Title
Universals and variation in language and thought: Concepts, communication, and semantic structure

Author
Carstensen, Alexandra

Publication Date
2016

Peer reviewed|Thesis/dissertation

Universals and variation in language and thought:
Concepts, communication, and semantic structure

by

Alexandra B Carstensen

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

PSYCHOLOGY

in the

Graduate Division
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Terry Regier, Chair
Thomas Griffiths
Mahesh Srinivasan
Eve Sweetser

Fall 2016

Universals and variation in language and thought:
Concepts, communication, and semantic structure

By

Alexandra B Carstensen

Abstract

Universals and variation in language and thought: Concepts, communication, and semantic structure

by

Alexandra B Carstensen

Doctor of Philosophy in Psychology

University of California, Berkeley

Terry Regier, Chair

Why do languages parcel human experience into categories in the ways they do, and to what extent do these categories in language shape our view of the world? Both language and nonlinguistic cognition vary across cultures, but not arbitrarily, suggesting that there may be universal constraints on how we talk and think. This dissertation explores the sources and consequences of universals and variation in language and thought in four parts.

The first study examines a major premise of the universalist view of cognition, that speakers of all languages share a universal conceptual space, which is partitioned by the categories in language. Previous research on color cognition supports this view; when English speakers successively pile-sort colors, their sorting recapitulates an independently proposed hierarchy of color semantics across languages (Boster, 1986). Here I extend that finding to the domain of spatial relations. Levinson et al. (2003) have proposed a hierarchy of spatial category differentiation, and I show that English speakers successively pile-sort spatial scenes in a manner that recapitulates that semantic hierarchy. This finding provides evidence for a specific hierarchy of spatial notions as a model of universals in conceptual structure, and suggests that universal patterns observed across languages reflect general cognitive forces that are available in the minds of speakers of a single language.

The second project of this dissertation demonstrates a process by which domain-specific conceptual universals and more general communicative pressures may shape categories in language, extending a previous account (Regier et al., 2015) of semantic universals and variation. In particular, I show that human simulation of cultural transmission in the lab produces systems of semantic categories that converge toward greater informativeness, in the domains of color and spatial relations. These findings suggest that larger-scale cultural transmission over historical time could have produced the diverse yet informative category systems found in the world's languages. This work supports the communicative efficiency account of semantic universals and variation and establishes a process through which categories in language become increasingly efficient and increasingly universal.

The third study extends the previous account of categories in language to cognition more broadly, showing that the same principles that govern efficient semantic systems also characterize nonlinguistic cognition. I provide an account of spatial cognition in which conceptual categories optimize the trade-off between informativeness (making for fine-grained and intuitively organized spatial categories) and simplicity (limiting the number of categories). I find that pile sorts made by speakers of diverse languages match this universal account more closely than they match the semantics of the sorter's native language. These results suggest that across languages, spatial cognition reflects universal pressures for efficient categorization, and observed universals in category structure and granularity result from these pressures.

The final project of this dissertation probes the role of language in online spatial reasoning, using linguistic interference to prevent participants from relying on language in solving a spatial task. In previous work, adult English speakers have been shown to use a spatial frame of reference that differs from that of nonhuman primates and toddlers (Haun et al., 2006), suggesting that learning the spatial frame of reference used in English may motivate a switch away from universal modes of spatial thought. I find that under linguistic interference, despite a sharp increase in error, adult English speakers fail to readopt the spatial frame of reference used by nonhuman primates and toddlers. This finding rules out the possibility that language affects spatial frames of reference online and accordingly argues against Kay and Kempton's (1984) account, which predicts a removable online role of language. This result raises the stakes of the debate over the role of language in nonlinguistic spatial frames of reference—either something other than language causes alignment between linguistic and nonlinguistic frames of reference, or language learning fundamentally restructures nonlinguistic spatial cognition in a way that is difficult to reverse.

The findings of this dissertation in the domain of space, taken together with parallels in other cognitive domains, reinforce an emerging consensus on the relation of language and thought, by which all people share a universal conceptual foundation that may be altered by language. The research here further elaborates this account, suggesting that universals and variation in both language and thought may derive to some extent from general principles of efficiency. At the same time, it challenges the generality of a classic formulation of this view (Kay & Kempton, 1984), motivating future research. In both complementing and challenging an emerging consensus on language and thought, this dissertation informs our view of language, a defining feature of human cognition, and contributes to a more complete understanding of the nature of thought.

