above-below*, *left-right*, and *front-back* 53

3.4.1 Method 54

3.4.1.1 Participants 54

3.4.1.2 Materials 55

3.4.1.3 Design and procedure 56

3.4.2 Results and discussion 57

3.4.2.1 Overview of key findings 57

3.4.2.2 Discrimination task 58

3.4.2.3 Picture description task 61

3.4.2.4 Summary and conclusions 61

3.5 Experiment 3: CP for *above* and *below* 62

3.5.1 Method 63

3.5.1.1 Participants 63

3.5.1.2 Materials 63

3.5.1.3 Design and procedure 64

3.5.2 Results and discussion 64

3.5.2.1 Overview of key findings 64

3.5.2.2 Discrimination task 65

3.5.2.3 Picture description task 68

3.5.2.4 Summary and conclusions 68

3.6 Experiment 4: CP for *above-below* vs. CP for *above* and *below* 70

3.6.1 Method 71

3.6.1.1 Participants 71

3.6.1.2 Materials 71

3.6.1.3 Design and procedure 72

3.6.2 Results and discussion 73

3.6.2.1 Overview of key findings 73

3.6.2.2 Discrimination task 74

3.6.2.3 Picture description task 78

3.6.2.4 Summary and conclusions 79

- 3.7 Experiment 5: CP for *left-right* vs. CP for *left* and *right* 80
  - 3.7.1 Method 81
    - 3.7.1.1 Participants 81
    - 3.7.1.2 Materials 81
    - 3.7.1.3 Design and procedure 83
  - 3.7.2 Results and discussion 83
    - 3.7.2.1 Overview of key findings 83
    - 3.7.2.2 Discrimination task 84
    - 3.7.2.3 Picture description task 88
    - 3.7.2.4 Summary and conclusions 88
- 3.8 Overall summary 89

4 General discussion **91**

- 4.1 When language is a window into the mind—and when it isn't 91
  - 4.1.1 Further contributions of the present investigation 92
  - 4.1.2 Limitations and benefits of dimensionality reduction 94
- 4.2 The inevitable Whorfian question 95
- 4.3 Extending the differentiation principle 97
- 4.4 Implications for related areas of research 99
  - 4.4.1 Spatial cognition through the lens of spatial language 99
  - 4.4.2 Categorical perception of spatial relations 101
- 4.5 In defense of concepts 102

References **104**

Appendix **121**

List of Tables

Table

1 Spatial prepositions used in Experiment 1 25

2 Clusters of prepositions yielded by *K*-means clustering analysis 33

3 Results of principal components analysis with varimax rotation 38

4 Component loadings from principal components analysis 39

5 Core members of clusters yielded by dimensionality reduction analyses 42

List of Figures

Figure

1 Patterns of relationships among words, concepts, and the world 4
2 Semantic clusters hypothesis 11
3 Scree plot from multidimensional scaling analysis 28
4 Multidimensional scaling solution of similarity data 29
5 Scree plot from *K*-means clustering analysis of MDS inter-item distances . . 30
6 Scree plot from *K*-means clustering analysis of original similarity data 32
7 Dendrogram from hierarchical clustering analysis 35
8 Scree plot from principal components analysis 37
9 Schematic diagram of spatial semantic structure 40
10 Stimuli used in Experiment 2 54
11 Examples of displays used in Experiment 2 55
12 Trial structure in Experiments 2–5 57
13 Results of Experiment 2 58
14 Results of Experiment 2 by category 60
15 Stimuli used in Experiments 3 and 4 63
16 Results of Experiment 3 65
17 Results of Experiment 3 by category 67
18 Additional stimuli used in Experiment 4 71
19 Results of Experiment 4 74
20 Results of Experiment 4 by specific category 77
21 Stimuli used in Experiment 5 82
22 Results of Experiment 5 84
23 Results of Experiment 5 by lexical category 87

Chapter 1: Introduction

1.1 Language as a window into the mind

Human cultures differ in a myriad of ways. Sometimes our differences can seem so vast as to imply incommensurate views of the world. Yet when we encounter the language of another culture, we are confronted with the reality that we are all, at some fundamental level, the same. Words like *meraki* (Greek for “doing something with soul, creativity, or love”), *toska* (Russian for “the sensation of great spiritual anguish”), and *jayus* (Indonesian for “a joke told so poorly that one cannot help but laugh”) capture ideas that may be difficult to articulate in our native tongue, but are undeniably universal. The world’s languages inevitably seem to reflect deep truths about the human mind and human experience.

For many cognitive scientists, the notion that language is a window into the mind is a guiding maxim (e.g., Chomsky, 1975; Lakoff, 1987; Pinker, 2007). Much research over the past forty years has been driven by the intuition that words pick out cognitively preindividuated chunks of experience—“intrinsically separate things” (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976, p. 383) or “natural partitions” (Gentner, 1982). This assumption implies that words directly reveal underlying concepts, which in turn reflect structure inherent in the world. But cross-linguistic work raises a challenge for this view: Languages differ dramatically in how they partition the world by name (Evans & Levinson, 2009; Malt & Wolff, 2010). To maintain the premise that words align with concepts, this linguistic diversity would have to be mirrored by corresponding conceptual diversity, with concepts only loosely connected to structure in the world. However, a number of recent findings are incompatible with this position: Where languages differ in their word meanings, conceptual differences are often lacking (e.g., Malt, Sloman, Gennari, Shi, & Wang, 1999; Munnich, Landau, & Dosher, 2001). Such evidence calls into question the notion that words are a direct route to concepts. In this dissertation, I examine how language might serve as a window into the mind despite the lack of alignment between words and concepts. In particular, I propose that similarities in meaning across multiple words, identifiable as cohesive clusters within the semantic structure of a domain, map onto prominent conceptual distinctions. I call this proposal the *semantic clusters hypothesis*. According to this hypothesis, language is a better reflection of the conceptual system at the level of clusters of words than at the level of words themselves. In the chapters that follow, I use the domain of space as a test bed for investigating the semantic clusters hypothesis. My ultimate conclusion will be that language can offer an illuminating window into the mind—if you know where to look.

1.2 Words and the world

The idea that word meanings are linked to structure in the world is perhaps most associated with the work of Rosch and colleagues in the 1970s. Rosch and Mervis (1975) noted that features in the world are not distributed in a random fashion, but instead tend to be correlated. For example, the features “has four legs,” “has fur,” “has a tail,” and “barks” frequently occur together, and objects possessing these features are called *dog*. Rosch et al. (1976) highlighted this kind of correlational structure as the driving force behind the so-called “basic” or preferred level of categorization for objects. According to Rosch et al., basic-level categories are the most differentiated: Members of the same basic-level category share many features, but have relatively few features in common with members of other categories (see also Mervis & Crisafi, 1982; Murphy & Brownell, 1985). Moreover, objects are typically named at the basic level, suggesting that our most frequently used words are those that capture salient structure in the world ("intrinsically separate things"; for a similar proposal, see Berlin, 1978).

Gentner (1982) expanded on this idea with her natural partitions hypothesis, which holds that common nouns pick out "highly cohesive collections of percepts" and that "children learning language have already isolated these cohesive packages… from their surroundings" (p. 324). This proposal implies that concepts mediate the relationship between words and the world: Structure in the world guides the concepts we form, and words map straightforwardly onto these concepts.[1] This pattern of relationships, which I call the *standard view*, is shown in Figure 1a. Much work on adult concepts and conceptual development seems to take this view as a given (e.g., Landau & Jackendoff, 1993; Mahon & Caramazza, 2009; Mandler, 2008; Murphy, 2002; Rogers & McClelland, 2004; Waxman & Gelman, 2010).

---

[1] The natural partitions hypothesis posits that structure in the world is less cohesive for the referents of relational terms, such as verbs and prepositions. As a consequence, the meanings of these terms are expected to show greater cross-linguistic variability than the meanings of nouns. Nevertheless, the hypothesis assumes that when the world presents salient structure to its observers, concepts and words will reflect this structure (for further discussion, see Gentner & Boroditsky, 2001; Malt, Gennari, & Imai, 2010).

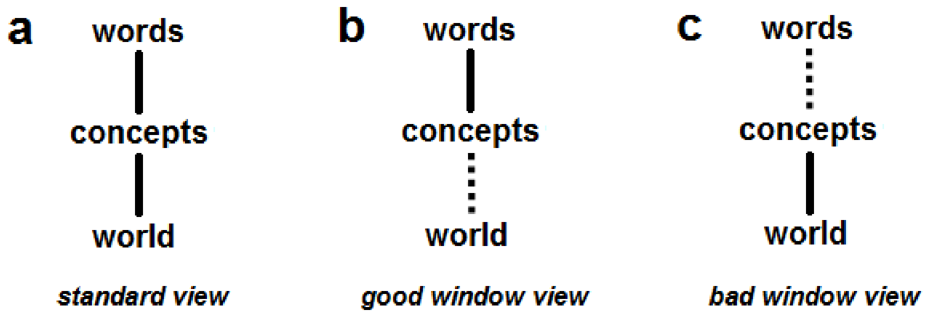

Figure 1. Patterns of relationships among words, concepts, and the world. (*A*) According to the standard view, both the relationship between words and concepts and the relationship between concepts and the world are tight. (*B*) The good window view holds that only the relationship between words and concepts is tight, while the relationship between concepts and the world is loose. (*C*) The bad window view suggests the opposite pattern. Portions of this figure were adapted from Figure 2 in Wolff and Holmes (2011).

1.3 The problem of linguistic diversity

If words reflect structure in the world, one might expect to find little divergence in the meanings of words across languages. In reality, though, cross-linguistic variation in word meaning is pervasive. Such variation has been documented in a wide range of domains, including color (Kay & Regier, 2003; Roberson, Davies, & Davidoff, 2000), household objects (Malt et al., 1999), body parts (Majid, Enfield, & van Staden, 2006), motion (Malt et al., 2010, 2011; Slobin, 1996a), spatial relations (Bowerman & Choi, 2001; Levinson et al., 2003), causal relations (Wolff, Jeon, & Li, 2009), and number (Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004), and in the semantic categories defined by grammatical morphemes such as gender markers (Sera et al., 2002; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005) and numeral classifiers (Saalbach & Imai, 2012). Variation occurs not only in domains for which structure in the world is unavailable or lacking in coherence (cf. Gentner, 1982; Gentner & Boroditsky, 2001), but also in perceptually rich domains labeled by common nouns (Majid et al., 2006; Malt et al., 1999). Moreover, languages do not vary only in the granularity with which they partition a given domain (cf. Berlin & Kay, 1969) or in the boundaries of their lexical categories (cf. Roberson et al., 2000); in some domains, different languages make cross-cutting distinctions (Bowerman & Choi, 2001; Malt et al., 1999).

This linguistic diversity challenges the standard view of the relationships among words, concepts, and the world (see Figure 1a). Assuming that structure in the world presents itself in much the same way across language groups, the existence of linguistic diversity implies that words do not necessarily reflect this structure. Nevertheless, part of the standard view—namely, the intuition that words provide a good window on concepts—could be maintained if the linguistic diversity were mirrored by corresponding conceptual diversity. Such an alignment would preserve the tight connection between words and concepts, but would imply that concepts are not isomorphic with structure in the world. The new pattern of relationships suggested by this proposal, which I call the *good window view*, is shown in Figure 1b.

How might speakers of different languages come to have different concepts? One possibility is that concepts shape words. As noted by Malt et al. (2010), certain needs, interests, or experiences might lead members of a culture to make relatively fine-grained conceptual distinctions within a domain and, in turn, to lexicalize those distinctions.[2] Another possibility is that words shape concepts. Under this scenario, synonymous with the Whorfian hypothesis (Whorf, 1956), the distinctions picked out by the words of one's native language come to be reflected in one's concepts. This process is perhaps best described from a developmental perspective: Young children, through the process of language acquisition, acquire the concepts reinforced by their native language; given that languages differ, children learning different languages will acquire different concepts (Imai & Gentner, 1997; McDonough, Choi, & Mandler, 2003).

1.4 Dissociations between words and concepts

The pattern of relationships suggested by the good window view—whether the product of concepts shaping words, words shaping concepts, or both—makes a strong empirical prediction: Speakers of different languages should differ in their performance on nonlinguistic cognitive tasks, in a manner that aligns with the lexical distinctions of their respective languages. This prediction has been supported by a number of recent findings from the Whorfian literature (for a review, see Wolff & Holmes, 2011). However, several other recent studies have shown that differences in word meaning across languages are not always paralleled by conceptual differences. Such findings challenge the idea that words are a good window on concepts.

---

[2] Several attested cases of linguistic diversity can be explained by such cultural factors (see Malt, 1995), but many cannot. For example, cross-linguistic variation in the relative use of different spatial frames of reference (e.g., *north*/*south* vs. *left*/*right*) is not readily predicted by ecological differences in the environments of different language groups (Majid, Bowerman, Kita, Haun, & Levinson, 2004). Thus, the possibility that concepts shape words cannot, by itself, account for the full range of documented linguistic diversity (Malt et al., 2010).

In one study, Malt et al. (1999) found that speakers of English, Spanish, and Chinese differed markedly in how they named a collection of common household containers (e.g., bottles, jars, boxes, etc.). For example, the objects that English speakers called *bottle* were named differently by Spanish speakers; some were called *frasco*, others *botella*, and still others *mamadera*. Notably, the Spanish speakers did not simply partition the English *bottle* category more finely: Some of the objects called *frasco* in Spanish were called *container* or *jar* in English, not *bottle*. However, despite this variation in naming, the three language groups showed remarkable agreement when sorting the containers on the basis of overall similarity. These results suggest that although speakers of English, Spanish, and Chinese carve up the container domain differently by name, they think about the objects in much the same way (see also Ameel, Storms, Malt, & Sloman, 2005).

The same kind of incongruity between words and the conceptual system has been observed in several other domains. Munnich et al. (2001) found that English, Japanese, and Korean speakers differed in their naming of various spatial locations, but not in their memory for the same locations. Similarly, Papafragou and colleagues showed that English and Greek speakers describe motion events differently, but remember and judge the similarity of the events comparably (Papafragou, Massey, & Gleitman, 2002; see also Gennari, Sloman, Malt, & Fitch, 2002), and allocate attention in similar ways when viewing such events (Papafragou, Hulbert, & Trueswell, 2008).[3] Analogous dissociations have been observed in studies investigating the conceptual correlates of grammatical gender (Sera et al., 2002; Vigliocco et al., 2005) and numeral classifier (Gao & Malt, 2009; Saalbach & Imai, 2007) systems.

The findings from these studies are incompatible with the good window view (and the standard view) because they show that distinctions captured by words are not always salient at the conceptual level. Convergence in nonlinguistic task performance across language groups implies that conceptual organization is more tied to structure in the world than to the lexical partitioning of the domain under investigation. In other words, these findings suggest the opposite pattern of relationships to that of the good window view: a tight connection between concepts and the world, and a loose connection between words and concepts. This inverted pattern, which I call the *bad window view*, is shown in Figure 1c.

1.5 Why are words a bad window on concepts?

Contrary to the other two views, the bad window view suggests that there is no simple alignment between words and concepts. Malt and colleagues (Malt et al., 1999, 2010; Malt & Sloman, 2004; Malt, Sloman, & Gennari, 2003) have identified several factors that might account for this mismatch, highlighting forces presumed to shape word meanings while leaving conceptual knowledge intact. One possible factor is the cultural

---

[3] In the Whorfian literature, it has sometimes been argued that tasks relying on memory and similarity judgments are poor indices of conceptual structure because they invite participants to adopt explicit, often linguistic, strategies (e.g., Pinker, 1994; Winawer et al., 2007). However, this criticism works against finding a dissociation between words and concepts: If participants adopt linguistic strategies, their performance on such tasks should align with their naming patterns (for further discussion of this issue, see Section 3.1).

needs, interests, and experiences of past, rather than current, speakers of a language. Given that words are passed down from one generation to the next, the word meanings of different languages may carry the imprint of past cultural differences even if present cultural conditions are comparable. Another factor is language contact. When one language acquires words from another, the boundaries of existing word meanings may shift to accommodate them. The invariably haphazard circumstances by which languages cross paths may lead to some degree of linguistic diversity. Yet another factor has to do with the structural characteristics of a language. Different languages have different syntactic and morphological systems, which may constrain the number of possible lexical distinctions in a domain (e.g., Malt et al., 2008). In short, word meanings may be subject to a host of historical and linguistic influences not directly tied to the conceptual system—or, as Gleitman and Papafragou (2005) put it, "a generous dollop of arbitrariness" (p. 638).

Although such factors might explain why selected word meanings fail to align with relevant conceptual knowledge, they seem unlikely to provide a complete account of the mismatch between words and concepts. There may be other, more fundamental reasons for this mismatch. A key observation, I believe, is that words serve a communicative function, and hence their meanings necessarily reflect the requirements for efficient communication (Malt et al., 1999). Among these requirements is the need to be maximally informative while simultaneously making the best use of limited cognitive resources (Fedzechkina, Jaeger, & Newport, 2012; Kemp & Regier, 2012). As a consequence of these pressures, word meanings will be well suited to communicative purposes, but sometimes at the expense of directly mirroring the conceptual system.

Consider, for example, a simple communication between speaker and listener, each of whom possesses concepts that reflect structure in the world, as assumed under the bad window view. To maximize the informativeness of the communication and to avoid wasting cognitive resources, the speaker might select words that are not redundant with the listener's preexisting concepts, but that provide additional information not as easily apprehended by our perceptual and cognitive systems. In this case, the speaker's words would make finer distinctions than those of the "natural partitions" (Gentner, 1982) that structure our conceptual knowledge, leading to a mismatch between words and concepts.

1.6 The semantic clusters hypothesis

Despite the evidence that words and concepts are dissociated, and the many possible factors driving this mismatch, the notion of language as a window into the mind need not be abandoned altogether. Indeed, the communicative pressures discussed above suggest that word meanings are not completely divorced from concepts; often, they may simply be more specific. This raises an interesting possibility: Words with similar meanings might, together, converge on the more global meaning associated with an underlying concept. On this view, individual words can be likened to snapshots of the same concept from different angles (see the left side of Figure 2). No single word will fully capture the concept, but by examining similarities in meaning across multiple words, the global meaning may emerge (see the right side of Figure 2). From this idea, hereafter referred to as the semantic clusters hypothesis, it follows that clusters of related words should map onto prominent conceptual distinctions.

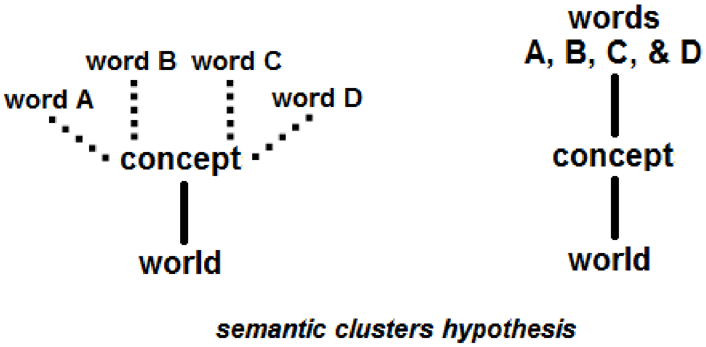

Figure 2. According to the semantic clusters hypothesis, individual words are only loosely connected to concepts, but clusters of words are tightly connected, providing a good window on concepts.

The idea that multiple words might provide insight into concepts is echoed by other recent approaches in the literature. These approaches scale the snapshot metaphor up to the level of entire languages: Each language is assumed to provide a snapshot of conceptual universals. Regier, Kay, and Khetarpal (2007) showed that the color naming systems of 110 different languages constitute near-optimal partitions of a standard perceptual color space, presumably reflecting universal constraints on color perception (see also Khetarpal, Majid, & Regier, 2009). Regier, Khetarpal, and Majid (2012) combined the naming patterns for spatial relations from nine historically distant languages to generate a semantic map, a graph-based representation of conceptual structure likewise presumed to be universal. Malt et al. (2011) pooled naming data from speakers of English, Dutch, Spanish, and Japanese for instances of human locomotion, and found that the aggregate captured the biomechanical discontinuity between walking and running gaits better than data from any of the individual languages.

These cross-linguistic approaches provide substantial insight into the kinds of conceptual distinctions that might be universally shared across languages and cultures. However, there is a notable limitation to these approaches. Although it is assumed that the distinctions identified through analyses of linguistic data constitute actual concepts possessed by speakers of individual languages, and that those concepts are more salient than other possible distinctions (e.g., those that are less cross-linguistically prevalent), these assumptions are not directly tested. The semantic clusters hypothesis differs from these approaches in suggesting a method for identifying candidate concepts using the words of a single language, and for assessing their salience empirically.[4]

1.6.1 The differentiation principle

In order to test the semantic clusters hypothesis, there must be a way of identifying clusters of words likely to align with concepts. The words of a domain may be grouped into categories at many different levels of abstraction, from the most general (e.g., all verbs expressing locomotion) to the most specific (e.g., verbs for furtive walking-type actions: *sneak*, *skulk*, *prowl*, etc.). At this point, it may be helpful to clarify a potential point of confusion. Those familiar with the concepts and categorization literature might believe that the level of abstraction associated with our most salient concepts has already been identified: the basic level (Rosch et al., 1976). However, studies investigating the basic level have focused exclusively on categories associated

---

[4] Hereafter, I use the terms “concept” and “conceptually salient” interchangeably. Conceptual distinctions may vary in their degree of salience, and salience will surely vary with context. It is unclear at what point a given distinction is sufficiently salient to be worthy of the designation “concept” (see Section 4.5 for discussion of how the present research may inform theories of the nature of concepts).

with individual words (e.g., *chair*, *piano*, *car*, etc.; cf. Murphy, 2002) and used linguistic tasks (e.g., object naming) to establish their psychological reality. Thus, the findings from such studies provide evidence for a privileged level of linguistic categorization (naming), but not of nonlinguistic conceptualization ("knowing")—a distinction that has only recently been fully appreciated (see Malt et al., 1999). Overlooked in the literature on basic level categorization are semantic categories that have no ready name (e.g., furtive walking-type actions). Because these multi-word categories are not explicitly coded in the semantic system, I will refer to them as *latent categories*.[5] The semantic clusters hypothesis predicts that certain latent categories will align better with concepts than the categories named by individual words, even basic-level ones.