Contents

Acknowledgments

I am deeply indebted to many fantastic scientists who have acted as inspiration, accomplice, and friend to me. Here I acknowledge a small number of the numerous people who have enriched my time at Berkeley.

My advisor, Terry Regier, has been instrumental to this work, and I thank him for his enthusiasm and constant support throughout my undergraduate and graduate education. Terry taught the first class I took on language and thought, and has had a hand in just about everything I've learned since. My research and my perspective on science have benefitted greatly from conversations with Terry—he is insightful, prudent, sometimes surprisingly opinionated, and always fun to talk with. He has been an inspiration in more ways than I can begin to elaborate here and for which I am forever grateful. Many thanks are also due to the other members of my committee: Tom Griffiths, whose inhuman memory and dry humor made for the most productive and entertaining meetings I could ask for; Mahesh Srinivasan, who helped me develop long-standing research interests in language learning into research projects; and Eve Sweetser, whose targeted questions and discerning criticism did a great deal to strengthen my work and prepare me for all past and future snake fights. I would also like to thank Rich Ivry, for advising me in my last years of undergrad and first years of grad school, setting a stellar example of work-life balance that still involves 3am emails, and never managing to conceal his unbridled enthusiasm for new ideas and knowledge.

I am also extremely lucky to have encountered George Lakoff, whose big ideas and compelling prose first inspired my interest in the nature of thought, the structure of language, and how it all gets that way. I owe many thanks to George for getting me started, and to Bryan Alvarez, for teaching me how much fun research can be. I am grateful to many wonderful colleagues and friends in the Language and Cognition Lab and CognAc, particularly Ellie Kon, Elisabeth Wehling, Kevin Holmes, Yang Xu, Peter Butcher, Dav Clark, Ryan Morehead, and Matt Crossley. I learned heaps, and had an extraordinary amount of fun variously working, living, and lifting heavy objects with the lot of them. I owe an additional debt of gratitude to the numerous dedicated undergraduate research assistants who helped me collect linguistic and experimental data across several languages and continents, especially Aaliyah Ichino, Katie Chen, Maggie Soun, Ana Cuevas, and Vanessa Matalon, who contributed data to the analysis in Chapter 3. That chapter is based on a published paper, and I thank Jing Xu and Cameron Smith for major contributions to that work, as well as Shubha Guha, Grace Neveu, Lev Michael, Asifa Majid, and Naveen Khetarpal for collecting and sharing data used in the paper. Additional thanks are due to Grace Neveu, Lev Michael, Asifa Majid, Naveen Khetarpal, and Sam Mchombo for contributing to the data and analyses in Chapter 4. The study in Chapter 5 was originally pioneered by Sophie Bridgers and would not exist without her major contributions. I thank Sophie for her work and thoughtful discussion, and also Mike Frank for generously sharing his experimental code. The work in this dissertation was supported by grants from the National Science Foundation through their Graduate Research Fellowship Program and the Spatial Intelligence and Learning Center, for which I am very grateful.

I owe countless thanks to Shaibya Dalal, Eduardo Joya, Bill Chen, Alana Carrara, and Rion Graham for years of love, friendship, and support, and their limitless patience with me the many times I rescheduled our plans to stay late at school. I am also hugely grateful to Caren Walker and Matt Vannucci for working alongside me on many of those late nights and weekends, and for making it fun. I thank Josh Abbott, who has in many ways defined my experience of graduate school and science, for sharing with me his profound sense of awe at the little wonders of the world, and his terrific Chihuahua, Meesh. Finally, I thank my mom and dad, for their unconditional love and support. They always said that when I grew up I would spend years and years working on one project, and of course, they were right.

Chapter 1

Introduction

Do speakers of different languages think differently or do we all share a single underlying conceptualization of the world around us? On one hand, if universals of thought exist independent of language, then these universals may reflect basic constraints on cognition. If this is the case, then what are these conceptual universals, where might they come from, and what principles do they follow? On the other hand, if thought varies across languages, this may indicate that the way we talk about the world alters the way we think about it. If so, then under what circumstances, and through what mechanisms does language change human cognition?

Here I review foundational research on the relation between language and thought and discuss several remaining gaps in our understanding of the questions above. I then describe my approach in addressing these gaps, which engages with an emerging consensus by evaluating several assumptions about universals in cognition and language and testing a classic view of language and thought. In particular, I present four projects, which test and characterize (1) specific universals in cognition, (2) general principles in language change, (3) general principles in thought, and (4) mechanisms by which language may influence thought. These projects, corresponding to the chapters of my dissertation, contribute evidence for universals in conceptual structure, an account of the forces that shape both language and cognition, and a challenge to the view that language actively influences cognition during thought.

1.1 Perspectives on language and thought

What is the nature of the relation between language and thought? Languages parcel human experience into categories that are picked out by words (e.g. green, up, snow) and features of grammar (plural, feminine, future). These categories vary considerably across languages, suggesting two opposing accounts of language and thought.

The *universalist* view holds that people perceive and construct the experienced world similarly, regardless of language. By this account, language acts to partition a universally shared conceptual space into varied but non-arbitrary semantic categories (Berlin & Kay, 1969; Jameson & D'Andrade, 1997; Levinson et al., 2003; Regier et al., 2007). Consistent with this view, there are recurring cross-linguistic tendencies that suggest a universal conceptual repertoire (Levinson et al., 2003; Khetarpal et al., 2009).

The other account, known as the Sapir-Whorf hypothesis, proposes that linguistic systems create categories which shape perception, causing speakers of different languages to conceptualize the world in fundamentally different ways (Whorf, 1956). This *relativist* view is

supported, in part, by evidence that variations in language align with systematic differences in thought: speakers of different languages solve mazes in opposite directions (Majid et al., 2004), pre-linguistic infants notice changes that adults overlook (Hespos & Spelke, 2004),[1] and routine distinctions made by speakers of some languages are ignored or impossible for others (Boroditsky & Gaby, 2010). The evidence that certain aspects of language correspond to divergence in basic cognition supports the idea that language plays an influential role in thought.

One proposal posits a reconciliation of these seemingly dissonant views, suggesting that there is a universal conceptual basis for the categorical distinctions that languages make, but that this conceptual foundation may be altered by language (e.g., Kay & Kempton, 1984; Hespos & Spelke, 2004; Regier & Kay, 2009). There is convergent support for this view across several domains (e.g., Hermer-Vazquez et al., 1999; Gilbert et al., 2006; Frank et al., 2008; Gilbert et al., 2008), but many important questions remain unanswered. If speakers of all languages share a common foundation of thought, what cognitive universals constitute this foundation? Are these cognitive universals apparent in the ways that languages parcel human experience into categories? More broadly, what pressures shape the universals and variation in category meanings, and through what processes do such pressures drive language change? Do these pressures parallel principles of nonlinguistic cognition? In what circumstances do the categories in language influence cognition directly, to what degree, and by what mechanisms?

This dissertation seeks to establish the nature of the emerging reconciliation between universalist and relativist views. The research in the following chapters informs all of the questions above, drawing on the tools and perspectives of three distinct but related traditions in the study of language and thought: semantic typology, language evolution, and linguistic relativity. In particular, I focus on spatial cognition as a testbed for this integrative approach. Spatial cognition provides an ideal domain for relating questions of language and thought across multiple perspectives because it is a foundational area of human cognition and a prominent topic in language. A substantial body of previous research in psychology has examined spatial reasoning, memory, and navigation as crucial components of cognition shared across cultures and species (e.g., Tolman & Honzik, 1930; Shepard & Metzler, 1971; Shepard & Cooper, 1982; Pick & Acredolo, 1983; Hermer & Spelke, 1994), yet complementary work in linguistics finds that human languages exhibit a large range of diverse spatial semantics (e.g., Bowerman & Pederson, 1992; Brown & Levinson, 1993; Levinson et al., 2003). Building on this previous work, I investigate the structure and relation of language and thought in the domain of space. Do universals in spatial cognition underlie variation across languages? What principles characterize universals and variation in spatial language and thought? What role does language play in spatial cognition?