Which latent categories align with concepts? Despite its limitations, the literature on basic level categorization offers a potentially powerful diagnostic. Recall that the basic level is defined as the most *differentiated*, maximizing similarity within categories relative to similarity across categories (Mervis & Crisafi, 1982; Murphy & Brownell, 1985; Rosch et al., 1976). This principle may be extended to latent categories: The most conceptually salient latent categories may be the most differentiated ones. Since the members of latent categories are words, differentiated latent categories will be comprised of words whose meanings are both highly similar to one another (i.e., specific) and highly dissimilar to the meanings of words in other categories (i.e., distinctive). According to this differentiation principle, the most conceptually salient latent categories of a domain will not be those with the most words (because they lack specificity) or the fewest words

---

[5] These categories can be distinguished from superordinate categories (e.g., furniture, musical instruments, vehicles), which typically refer to an entire domain. Latent categories may be said to fall somewhere between the superordinate and basic levels in the traditional taxonomy (Rosch et al., 1976), though this is less clear in non-object domains.

(because they lack distinctiveness), but those that achieve an optimal tradeoff between specificity and distinctiveness (Murphy & Brownell, 1985). Under this principle, individual word meanings will be low in differentiation relative to latent categories because they are the most specific but least distinctive categories of a domain.

1.6.2 Testing the hypothesis

There are two steps to testing the semantic clusters hypothesis. First, a measure of the semantic structure of a domain must be obtained. From this structure, clusters of words may be identified, some of which constitute differentiated latent categories. Second, the conceptual salience of these categories must be assessed. Evidence for conceptual salience would come from showing that the categories play a role in cognitive processes unrelated to language.

The first step requires a measure of the similarities among all of the words in the domain of interest.[6] The simplest method of collecting similarity data is to present people with two words at a time and have them judge how similar the words are to each other. However, this method is impractical with a large inventory of words given the large number of pairwise combinations. An alternative method is to have people sort words into groups based on the similarity of their meanings (Rosenberg & Kim, 1975; Wolff & Song, 2003). Words with similar meanings will tend to be grouped together more often than those with dissimilar meanings. These co-occurrences could be

---

[6] The term "domain" often goes undefined in the literature. In this research, I adopt the following definitions: At the conceptual level, a domain may be viewed as a body of knowledge about some class of attributes, entities, relations, or actions in the world (e.g., color, body parts, spatial relations, etc.). At the semantic level, a domain may be considered the set of words typically used to describe the items in the class, and the similarities among the meanings of these words.

combined across participants to construct a similarity matrix, and this matrix could then be analyzed using dimensionality reduction techniques (e.g., multidimensional scaling, *K*-means clustering, hierarchical clustering, principal components analysis). Such techniques would be used to identify clusters of words, including differentiated latent categories.

The second step, assessing the conceptual salience of these categories, requires examining how the categories might factor into nonlinguistic processing. One way of establishing 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 (Harnad, 1987). This phenomenon, known as categorical perception (CP), occurs across a wide range of domains (Goldstone & Hendrickson, 2010). In the case of latent categories, CP could be tested by presenting people with visual displays showing multiple items from the domain of interest, and having them discriminate among these items. On some trials, the items would be referents of words from different latent categories, and on other trials, the items would be referents of words from the same latent category. CP would be indicated by superior discrimination on between-category than within-category trials. Such an effect would suggest that latent categories are spontaneously accessed even when people are not using language. As a critical test of the semantic clusters hypothesis, the conceptual salience of more versus less differentiated latent categories could be assessed by comparing their respective CP effects. Support for the hypothesis, and for the differentiation principle underlying it, would come from stronger CP effects for more than less differentiated categories, including individual word meanings.

1.7 Space as a test bed

In the present research, I adopt 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. Despite striking differences in the encoding of spatial relations across languages (Bowerman & Choi, 2001; Feist, 2000, 2008; Levinson et al., 2003), analogous variation on nonlinguistic spatial tasks is not always observed (Khetarpal, Majid, Malt, Sloman, & Regier, 2010; Li & Gleitman, 2002; Munnich et al., 2001). The spatial domain thus offers an intriguing test bed for exploring how language might provide a window into the mind even when words do not align with concepts. In addition, the nonlinguistic conceptualization of spatial relations is more amenable to direct testing than, for example, concepts of mental states or possession, which are arguably inseparable from their linguistic instantiations (Feist, 2000; Gentner & Goldin-Meadow, 2003a).

1.7.1 Overview of the dissertation and empirical predictions

Chapters 2 and 3 present five experiments testing the semantic clusters hypothesis in the spatial domain with native English-speaking participants. The experiment in Chapter 2 examines the semantic structure of the spatial domain. Participants were asked to sort a large inventory of spatial prepositions into groups based on the similarity of their meanings, and dimensionality reduction techniques were used to identify clusters of words (Experiment 1). Some of these clusters, or latent categories, are shown to be more differentiated than others. The four experiments in Chapter 3 examine the conceptual salience of these categories, using CP as a diagnostic. CP effects for latent categories were assessed on their own (Experiments 2 and 3) and in comparison to other categories higher or lower in differentiation, including other latent categories (Experiment 4) and categories associated with individual words (Experiment 5).

The semantic clusters hypothesis makes two predictions about the conceptual salience of the various categories, as indicated by CP. First, latent categories that are observed across several different dimensionality reduction methods, and hence reflect key aspects of semantic structure, should align with conceptual distinctions (i.e., elicit CP effects). Second, more differentiated latent categories should be *more* conceptually salient (i.e., elicit *stronger* CP effects) than less differentiated categories, whether they be other latent categories or categories associated with individual words. These predictions are supported by the findings of the experiments in Chapter 3.

In Chapter 4, I conclude with a general discussion of how the present investigation may further understanding of the language-thought interface. In particular, I suggest that an approach focusing on *macrosemantics*, the study of meaning beyond the level of individual words, may provide a fresh perspective to ongoing debates on the relationship between language and thought.

Chapter 2: Inferring semantic structure

2.1 The semantics of space

The initial phase in testing the semantic clusters hypothesis is to map out the semantic structure of the domain of interest. In the spatial domain, inferring semantic structure involves examining the limited set of words a language has for describing the infinite possible configurations between physical objects in space. These sets of words include spatial prepositions in English, spatial verbs in Korean, locative adpositions in Turkish, and relational nouns in Japanese (Choi & Bowerman, 1991; Feist, 2000). Each of these terms describes a spatial relation between two or more entities, typically one located entity (*figure*) and one reference entity (*ground*). The meaning of a given spatial term may be considered the set of configurations between figure and ground to which the term applies (Feist, 2000).

Much attention has been devoted to characterizing the sets of figure-ground configurations linked to specific spatial terms. Most approaches have focused on abstract attributes of physical scenes described by the terms (see Feist, 2000), including geometry (Herskovits, 1986; Landau & Jackendoff, 1993; Talmy, 1983), function (Coventry & Garrod, 2004; Vandeloise, 1991), and qualitative physical characteristics such as animacy (Bowerman & Choi, 2001; Talmy, 1988). Cross-linguistic work has examined similarities and differences in the relative weighting of these attributes (and specific values of the attributes such as containment, contact, and support) across a variety of languages (Bowerman & Pederson, 1992; Feist, 2000, 2008; Levinson et al., 2003). A widely used elicitation tool in cross-linguistic studies is Bowerman and Pederson's

(1992) Topological Relations Picture Series (TRPS), a set of line drawings depicting a range of configurations between two objects (e.g., a cup on a table, a picture on a wall). The TRPS materials were specifically designed to elicit words for topological spatial relations (i.e., relations between objects that are contiguous, coincident, or close in proximity), largely ignoring "projective" relations, or the angle or direction of one object relative to another with respect to a specific frame of reference (e.g., "X is *in front of* / *behind* / *to the left of* / *to the right of* Y"; see Levinson, 2001).

On account of the extensive use of the TRPS materials, much is known about how different languages describe topological spatial relations, including rather subtle cross-linguistic differences in the meanings of terms like English *in* and *on* (Bowerman & Choi, 2001; Bowerman & Pederson, 1992; Feist, 2000, 2008; Khetarpal et al., 2009, 2010; Levinson et al., 2003; Pacer, Carstensen, & Regier, 2012; Regier et al., 2012; Xu & Kemp, 2012). However, topological terms comprise only a subset of the spatial terms of a language. More global aspects of the semantic structure of spatial relations, such as how the meanings of topological terms relate to those of projective terms, have not been examined. What is missing, in other words, is a bird's-eye view of spatial semantic structure as a whole.

This kind of wide-ranging perspective on semantic structure has been provided in other domains. Relationships among the meanings of words, and clusters of words, have been identified for such diverse semantic classes as emotion terms (Bush, 1973; Russell, 1980, 1983; Watson, Clark, & Tellegen, 1984), animal names (Henley, 1969; Rips, Shoben, & Smith, 1973), body parts (Carroll, 1976; Shepard, 1980), occupations (Sattath & Tversky, 1977), kinship terms (Rosenberg & Kim, 1975), and interpersonal verbs (Au, 1986). In all of these investigations, dimensionality reduction techniques were applied to semantic similarity data to discover the major dimensions underlying the meanings of words in the domain of interest, including global meanings shared by multiple words. For example, animal terms divide into separate clusters for herbivores and carnivores (Henley, 1969), and body part terms divide into clusters delineated by joints (Carroll, 1976; Shepard, 1980; see also Majid et al., 2006). These findings suggest that dimensionality reduction may also be useful for inferring the semantic structure of the spatial domain, and specifically for identifying clusters of spatial terms that may align with conceptual distinctions.

2.2 Dimensionality reduction methods

There are a number of methods for reducing the dimensionality of complex data. For the purposes of testing the semantic clusters hypothesis, it is critical that any clusters of words identified by means of dimensionality reduction be discernible across multiple methods. This would indicate that the clusters reflect genuine semantic structure, and are not merely an artifact of a particular method. At the same time, different methods often reveal complementary aspects of the underlying structure of a data set (Shepard, 1980), suggesting that the results from a single method might prove informative even if they stand in contrast to the results from other methods. With these interpretive considerations in mind, multiple methods were utilized in Experiment 1 to infer the semantic structure of spatial relations. Here I briefly describe the four methods used.

Multidimensional scaling (MDS) provides a spatial representation of the similarities among a group of items (Hout, Papesh, & Goldinger, 2012). Specifically, MDS takes pairwise estimates of similarity as input and constructs a representation of these items in a low-dimensional space. Within this space, the Euclidean distances between points match the observed (dis)similarities between items as closely as possible. The number of dimensions is specified by the researcher; adding more dimensions produces a better statistical fit, but makes the solution harder to interpret. A key advantage of MDS over other methods is that it permits visual appreciation of the relational structure of a data set, facilitating identification of underlying dimensions and clusters of items along those dimensions.

K-*means clustering* (KMC) offers no visual representation of the data, but may be considered a more direct method for identifying clusters of items. Given similarity data for N items, KMC partitions the items into K clusters such that the clusters are externally isolated and internally cohesive (Steinley, 2006). The means of the clusters define cluster membership: Each item belongs to the cluster whose mean is closest to it in Euclidean distance. As with the number of dimensions in MDS, the value of K is prespecified; the optimal choice of K strikes a reasonable balance between maximum compression of the data ($K = 1$) and maximum veridicality ($K = N$). Assuming that the appropriate value of K is selected, KMC reveals the major relational distinctions among a group of items.

In contrast to KMC, *hierarchical clustering* (HC) does not provide a single partitioning of a data set, but instead builds a hierarchy of clusters (Johnson, 1967). In the agglomerative approach to HC known as Ward’s method (Ward, 1963), each item is initially treated as its own cluster, and pairs of clusters are successively merged in such a way as to minimize the increase in within-cluster variance. When merged clusters in turn merge with each other, the result is a hierarchical structure. This structure may be visualized in a *dendrogram*, or tree diagram, which reveals the relations among clusters of different granularities. Thus, HC may be useful for comparing the relative prominence of different levels of structure within a domain.

Finally, *principal components analysis* (PCA) seeks to combine items into a small set of variables (principal components) in an orthogonal, linear fashion, maximizing the amount of variance explained (Jolliffe, 2002). The first principal component explains the most variance, and each successive component explains the largest possible amount of the remaining variance, with the constraint that it be orthogonal to the preceding components. The output of PCA includes component loadings, which indicate how much of the variance for a particular item is explained by each component. To the extent that items load primarily onto a single component, the components may be said to capture important dimensions in the structure of the data.

Given the many differences among the four methods, it is useful to consider what pattern of results would constitute convergence versus divergence across methods. Under a scenario of total convergence, spatial words would form tight clusters in an MDS solution, these same clusters would be evident in KMC and HC analyses, and items from different clusters would load onto different principal components in PCA. These results would imply that the clusters reflect key aspects of spatial semantic structure. Under a scenario of total divergence, the different methods would produce cross-cutting distinctions, with the clusters yielded by one method differing substantially from those of other methods. It might also be difficult to decipher the common meaning shared by the words within a given cluster. In this case, general conclusions about spatial semantic structure would be elusive.

Between these two extremes, and perhaps more likely than either one given that there are both similarities and differences in the algorithms and underlying approaches of the different methods (Everitt & Rabe-Hesketh, 1997; Hout et al., 2012; Tversky, 1975), is the scenario that the clusters will be substantially similar, but not fully identical, across methods. Possible differences include some methods yielding more fine-grained clusters than others and certain words shifting cluster membership depending on the method used. Notably, these differences in the granularity and boundaries of the clusters might ultimately prove useful for testing the predictions of the semantic clusters hypothesis. Clusters of different granularities, for example, may differ in their degree of differentiation; according to the hypothesis, the more differentiated clusters should be more conceptually salient. The varying degrees of convergence and divergence across methods were considered in interpreting the results of Experiment 1.

2.3 Experiment 1: Sorting prepositions

The goal of Experiment 1 was to infer the semantic structure of spatial relations. Participants sorted spatial prepositions into groups based on the similarities among their meanings, and a similarity matrix was derived from the sorting data. This matrix was then analyzed using the four dimensionality reduction techniques described above: MDS, KMC, HC, and PCA. The clusters of words emerging from these analyses constitute candidate concepts in the spatial domain.

2.3.1 Method

2.3.1.1 Participants

Sixty-three Emory University undergraduates (47 female) participated for course credit or payment. All participants were native English speakers. One female participant was excluded for not following instructions.

2.3.1.2 Materials

An inventory of English spatial prepositions was assembled by adapting a comprehensive list of prepositions originally compiled by Landau and Jackendoff (1993). Forty-two prepositions were selected from Landau and Jackendoff's list, omitting archaic or technical prepositions (e.g., *betwixt*, *without*), intransitive prepositions (e.g., *apart*, *downstairs*), and prepositions with non-spatial (e.g., *ago, despite*) or predominantly metaphorical (e.g., *in line with*) meanings. Prepositions requiring a phrasal verb construction (e.g., *through*, as in "pierce through") were also omitted because participants might find it difficult to judge their meaning independently. The resulting inventory, shown in Table 1, may be considered a relatively complete list of commonly used spatial prepositions in English.[7]

---

[7] The preposition *next to,* absent from Landau and Jackendoff's (1993) list, was also omitted here.

Table 1

Spatial Prepositions Used in Experiment 1

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

* Excluded from analyses (see Section 2.3.1.3).

Each of the prepositions was printed at the top of a 4 in. × 6 in. index card. Below each term were two sentences representing prototypical spatial usages of the term. The example sentences were selected from dictionaries and chosen on the basis of their conciseness, concreteness, and uniqueness with respect to the other example sentences. In some cases, the original sentences were modified to ensure that tense and sentence length were relatively uniform across sentences and to avoid the use of phrasal verbs. The preposition in each sentence was printed in bold. The sentences are provided in the Appendix.

2.3.1.3 Procedure

The experiment consisted of two phases. In the first phase, participants were presented with the stack of index cards and were asked to write a definition for each preposition on the basis of the two example sentences. The purpose of this task was to encourage participants to think relatively deeply about the meanings of the prepositions. The order of the cards was randomized differently for each participant. For a subset of participants, the preposition *against* was inadvertently omitted from the stack of cards; as a result, this term was excluded from analyses.[8]

In the second phase, participants were asked to sort 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.

2.3.2 Results and discussion

The number of groups into which participants divided the prepositions ranged from 5 to 29 ($M = 14.1$, $SD = 5.8$). To reveal any systematic patterns underlying this apparent variability, the sorting data were analyzed using MDS, KMC, HC, and PCA. The raw data were first converted into a pairwise similarity matrix for use as input to each analysis. The similarity between each pair of prepositions was 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 of the participants grouped *above* and *below* together, the similarity between them would be 31 ÷ 62 = .5, and so on.

---

[8] Analyses of the complete sorting data of participants for whom *against* was included ($N = 34$) converged with those of the full data set. The term *against* tended to be grouped with the terms in the *left-right* cluster (see Section 2.3.2).

To preview the results, the various dimensionality reduction methods produced broadly similar representations of spatial semantic structure. Each method partitioned the prepositions into a small number of clusters (or components, in the case of PCA), capturing significant variance in participants’ sorts. The resulting clusters were essentially the same across methods, with some variation in their granularity and boundaries. Below I present the results from each analysis in turn, followed by some general conclusions about spatial semantic structure suggested by considering the body of results collectively.

2.3.2.1 Multidimensional scaling

The similarity matrix was submitted as input to a MDS algorithm, ALSCAL (ordinal model; Takane, Young, & De Leeuw, 1977), and solutions of various dimensionalities were generated. The appropriate number of dimensions was determined by comparing the stress value—a measure of the degree of fit between the estimated inter-item distances in the MDS solution and the input distances (i.e., pairwise dissimilarities)—at the different dimensionalities (see Hout et al., 2012). Figure 3 shows a scree plot, with stress values plotted as a function of dimensions.

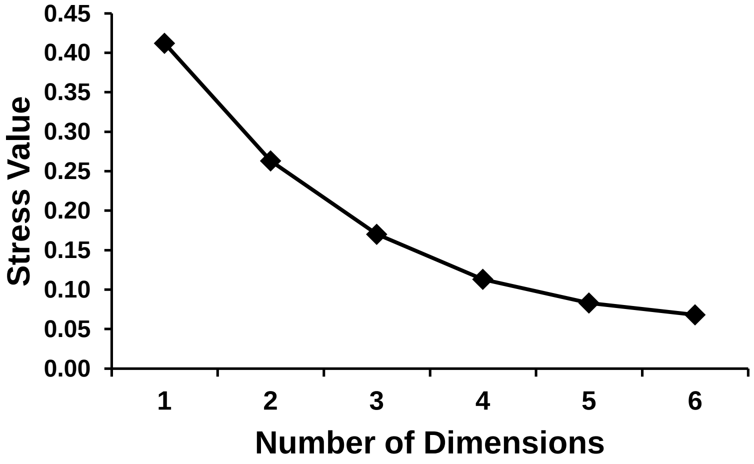

Figure 3. Scree plot showing stress value as a function of the number of dimensions in the MDS analysis. Note that there is no clear "elbow" in the graph.

According to the "elbow" method, the appropriate number of dimensions is indicated by the stress value at which additional dimensions no longer improve the fit substantially (Hout et al., 2012; Thorndike, 1953). There is no obvious elbow in the scree plot, but the largest decline in stress values occurs between 1 and 2 dimensions. As the goal of the MDS analysis was to generate a visual representation of the data permitting identification of clusters of words, the two-dimensional solution (stress value = .26; $R^2 = .66$) is presented in Figure 4.[9]

---

[9] The three-dimensional solution (stress value = .17; $R^2 = .79$) provided little additional information. The third dimension could be interpreted as reflecting a distinction between spatial terms encoding metric (e.g., *far from*, *near*) versus nonmetric (e.g., *above*, *to the left of*) information, but this dimension also distinguishes the four clusters in Figure 4 reasonably well.

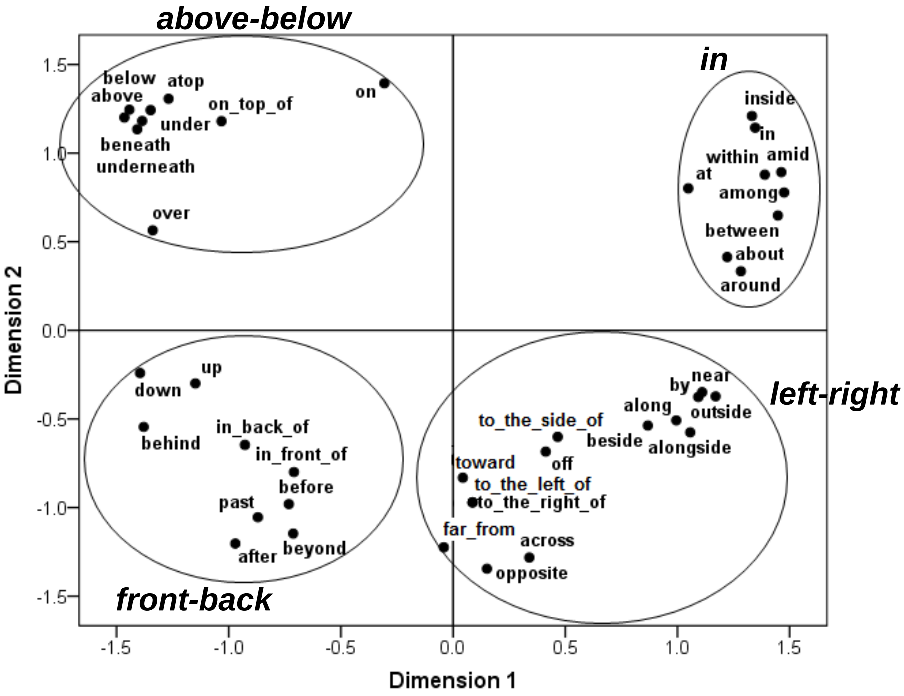

Figure 4. Multidimensional scaling solution of the similarity data, with *K*-means clusters added to the solution to aid interpretation. Clusters are labeled for descriptive purposes.