Perspectives on spatial language and cognition

Many questions on the relation between spatial language and thought are grounded in the observation that spatial meanings vary across languages. Hence, the basis for this line of inquiry is found in semantic typology, a subfield of linguistics which seeks to survey and catalog the

[1] While this study is often cited in support of universalist views, it would be difficult to interpret its findings as supporting an exclusively universalist view of thought, as the study suggests a striking difference in cognition between infants and adults (and implies a parallel difference between adult speakers of English and Korean). The authors interpret their findings as evidence that conceptual representations are universal and established before language learning, but nonetheless attribute the observed difference in cognition to differing linguistic experience,

variety of meanings picked out by categories in language, and the larger systems of meaning that these categories form. Semantic typology presents a unique perspective on structured meaning in language, and a natural starting point for studies of language and thought across many domains of cognition. To the extent that language reflects cognition, the models of semantic structure developed by typologists afford a rich set of hypotheses about analogous conceptual structures (e.g., Croft, 2003; Haspelmath et al., 2005; of particular relevance for this dissertation: Kay & McDaniel, 1978; Talmy, 1983; Levinson et al., 2003). The first project of this dissertation, in Chapter 2, explores this application of semantic typology, translating a model of spatial semantics into a hypothesis on the structure of spatial cognition to evaluate whether universals in spatial cognition underlie the variation in language.

Building broadly on traditions in typology, a recent movement in cognitive science seeks to explain linguistic structure through accounts that emphasize the function of language as a communicative system (e.g., Piantadosi et al., 2011; 2012; Fedzechkina et al., 2012; Gibson et al., 2013). Applied within semantic typology, this approach characterizes patterns in the semantic structure of language as driven by general principles of efficiency from an information theoretic framework (Regier et al., 2007; Baddeley & Attewell, 2009; Kemp & Regier, 2012). One such study demonstrates that these principles of communication can be used to explain universals and variation in spatial category systems across languages (Khetarpal et al., 2013), simultaneously motivating universal tendencies and providing constraints on variation.

While this approach, and those grounded in semantic typology more generally, contribute many observations about the patterns of meaning in language—and corresponding hypotheses about cognition—they leave open questions about how these patterns arise. To evaluate the processes that shape systems of meaning, an account of the relation between language and thought must also be informed by studies of language evolution. Research in this area considers the mechanisms and processes of language change, including direct constraints on language produced by the dynamics of language transmission (e.g., informational bottlenecks as limitations on holistic systems of meaning; Kirby, 2002) and indirect influences imposed by cognitive biases in learning and memory (e.g., Kalish et al., 2007). It provides computational and empirical tools for assessing the selective pressures that shape semantic structure (e.g., Kirby et al., 2008) and identifying implicit cognitive biases (Kalish et al., 2007) through the experimental paradigm of iterated learning. Consequently, simulations of language learning and related techniques can be used to uncover the general principles that contribute to the structure of language and cognition[2]. This approach informs universal theories by providing empirical grounding for specific universal features and general principles. In this vein, Chapter 3 employs a simulated language learning paradigm to demonstrate pressures in language transmission that give rise to efficient systems of spatial categories. Additionally, this chapter shows that these pressures are consistent with a previous information-theoretic account (Khetarpal et al., 2013) of universals and variation in spatial semantics. The success of this account in characterizing both the typological patterns and evolution of spatial semantics lays the groundwork for an analogous formulation of cognitive principles.

Chapter 4 evaluates whether this information-theoretic account of structure in linguistic meanings further characterizes thought, providing a bridge between accounts of language and

[2] Studies of language evolution include analysis of historical corpora, phylogenetic approaches to language, and many other techniques that are not addressed here. This discussion highlights laboratory simulations of language change and transmission as an approach that is central to recent work on the evolution of semantic systems and of particular relevance to this dissertation.

cognition. For this purpose, it is necessary to assess nonlinguistic cognition more directly; work within semantic typology and language evolution may reveal structure imposed by language, but such structure need not correspond to the architecture of nonlinguistic cognition. The study in Chapter 4 uses a nonlinguistic task to compare behavior across speakers of semantically diverse languages, and finds support for the idea that general principles of efficiency account for the structure of thought in addition to that of language.

Any evaluation of universalist and relativist views of language and thought hinges on the nature of thought, which may be assessed broadly in comparative studies across languages, as in Chapter 4, but also in focused experiments on linguistic relativity. Convergent data across a range of cognitive experiments is necessary to establish the roles of language-specific and universal forces in shaping the concepts and processes we use to represent the world around us. Cognitive and linguistic tasks enable assessments of both the baseline nature of nonlinguistic cognition and cases in which language plays an active, causal role in defining or influencing thought (e.g., Hermer-Vazquez et al., 1999; Frank et al., 2008; Gilbert et al., 2008). Following work in this area, Chapter 5 tests a mechanism through which variation in spatial language may influence spatial cognition, and finds evidence that challenges a classic account of language and thought in the spatial domain.

Answering the broad question of how language and thought relate is the primary project of an entire field within cognitive science, and this central question is composed of a constellation of related inquiries that are often posed and answered by largely separate approaches. While these questions are often treated independently by research programs in semantic typology, language evolution, and linguistic relativity, each of these projects informs and constrains the others. The research in this dissertation synthesizes the major ideas and methodologies of these programs to inform a more cohesive account of the relation between language and thought in the domain of spatial cognition.

In the following section, I will review the tradition of examining questions in language and thought on a domain-by-domain basis, and the motivation for doing so in two foundational domains: color and space. This dissertation follows that tradition, building on a broad foundation of basic research in spatial cognition to test the generality of accounts and principles established in the color domain. After reviewing some of the foundational work and the general motivations for studying cognition in the domain of space, I will pose several open questions and outline, in brief, my research addressing them. These studies, corresponding to the chapters of my dissertation, integrate ideas and approaches across the programs of semantic typology, language evolution, and linguistic relativity. These chapters will, in turn, address universals in the structure of spatial meanings in language; evaluate accounts of these typological universals as models of spatial cognition; examine the general principles that guide semantic structure, language change, and conceptual structure; and assess the role of language in determining how we conceive of the world around us.

1.2 Foundational topics in language and thought: Color and space

"Thinking is most mysterious, and by far the greatest light upon it that we have is thrown by the study of language."

Benjamin Lee Whorf, 1956

There are a wide variety of opposing perspectives on language and thought, and disagreements are common within philosophical traditions and even within individuals across time. For instance, Bertrand Russell believed that "language serves…to make possible thoughts which could not exist without it" (Russell, 1948) whereas Wittgenstein, his student, initially argued that "language disguises thought," imposing limits on expressible meaning and on philosophy more broadly (Wittgenstein, 1922), but later took a staunchly pragmatic view of "ordinary language," which emphasized its informativeness and utility in context (Wittgenstein, 1953). These views in turn provide a stark contrast to earlier notions of platonic realism in which the categories of language refer to abstract kinds, properties, and relations that are innate and universal in thought. The diversity (and mutability) of classic views on the nature and relation of language and thought underscore the need for specific, falsifiable, and empirically-driven accounts.

Modern approaches in cognitive science answer this challenge by translating abstract philosophical views into theories with concrete predictions on a domain-by-domain basis. The domains of color and space represent two major testbeds in this tradition. Here I will review previous work studying the relation of language and thought within these domains. I include additional reviews of relevant prior work from these domains in each dissertation chapter.

Color cognition and language

Much of the early experimental work on language and thought addressed questions within the domain of color, which serves as an ideal test case for several reasons. From a typological perspective, color is appealing because color words occur across a wide range of languages and the categories picked out by these words vary in interesting ways (e.g., Heider & Olivier, 1972; Davidoff et al., 1999). Color stimuli are also relatively easy to create and use in linguistic elicitation tasks in the field, enabling large-scale surveys of color vocabularies, such as the World Color Survey, which documented color naming in 110 unwritten languages (Cook et al., 2005). The availability of this typological data on color has paved the way for cross-linguistic comparison and evolutionary accounts of universals in color semantics (e.g. Kay & McDaniel, 1978).