To help identify clusters, the inter-item distances in the MDS solution were combined into a pairwise similarity matrix. This new matrix was then submitted as input to a series of KMC analyses, using different values of *K*. The scree plot in Figure 5, showing within-cluster variance plotted against the number of *K*-means clusters, shows a clear elbow at 4 clusters. This elbow indicates that partitioning the MDS space into four clusters strikes a reasonable balance between minimizing within-cluster variance and reducing the data to a small set of interpretable clusters. These clusters were marked on the MDS solution. As indicated in Figure 4, I will refer to the clusters as *above-below*, *front-back*, *left-right*, and *in*.[10] Notably, the *above-below*, *front-back*, and *left-right* clusters each contain words that are essentially opposite in meaning. This suggests that the clusters cannot be reduced to individual word meanings, but instead capture more global semantic distinctions.

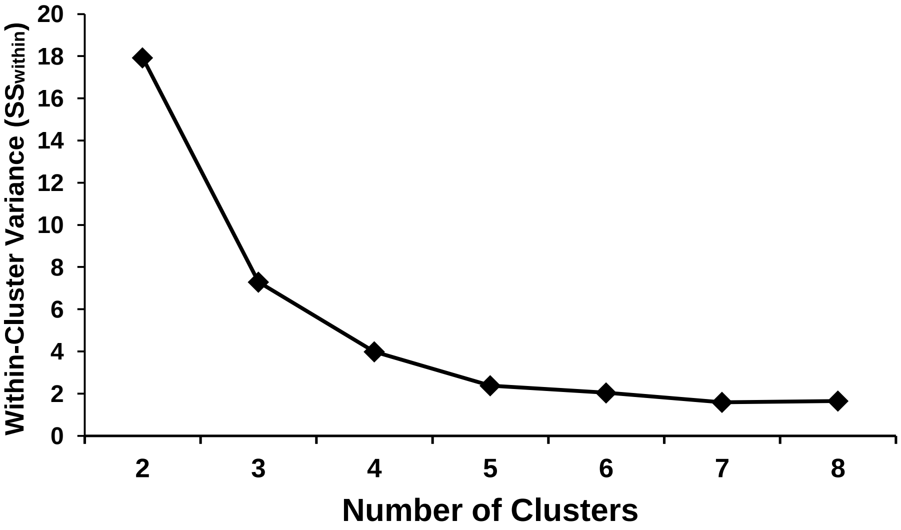

Figure 5. Scree plot showing within-cluster variance as a function of number of clusters (i.e., *K*) in the KMC analysis of MDS inter-item distances. An elbow occurs at 4, indicating that minimal further reduction in within-cluster variance is obtained with additional clusters.

In addition to identifying clusters within the two-dimensional MDS space in Figure 4, the dimensions themselves may also be interpreted. These dimensions seem to capture broad distinctions in how spatial relations are structured in the world. The y-axis

---

[10] An analogous HC analysis of the MDS inter-item distances yielded the same four major clusters.

reflects the distinction between topological relations and projective relations (Levinson, 2001). Most of the prepositions in the *above-below* and *in* clusters refer to relations between contiguous objects, whereas most of the prepositions 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 (i.e., objects in front of the body or above ground level are easier to perceive than objects behind the body or below ground level, respectively), but that the *left-right* axis is perceptually symmetric (e.g., objects to the left and right of the body are perceived about equally well; Clark, 1973; Franklin & Tversky, 1990). The *in* cluster does not appear to fit this framework; relations of containment and proximity are not readily characterized in terms of symmetry. However, the *in* cluster is also located at the extreme end of the x-axis, past most of the terms in the *left-right* cluster. This suggests that the *in* and *left-right* clusters are not as similar to each other as the *above-below* and *front-back* clusters are on whatever dimension is captured by the x-axis. Of course, given that the two-dimensional solution may not represent the optimal fit for the data, there is a danger in overinterpreting these dimensions.

In sum, the results from the MDS analysis point to the existence of four global clusters of spatial prepositions. As will be shown, however, the other dimensionality reduction methods deviate from MDS in dividing some of these clusters into more specific groupings, suggesting multiple levels of spatial semantic structure.

2.3.2.2 K*-means clustering*

In the previous analysis, KMC was used to help identify clusters within the MDS solution. A separate series of KMC analyses were conducted on the original similarity matrix, again using different values of *K*. This analysis, unlike the previous one, yielded a scree plot in which within-cluster variance levels off at five clusters, not four (see Figure 6).

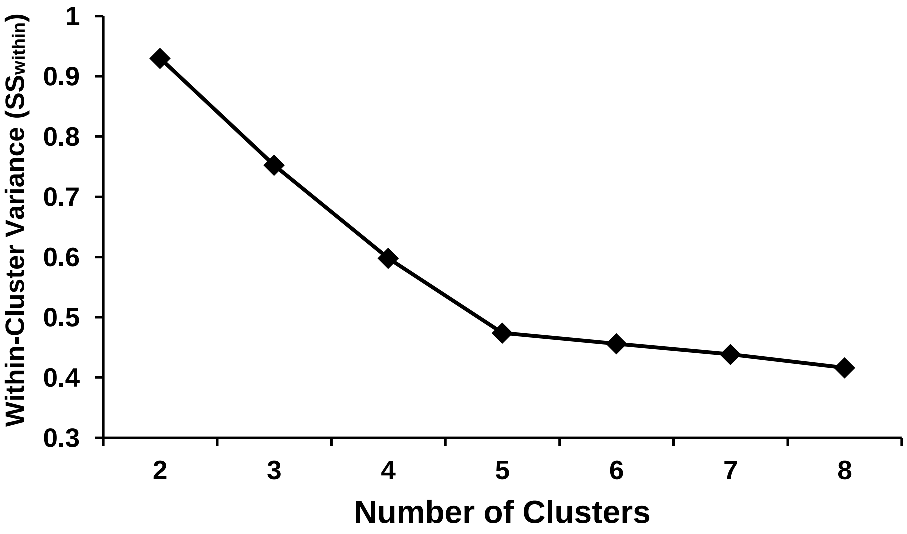

Figure 6. Scree plot showing within-cluster variance as a function of the number of clusters in the *K*-means clustering analysis of the original similarity data. An elbow occurs at 5, indicating that minimal further reduction in within-cluster variance is obtained with additional clusters.

Table 2 shows the assignment of prepositions to the five clusters. The most obvious difference between these clustering results and those of MDS is that there are separate clusters for terms referring to *above* and *below* relations, rather than a single *above-below* cluster. Also, six of the terms in the MDS *left-right* cluster (*across, far from*, *off*, *opposite*, *outside,* and *toward*) are in the KMC *front-back* cluster. The results from the two methods are otherwise identical, indicating substantial convergence. Interestingly, the terms assumed to be most defining of the clusters (e.g., *to the left of*, *to the right of*, *in front of*, *in back of*, etc.) are stable across methods, suggesting that they may constitute prototypical exemplars of the global meanings associated with the clusters. That MDS and KMC yielded essentially the same clusters, notwithstanding differences in granularity and cluster membership, implies that these clusters capture key aspects of spatial semantic structure.

Table 2

Clusters of Prepositions Yielded by K*-means Clustering Analysis*

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

Note. *K* = 5; clusters are labeled for descriptive purposes.

2.3.2.3 Hierarchical clustering

A HC analysis of the original similarity matrix using Ward's method yielded the dendrogram shown in Figure 7. Unlike the previous analyses, this representation of semantic structure points to the hierarchical organization of spatial terms, with clusters at multiple levels of granularity. The horizontal lines in the dendrogram (x-axis) represent the dissimilarity between clusters; longer lines indicate more global distinctions within the domain. Starting from the right side of the dendrogram, the first partitioning of the domain (x = 25) isolates the *above-below* cluster from all other terms. The next partitioning (x = 22) divides the *above-below* cluster into separate *above* and *below* clusters. Subsequent partitionings distinguish *left-right* from *front-back* and *in* (x = 20), *front-back* from *in* (x = 15), and *front* from *back* (x = 4).

At the terminal end of the dendrogram, the most cohesive proximal clusters into which the individual words fall are essentially the same five clusters suggested by KMC, except that the *front-back* cluster is divided further. However, given that this partitioning occurs much farther downstream than the other major partitionings of the domain, the distinction between *front* and *back* may be relatively less prominent than the other distinctions. Membership within the five main clusters in HC is also virtually identical to that of KMC. The only exception is that the term *outside,* part of the KMC *front-back* cluster (and the MDS *left-right* cluster), is in the HC *in* cluster. These differences across methods suggest that *outside* is not well captured by any of the clusters.

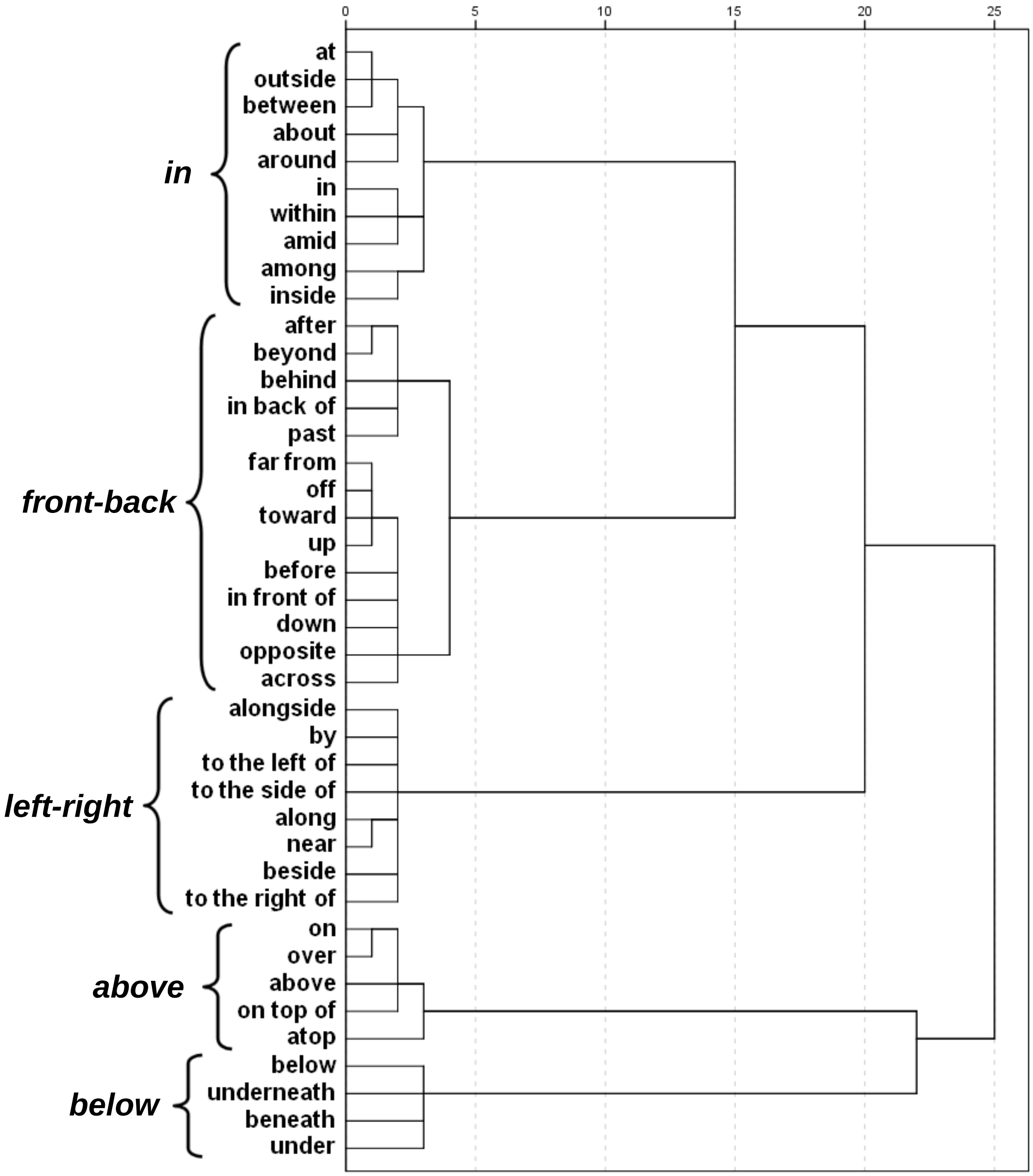

Figure 7. Dendrogram showing the results from the hierarchical clustering analysis. The length of the horizontal lines (x-axis) represents the dissimilarity between clusters. Clusters are labeled for descriptive purposes.

In sum, the results from the HC analysis are almost entirely congruent with those of the KMC analysis. The HC analysis also provides some insight into differences in the results produced by the MDS and KMC methods. Whereas MDS yielded a single *above-below* cluster and KMC yielded separate *above* and *below* clusters, HC suggests that both sets of clusters exist, but at different levels in the hierarchical structure of the domain. One reason why MDS may not have distinguished *above* and *below* is that the MDS solution attempts to account for the largest differences in the entire data set (Murphy, 2002). The *above* and *below* terms were extremely dissimilar from all other terms in participants' sorts, and hence were as far apart from them as possible in the solution. As a consequence, the *above* and *below* terms were close together despite also being quite dissimilar from each other, as revealed by KMC and HC.

2.3.2.4 Principal components analysis

The original similarity matrix was submitted as input to PCA. As a first step, a scree plot was generated to determine the number of principal components to extract. The plot in Figure 8 shows the eigenvalue (i.e., the variance explained by a given component) for each of 41 possible components for the 41 prepositions. There is a clear elbow at 5, suggesting that only the first five principal components explain significant variance.

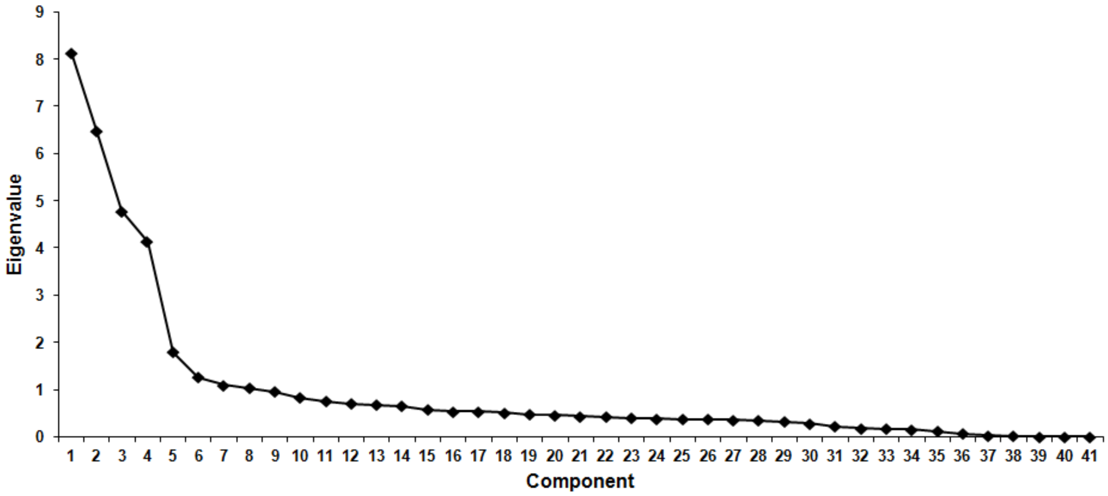

Figure 8. Scree plot of eigenvalues for each of the components in the principal components analysis. An elbow occurs at 5, indicating that minimal further variance is explained with additional components.

PCA was performed on the similarity matrix to extract just these five components, followed by a varimax rotation to maximize the distinctions among the components (see Jolliffe, 2002). Table 3 shows the results of the analysis. The five rotated components each had eigenvalues greater than 3 and together accounted for 61.7% of the variance in the data.[11]

---

[11] Three additional components had eigenvalues greater than 1, another common threshold for deciding how many components to extract (Jolliffe, 2002). However, these components accounted for only 8.3% additional variance and were less easily interpreted than the first five.

Table 3

Results of Principal Components Analysis with Varimax Rotation

Component	Eigenvalue	% Variance Explained	Cumulative % Variance Explained
1	6.56	16.0%	16.0%
2	5.33	13.0%	29.0%
3	5.33	13.0%	42.0%
4	4.85	11.8%	53.8%
5	3.25	7.9%	61.7%

Table 4 shows the loadings of the 41 prepositions onto the five principal components, with the highest loading for each preposition marked in bold. Some of the prepositions had cross-loadings of .3 or higher on additional components (marked in italics), indicating that their meanings are not well captured by a single component. An examination of the terms without high cross-loadings suggests the following labels for the components: *left-right*, *above*, *below*, *back*, and *front*. As these labels imply, the components match several of the clusters yielded by the previous analyses. Notably absent is the *in* cluster; all of the terms from this cluster, including *outside,* loaded onto multiple components rather than forming their own component (e.g., *in*, *inside,* and *within* have loadings of higher than .3 on all five components). Of the remaining terms, all but two loaded highest on the component corresponding to its same-name KMC/HC cluster (with the *front* and *back* components regarded as subclusters within the *front-back* cluster yielded by the other methods). The exceptions, *up* and *down*, were both in the *front-back* cluster in KMC and HC, but loaded highest on the *above* and *below* components, respectively; however, each also cross-loaded on the *back* component.

Table 4

Component Loadings from Principal Components Analysis

			Component		
Preposition	**1. *left-right***	**2. *above***	**3. *below***	**4. *back***	**5. *front***
beside	**.86**	.09	.09	-.06	.03
along	**.84**	.09	.09	-.05	.12
alongside	**.83**	.11	.09	-.05	.07
by	**.82**	.13	.12	-.08	.09
to the side of	**.80**	.08	.02	-.10	-.15
near	**.79**	.16	.18	-.05	.21
to the right of	**.73**	.07	.00	-.09	-.23
to the left of	**.73**	.07	.00	-.09	-.23
over	-.21	**-.87**	.03	-.01	.03
on	-.15	**-.84**	.05	-.20	.09
above	-.15	**-.82**	.01	-.13	.04
on top of	-.13	**-.81**	.02	-.16	.04
atop	-.15	**-.80**	.01	-.16	.04
up	-.17	**-.75**	.02	*.32*	-.07
under	-.19	.08	**-.80**	-.13	.11
underneath	-.19	.06	**-.80**	-.13	.11
beneath	-.19	.07	**-.80**	-.13	.11
below	-.19	.07	**-.80**	-.13	.11
down	-.28	.07	**-.77**	*.31*	.01
outside	-.22	*.47*	**.57**	-.11	-.11
at	-.10	.05	**.51**	*-.48*	.18
between	-.26	*.42*	**.49**	*-.37*	*.43*
within	*-.30*	*.36*	**.44**	*-.37*	*.34*
inside	*-.31*	*.33*	**.41**	*-.35*	*.33*
in	*-.31*	*.33*	**.41**	*-.36*	*.33*
after	-.16	.06	-.02	**.82**	-.18
far from	-.21	.06	.19	**.77**	-.02
beyond	-.21	-.01	.06	**.75**	-.10
past	-.21	.04	.02	**.74**	-.13
off	*.42*	.11	.26	**.67**	.04
behind	-.09	.07	-.17	**.55**	-.20
in back of	.02	.09	-.14	**.48**	-.29
in front of	.11	.00	.03	.10	**-.60**
toward	.16	.05	.23	*.31*	**-.57**
before	-.02	.04	.06	.25	**-.51**
opposite	-.05	.16	.15	.01	**-.41**
across	-.05	.15	.15	.02	**-.35**
around	*.40*	.25	.25	-.05	**.52**
about	.27	*.31*	*.37*	-.15	**.51**
among	-.14	*.35*	*.40*	-.26	**.48**
amid	-.17	*.34*	*.40*	-.26	**.47**

Note. For each preposition, the highest component loading is in bold and other component loadings above .3 or below -.3 are in italics. Components are labeled for descriptive purposes.

In sum, the results from PCA are similar to those of the other methods. Additionally, PCA highlights the distinction between *front* and *back*, evident to a lesser degree in the HC analysis, and suggests that the *in* cluster may be relatively less cohesive than the other clusters.

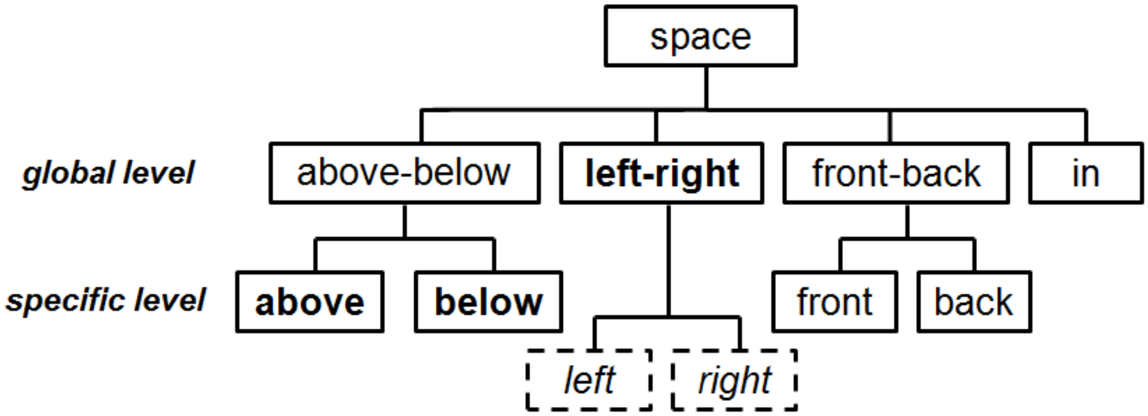

Figure 9. Schematic diagram of spatial semantic structure, as suggested by the collective results of Experiment 1. The clusters in bold are more differentiated than their immediate superordinate (*above-below*) or subordinates (*left*, *right*), and hence are predicted to be more conceptually salient. The dashed lines indicate that *left* and *right* are individual word meanings, not clusters of words.