These formulations of structure in language have in turn proved useful as models of cognitive universals (Boster, 1986). As with color language, color perception and cognition are broadly accessible to psychophysical and experimental tools within vision science and psychology, providing for easy comparison between color language and thought (e.g., color memory in Heider & Olivier, 1972). Color cognition is also noteworthy because it provides a particularly strong test of linguistic relativity by assessing thought in a domain with a firm perceptual basis, which is—at least initially in development—fully independent of language, as the early development of color perception proceeds without input from language. Furthermore, the grounding of color in measurable, low-level perception provides for relatively objective frameworks (e.g. standard models of perceptual color space, such as CIELAB; Wyszecki & Stiles, 1967) that can be used to identify distortions of perceptual space in cognition. Kay and

Kempton (1984) exploited this feature to show that the perceived similarities between colors are biased by color categories in language, and that this bias disappears when linguistic strategies are interrupted (see also Roberson & Davidoff, 2000; Gilbert et al., 2006; Winawer et al., 2007). This finding suggests that speakers of all languages share a universal conceptual space which is independent of language, but that language, when engaged, can affect perception in this space. Kay and Kempton proposed specifically that cognition has two tiers: one that corresponds to universals of perception, and a second, removable tier in which the categories of language overlay this universal space. This experiment served to establish a foundational paradigm for studies of linguistic relativity and advance the two-tiered view of cognition as a unified version of the universalist and relativist accounts.

These, and many other studies in the domain of color, have produced accounts of universals (e.g., Berlin & Kay, 1969; Heider, 1972; Heider & Olivier, 1972; Kay & McDaniel, 1978; Kay & Regier, 2003; Lindsay & Brown, 2006; Lindsay & Brown, 2009), variation (Roberson et al., 2000; Roberson et al., 2005), and structure in language and cognition (Brown & Lenneberg, 1954; Regier et al., 2007; Regier et al., 2009; Taylor et al., 2013; Holmes & Regier, 2016; Cibelli et al., 2016; Abbott et al., 2016), and characterized a small but significant role of language in shaping color cognition (e.g., Kay & Kempton, 1984; Roberson & Davidoff, 2000; Gilbert et al., 2006; Winawer et al., 2007; Roberson et al., 2008). These findings in color serve as the basis for a budding consensus on language and thought in this domain, by which universal conceptual foundations underlie and constrain the categories in language, but language may also alter cognition. On this consensus view, the influence of language on cognition is impermanent, and can be interrupted to reveal the universal bedrock of cognition.

Generalizing accounts of language and thought: The case of space

Support for the universals-and-language account comes from several other domains, including number (Frank et al., 2008), biological kinds (Gilbert et al., 2008), and spatial cognition (Hermer-Vazquez et al., 1999), but the evidence in these domains is limited and conflicting, presenting less of a consensus (cf. Butterworth et al., 2008; Holmes & Wolff, 2012; Ratliff & Newcombe, 2008, respectively). This may be due, in part, to the challenges associated with obtaining and interpreting evidence in domains that are less amenable to linguistic elicitation and cognitive experimentation, and also less directly grounded in perception.

However, the obstacles that make these domains more challenging to assess also afford interesting complexity and stronger tests of generality for accounts founded on color. Spatial cognition in particular provides a compelling contrast to color for several reasons, discussed below: the spatial domain has cross-cultural significance, an abstract nature, and diverse subdomains.

With respect to the cross-cultural significance of space, spatial vocabulary, like color, is ubiquitous across languages—but unlike color, which can be seen as trivial (see Kuschel & Monberg, 1974: “We don’t talk much about colour here”), the domain of space has clear relevance to critical human needs and activities like foraging, navigation, and construction.

Spatial cognition is also more abstract than color, in that space has grounding in visual, auditory, and haptic perception, and this multimodal quality obscures perceptual accounts and leaves conceptual representations of space underdetermined. This property situates space slightly further along the abstraction gradient, providing an ideal next step in abstracting accounts from color. Studies focused on the abstract properties of spatial reasoning have been foundational to

cognitive science as a field, and established major paradigms and principles of psychology. In particular, research on navigation in rats (Tolman & Honzik, 1930) inspired the paradigm shift from behaviorism to cognitive psychology and spurred the cognitive revolution by demonstrating the existence of cognitive maps and consequently mental representations. Mental rotation studies provided initial evidence for mental simulation (Shepard & Metzler, 1971), and led to the proposal of mental imagery (Shepard & Cooper, 1982), now a conventional view of representation. Related lines of inquiry continue to stimulate influential research on spatial navigation (Wolbers & Hegarty, 2010) and on mental simulation, as a mechanism of language comprehension (Zwaan, 2003), and as a source of inferences about the physical world (Battaglia et al., 2013), further demonstrating the value of assessing cognition in the relatively abstract domain of space. The central role of space in classic debates across psychology and cognitive science is a testament to the versatility of the domain in providing the means to operationalize and test many diverse questions.