2.3.2.5 General conclusions

Analyses of the similarity data using four different dimensionality reduction methods converged on broadly similar representations of spatial semantic structure. Each method also provided unique information. The schematic diagram in Figure 9 summarizes some general conclusions that may be drawn from considering this body of results collectively. Spatial terms appear to divide into four global clusters: *above-below*, *left-right*, *front-back*, and *in* (MDS); however, these clusters may differ in their relative prominence (HC; not represented in Figure 9). The global *above-below* cluster divides into specific *above* and *below* clusters (KMC, HC, PCA) and the global *front-back* cluster divides into specific *front* and *back* clusters (HC, PCA), but the former distinction may be more prominent than the latter (KMC, HC). The global *left-right* cluster does not divide further (KMC, HC, PCA), suggesting that the distinction between *left* and *right* is merely a lexical one. Finally, the global *in* cluster is relatively less cohesive than the others, as many of its terms are similar in meaning to terms in other clusters (PCA).

2.3.2.6 Correspondence with individual data

A potential concern with drawing general conclusions of this sort is that the aggregated results from multiple dimensionality reduction methods, each itself a form of aggregation, may be relatively far removed from any individual's semantic knowledge about spatial relations. In other words, the diagram in Figure 9, though a reasonable approximation of spatial semantic structure given the collective results, might be a poor reflection of individual participants' representation of this structure. One way to address this concern is to examine the degree of correspondence between the clusters of words in the collective structure and the groups of prepositions formed by individual participants. If the sorting behavior of most participants is consistent with the clusters in Figure 9, this would suggest that the dimensionality reduction methods captured shared tendencies in how semantic knowledge about space is organized.

To measure the degree of correspondence between participants' sorts and the collective structure, the clusters were first narrowed down to their "core members," defined as prepositions assigned to the same global cluster across all four dimensionality reduction methods. This excluded all prepositions assigned to the *in* cluster by any method, along with six additional prepositions, leaving 24 core members of the global *above-below*, *left-right*, and *front-back* clusters. These prepositions are listed in Table 5.

Table 5

Core Members of Clusters Yielded by Dimensionality Reduction Analyses

		Cluster		
above-below		***left-right***	***front-back***	
above	***below***		***front***	***back***
above	below	along	before	after
atop	beneath	alongside	in front of	behind
on	under	beside		beyond
on top of	underneath	by		in back of
over		near		past
		to the left of		
		to the right of		
		to the side of		

Note. Clusters are labeled for descriptive purposes.

Next, individual participants' raw sorting data were examined for "violations," groups of prepositions containing core members from different global clusters. Of the 384 groups with at least two core members at this level, only 47 were violations (12.2%), an average of 0.8 violations per participant (*SD* = 1.1). The majority of the violations were groupings of *front-back* terms with *above-below* (42.6%) or *left-right* (34.0%)

terms; *above-below* and *left-right* terms were rarely grouped together (14.9%).[12] This pattern is consistent with the MDS solution in Figure 4, in which the *front-back* cluster is closer to the *above-below* and *left-right* clusters than the latter two clusters are to each other. Overall, the results suggest that individual participants honored the global level of the collective structure to a considerable degree in their sorts.

The same analysis was conducted at the specific level of the collective structure. As indicated in Table 5, the 16 core members of the global *above-below* and *front-back* clusters are also members of the specific *above*, *below*, *front,* and *back* clusters. At this level of structure, violations constitute groups containing core members from both the *above* and *below* clusters or both the *front* and *back* clusters. Of the 278 groups with at least two core members at this level, only 27 were violations (9.7%), an average of 0.4 violations per participant (*SD* = 0.8). Almost all of these violations were groupings of *front* terms with *back* terms (85.2%); *above* and *below* terms were rarely grouped together (11.1%).[13] This dissociation is reflected in the results from KMC, in which *above* and *below,* but not *front* and *back,* divided into separate clusters, and in the results from HC, in which the distinction between *above* and *below* was one of the most prominent of the domain. Nevertheless, as with the global level, individual participants honored the specific level of the collective structure to a considerable degree.

Interestingly, none of the dimensionality reduction methods divided the global *left-right* cluster into specific *left* and *right* clusters, and this was likewise reflected in

---

[12] The remaining violations (4 of 47; 8.5%) were groupings of terms from all three global clusters.

[13] One participant formed a single group that included core members from the *front*, *back*, *above*, and *below* clusters (3.7% of violations at this level).

individual sorts: 49 of 61 participants (79.0%) grouped *to the left of* and *to the right of* together. These results suggest that *left* and *right*, though distinguished at the lexical level, are regarded as essentially interchangeable in relation to the rest of the terms in the domain. One reason why the *left-right* cluster may be treated differently is that most of the core members of this cluster are relatively non-specific; *beside* and *to the side of*, for example, denote relations of adjacency (often, though not exclusively, along the horizontal axis), but do not specify the side of the figure with respect to the ground. *To the left of* and *to the right of* are more specific, and may be regarded as highly similar to each other for this reason. In contrast, the *above-below* and *front-back* clusters consist of multiple terms with specific yet opposing meanings (e.g., *above*, *atop*, *over* vs. *below*, *beneath*, *under*), and hence the clusters may be divided further on this basis.

2.3.2.7 Relative differentiation of the clusters

The preceding analyses suggest that individual English speakers' semantic knowledge about space may resemble, to a significant degree, the structure shown in Figure 9. Critically, however, the goal of Experiment 1 was not merely to characterize semantic structure, but also to identify which parts of this structure may align with conceptual distinctions. Recall that the semantic clusters hypothesis predicts that the clusters of words most likely to be conceptually salient are those that are the most differentiated, maximizing within-cluster similarity relative to between-cluster similarity. To identify such clusters, differentiation scores were computed for the clusters in Table 5 based on the similarities among their core members, derived from the original similarity matrix. Given the relatively large number of violations for the global *front-back* and specific *front* and *back* clusters, differentiation analyses focused on the other clusters, for which clearer predictions could be made. Following the literature on basic level categorization, the differentiation score for a given cluster was defined as the average pairwise similarity of core members of the cluster with each other, minus the average pairwise similarity of core members of the cluster with core members of other clusters at the same level of structure (cf. Mervis & Crisafi, 1982). Accordingly, differentiation scores fell between 0 and 1, with higher scores indicating greater differentiation.

The differentiation score of the *above-below* cluster was .36, considerably lower than those of its subordinates, *above* (.64) and *below* (.90).[14] The differentiation score of the *left-right* cluster was .43. As the subordinates of this cluster (*left* and *right*) are comprised of just a single member, it is not possible to calculate a differentiation score for them. However, it is safe to assume that these categories are less differentiated than the *left-right* cluster: Their between-cluster similarity (i.e., the similarity between *to the left of* and *to the right of*) was .79 and their within-cluster similarity (e.g., the similarity between multiple terms meaning ‘left’, if any existed) would presumably be higher still, so the differentiation scores of these categories (within minus between) should be much lower than that of the *left-right* cluster. Thus, the differentiation principle makes opposing predictions about the relative conceptual salience of the global *above-below* and *left-right* clusters: Whereas the former should be *less* conceptually salient than its

---

[14] These scores cannot be statistically compared because they are not independent.

subordinates, the latter should be *more* conceptually salient than its subordinates (see Figure 9).[15] These predictions are tested in the experiments reported in Chapter 3.

2.4 Summary

The results of Experiment 1 provide a bird's-eye view of the semantic structure of spatial relations, identifying distinctions that cut across multiple words in the domain at different levels of granularity. Within this structure, certain clusters of words were shown to be more differentiated than others. According to the semantic clusters hypothesis, these differentiated clusters offer a window into the conceptualization of space. The experiments in Chapter 3 test this hypothesis by examining whether and to what extent the clusters factor into people's nonlinguistic thinking about spatial relations.

---

[15] These predictions remain unchanged if the *front-back* cluster is included in the calculation of between-cluster similarity for the *above-below* and *left-right* clusters. Differentiation analyses comparing the *front-back* cluster to its subordinates yielded mixed results: the differentiation score of the *front-back* cluster (.27) was considerably lower than that of the *front* cluster (.60), but comparable to that of the *back* cluster (.31).

Chapter 3: Assessing conceptual salience

3.1 Interpreting semantic structure

In using dimensionality reduction techniques like those employed in Experiment 1 to infer semantic structure, many researchers have assumed that the results reflect conceptual, not merely semantic, representations (e.g., Bush, 1973; Deese, 1962; Henley, 1969; Romney & d'Andrade, 1964; Watson et al., 1984). Deese (1962), for example, suggested that relations among word meanings "derive in whole or part from the structures or categories of the human mind" (p. 174). This assumption is sensible under the standard view of the relationships among words, concepts, and the world (see Figure 1a), or even from a Whorfian perspective (see Figure 1b). However, given the evidence suggesting a looser connection between words and concepts than either of these views can accommodate (see Figure 1c and Sections 1.4-1.5), language can no longer be relied on as a reliable guide to the conceptual system (see Malt et al., 2010, 2011). Whereas previous researchers might have concluded that the results of Experiment 1 constitute the conceptual organization of spatial relations, it is now clear that such organization must be verified independently.

The need to exercise caution in the interpretation of semantic structure was anticipated by a handful of researchers prior to the recent wave of research demonstrating dissociations between words and concepts. Using MDS, Russell (1980) found that the meanings of English emotion words (e.g., *happy*, *afraid*) fell along two key dimensions, pleasure-displeasure and degree of arousal. In subsequent work, Russell and colleagues sought evidence that this semantic structure reflects “a process fundamental to the human conception of emotion” (Russell & Bullock, 1985, p. 1290). They showed that the same structure can be derived from emotion words in other languages (Russell, 1983; see also Watson et al., 1984) and from facial expressions corresponding to emotion words (Russell & Bullock, 1985), even in preschool children who are not yet proficient with such words (Russell & Bullock, 1985, 1986). Though suggestive, however, such findings do not provide conclusive evidence for the conceptual status of the semantic structure. Different languages may share similar semantic structure for reasons independent of nonlinguistic conceptualization (e.g., Kemp & Regier, 2012), subjective similarity judgments for nonlinguistic stimuli, including facial expressions, may be influenced by whether the stimuli share the same name (Pinker, 1994; Winawer et al., 2007), and developmental evidence does not speak directly to the nature of the conceptual system in mature language users (Holmes & Wolff, 2012a).

Such observations imply that verifying the conceptual status of semantic structure requires a different kind of evidence than has been provided previously. Here the literature on the Whorfian hypothesis may offer some useful lessons. For decades, the claim that language influences thought suffered from a dearth of convincing empirical support: Languages were shown to differ, but these differences alone are insufficient evidence that their speakers think differently (Murphy, 2002; Pinker, 1994). Even when speakers of different languages were shown to perform differently on ostensibly nonlinguistic tasks (e.g., judging the similarity of colors; Kay & Kempton, 1984), the results were criticized because they could reflect a deliberate strategy on the part of participants to follow the semantic distinctions of their language (Pinker, 1994; Winawer et al., 2007). More recently, however, a number of studies have provided compelling demonstrations of cross-linguistic differences in thinking, using tasks that minimize the explicit use of linguistic strategies (see Wolff & Holmes, 2011). In the remaining experiments, I employ such tasks within a single language group to assess the conceptual salience of the spatial semantic structure identified in Experiment 1, testing the specific predictions of the semantic clusters hypothesis.

3.2 Categorical perception as a diagnostic of conceptual salience

According to the semantic clusters hypothesis, some of the clusters in spatial semantic structure should align particularly well with nonlinguistic spatial concepts. Investigating this possibility requires a method of assessing the potential involvement of the clusters in nonlinguistic thinking, ideally one that avoids the criticisms associated with the use of explicit tasks. Such a method is provided by tasks testing for categorical perception (CP), the superior ability to discriminate items from different categories compared to the same category (Goldstone & Hendrickson, 2010; Harnad, 1987). An advantage of CP over other candidate methods of assessing conceptual salience is that people are generally not aware that their perceptual judgments may be influenced by the categories they possess. Unlike subjective tasks in which participants are free to base their judgments on any property of the items presented, including their category membership, CP tasks involve simple, objective judgments about the immediate visual appearance of the items.[16] When such judgments are nevertheless shown to be influenced

---

[16] Though initially studied in the domain of speech perception (e.g., Liberman, Harris, Hoffman, & Griffith, 1957), CP has since been shown for a variety of visual stimuli.

by the categories of the items—as has been demonstrated for such diverse stimuli as shapes (Goldstone, 1994), colors (Özgen & Davies, 2002), faces (Kikutani, Roberson, & Hanley, 2010), household objects (Newell & Bülthoff, 2002), and fur patterns on cattle (Goldstein & Davidoff, 2008)—this provides compelling evidence for the conceptual salience of the categories.

The clusters of words in spatial semantic structure may be regarded as latent categories—categories for which there is no common label shared by all of the words in a cluster (and the spatial relations to which they refer), at least at the global level. CP for such categories would be revealed by superior discrimination of spatial relations encoded by words from different categories compared to spatial relations encoded by words from the same category, over and above any perceptual differences between the depicted relations. CP effects of this sort would suggest that people spontaneously access latent categories during the visual perception of spatial relations, even when they are not using language. In other words, such effects would imply that the categories capture important conceptual distinctions in the spatial domain.

3.2.1 Words and CP

The possibility that latent categories will give rise to CP effects might seem unlikely because some have argued that CP is driven by lexical codes—representations of the labels of individual words (Regier & Kay, 2009; Roberson & Hanley, 2010). Support for this idea comes from studies showing that CP effects for lexical categories (e.g., *blue* vs. *green*, *dog* vs. *cat*) are lateralized to the left hemisphere (Drivonikou et al., 2007; Gilbert, Regier, Kay, & Ivry, 2006, 2008; Roberson, Pak, & Hanley, 2008; but see

Franklin, Catherwood, Alvarez, & Axelsson, 2010), the side of the brain dominant for language (Hellige, 1993) and presumably where lexical codes are stored. Left-lateralized CP effects have been interpreted as evidence for the online meddling of lexical codes in perceptual processing—a kind of Whorfian effect in which contrastive labels accentuate perceptual differences but common labels minimize these differences (Regier & Kay, 2009; Roberson & Hanley, 2010; see also Lupyan, 2012). Recently, however, this interpretation has been called into question. Holmes and Wolff (2012a) found that unlabeled categories of novel objects also elicited left-lateralized CP, suggesting that the representations driving CP may be non-lexical in nature (see also Franklin et al., 2010; Kosslyn et al., 1989).

Given that CP does not depend on lexical codes, CP effects might also be observed for other unlabeled categories, including latent ones. Indeed, CP for many latent categories, if found, could not be attributed to lexical codes. Consider the global *above-below*, *left-right*, and *front-back* clusters (hereafter, categories) observed in Experiment 1. As noted above, CP for these categories would be revealed by superior discrimination of between-category pairs of spatial relations (e.g., *front* vs. *above*) compared to within-category pairs (e.g., *front* vs. *back*). Because the items in both pairs have contrastive labels, the labels alone cannot be the driving force behind any observed CP effects. CP for these categories would thus imply the influence of unlabeled categories on visual perceptual processing, a non-lexical effect like that observed by Holmes and Wolff (2012a).

3.3 Using CP to test the semantic clusters hypothesis

Experiments 2 through 5 used CP tasks to test the predictions of the semantic clusters hypothesis. The first prediction was that latent categories would be conceptually salient. Experiments 2 and 3 tested this prediction for categories at different levels of the spatial semantic structure inferred in Experiment 1 (see Figure 9). CP effects for the global *above-below*, *left-right*, and *front-back* categories were tested in Experiment 2, and CP effects for the specific *above* and *below* categories were tested in Experiment 3. Such effects would suggest that the latent categories found most consistently across dimensionality reduction methods align with prominent conceptual distinctions.

The second prediction, based on the differentiation principle underlying the semantic clusters hypothesis, was that more differentiated latent categories would be more conceptually salient than less differentiated categories. This prediction was tested by comparing CP effects for different sets of categories. Two alternatives to the differentiation principle were considered: a *simple prevalence view*, the possibility that global categories would be more conceptually salient than specific categories by virtue of containing more words, and a *lexical superiority view*, the possibility that categories associated with individual words (i.e., lexical categories) would be more conceptually salient than latent categories by virtue of being explicitly coded in language. Experiment 4 tested the differentiation principle against the simple prevalence view by comparing CP effects for the *above* and *below* categories (more differentiated but containing fewer words) and the *above-below* category (less differentiated but containing more words). Experiment 5 tested the differentiation principle against the lexical superiority view by comparing CP effects for the *left-right* category (more differentiated but latent) and the *left* and *right* categories (less differentiated but lexical). To preview the results, both predictions of the semantic clusters hypothesis were supported, with the differentiation principle providing a better explanation of the results than either of the alternative views.

3.4 Experiment 2: CP for *above-below, left-right,* and *front-back*

In Experiment 2, participants were presented with multiple pictures of spatial relations from the global *above-below*, *left-right*, and *front-back* categories.[17] Participants were asked to decide whether the pictures were perceptually identical or one of the pictures (the target) was different from the others (the distractors). On the latter trials, the target was either from the same latent category as the distractors (within-category trials) or a different latent category (between-category trials). CP would be indicated by faster or more accurate performance on between-category than within-category trials. Given recent findings suggesting that CP effects are stronger in the left hemisphere and sometimes entirely absent in the right (e.g., Gilbert et al., 2006; Holmes & Wolff, 2012a), the location of the target was varied across trials to accommodate the possibility that any observed CP effects would be left-lateralized. Because input to each side of the visual field is initially processed by the contralateral hemisphere (Hellige, 1993), left-lateralized CP would be indicated by stronger CP effects when the target is presented in the right visual field (RVF) than in the left visual field (LVF).

---

[17] The *in* category was not included because it was found to be less cohesive than the other global categories in Experiment 1, and because it was the only global category that did not contain terms with opposing meanings, making CP more difficult to assess.

3.4.1 Method

3.4.1.1 Participants

Twenty-two Emory University undergraduates (18 female) participated for course credit or payment. All participants were native English speakers, right-handed, and had normal or corrected-to-normal vision. Four participants (all female) were excluded, 3 for low accuracy (less than 65% correct on test trials showing non-identical spatial relations) and 1 for a mean reaction time (RT) greater than 2.5 standard deviations above the mean for all participants.

Figure 10. 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*.

3.4.1.2 Materials

The materials were 12 pictures of a bird and an airplane. As shown in Figure 10, each picture displayed the bird and the airplane from one of three perspectives (front view, side view, or top view). There were four pictures from each perspective, each showing the bird in a different position with respect to the airplane. The distance between the bird and the airplane, as determined by their closest edges, was the same for all positions. The pictures were created using the graphics package Autodesk 3D Studio Max 2010.

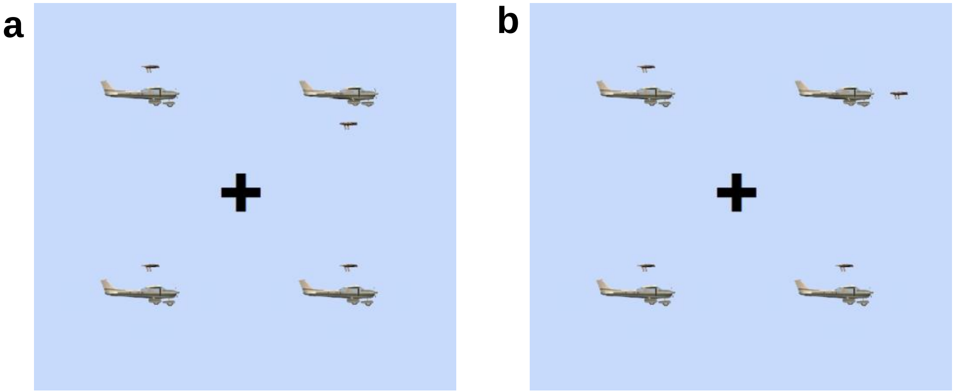

Figure 11. Examples of displays used in Experiment 2. (*A*) within-category trial (*below* target, *above* distractors). (*B*) between-category trial (*front* target, *above* distractors).

In the discrimination task, each display consisted of a fixation marker surrounded by four pictures, all from the same perspective (see Figure 11). In each picture, the center of the airplane subtended 11.5° (h) × 12.8° (v) visual angle.

3.4.1.3 Design and procedure

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

On half of the test trials in each block, the four pictures in each display were identical (“same” trials). This resulted in four unique “same” displays in each block, with each display presented 24 times. On the other half of the test trials, three pictures (distractors) were identical and the fourth picture (target) was different (“different” trials). There were two kinds of “different” trials: (a) within-category trials, in which the target and distractors were from the same latent category; and (b) between-category trials, in which the target and distractors were from different latent categories (see Figure 11). Across “different” trials, each picture served as the target at all four positions in the display, resulting in 48 unique “different” displays in each block (16 within-category, 32 between-category), each presented twice. All trials were presented in random order.

Figure 12 shows the trial structure used in this and all subsequent experiments. On each trial, 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. Each display appeared for only 200 ms, ensuring that the information presented to each visual field was initially processed by the contralateral hemisphere. Participants received feedback on their accuracy and response time after each practice trial; no feedback was given on test trials.

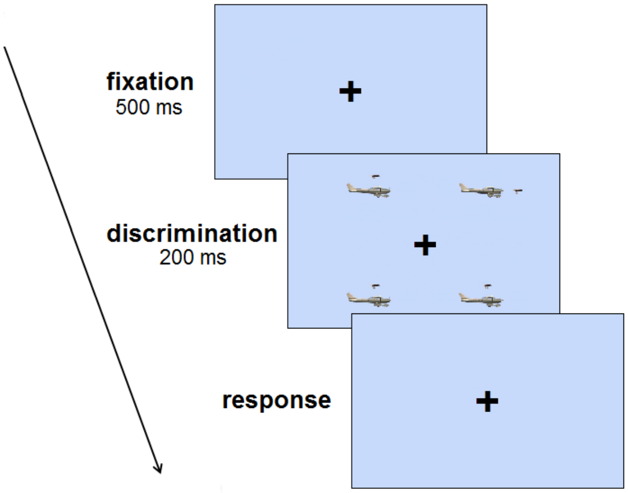

Figure 12. Trial structure in Experiments 2–5. After the discrimination display disappeared, a fixation marker remained on the screen until participants made a response.

Following the discrimination task, participants were presented with a randomly ordered stack of index cards, each showing one of the 12 pictures. For each picture, participants wrote a brief description of the locations of the bird and the airplane relative to each other. This task served as a manipulation check, verifying that participants interpreted the pictures as instances of the spatial relations they were intended to represent. The entire experiment took approximately 20 minutes to complete.

3.4.2 Results and discussion

3.4.2.1 Overview of key findings

As shown in Figure 13, the categories *above-below*, *left-right*, and *front-back* elicited CP effects: Participants were faster to discriminate spatial relations from different categories than from the same category. Notably, these effects were found only in the RVF, indicating that CP was lateralized to the left hemisphere, as in previous studies (e.g., Gilbert et al., 2006, 2008; Holmes & Wolff, 2012a).

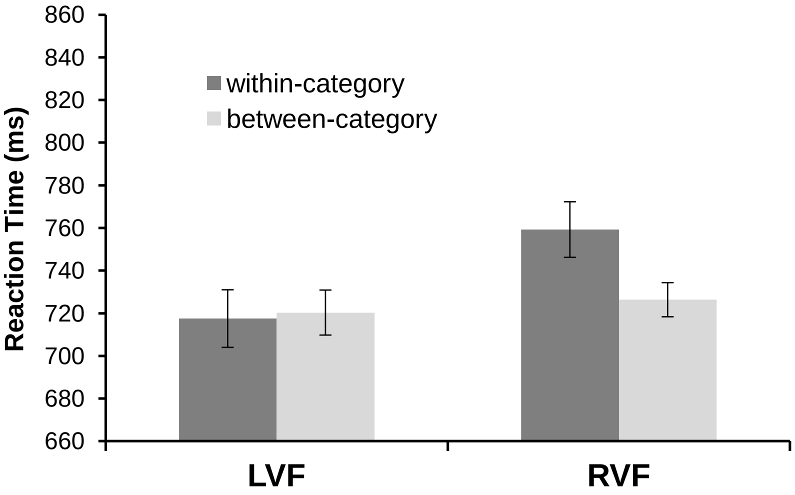

Figure 13. Results of Experiment 2 (*N* = 18). Error bars are 95% within-subjects confidence intervals (Loftus & Masson, 1994). LVF = left visual field; RVF = right visual field.

3.4.2.2 Discrimination task

Mean accuracy on the discrimination task 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 effects could be assessed. Trials in which participants responded incorrectly (12.2%) or in which reaction time (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-category vs. between-category) repeated-measures analysis of variance (ANOVA) on RT for the remaining trials yielded significant main effects of visual field, *F*(1, 17) = 8.83, *p* = .009, and category relation, *F*(1, 17) = 8.53, *p* = .01. The latter effect, with faster discrimination on between-category than within-category trials, is indicative of CP. However, there was also a significant interaction between visual field and category relation, *F*(1, 17) = 11.60, *p* = .003. The between-category advantage occurred only when the target appeared in the RVF, *t*(17) = 5.17, *p* < .0001, *d* = 1.22; there was no such advantage for LVF targets, *t*(17) = .34, *p* > .7 (see Figure 13). This interaction demonstrates that CP was left-lateralized. An analogous ANOVA on the accuracy data yielded no significant effects (*p*s > .05), implying that there was no speed-accuracy tradeoff.

The RT interaction rules out an alternative explanation for the CP effect, namely that the between-category advantage was the result of differences in perceptual similarity across the two kinds of trials (despite attempts to minimize such differences when designing the materials). Because any differences in similarity would presumably elicit CP-like effects in both visual fields (Gilbert et al., 2006, 2008), the interaction suggests that the within-category pictures were not more similar than the between-category pictures. The pattern of the results in the LVF may thus be regarded as reflecting baseline discrimination of the spatial relations under investigation, against which performance in the RVF may be compared. Interestingly, comparisons of performance in the two visual fields showed that the CP effect was one of interference, but not facilitation: Within-category trials were slower for RVF than LVF targets, *t*(17) = 3.63, *p* = .002, *d* = .86, but between-category trials were no faster for RVF than LVF targets, *t*(17) = .84, *p* > .4 (see Figure 13).

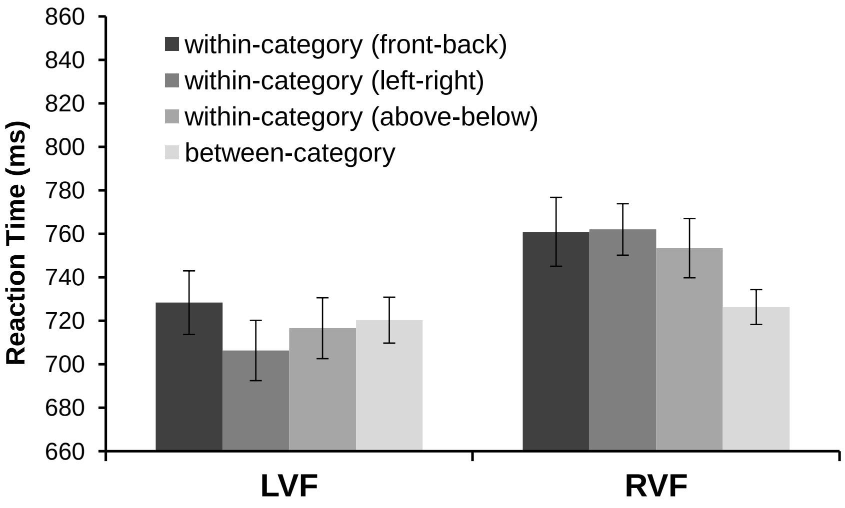

Figure 14. RT results of Experiment 2 by category. Error bars are 95% within-subjects confidence intervals.

Additional RT analyses showed that each of the three latent categories elicited reliable left-lateralized CP effects. Within-category trials were divided according to the category membership of the target and distractors: *above-below* (i.e., *above* target and *below* distractors, or *below* target and *above* distractors), *left-right*, and *front-back*. Discrimination was faster for RVF targets on between-category trials than on each of the three kinds of within-category trials (*above-below*: $t(17) = 2.17$, $p = .04$, $d = .51$; *left-right*: $t(17) = 2.66$, $p = .02$, $d = .63$; *front-back*: $t(17) = 2.13$, $p = .05$, $d = .50$; see Figure 14). For LVF targets, none of these differences reached significance (ps > .2). These findings demonstrate that all of the categories, not just one or two of them, contributed to the overall left-lateralized CP effect.

3.4.2.3 Picture description task

The results of the picture description task confirmed that the stimuli were good examples of the spatial relations they were intended to represent. 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 six of the pictures, there was 100% agreement. The four *front-back* pictures were most likely to elicit descriptions that included prepositions not from the intended category. Six of the 18 participants consistently described these pictures using horizontal or vertical terms (e.g., “the bird is to the right of the plane” for the *front* picture in Figure 10b), implying that they viewed the pictures two-dimensionally rather than three-dimensionally. Nevertheless, the majority of participants gave *front-back* terms for these pictures, suggesting that the CP effects observed for the *front-back* category are distinct from those for the other two categories.[18] In addition, none of the 216 total descriptions included global terms for the relations (e.g., “horizontal” for pictures in the *left-right* category), suggesting that the CP effects were not driven by lexical representations.

3.4.2.4 Summary and conclusions

The results of Experiment 2 provide evidence for the conceptual salience of the global *above-below*, *left-right*, and *front-back* categories. The findings imply that certain clusters of words align with conceptual distinctions that are spontaneously accessed in

---

[18] The *front-back* pictures from the top view (see Figure 10c) could also be interpreted as showing a side view (i.e., with the airplane flying upward in space). Under this interpretation, the use of vertical terms for the relations would not be inaccurate even from a three-dimensional perspective. That the majority of participants gave *front-back* terms for these pictures suggests, however, that the pictures were nevertheless perceived as showing different relations than the *above-below* and *left-right* pictures.

nonlinguistic contexts, influencing simple, objective perceptual decisions. The next experiment examined whether this kind of alignment with the conceptual system is limited to global latent categories, or would also be observed for categories at the specific level of semantic structure identified in Experiment 1.

3.5 Experiment 3: CP for *above* and *below*

Experiment 3 tested CP for the subordinates of the global *above-below* category, namely the specific *above* and *below* categories.[19] Because the individual words within these categories are essentially synonymous (e.g., *below*, *beneath*, *under*, and *underneath*), within-category trials consisted of two different exemplars of the same spatial relation (e.g., the target showed a *below* relation, and the distractors showed a different *below* relation). Between-category trials consisted of exemplars of different spatial relations (e.g., *below* target, *above* distractors). As in the previous experiment, CP would be indicated by superior performance on between-category than within-category trials, possibly only in the left hemisphere (RVF).

---

[19] The *front* and *back* categories were not tested because they were less consistently observed across dimensionality reduction methods in Experiment 1, and because the *front-back* stimuli in Experiment 2 were sometimes interpreted two-dimensionally (i.e., as *above-below* or *left-right*).

3.5.1 Method

3.5.1.1 Participants

Twenty-two participants (16 female) from the same population sampled in Experiment 2 took part in this experiment. Four participants (3 female) were excluded for low accuracy.

Figure 15. Stimuli used in Experiments 3 and 4. (*A*) *above-far* and *above-near*. (*B*) *below-far* and *below-near*.

3.5.1.2 Materials

The materials were adapted from those of the previous experiment. Four new pictures showing the bird and the airplane from the front view were created (see Figure 15). Two pictures showed an *above* relation, and the other two pictures showed a *below* relation. There were *near* and *far* versions of each kind of relation, representing different exemplars from the same category. The distance between the bird and the airplane was the same in the *above-near* and *below-near* pictures, and in the *above-far* and *below-far* pictures. As in Experiment 2, each display in the discrimination task consisted of a fixation marker surrounded by four pictures. Because the pictures were designed for use in the next experiment (which included *left-right* relations) as well, they were wider than those of Experiment 2. In each picture, the center of the airplane subtended 17.9° (h) × 13.3° (v) visual angle.

3.5.1.3 Design and procedure

There were 24 practice trials and 192 test trials, presented in a single block. As in Experiment 2, the test trials were divided equally into "same" and "different" trials, and "different" trials were divided into within-category (32 total) and between-category (64 total) trials. All other aspects of the design and procedure were identical to those of Experiment 2.

3.5.2 Results and discussion

3.5.2.1 Overview of key findings

As shown in Figure 16, the categories *above* and *below* elicited CP effects. As with the global categories examined in Experiment 2, faster discrimination of spatial relations from different categories (in this case, *above* vs. *below*) than the same category (*above* vs. *above, below* vs. *below*) occurred only in the RVF, indicating that CP was left-lateralized.

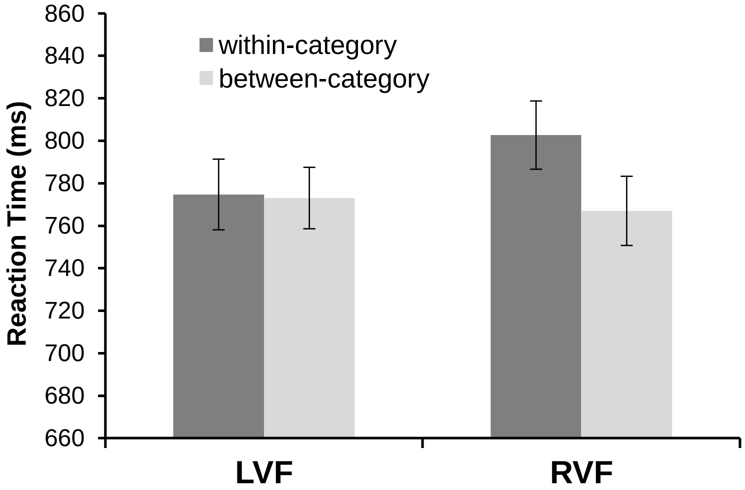

Figure 16. Results of Experiment 3 (*N* = 18). Error bars are 95% within-subjects confidence intervals.

3.5.2.2 Discrimination task

Mean accuracy on the discrimination task was 82.5% (*SD* = 8.2), with no difference between “same” and “different” trials (*p* > .8). As in the previous experiment, “different” trials in which participants responded incorrectly (17.3%) or in which RT was greater than 2.5 standard deviations from individual means (3.0%) were excluded. A 2 (visual field) × 2 (category relation) ANOVA was conducted on the RT data for the remaining trials. Although neither main effect reached significance (visual field: *F*(1, 17) = 2.69, *p* = .12; category relation: *F*(1, 17) = 2.91, *p* = .11), there was a significant interaction, *F*(1, 17) = 4.66, *p* = .05. Participants were faster to discriminate RVF targets on between-category than within-category trials, *t*(17) = 2.59, *p* = .02, *d* = .61, but no such difference was observed for LVF targets, *t*(17) = .13, *p* > .8, indicating left-lateralized CP (see Figure 16). As in the previous experiment, the categories interfered with, but did not facilitate, performance: Within-category trials were slower for RVF than LVF targets, *t*(17) = 2.55, *p* = .02, *d* = .60, but between-category trials were no faster for RVF than LVF targets, *t*(17) = .63, *p* > .5.

An analogous ANOVA on the accuracy data yielded a significant main effect of category relation, *F*(1, 17) = 12.63, *p* = .002, *d* = .84, but no main effect of visual field or interaction (*p*s > .2). Accuracy was higher on between-category trials (*M* = 85.9%, *SD* = 7.4) than within-category trials (*M* = 76.8%, *SD* = 12.6), consistent with the RT results and implying that there was no speed-accuracy tradeoff. Although the accuracy results suggest that perceptual differences may have led to an overall benefit for between-category over within-category trials,[20] lateralization of the RT effect suggests that participants also accessed the category distinctions among the stimuli when completing the task.

Additional RT analyses revealed that both the *above* and *below* categories elicited reliable CP effects, but that these effects were left-lateralized only in the case of the *above* category. Within-category trials were divided according to the category membership of the target and distractors, *above* or *below*. For RVF targets, between-category trials were faster than both within-category *above* trials, *t*(17) = 2.14, *p* = .05, *d* = .50, and within-category *below* trials, *t*(17) = 2.08, *p* = .05, *d* = .49. For LVF targets, between-category trials were faster than within-category *below* trials, *t*(17) = 2.34, *p* =

---

[20] Although the distance between the bird and the airplane was the same in the two *near* pictures and in the two *far* pictures, the distance between the position of the bird across pictures on between-category trials (e.g., *above-near* and *below-far*) was greater, on average, than that on within-category trials (e.g., *above-near* and *above-far*). Thus, the stimuli on the latter trials may have been more difficult to discriminate.

.03, *d* = .55, but were marginally slower than within-category *above* trials, *t*(17) = 2.06, *p* = .06 (see Figure 17).

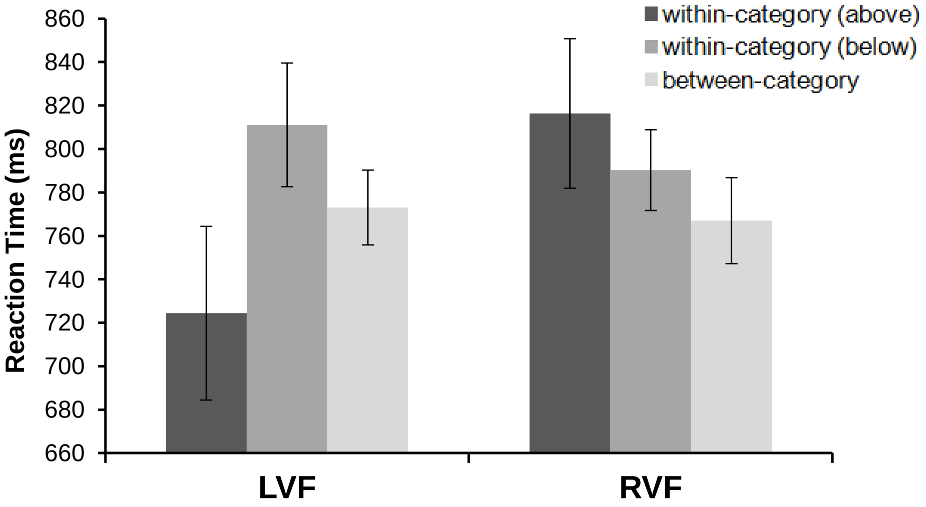

Figure 17. Results of Experiment 3 by category. Error bars are 95% within-subjects confidence intervals.

These findings suggest that the overall left-lateralized CP effect, collapsing across the two kinds of within-category trials, was driven by the *above* category. The lack of lateralization for the *below* category is consistent with the possibility that the within-category *below* pictures were more perceptually similar to each other than were the between-category pictures—and hence that the CP effect for the *below* category was driven by differences in perceptual similarity, rather than category membership. However, this possibility seems unlikely because the perceptual similarity between the *below-far* and *below-near* pictures was the same as that between the *above-far* and *above-near* pictures (see Figure 15), so it is unclear why similarity would govern performance only in the case of the *below* category. More likely is that the *below* category gave rise to particularly strong CP effects, evident in both visual fields. This latter possibility is consistent with the differentiation principle: The *below* category had a higher differentiation score than the *above* category in Experiment 1 (see Section 2.3.2.7).

3.5.2.3 Picture description task

The results of the picture description task confirmed that the stimuli were good examples of *above* and *below* relations. Across the four pictures, 95.8% of the descriptions included prepositions from the intended category, and the remaining descriptions were reasonable interpretations of the pictures (e.g., “the bird is far south of the plane” for the *below-far* picture). Importantly, none of the 72 total descriptions included prepositions from a different category than the intended one. Most participants used modifiers to distinguish between the *near* and *far* exemplars of each relation (e.g., “directly below” and “far below” for the *below-near* and *below-far* pictures, respectively), but the preposition was usually the same for the two pictures.

3.5.2.4 Summary and conclusions

The results of Experiment 3 provide evidence for the conceptual salience of the *above* and *below* categories. Latent categories align with concepts not only at the global level of spatial semantic structure, as was shown in the previous experiment, but also at the specific level. Together, the findings of Experiments 2 and 3 suggest that several different groups of spatial words capture distinctions that factor into the nonlinguistic processing of spatial relations.

Despite these contributions, the findings thus far do not speak to which latent categories figure most prominently in nonlinguistic spatial processing. One possibility, suggested by analogy with the cross-linguistic approaches to inferring conceptual universals discussed in Chapter 1 (see Section 1.6), is that global categories are more conceptually salient than specific ones. In those approaches, combining data from many, as opposed to few, languages will invariably yield a better approximation of universal conceptual structure (e.g., Regier et al., 2012). Within a single language, global categories might be more conceptually salient than specific categories for a similar reason, namely that they capture meanings shared by many words: The more words in the category, the better these words may approximate the underlying concept. When a given aspect of experience is captured by many words, this implies that people will attend to it across a large number of communicative contexts ("thinking for speaking"; Slobin, 1996b), and possibly even when they are not using language (Boroditsky, Schmidt, & Phillips, 2003).

In contrast to this simple prevalence view, the semantic clusters hypothesis predicts that conceptually salient categories are not necessarily those with more words, but rather those that are more differentiated. Some global categories might be less differentiated than specific categories because they group together words that are quite disparate in meaning, resulting in low within-category similarity. In Experiment 1, the specific *above* and *below* categories were shown to be more differentiated than the global *above-below* category. According to the simple prevalence view, *above-below* should nevertheless be more conceptually salient than *above* and *below*. The semantic clusters hypothesis makes the opposite prediction. Experiment 4 tested these competing predictions by comparing the respective CP effects of the two sets of categories.

3.6 Experiment 4: CP for *above-below* vs. CP for *above* and *below*

In Experiment 4, CP effects for the global *above-below* category and the specific *above* and *below* categories were pitted against each other. To accomplish this, the experiment included all three kinds of trials used in the previous two experiments. Participants were presented with spatial relations from the same specific category (within-specific trials; i.e., *above* vs. *above* or *below* vs. *below*), from different specific categories but the same global category (between-specific/within-global trials; i.e., *above* vs. *below*), and from different global categories (between-global trials; in this case, *above-below* vs. *left-right*, with the latter serving as a comparison global category). CP effects for the *above-below* category would be indicated by superior performance on between-global than between-specific/within-global trials, whereas CP effects for the *above* and *below* categories would be indicated by superior performance on between-specific/within-global than within-specific trials.[21] The magnitudes of the respective CP effects were compared to determine which set of categories is more conceptually salient.

---

[21] The spatial relations on within-specific trials were also from the same global category, and the spatial relations on between-global trials were also from different specific categories. However, CP effects are more easily interpreted when the trial types are minimal pairs, as in the comparisons described above.

3.6.1 Method

3.6.1.1 Participants

Thirty-seven participants (21 female) from the same population sampled in the previous two experiments took part in this experiment. Eleven participants (6 female) were excluded, 10 for low accuracy and 1 for a mean RT greater than 2.5 SD above the mean for all participants.