Beyond its abstract nature, the spatial domain offers a rich and complex testbed in the form of several distinct subdomains, and surveys of semantic structure and nonlinguistic cognition in these areas lay the groundwork for asking targeted questions about the relation of language and thought. These subdomains of spatial cognition include spatial orientation (e.g., Pick & Acredolo, 1983; Hermer & Spelke, 1994), topological spatial relations (e.g., Talmy, 1983; Bowerman & Pederson, 1992), and spatial frames of reference (e.g., Brown & Levinson, 1993), each of which has served as the basis for illuminating lines of research and debate on language and thought. Spatial orientation, for instance, affords a primary topic of debate between nativist and empiricist accounts of the sources, development, and representation of spatial knowledge, with each view seeking to explain how children and adults use information about their surroundings to encode and locate positions in space (e.g., Hermer & Spelke, 1994; Newcombe & Ratliff, 2007). In another vein, the subdomain of spatial topology, which concerns relative locations often picked out by prepositions like "in" and "on," has contributed to surveys of semantic structure and the finding that spatial topological systems vary considerably across languages (e.g. Levinson et al., 2003), which in turn enables tests of linguistic relativity (Hespos & Spelke, 2004; Khetarpal et al., 2010). Similarly, the area of spatial frames of reference, which examines how people establish and use coordinate systems with contrasting notions like "north" or "left", has supported typological research (e.g., Levinson, 1996), a spirited debate on linguistic relativity (e.g., Levinson et al., 2002; Li & Gleitman, 2002), and cross-species comparisons of cognition (Haun et al., 2006).

Across these topics of research, and others, the domain of spatial cognition provides a rich testbed for a diversity of questions, many of which inform central questions on language, cognition, and the relation between them. In space, as in color, the evidence for universalist and relativist accounts of language and cognition is mixed: recurring tendencies in spatial cognition across semantically diverse languages support the universalist view (e.g., Levinson et al., 2003; Khetarpal et al., 2009; Khetarpal et al., 2013), but at the same time, linguistic diversity often heralds divergence in spatial thought (e.g., Hermer & Spelke, 1996; Levinson, 1996; Levinson et al., 2002; Majid et al., 2004). This dissertation seeks to further establish the nature of the relation between universal and language-specific forces in the domain of spatial cognition. Building on work in the color domain, the following studies test the generality of several key assumptions of the universalist and relativist views in the domain of space. In the following section, I outline the goals and corresponding projects of this dissertation in the spatial domain, which test and

characterize (1) specific universals in cognition, (2) general principles in language change, (3) general principles in thought, and (4) mechanisms by which language may influence thought.

1.3 Goals of the dissertation

This dissertation characterizes the relation of language and thought in the spatial domain. The following chapters engage cross-cutting questions from semantic typology, language evolution, and linguistic relativity to explain universals and variation in language, cognition, and interactions of the two. In particular, these chapters will address four open questions on the nature of semantic and conceptual universals, and the role of language in spatial cognition:

Chapter 2 Universals in conceptual structure: What influences underlie universal tendencies in language—is there a universal conceptual space for spatial relations that speakers respect across languages?

Chapter 3 Universal forces in language change: What processes shape semantic systems in language? Do changes in the semantics of spatial categories follow principles of informative communication?

Chapter 4 Universal principles of cognition: Do cognitive universals follow general principles? In particular, does spatial cognition follow the same principles that govern semantic systems across languages?

Chapter 5 Language in cognition: What causes cognition to vary in line with language? Does language-like spatial reasoning depend on access to language?

Below, I briefly outline the research projects corresponding to each question and preview my findings.

Chapter 2: Do universals of language reflect a universal conceptual space?

The first line of work explores the relation between universals in language and thought. The major premise of the universalist view is that speakers of all languages share a universal conceptual space, which is partitioned by the categories in language. Previous findings support this account in the domain of color. Specifically, systems of color categories across languages appear to follow a universal semantic hierarchy (Kay & McDaniel, 1978), and these universals in color language correspond closely to universals in color cognition (Boster, 1986).