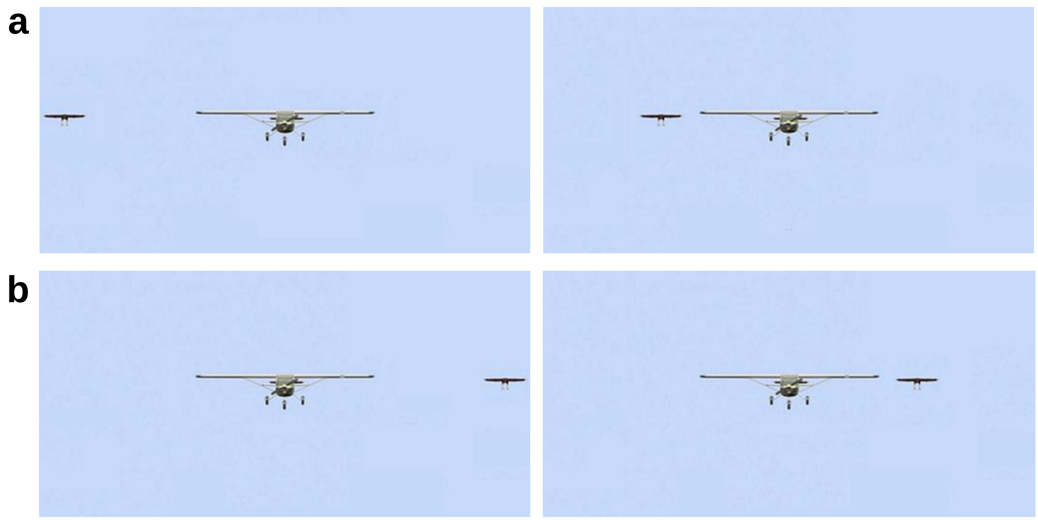

Figure 18. Stimuli showing *left* and *right* relations used in Experiment 4. (*A*) *left-far* and *left-near*. (*B*) *right-far* and *right-near*.

3.6.1.2 Materials

The materials included the same four pictures used in Experiment 3 (see Figure 15), plus an additional four pictures showing *left* and *right* relations from the front view (see Figure 18). As with *above* and *below*, there were *near* and *far* versions of the new pictures, representing different exemplars from the same category. The distance between the bird and the airplane was the same in the *left-near* and *right-near* pictures, and in the *left-far* and *right-far* pictures. The discrimination task included the same kinds of displays used in the previous experiment. Because the experiment was designed to assess CP effects for the *above-below* category and its subordinates, ensuring the discriminability of the various combinations of *left* and *right* relations was not essential; the *left* and *right* pictures were included to provide between-global trials for the *above-below* category. As a consequence, the distance between the position of the bird in the *near* and *far* versions of the *left* and *right* pictures was relatively small (though as large as possible using pictures from the front view), making discrimination more difficult on within-specific trials with these pictures than on the other kinds of trials.

3.6.1.3 Design and procedure

There were 24 practice trials and 384 test trials, presented in a single block. As in the previous experiments, the test trials were divided equally into “same” and “different” trials. There were three kinds of “different” trials: (a) within-specific trials, in which the target and distractors were from the same specific category (e.g., *above* target, *above* distractors); (b) between-specific/within-global (hereafter, between-specific) trials, in which the target and distractors were from different specific categories but the same global category (e.g., *above* target, *below* distractors); and (c) between-global trials, in which the target and distractors were from different global categories (e.g., *above* target, *left* distractors).

The eight pictures and four possible positions of the target in each display combined for 224 unique displays (32 within-specific, 64 between-specific, 128 between-global). Given the small number of possible within-specific trials relative to the other trial types and the need to keep the experiment to a practical duration, 64 trials of each type were presented. This meant that for a given participant, all possible within-specific displays were presented twice each, all possible between-specific displays were presented once each, and half of the possible between-global displays were presented once each. The between-global displays were selected such that the number of trials with LVF and RVF targets was the same. The entire experiment took approximately 30 minutes to complete. All other aspects of the design and procedure were identical to those of the previous two experiments.

3.6.2 Results and discussion

3.6.2.1 Overview of key findings

The results support the semantic clusters hypothesis over the simple prevalence view in showing that more differentiated categories elicit stronger CP effects than less differentiated categories. As shown in Figure 19, discrimination was faster when spatial relations came from different specific categories (i.e., *above* vs. *below*) than from the same specific category (i.e., *above* vs. *above*, or *below* vs. *below*). No such difference was observed for spatial relations from different global categories (i.e., *above-below* vs. *left-right*) compared to the same global category (i.e., *above* vs. *below*). Thus, when pitted against each other, CP effects were observed for the more differentiated *above* and *below* categories, but not the less differentiated *above-below* category, even though the latter contains more words. The same general pattern of results was observed in both visual fields, indicating that CP was less lateralized than in the previous experiments (see below for discussion of this point).

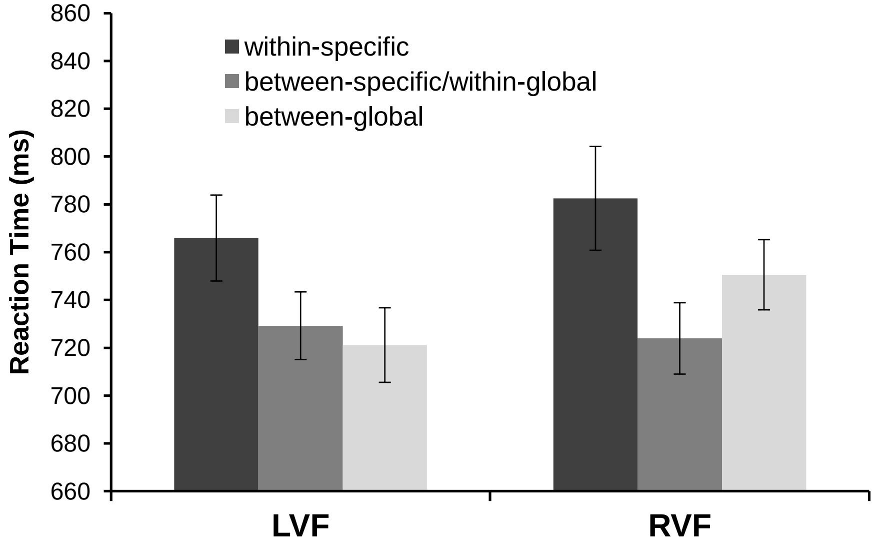

Figure 19. Results of Experiment 4 (*N* = 26). Error bars are 95% within-subjects confidence intervals.

3.6.2.2 Discrimination task

Mean accuracy on the discrimination task was 82.7% (*SD* = 4.7). Accuracy was higher on “same” (*M* = 85.8%, *SD* = 7.3) than “different” trials (*M* = 79.5%, *SD* = 7.9), a difference that can be attributed to poorer performance on within-specific *left* and *right* trials (*M* = 55.6%, *SD* = 21.4) than on the other types of “different” trials, all of which had a mean accuracy greater than 70%.

To test the predictions of interest, subsequent analyses focused on “different” trials in which the target or distractors, or both, came from the *above-below* category or its subordinates. As in the previous experiments, trials in which participants responded incorrectly (16.6%) or in which RT was greater than 2.5 standard deviations from individual means (2.7%) were excluded. A 2 (visual field) × 3 (category relation: within-specific vs. between-specific vs. between-global) ANOVA was conducted on the RT data for the remaining trials. This analysis yielded a significant main effect of category relation, $F(2, 50) = 17.34$, $p < .0001$, but neither the main effect of visual field, $F(1, 25) = 3.32$, $p = .08$, nor the interaction, $F(2, 50) = 1.87$, $p = .16$, reached significance. Pairwise comparisons of the different trial types revealed that discrimination was faster on between-specific ($M = 727$ ms, $SD = 106$) than within-specific ($M = 775$ ms, $SD = 123$) trials, $t(25) = 5.86$, $p < .0001$, $d = 1.15$, but that there was no significant difference on between-global ($M = 736$ ms, $SD = 106$) relative to between-specific (within-global) trials, $t(25) = 1.35$, $p = .19$. Given that the CP effects in the previous experiments were observed predominantly in the RVF, comparisons of just the trials with RVF targets were also conducted. In these comparisons, the same pattern as for the overall data was observed for between-specific and within-specific trials, $t(25) = 4.48$, $p < .0001$, indicative of CP for the specific *above* and *below* categories. However, unlike for the overall data, discrimination was significantly faster on between-specific than between-global trials, $t(25) = 2.42$, $p = .02$ (see Figure 19), a difference that works against CP for the global *above-below* category.

Although these results point to CP effects for the specific categories in both visual fields, the results of the previous experiments suggest that performance in the LVF may serve as a baseline for interpreting the pattern of results in the RVF. Comparisons of performance on trials with LVF versus RVF targets revealed that discrimination was slower in the RVF than the LVF on between-global trials, $t(25) = 2.80$, $p = .01$, $d = .55$

(see Figure 19). No such difference was observed for either within-specific or between-specific trials (*p*s > .2). These results suggest that the absence of CP for the global *above-below* category was, if anything, accentuated in the RVF relative to baseline apprehension of the perceptual properties of the stimuli.

An analogous 2 × 3 ANOVA on the accuracy data yielded significant main effects of visual field, $F(1, 25) = 5.37$, $p = .03$, $d = .45$, and category relation, $F(2, 50) = 34.90$, $p < .0001$, but no interaction, $F(2, 50) = .56$, $p > .5$. Accuracy was higher for LVF (M = 83.7%, SD = 7.7) than RVF (M = 80.3%, SD = 8.6) targets, perhaps reflecting the right hemisphere advantage for spatial processing (Ratcliff, 1979). To examine the main effect of category relation, pairwise comparisons of the different trial types (across both visual fields) were conducted. Consistent with the RT results, accuracy was higher on between-specific (M = 85.3%, SD = 7.9) than within-specific (M = 72.2%, SD = 12.1) trials, $t(25) = 5.89$, $p < .0001$, $d = 1.14$, but the difference in accuracy on between-global (M = 88.1%, SD = 7.4) relative to between-specific (within-global) trials did not reach significance, $t(25) = 1.77$, $p = .09$. As in the previous experiments, these results imply that there was no speed-accuracy tradeoff.

The results from the preceding analyses are indicative of CP for the specific *above* and *below* categories, but provide no evidence of CP for the global *above-below* category. However, given that the CP effects for the specific categories were not lateralized, it remains possible that these effects were driven by differences in perceptual similarity, not category membership. Additional RT analyses do not support this possibility. Within-specific trials were divided according to the category membership of the target and distractors, *above* or *below*. For RVF targets, between-specific trials were faster than both within-specific *above* trials, *t*(25) = 2.13, *p* = .04, *d* = .42, and within-specific *below* trials, *t*(25) = 2.44, *p* = .02, *d* = .48. For LVF targets, between-specific trials were faster than within-specific *below* trials, *t*(25) = 3.35, *p* = .003, *d* = .66, but not within-specific *above* trials, *t*(25) = .95, *p* > .3 (see Figure 20). These results replicate Experiment 3 in showing left-lateralized CP for the *above* category (though this effect was not sufficient to produce an overall left-lateralized CP effect), but CP in both visual fields for the *below* category. Because there is no obvious reason why only the latter effects would be governed by perceptual similarity, the results for both the *above* and *below* categories would seem to reflect genuine CP effects (see Section 3.5.2.2).

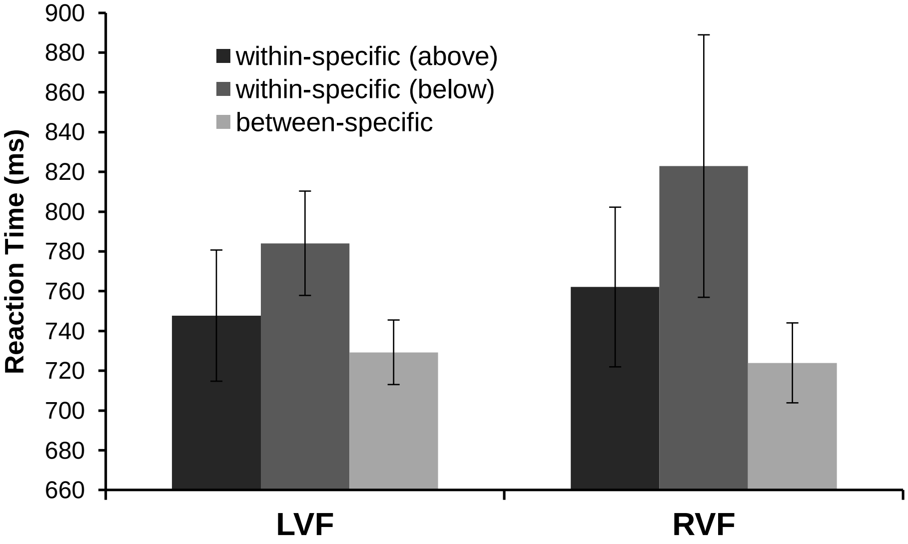

Figure 20. Results of Experiment 4 by specific category. Error bars are 95% within-subjects confidence intervals. The scale of the y-axis is different from that in previous figures to accommodate the larger error bars when breaking down the results by category.

The lack of CP for the global *above-below* category stands in contrast to the results of Experiment 2, in which a left-lateralized CP effect was observed for this category. Unlike that experiment, the present one included trials requiring discriminations within the specific *above* and *below* categories. These relatively fine-grained discriminations may have rendered the more global distinction between *above-below* and *left-right* less salient than when considered in isolation. Taken together, the results from Experiments 2 and 4 suggest that the *above-below* category, though conceptually salient, is trumped in salience by its more fine-grained subordinates, *above* and *below*. This pattern of results is consistent with the differentiation scores of the respective categories.

3.6.2.3 Picture description task

The results of the picture description task were similar to those of the previous experiments. Across the eight pictures, 97.6% of the descriptions included prepositions from the intended category, and the remaining descriptions were either reasonable interpretations of the pictures (e.g., “very close, parallel” for the *right-near* picture) or did not include relational information (e.g., “safe distance” for the *right-far* picture). None of the 208 total descriptions included prepositions from a different category than the intended one. As in the previous experiment, most participants used modifiers to distinguish between the *near* and *far* exemplars of each relation, but used the same preposition for the two exemplars.

3.6.2.4 Summary and conclusions

The results of Experiment 4 are consistent with the semantic clusters hypothesis, but not the simple prevalence view. More differentiated latent categories, despite containing fewer words, were found to be more conceptually salient than less differentiated latent categories. Although the previous experiments showed that both the global *above-below* category (Experiment 2) and the specific *above* and *below* categories (Experiment 3) elicit CP effects, here the latter categories were found to elicit stronger CP effects than the former. Indeed, CP effects for *above-below* were entirely absent when more fine-grained discriminations within the *above* and *below* categories were included in the task.

There is an alternative explanation for the findings of Experiment 4. Although it has been assumed to this point that the CP effects tested in this experiment were those for the latent *above* and *below* categories identified in Experiment 1, it is also possible that these effects reflect CP for the individual words *above* and *below*. This possibility is suggested by the observation that the words within the latent *above* and *below* categories are essentially synonymous, and by the finding that the different exemplars of the *above* and *below* relations in Experiment 4 were usually labeled with the same preposition in the picture description task. If the observed CP effects are for lexical rather than latent categories, this would suggest that categories associated with individual words might be more conceptually salient than latent categories, regardless of differentiation. Moreover, such an effect of lexical superiority would imply that words provide a better window on concepts than suggested by the semantic clusters hypothesis.

It is impossible to rule out this lexical superiority view based on the present findings because differentiation and category status were confounded in Experiment 4. The more differentiated categories, *above* and *below*, were also those that might be considered lexical rather than latent. Experiment 5 was designed to tease apart these factors. The conceptual salience of the global *left-right* category was tested relative to that of its subordinates, *left* and *right*. In Experiment 1, *left-right* was shown to be more differentiated than its subordinates. There was also no evidence of latent *left* and *right* categories, suggesting that the distinction between *left* and *right* is purely lexical. According to the lexical superiority view, these lexical categories should be more conceptually salient than the global *left-right* category. Because *left-right* is more differentiated, however, the semantic clusters hypothesis makes the opposite prediction. Experiment 5 tested these competing predictions by comparing the respective CP effects of the two sets of categories.

3.7 Experiment 5: CP for *left-right* vs. CP for *left* and *right*

Similar to the previous experiment, Experiment 5 pitted CP effects for the global *left-right* category against those for its subordinates, the lexical categories *left* and *right*. The experiment included essentially the same three kinds of trials as in Experiment 4, except with lexical categories rather than specific (latent) ones. Participants were presented with spatial relations from the same lexical category (within-lexical trials; i.e., *left* vs. *left* or *right* vs. *right*), from different lexical categories but the same global category (between-lexical/within-global trials; i.e., *left* vs. *right*), and from different global categories (between-global trials; in this case, *left-right* vs. *front-back*, with the latter serving as a comparison global category). CP effects for the *left-right* category would be indicated by superior performance on between-global than between-lexical/within-global trials, whereas CP effects for the *left* and *right* categories would be indicated by superior performance on between-lexical/within-global than within-lexical trials. The magnitudes of the respective CP effects were compared to determine which set of categories is more conceptually salient. Previous research has shown that CP for lexical categories is left-lateralized (Gilbert et al., 2006, 2008; Holmes & Wolff, 2012a), and the previous experiments in the present investigation suggest that CP for latent categories may likewise be more robust in the left hemisphere. Thus, in the present experiment, special attention was paid to interpreting the pattern of results obtained in the RVF.

3.7.1 Method

3.7.1.1 Participants

Twenty-six participants (19 female) from the same population sampled in the previous three experiments took part in this experiment. Four participants (2 female) were excluded for low accuracy.

3.7.1.2 Materials

The materials were the eight pictures in Figure 21, showing four possible relations between the bird and the airplane from the top view: *left*, *right*, *front*, and *back*, with *near* and *far* versions of each relation. The top view, rather than the front view of the previous experiment, was chosen because the width of the airplane was smaller from this view. This allowed for a greater distance between the position of the bird in the *near* and *far* versions of the *left* and *right* pictures, for which discrimination was extremely poor in Experiment 4. As a result, *front-back* (rather than *above-below*) served as the comparison category on between-global trials. Some participants in Experiment 2 interpreted *front-back* pictures two-dimensionally. With the present materials, such an interpretation would lead to *front* and *back* pictures being treated as *above* and *below*, respectively, and hence still in a different global category than the *left* and *right* pictures.

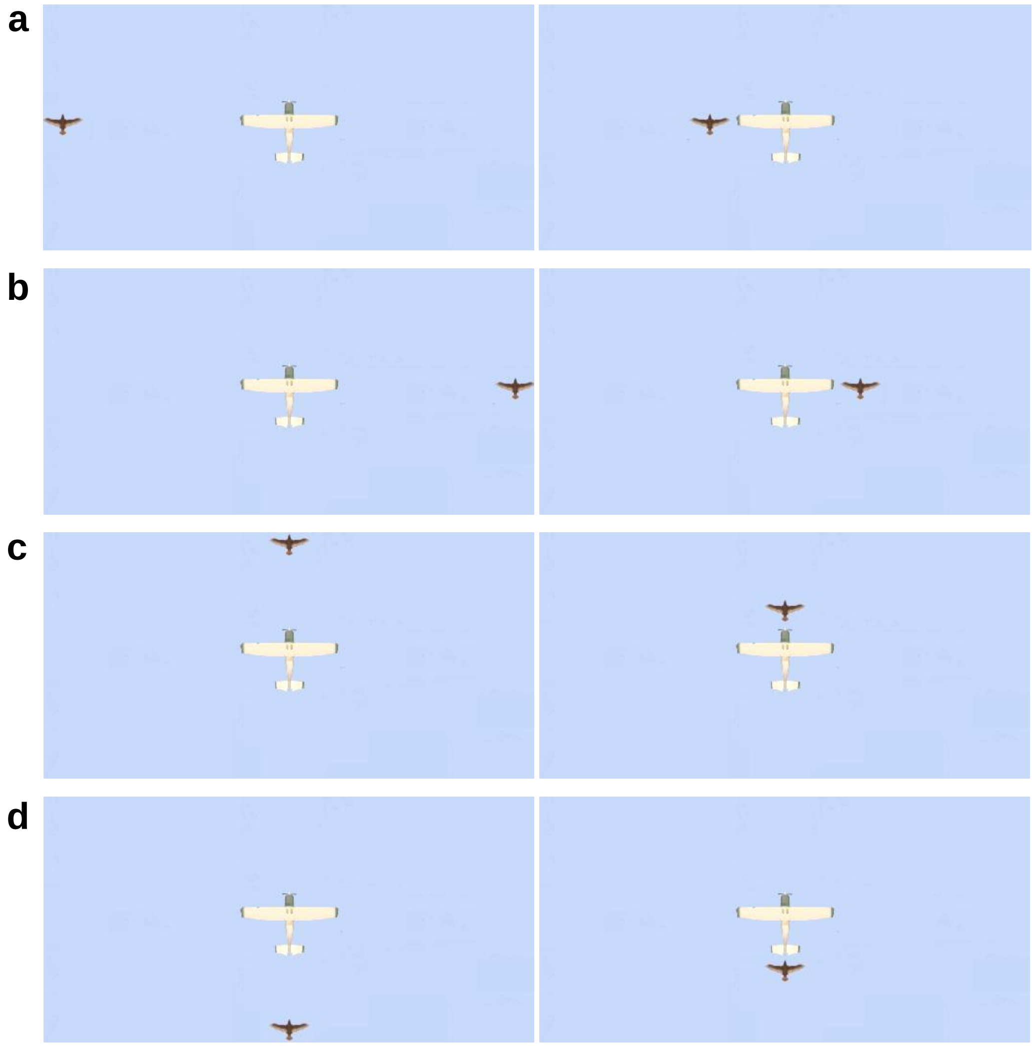

Figure 21. Stimuli used in Experiment 5. (*A*) *left-far* and *left-near*. (*B*) *right-far* and *right-near*. (C) *front-far* and *front-near*. (D) *back-far* and *back-near*.

The distance between the bird and the airplane was the same in the *left-near* and *right-near* pictures, in the *left-far* and *right-far* pictures, and in their *front* and *back* counterparts. The discrimination task included the same kinds of displays used in the previous two experiments.

3.7.1.3 Design and procedure

The design was similar to that of the previous experiment, with three kinds of "different" trials. Because *left* and *right* are lexical rather than specific (latent) categories, the within-specific and between-specific trial types were renamed within-lexical and between-lexical, respectively. Thus, the experiment consisted of within-lexical, between-lexical (within-global), and between-global trials. All other aspects of the design and procedure were identical to those of Experiment 4.

3.7.2 Results and discussion

3.7.2.1 Overview of key findings

The results support the semantic clusters hypothesis over the lexical superiority view in showing that more differentiated latent categories elicit stronger CP effects than less differentiated lexical categories, at least in the RVF. As shown in Figure 22, discrimination in the RVF was faster when spatial relations came from different global categories (i.e., *left-right* vs. *front-back*) than from the same global category (i.e., *left* vs. *right*). No such difference was observed for spatial relations from different lexical categories (same global category) compared to the same lexical category (i.e., *left* vs. *left*, or *right* vs. *right*). Thus, when pitted against each other, CP effects were observed for the more differentiated *left-right* category, but not the less differentiated *left* and *right* categories, even though the latter categories are lexical. The pattern of results in the LVF differed from that in the RVF, a point to be elaborated on below.

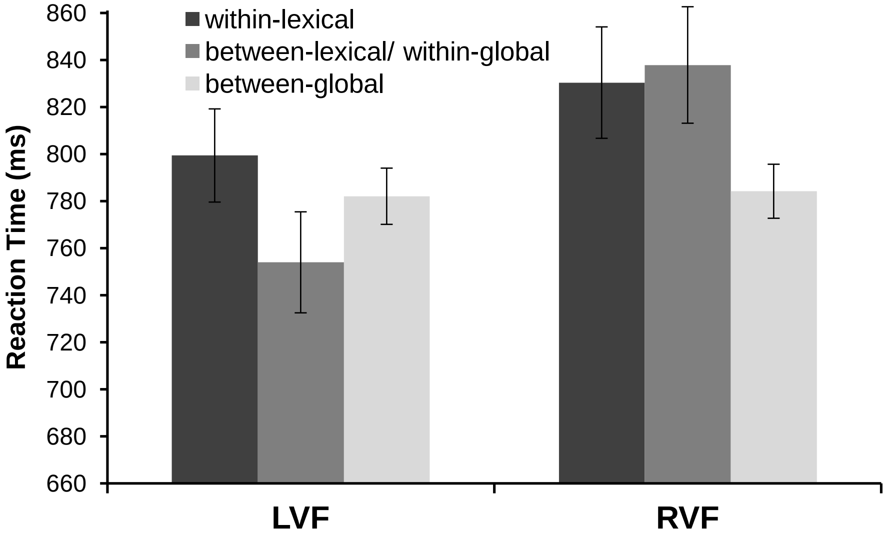

Figure 22. Results of Experiment 5 (*N* = 22). Error bars are 95% within-subjects confidence intervals.

3.7.2.2 Discrimination task

Mean accuracy on the discrimination task was 85.4% (*SD* = 6.9), with no difference between “same” and “different” trials (*p* > .8). To test the predictions of interest, subsequent analyses focused on “different” trials for which either the target or distractors (or both) came from the *left-right* category or its lexical subordinates. As in the previous experiments, trials in which participants responded incorrectly (9.9%) or in which RT was greater than 2.5 standard deviations from individual means (3.9%) were excluded. A 2 (visual field) × 3 (category relation) ANOVA on the RT data for the remaining trials yielded significant main effects of visual field, $F(1, 21) = 12.66$, $p = .002$, and category relation, $F(2, 42) = 10.09$, $p = .0002$, and a significant interaction, $F(2, 42) = 6.86$, $p = .003$. Given this interaction, pairwise comparisons of the different trial types were conducted separately for each visual field. For RVF targets, discrimination was faster on between-global than between-lexical (within-global) trials, $t(21) = 4.03$, $p = .0006$, $d = .86$, indicative of CP for the global *left-right* category, but not on between-lexical relative to within-lexical trials, $t(21) = .50$, $p > .6$, providing no evidence of CP for the lexical *left* and *right* categories. A different pattern was observed for LVF targets: Discrimination was faster on between-lexical than on both within-lexical trials, $t(21) = 3.70$, $p = .001$, $d = .79$, and between-global trials, $t(21) = 2.47$, $p = .02$, $d = .53$ (see Figure 22).

These different patterns of results in the two visual fields may, in part, be due to differences in perceptual similarity among the different trial types. As suggested previously, performance in the LVF may reflect baseline apprehension of the perceptual properties of the stimuli, a pattern against which performance in the RVF may be interpreted. Comparisons of performance on trials with LVF versus RVF targets revealed that discrimination was slower in the RVF than the LVF on between-lexical trials, $t(21) = 4.07$, $p = .0005$, $d = .87$, a difference leading to weaker CP effects for the *left* and *right* categories but stronger CP effects for the *left-right* category (see Figure 22). No such difference was observed for either within-lexical or between-global trials (ps > .1). These results suggest the existence of CP effects in the RVF over and above significant differences in baseline discrimination across the different kinds of trials.

This characterization of LVF performance as reflecting baseline discrimination is supported by analyses of the accuracy data. An analogous 2 × 3 ANOVA on these data yielded significant main effects of visual field, *F*(1, 21) = 6.37, *p* = .02, *d* = .55, and category relation, *F*(2, 42) = 11.59, *p* < .0001, but no interaction between the two factors, *F*(2, 42) = .32, *p* > .7. As in Experiment 4, the main effect of visual field revealed that accuracy was higher for LVF (*M* = 91.6%, *SD* = 6.6) than RVF (*M* = 88.0%, *SD* = 8.2) targets, consistent with the right hemisphere advantage for spatial processing (Ratcliff, 1979). Pairwise comparisons of the different trial types (across both visual fields) showed that accuracy was significantly higher on between-lexical (*M* = 93.0%, *SD* = 5.6) than within-lexical (*M* = 84.2%, *SD* = 11.8) trials, *t*(21) = 3.68, *p* = .001, *d* = .78, and marginally higher on between-lexical than between-global (*M* = 91.6%, *SD* = 6.7) trials, *t*(21) = 1.84, *p* = .08. This pattern of results, consistent with that observed for the RT data in the LVF (and thus implying no speed-accuracy tradeoff), suggests that the spatial relations on between-lexical trials (i.e., *left* vs. *right*) were less perceptually similar than the spatial relations on the other two kinds of trials, producing an overall benefit in discrimination. In contrast, between-lexical trials were much slower in the RVF, suggesting that CP eliminated the baseline advantage for these trials evident in the accuracy data and the LVF RT data.

Additional RT analyses were likewise consistent with this interpretation. Within-lexical trials were divided according to the category membership of the target and distractors, *left* or *right*. For RVF targets, between-lexical trials were no faster than within-lexical *left* or within-lexical *right* trials (*p*s > .4), consistent with the overall pattern in the RVF. For LVF targets, between-lexical trials were faster than both within-lexical *left* trials, *t*(21) = 2.08, *p* = .05, *d* = .44, and within-lexical *right* trials, *t*(21) = 3.39, *p* = .003, *d* = .72 (see Figure 23), consistent with the overall pattern in the LVF. Thus, in both visual fields, the results for each lexical category (*left* and *right*) mirror the overall results when collapsing across the two categories.

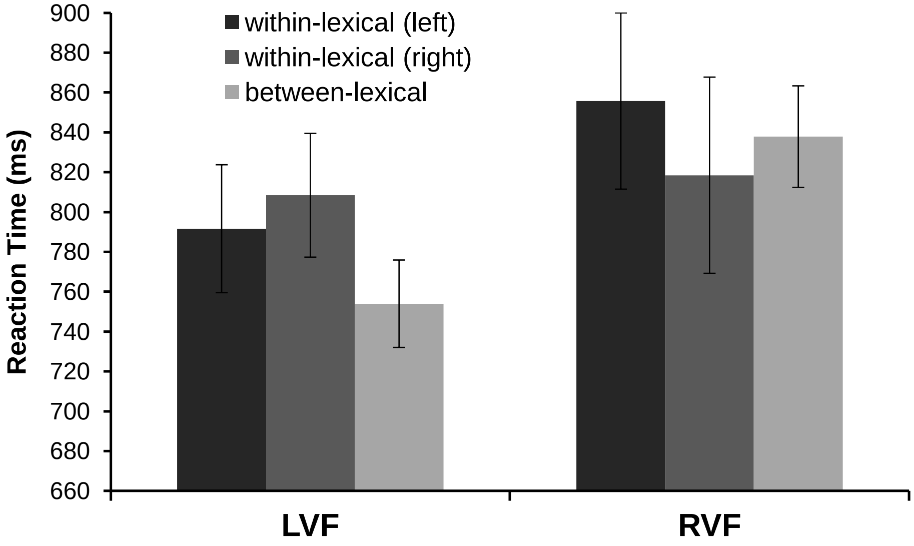

Figure 23. Results of Experiment 5 by lexical category. Error bars are 95% within-subjects confidence intervals. The scale of the y-axis is different from that in previous figures to accommodate the larger error bars when breaking down the results by category.

The present results replicate those of Experiment 2 in showing CP in the RVF for the *left-right* category. Thus, CP for *left-right* was found not only when examined in isolation, but also when pitted against its lexical subordinates. This pattern of results is consistent with the differentiation scores of the respective categories.

3.7.2.3 Picture description task

The results of the picture description task were similar to those of the previous experiments. The description of one participant was lost due to experimenter error. Across the eight pictures, 77.4% of the remaining descriptions included prepositions from the intended category. For the *left* and *right* pictures, there was 98.8% agreement. For the *front* and *back* pictures, nine of the 21 participants for whom descriptions were analyzed consistently used vertical terms (e.g., “the bird is far above the plane” for the *front-far* picture in Figure 21c), implying that they viewed the pictures two-dimensionally rather than three-dimensionally. However, as noted above, such an interpretation still treats these pictures as members of a different global category (*above-below*) than the *left* and *right* pictures. As in the previous experiments, most participants used modifiers to distinguish between the *near* and *far* exemplars of each relation.

3.7.2.4 Summary and conclusions

The results of Experiment 5 are consistent with the semantic clusters hypothesis, but not the lexical superiority view. More differentiated latent categories were found to be more conceptually salient than less differentiated lexical categories, inconsistent with the idea that prominent conceptual distinctions are explicitly coded in language. These findings add to the existing body of evidence showing dissociations between words and concepts (see Section 1.5).

Taken together, the results of Experiments 4 and 5 suggest that the differentiation principle provides a better account of conceptual salience than either the simple prevalence view or the lexical superiority view. In Experiment 4, CP was found to be stronger for the specific *above* and *below* categories than the global *above-below* category, inconsistent with the simple prevalence view. In Experiment 5, CP was found to be stronger for the global *left-right* category than the lexical *left* and *right* categories, inconsistent with the lexical superiority view (and implying that the CP effects for *above* and *below* in Experiment 4 were for latent rather than lexical categories). Only the differentiation principle can explain the results from both experiments, suggesting that this principle—and the semantic clusters hypothesis more generally—may serve as a useful rubric for determining which semantic categories are more prominent than others at the conceptual level.

3.8 Overall summary

Four experiments were conducted to test the predictions of the semantic clusters hypothesis. The prediction that latent categories would be conceptually salient was supported by the results of Experiments 2 and 3. Several latent categories at different levels of spatial semantic structure, global and specific, were found to elicit CP effects. The prediction that more differentiated latent categories would be more conceptually salient than less differentiated categories was supported by the results of Experiments 4 and 5. Latent categories, whether global or specific, were found to elicit stronger CP effects than other categories if they were more differentiated. This set of results was shown to be incompatible with two alternative hypotheses, simple prevalence and lexical superiority. Altogether, the results of these four experiments provide compelling support for the semantic clusters hypothesis. The findings indicate that, at least in the spatial domain, clusters of words align with distinctions that figure prominently in the nonlinguistic processing of spatial relations. They also suggest that the more differentiated these clusters are—that is, the greater the similarity of the words within a cluster compared to the words in different clusters—the greater the role of the corresponding conceptual distinctions in nonlinguistic thinking.

Chapter 4: General discussion

4.1 When language is a window into the mind—and when it isn't

Recent research on the language-thought interface has led to a paradox. Although language is often viewed as a rich source of evidence about the human mind, a number of studies have revealed striking dissociations between word meanings and conceptual knowledge. The present research offered 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 should map onto prominent conceptual distinctions, with highly differentiated clusters being the most conceptually salient.

The results of five experiments in the spatial domain provide support for this hypothesis. In Experiment 1, the semantic structure of spatial relations was inferred from similarities among the meanings of spatial prepositions, which formed clusters at different levels of granularity. In Experiments 2 and 3, several of these clusters were shown to align with conceptual distinctions influential in the nonlinguistic processing of spatial relations. In Experiments 4 and 5, clusters that were more differentiated at the semantic level were also found to be more salient at the conceptual level. Together, these findings suggest that words, despite often failing to reveal unique concepts, may nevertheless have much to tell us about the nature of the conceptual system. The findings indicate that a broader perspective on word meaning may be necessary to capitalize on this potential. Taking into consideration the macrosemantics of a domain, elements of meaning shared by many words, suggests that language and the conceptual system may share a common underlying structure, even though the particulars of the two systems may vary considerably. This common structure may often be obscured from view when focusing exclusively on the nuances that distinguish individual word meanings. Thus, the findings highlight the importance of considering multiple levels of representational structure for understanding the relationship between language and thought. A seemingly loose connection from the perspective of words may belie a much tighter relationship at deeper levels of structure.

4.1.1 Further contributions of the present investigation

In addition to illuminating the nature of the language-thought interface, the present investigation offers a possible template for conducting rigorous research on concepts. As discussed in Chapter 3 (see Section 3.1), early work on the Whorfian hypothesis was criticized for taking the existence of cross-linguistic differences in word meaning as evidence for conceptual differences. Yet much modern research on the conceptual system also conflates words with concepts, implicitly subscribing to what I have been calling the standard view (see Figure 1a). In many studies of real-world concepts (as opposed to artificial, experimenter-defined ones), participants are asked to name, recall, list features for, judge the similarity of, or make inductive inferences about words, and resulting patterns of performance (and increasingly, neural activity; see Mahon & Caramazza, 2009) are interpreted as reflecting underlying concepts (for examples, see Medin, Lynch, & Solomon, 2000; Murphy, 2002). However, if words and concepts are at least partially dissociated, such an interpretation is unwarranted (Malt et al., 2011). Even ostensibly nonlinguistic measures of concepts (e.g., sorting pictures) are vulnerable to this concern if participants access their lexical knowledge while completing them (Pinker, 1994; Winawer et al., 2007).

To avoid these pitfalls, the present research utilized nonlinguistic tasks that minimized the involvement of linguistic representations, and that were designed such that evidence for the candidate concepts under investigation could not be attributed to the meanings of individual words (but see Section 3.6.2.4 for some caveats). Future research on concepts might benefit from employing such tasks to verify that the representations being studied exist in a form independent of language. Moreover, using such tasks in tandem with a macrosemantic approach to meaning, as exemplified in the present research, would expand the range of candidate concepts to be examined. Of course, perceptual tasks are ill-suited to studying domains for which the entities are not directly observable (e.g., mental states) and to getting at aspects of conceptualization that depend on elaborate background knowledge (e.g., naïve theories; Murphy & Medin, 1985). Nevertheless, such tasks may provide substantial insight into the kinds of distinctions that figure most prominently in people's everyday nonlinguistic thinking and how the associated representations are organized within the conceptual system.

Why has the word-concept conflation been so hard to shake? One likely reason is that many researchers studying concepts may be unaware of the extent of linguistic diversity and its implications for this assumption. Yet even when such diversity is taken into consideration, it seems that the most common solution is to endorse the good window view (see Figure 1b), preserving the word-concept alignment by positing that speakers of different languages have different concepts (Gentner & Goldin-Meadow, 2003b; Gumperz & Levinson, 1996). The kind of logic typically cited for this view is that "[e]very linguistic distinction must be supported by the relevant conceptual distinctions, perceptual acuities, and mental algorithms" (Levinson, 1996, p. 374). At some level, this claim must be true because otherwise it would be impossible to use one's language properly; as Landau and Jackendoff (1993) note, "whatever we can talk about we can also represent" (p. 217). However, it is far from a given that the representations underlying language use are the same as those supporting nonlinguistic thinking, and a growing body of research suggests that they are not (see Section 1.4). The present research illustrates how the insight that words and concepts differ may lead to the discovery of deeper, and arguably more interesting, ways in which language and the conceptual system may nevertheless be connected.

4.1.2 Limitations and benefits of dimensionality reduction

A macrosemantic approach can only go so far in characterizing the meanings of words. Applying dimensionality reduction methods to linguistic data inevitably leads to information loss, and the results may capture little of what it intuitively means to know a word (Murphy, 2002). For example, showing that *beside* is a member of the *left-right* cluster does not tell us very much about its meaning. Moreover, some of the words examined in Experiment 1 (e.g., *outside*) did not fall neatly into any of the clusters that emerged from the analyses. Critically, however, the purpose of the macrosemantic approach is not to provide a complete account of lexical representation, but rather to identify recurring patterns across many words that may be linked to conceptual knowledge. As illustrated in Experiment 1, this approach can reveal how words that one might not have expected to go together (e.g., words with opposite meanings, as in the *above-below*, *left-right*, and *front-back* clusters) may in fact be represented as a cohesive semantic unit. The conceptual salience of these units suggests that the statistical algorithms at work in dimensionality reduction may mirror a kind of abstraction process that regularly occurs in the human mind. The present findings also underscore the value of using multiple dimensionality reduction methods to infer semantic structure. The results from any single method present a skewed view of this structure, whereas multiple methods may provide both shared and complementary information.

4.2 The inevitable Whorfian question

The present findings demonstrate how language may serve as a window into the mind, but they do not speak directly to the question that has dominated research on the language-thought interface for much of the past half century, namely whether language shapes thought (see Wolff & Holmes, 2011). The spatial semantic structure of English, though shown to be conceptually salient for English speakers, is not necessarily universal. In principle, languages could differ in the clusters formed by their spatial terms, and these differences could lead to differences in nonlinguistic spatial thinking.

On the one hand, such differences might seem unlikely because clusters of words presumably reflect, and may be constrained by, structure in the world to a much larger degree than individual words (see Malt et al., 2008). Languages will surely differ in the number of terms within the *above-below*, *left-right*, and *front-back* clusters, but it seems doubtful that any of these clusters would be entirely absent from a language, simply because the spatial relations they capture seem so fundamental to our experience in the world (Clark, 1973; Franklin & Tversky, 1990). On the other hand, many striking cross-linguistic differences in spatial word meanings have already been documented, suggesting the possibility of deeper differences in spatial semantic structure. For example, Korean lexicalizes the distinction between tight-fit (*kkita*) and loose-fit (*nehta/nohta*), collapsing over the English distinction between containment (*in*) and support (*on*; Bowerman & Choi, 2001). It is unclear where the Korean terms would fall in the English structure, or whether they could even be accommodated at all. Similarly, many languages make little or no use of relative spatial terms like *to the left of* and *to the right of*, relying instead on allocentric spatial coordinates (e.g., north/south/east/west; Majid et al., 2004). Spatial terms in such languages might conceivably form quite different clusters from those observed in English.

If differences in spatial semantic structure are observed across languages, an investigation of corresponding conceptual differences would provide a particularly strong test of the semantic clusters hypothesis. Though agnostic with respect to the universality of semantic structure, the hypothesis implies that clusters of words in a given language should prove conceptually salient even if they are not found cross-linguistically. Speakers of languages with different clusters would thus be expected to show different patterns of performance on CP tasks. Consider, for example, a hypothetical language in which *left-right* terms divide into separate *left* and *right* clusters. CP effects for the distinction between *left* and *right* should be stronger in speakers of this language than in English speakers, for whom the distinction is purely a lexical one. This kind of Whorfian effect would be unlike any previously reported in the literature in that it would be driven by categories not explicitly encoded in the semantic system (and of which many languages users may have no conscious awareness). Probing the existence of such effects may represent the next frontier in research on the Whorfian hypothesis.

4.3 Extending the differentiation principle

The findings from Experiments 4 and 5 highlight differentiation, the principle underlying the semantic clusters hypothesis, as a useful metric for predicting which semantic categories will align with concepts. Following research on basic level categorization, differentiation was operationalized in terms of both specificity and distinctiveness, corresponding to within-category similarity and between-category dissimilarity, respectively (Mervis & Crisafi, 1982; Murphy & Brownell, 1985). In future work, it would be informative to examine the relative predictive power of these two properties. For the spatial categories examined here, specificity played a larger role simply because distinctiveness was uniformly high; the core members of different clusters were rarely grouped together. In other domains, particularly those for which cross-classification is common (e.g., foods: meats/vegetables vs. breakfast foods/dinner foods; Ross & Murphy, 1999; see also Barsalou, 1983), the two properties might be weighted differently. In the food domain, for example, a “meat” cluster might be low in distinctiveness relative to a “breakfast foods” cluster because certain terms might reasonably be considered core members of both clusters (e.g., *sausage*). Greater variability in the distinctiveness of different clusters would allow for examining the importance of distinctiveness in predicting conceptual salience, and for validating the differentiation principle more generally.

Differentiation may also provide an explanation for why evidence supporting the Whorfian hypothesis is often hard to come by. In many of the domains for which potential Whorfian effects have been investigated, lexical categories may be relatively low in differentiation. Grammatical gender categories, for example, pick out groups of entities that differ dramatically: Clocks, forks, and books have little in common perceptually or functionally, but all are masculine in Spanish; magazines, despite their relative similarity to books, are feminine.[22] Thus, it may not be especially surprising that evidence for the conceptual salience of these categories is, at best, mixed (Sera et al., 2002; Vigliocco et al., 2005). Even categories named by common nouns may not be highly differentiated. Chairs and sofas, bottles and jars, and bowls and plates may be much more similar to each other than their contrastive labels would imply (Malt et al., 1999, 2003, 2011). These observations are consistent with the possibility that word meanings draw finer distinctions than are provided by structure in the world (see Sections 1.5-1.6). Many lexical categories, though highly specific, may not be especially distinctive, and hence will serve as a relatively poor reflection of how the entities to which they refer are conceptualized. Perhaps more likely candidates for conceptual salience are natural kind categories, such as plants and animals, which pick out scientifically distinct species with many correlated properties (Berlin, 1978; Malt, 1995). However, given that the world presents such strong structure for plants and animals, there may, as a consequence, be relatively little linguistic diversity in these domains, and hence little potential for Whorfian effects. This analysis suggests that at the level of individual

---

[22] Here I am referring to the similarities among entities in the world, rather than the similarities among words, which were used to calculate differentiation scores for latent categories in Chapter 2. The notion of differentiation is essentially the same for the two sets of items.

words, the Whorfian hypothesis faces a considerable uphill battle: When lexical categories are low in differentiation, they may fail to align with concepts, but when they are highly differentiated, they may align with concepts that are shared across different language groups.

4.4 Implications for related areas of research

Although the present investigation used space strictly as a test bed for investigating the semantic clusters hypothesis, the findings may inform research on spatial cognition. Similarly, although CP was employed as a tool for assessing conceptual salience, the findings on CP for spatial categories may have implications for the study of CP more generally. Here I discuss the contributions of the present findings to these related areas of research.

4.4.1 Spatial cognition through the lens of spatial language

Much research on spatial cognition has focused on egocentric representations of space based on body-centered coordinates (Vallar et al., 1999; Wang & Spelke, 2002). There is also considerable evidence that space is represented allocentrically, with cognitive maps encoding the surrounding environment for navigation purposes (O’Keefe & Nadel, 1978; Poucet, 1993). A recent proposal by Jeffery, Jovalekic, Verriotis, and Hayman (2012) posits that allocentric representations of three-dimensional space are “bicoded,” with the vertical plane represented qualitatively differently from the horizontal plane (i.e., the plane of navigation). According to Jeffery et al., organisms’ greater experience with locomotion along the horizontal plane leads to relatively fine-grained horizontal representations and coarser, non-metric vertical representations.

Intriguingly, the present findings, both linguistic and nonlinguistic, point to spatial representations that may resemble those proposed by Jeffery et al. (2012). The distinction between vertical and horizontal is evident in the spatial semantic structure inferred in Experiment 1, particularly in that the first major cut of the domain (suggested by the HC analysis) was between vertical terms and all other prepositions (including *left-right* and *front-back* terms, both referring to the horizontal plane). Differences in the granularity of horizontal and vertical representations are suggested by the observation that prepositions encoding metric information (e.g., *near*, *far from*) clustered exclusively with horizontal terms, and that only vertical terms divided into discrete, differentiated clusters (i.e., *above* vs. *below*). Additionally, the vertical *above* and *below* clusters gave rise to CP effects (Experiments 3 and 4), implying non-metric representations, but there was no evidence that the horizontal plane is perceived categorically (i.e., *left* vs. *right*; Experiment 5). Although these findings are consistent with Jeffery et al.'s proposal, they suggest an alternative explanation for the differences between horizontal and vertical representations. Given that spatial prepositions often refer to static spatial configurations, not necessarily motion through space, bicoded representations may be a property of spatial perception in general, rather than being tied specifically to navigation (Holmes & Wolff, 2012b). These observations illustrate how exploring the language-thought interface may inform research on other aspects of cognition for which linguistic evidence is not traditionally considered (Wolff & Malt, 2010).

4.4.2 Categorical perception of spatial relations

Although CP has been demonstrated for many different kinds of visual stimuli, the present findings are the first to show that CP can result from categories defined by the relations among stimuli, rather than properties of the stimuli themselves (e.g., color, shape). For all of the spatial categories examined here, the objects in the different categories (i.e., the bird and the airplane) were identical; what differed were the locations of the objects with respect to each other. The observation of CP for relational categories suggests that the packaging of relational information into discrete units, a fundamental problem in early word learning (Gentner, 1982; Parish-Morris, Pruden, Ma, Hirsh-Pasek, & Golinkoff, 2010), may be grounded in the workings of our perceptual systems.

The CP effects observed in the present experiments were predominantly lateralized to the left hemisphere. The phenomenon of left-lateralized CP, first discovered by Gilbert et al. (2006) in the color domain and since extended to other kinds of visual categories and stimuli (e.g., Gilbert et al., 2008; Holmes & Wolff, 2012a; Zhou et al., 2010), has recently been contested by several non-replications (Brown, Lindsey, & Guckes, 2011; Witzel & Gegenfurtner). The present findings provide support for the generality of left-lateralized CP effects, but they also challenge the prevailing Whorfian interpretation of these effects. In Experiments 2 and 5, spatial relations were discriminated faster in the RVF (left hemisphere) when they came from different global clusters than the same global cluster, even though the relations had different labels in both cases. These findings are consistent with Holmes and Wolff's (2012a) proposal that CP effects are driven by nonlinguistic categorical representations, not lexical codes.

4.5 In defense of concepts

Throughout this dissertation, I have been referring to concepts in the way that they have traditionally been characterized in the field; that is, as “mental representations of classes of things” (Murphy, 2002, p. 5), presumably taking the form of “discrete, bounded, and stable units of knowledge stored in long-term memory” (Malt et al., 2011, p. 524). This view of concepts has been challenged on various grounds, with alternative characterizations emphasizing the dynamic and context-sensitive nature of knowledge representation (Barsalou, 1987), the primacy of perceptual information in concept learning (Smith & Jones, 1993), and the heterogeneity of the processes governing concept use (Machery, 2009). Recently, the traditional view of concepts has been challenged on linguistic grounds. Malt et al. (2010, 2011) argued that because words do not necessarily map onto discrete chunks of nonlinguistic knowledge, it may be inaccurate to view the conceptual system as possessing any such chunks or inherent boundaries. For this reason, Malt et al. advocated discarding the notion of concepts altogether, arguing that nonlinguistic knowledge may be better characterized in terms of more fine-grained components or features (e.g., manner, path, goal, end)—what may be considered *micro*semantic aspects of meaning (see also Pacer et al., 2012; Xu & Kemp, 2010). However, such microunits, or “primitives,” though smaller and less confusable with word meanings, seem no less “discrete, bounded, and stable” than the traditional concepts they are intended to replace. The present research, in providing evidence for conceptually salient global meanings, implies the existence of similarly discrete units of knowledge at a coarser level of semantic structure. These observations suggest that looking beyond word meanings, whether above or below the level of the word, may bring us closer to discovering concepts, in their traditional sense, than ever before.

References

Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language, 53*, 60–80.

Au, T. K.-F. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language, 25*, 104–122.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101-140). New York, NY: Cambridge University Press.

Berlin, B. (1978). Ethnobiological classification. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 9–26). Hillsdale, NJ: Lawrence Erlbaum.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–79). Cambridge, MA: MIT Press.

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge, UK: Cambridge University Press.

Bowerman, M., & Pederson, E. (1992). Cross-linguistic studies of spatial semantic organization. *Annual Report of the Max Planck Institute for Psycholinguistics 1992*, 53–56.

Brown, A. M., Lindsey, D. T., & Guckes, K. M. (2011). Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision*, *11*. doi: 10.1167/11.12.2

Bush, L. E. (1973). Individual differences multidimensional scaling of adjectives denoting feelings. *Journal of Personality and Social Psychology*, *25*, 50–57.

Carroll, J. D. (1976). Spatial, non-spatial and hybrid models for scaling. *Psychometrika*, *41*, 439–463.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*, 83–121.

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

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–63). New York: Academic Press.

Coventry, K. R., & Garrod, S. C. (2004). *Saying, seeing and acting: The psychological semantics of spatial prepositions*. New York, NY: Psychology Press.

Deese, J. (1962). On the structure of associative meaning. *Psychological Review*, *69*, 161–175.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences*, *104*, 1097–1102.

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*, 429–492.

Everitt, B. S., & Rabe-Hesketh, S. (1997). *The analysis of proximity data*. London, UK: Arnold.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences, 109*, 17897–17902.

Feist, M. I. (2000). *On* in *and* on*: An investigation into the linguistic encoding of spatial scenes*. (Doctoral dissertation.) Northwestern University, Evanston, IL.

Feist, M. I. (2008). Space between languages. *Cognitive Science, 32*, 1177–1199.

Franklin, A., Catherwood, D., Alvarez, J., & Axelsson, E. (2010). Hemispheric asymmetries in categorical perception of orientation in infants and adults. *Neuropsychologia*, *48*, 2648–2657.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, *119*, 63–76.

Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, *24*, 1124–1179.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*, 49–79.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development, volume 2: Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum Associations.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge, UK: Cambridge University Press.

Gentner, D., & Goldin-Meadow, S. (2003a). Whither Whorf. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 3-14). Cambridge, MA: MIT Press.

Gentner, D., & Goldin-Meadow, S. (Eds.). (2003b). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.

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

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language, 105,* 91–98.

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & B. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 633–661). Cambridge, UK: Cambridge University Press.

Goldstein, J., & Davidoff, J. (2008). Categorical perception of animal patterns. *British Journal of Psychology*, *99*, 229–243.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*, 178–200.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 69–78.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*, 496–499.

Gumperz, J. J., & Levinson, S. C. (Eds.). (1996). *Rethinking linguistic relativity*. Cambridge, UK: Cambridge University Press.

Harnad, S. (1987). Introduction: Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1–25). Cambridge, UK: Cambridge University Press.

Hellige, J. B. (1993). *Hemispheric asymmetry: What's right and what's left*. Cambridge, MA: Harvard University Press.

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior*, *8*, 176–184.

Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions of English*. Cambridge, UK: Cambridge University Press.

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

Holmes, K. J., & Wolff, P. (2012b, to appear). Spatial language as a window on representations of 3D space: Commentary on Jeffery et al. *Behavioral and Brain Sciences*.

Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2012). Multidimensional scaling. *Wiley Interdisciplinary Reviews: Cognitive Science*. doi: 10.1002/wcs.1203

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, *62*, 169–200.

Jeffery, K. J., Jovalekic, A., Verriotis, M., & Hayman, R. (2012, in press). Navigating in a 3D world. *Behavioral and Brain Sciences*.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

Jolliffe, I. T. (2002). *Principal component analysis, second edition*. New York, NY: Springer.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*, 65–79.

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, *100*, 9085–9089.

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

Khetarpal, N., Majid, A., & Regier, T. (2009). Spatial terms reflect near-optimal spatial categories. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2396–2401). Austin, TX: Cognitive Science Society.

Khetarpal, N., Majid, A., Malt, B., Sloman, S., & Regier, T. (2010). Similarity judgments reflect both language and cross-language tendencies: Evidence from two semantic domains. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 358–363). Austin, TX: Cognitive Science Society.

Kikutani, M., Roberson, D., & Hanley, J. R. (2010). Categorical perception for unfamiliar faces: Effect of covert and overt face learning. *Psychological Science, 21,* 865–872.

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: What categories reveal about the mind*. Chicago, IL: 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. (1996). Language and space. *Annual Review of Anthropology, 25,* 353–382.

Levinson, S. C. (2001). Space, linguistic expression of. In N. J. Smelser & P. Baltes (Eds.), *International encyclopedia of social and behavioral sciences, vol. 22* (pp. 14749–14752). Amsterdam: Elsevier Science.

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.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology, 3*. doi:10.3389/fpsyg.2012.00054

Machery, E. (2009). *Doing without concepts*. New York, NY: Oxford University Press.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51.

Mandler, J. M. (2008). On the birth and growth of concepts. *Philosophical Psychology*, *21*, 207–230.

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.

Majid, A., Enfield, N. J., & van Staden (2006). Parts of the body: Cross-linguistic categorisation. *Language Sciences*, *28*, 137–147.

Malt, B. C. (1995). Category coherence in cross-cultural perspective. *Cognitive Psychology*, *29*, 85–148.

Malt, B. C., Ameel, E., Gennari, S., Imai, M., Saji, N., & Majid, A. (2011). Do words reveal concepts? In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting 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* (pp. 29–57). New York, NY: Oxford University Press.

Malt, B. C., Gennari, S., Imai, M., Ameel, E., Tsuda, N., & Majid, A. (2008). Talking about walking: Biomechanics and the language of locomotion. *Psychological Science, 19,* 232–240.

Malt, B. C., & Sloman, S. A. (2004). Conversation and convention: Enduring influences on name choice for common objects. *Memory & Cognition*, *32*, 1346–1354.

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., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language*, *49*, 20–42.

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

Mandler, J. (2008). On the birth and growth of concepts. *Philosophical Psychology*, *21*, 207–230.

McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology, 46*, 229–259.

Medin, D. L., Lynch, E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual Review of Psychology, 51*, 121–147.

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development, 53*, 258–266.

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

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 70–84.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.

Newell, F. N., & Bülthoff, H. H. (2002). Categorical perception of familiar objects. *Cognition*, *85*, 113–143.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford, UK: Clarendon Press.

Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, *131*, 477–493.

Pacer, M., Carstensen, A., & Regier, T. (2012). Grounding spatial language in non-linguistic cognition: Evidence for universal *and* relative spatial semantics in thought. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 827–832). Austin, TX: Cognitive Science Society.

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

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, *84*, 189–219.

Parish-Morris, J., Pruden, S. M., Ma, W., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). A world of relations: Relational words. In B. C. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 219–242). Oxford, UK: Oxford University Press.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*, 499–503.

Pinker, S. (1994). *The language instinct*. New York, NY: William Morrow and Company, Inc.

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

Poucet, B. (1993). Spatial cognitive maps in animals: New hypotheses on their structure and neural mechanisms. *Psychological Review*, *100*, 163–182.

Ratcliff, G. (1979). Spatial thought, mental rotation and the right cerebral hemisphere. *Neuropsychologia*, *17*, 49–54.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences, 13*, 439–446.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*, 1436–1441.

Regier, T., Khetarpal, N., & Majid, A. (2012, in press). Inferring semantic maps. *Linguistic Typology*.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1–20.

Roberson, D., Davies, I. R. L., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*, 369–398.

Roberson, D., & Hanley, J. R. (2010). Relatively speaking: An account of the relationship between language and thought in the color domain. In B. C. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 183–198). New York, NY: Oxford University Press.

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition, 107*, 752–762.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Romney, A. K., & d’Andrade, R. G. (1964). Cognitive aspects of English kin terms. *American Anthropologist*, *66*, 146–170.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

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.

Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research, 10,* 489–502.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, *38,* 495–553.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.

Russell, J. A. (1983). Pancultural aspects of the human conceptual organization of emotions. *Journal of Personality and Social Psychology, 45,* 1281–1288.

Russell, J. A., & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. *Journal of Personality and Social Psychology*, *48*, 1290–1298.

Russell, J. A., & Bullock, M. (1986). On the dimensions preschoolers use to interpret facial expressions of emotion. *Developmental Psychology*, *22,* 97–102.

Saalbach, H., & Imai, M. (2007). Scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General*, *136*, 485–501.

Saalbach, H., & Imai, M. (2012). The relation between linguistic categories and cognition: The case of numeral classifiers. *Language and Cognitive Processes*, *27*, 381–428.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, *42*, 319–345.

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodríguez, W., & Poulin-Dubois, D. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, *131*, 377–397.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*, 390–398.

Slobin, D. I. (1996a). Two ways to travel: Verbs of motion in English and Spanish. In M. S. Shibatani & S. A. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (pp. 195–220). Oxford, UK: Clarendon Press.

Slobin, D. I. (1996b). From “thought and language” to “thinking for speaking.” In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, UK: Cambridge University Press.

Smith, L. B., & Jones, S. S. (1993). Cognition without concepts. *Cognitive Development*, *8*, 181–188.

Steinley, D. (2006). *K*-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*, 1–34.

Takane, Y., Young, F. W., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, *42*, 7–67.

Talmy, L. (1983). How language structures space. In H. Pick & L. Acredolo (Eds.), *Spatial orientation: Theory, research, and application*. New York: Plenum Press.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*, 49–100.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*, 267–276.

Tversky, A. (1975). Features of similarity. *Psychological Review*, *84*, 327–352.

Vallar, G., Lobel, E., Galati, G., Berthoz, A., Pizzamiglio, L., & Le Bihan, D. (1999). A fronto-parietal system for computing the egocentric spatial frame of reference in humans. *Experimental Brain Research*, *124*, 281–286.

Vandeloise, C. (1991). *Spatial prepositions: A case study from French*. Chicago, IL: University of Chicago Press.

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology: General*, *134*, 501–520.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences*, *6*, 376–382.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236–244.

Watson, D., Clark, L. A., & Tellegen, A. (1984). Cross-cultural convergence in the structure of mood: A Japanese replication and a comparison with U.S. findings. *Journal of Personality and Social Psychology*, *47*, 127–144.

Waxman, S. R., & Gelman, S. A. (2010). Different kinds of concepts and different kinds of words: What words do for human cognition. In D. Mareschal, P. C. Quinn, &

S. E. G. Lea (Eds.), *The making of human concepts* (pp. 99–130). New York, NY: Oxford University Press.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). Cambridge, MA: MIT Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*, 7780–7785.

Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, *11*. doi:10.1167/11.12.16

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 253–256.

Wolff, P., Jeon, G.-H., & Li, Y. (2009). Causers in English, Korean, and Chinese and the individuation of events. *Language and Cognition*, *1*, 167–196.

Wolff, P., & Malt, B. C. (2010). The language-thought interface: An introduction. In B. C. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 3–15). Oxford, UK: Oxford University Press.

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

Xu, Y., & Kemp, C. (2010). Constructing spatial concepts from universal primitives. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 346–351). Austin, TX: Cognitive Science Society.

Zhou, K., Mo, L., Kay, P., Kwok, V. P. Y., Ip, T. N. M., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences*, *107*, 9974–9978.

Appendix

Sentences used in Experiment 1

Preposition	*Sentences*
about	She explored the rivers and streams **about** the estate. You'll find him somewhere **about** the office.
above	The apartment is **above** a restaurant. There is a cool spring **above** the timberline.
across	Anyone from the houses **across** the road could see him. The butcher's shop is **across** the street.
after	Z is **after** Y in the alphabet. They are in line one **after** another.
against	The ladder **against** the wall is for sale. The branches of the tree are **against** the window.
along	He lives in the house **along** the river. There's a post-box somewhere **along** this street.
alongside	Much of the industry was located **alongside** rivers. The road is **alongside** the pipeline.
amid	Over there is a tiny bungalow **amid** clusters of trees. There was a single dark bird **amid** a flock of white pigeons.
among	The forest contains a pine tree **among** cedars. There were ducks **among** the geese.
around	There were flowers **around** the tree. He admired the sash **around** her waist.
at	The dot is **at** the center of the page. The children are **at** the table.
atop	**Atop** a sheet of paper is an envelope. The house is **atop** a cliff overlooking the ocean.

before	They stopped **before** a large white villa. She was **before** me in the queue.
behind	The broom is **behind** the door. The recording machinery is **behind** the screens.
below	The boat is **below** the surface of the water. She hurt her leg **below** the knee.
beneath	We had a picnic **beneath** a large tree. **Beneath** this floor there's a cellar.
beside	The man **beside** her was wearing a brown suit and hat. Their house is **beside** a small lake.
between	The office has two desks with a table **between** them. There are fences **between** all the houses.
beyond	**Beyond** those hills there is a river. He pointed to a spot **beyond** the trees.
by	They have a house **by** the lake. The lamp was **by** the door.
down	Their house is halfway **down** the hill. There is mustard **down** the front of your shirt.
far from	I live **far from** Chicago. The cat strayed **far from** home.
in	My mother is **in** bed. The letter is **in** the wastebasket.
in back of	My dad demolished an old shed **in back of** his barn. The garage is **in back of** their yard.
in front of	The lawn is **in front of** the house. She is **in front of** her mirror.
inside	Her hands were **inside** her pockets. A radio was playing **inside** the apartment.
near	There are several beaches **near** here. The parking lot was **near** the sawmill.

off	The artery **off** the heart was blocked. The house is a mile **off** the coast.
on	The vase is **on** the table. There is a lot of frosting **on** the cake.
on top of	The books are **on top of** the desk. The trays are one **on top of** another.
opposite	He lives in the house **opposite** mine. The school is **opposite** a park.
outside	We waited **outside** the house. There was a boy **outside** the door.
over	He looked at himself in the mirror **over** the fireplace. **Over** the hill is a small village.
past	The house is a mile **past** the first stoplight. She turned left just **past** the stairs.
to the left of	He is **to the left of** her. The fireplace is **to the left of** the desk.
to the right of	**To the right of** the gas station is a library. She is **to the right of** the bed.
to the side of	The light is **to the side of** the car's rear window. There is a small shed **to the side of** the house.
toward	His back was **toward** me. The painting is down the corridor **toward** the foyer.
under	Your pencil is **under** the chair. The hot plate is **under** an insulated lid.
underneath	He lives in the apartment **underneath** mine. The table was **underneath** the olive trees.
up	Their house is **up** the road. Her office is **up** the hall on the right.
within	They live **within** the city limits. Everything I need is **within** a few miles of my apartment.