I evaluate the generality of this finding in the case of spatial relations. Levinson et al. (2003) assessed topological spatial relation names across languages and proposed a hierarchy of spatial notions underlying the diversity of semantic systems. I test this semantic hierarchy as a model of cognition by asking English speakers to successively pile sort spatial scenes into increasingly finer categories. The universalist view holds that commonalities in language result from languages partitioning a shared conceptual space. By this account, the semantic universals in Levinson et al.'s hierarchy are caused by cognitive universals, and so will closely match the pile sorts created by speakers of any language. Consistent with this prediction, the pile sorts made by English speakers recapitulate the proposed model of universal semantics. This finding provides support for the proposal that all people share a largely universal conceptual space for topological spatial relations. Additionally, it presents evidence for a specific hierarchy of spatial

notions in cognition and in doing so extends Boster's (1986) demonstration in color to the novel domain of spatial relations.

Chapter 3: How do evolutionary pressures shape the universals in language?

The previous chapter provides an explanation of linguistic universals as reflections of a particular, domain-specific hierarchy of spatial notions. Other work has shown that principles of *communicative efficiency* can explain linguistic universals in a *domain-general* way by appealing to general qualities like similarity and simplicity rather than the specific semantic features of a domain (e.g., Kemp & Regier, 2012; Regier et al., 2015). In particular, Khetarpal et al. (2009, 2013) have shown that spatial systems in language are near-optimally informative by this account. However, in a commentary on this line of work, Stephen Levinson (2012) pointed out that although this research explains cross-language semantic variation in communicative terms, it does not tell us "where our categories come from" (p. 989); that is, it does not establish what *process* gives rise to the diverse attested systems of informative categories. Chapter 3 addresses that challenge.

I show that human simulation of cultural transmission in the lab produces systems of semantic categories that converge toward greater informativeness, in the domains of color and spatial relations. Moreover, as these simulated languages become more informative, they increasingly resemble the semantic structure of natural human languages. These findings suggest that (a) language users are biased toward efficient linguistic systems, and (b) larger-scale cultural transmission over historical time could have produced the diverse yet informative category systems found in the world's languages. More broadly, this result may indicate that nonlinguistic cognition (revealed through learning biases in this task) respects the same general principles of efficiency observed in semantics across languages.

Chapter 4: Do principles of efficiency explain universals in cognition?

Here, I ask whether the principles that govern efficient semantic systems, as explored in the previous chapter, also characterize nonlinguistic cognition. Inspired by an earlier treatment of color memory (Lantz and Stefflre, 1964), this question adopts a view of thought as self-directed communication. By this account, memory can be seen as a process of transmitting information to oneself over time with the brain as a channel, analogous to transmitting information to others through language. This perspective predicts that categories in cognition would follow the same general principles as those in language.

To evaluate this view of thought, I follow an earlier study by Khetarpal et al. (2010), which assessed language and cognition in speakers of English and Dutch via pile sorting of spatial stimuli, and identified small language-specific tendencies and robust universals in pile sorting. First, to provide a stronger test of these universalist findings, I reproduce the previous study in two new languages, Chichewa and Máíhɨki, which lend greater linguistic and cultural diversity to the study population. I find that speakers of all four languages, despite large differences in the granularity of their linguistic spatial systems, make pile sorts with similarly fine granularity. Next, I test whether this universal tendency reflects a drive for efficiency in cognition, analogous to that found in language. I present an account of spatial cognition in which conceptual categories maximize the trade-off between informativeness (making for fine-grained and intuitively organized spatial categories) and simplicity (limiting the number of categories).

Consistent with this view, I find that pile sorts made by speakers of Chichewa and Máíhɨki match this universal account more closely than they match the semantics of their native languages.

Chapter 5: Does language play an active role in thought?

The first three chapters characterize universal tendencies in semantic systems, simulated language evolution, and nonlinguistic cognition. They suggest a universalist account in each case, by which cognition is shared across languages, and this shared representation projects commonalities onto the structure of semantic systems, simulated languages, and pile sorts. In all of these cases, the universal tendencies in category structure reflect unseen universals in cognition. Accounts of linguistic relativity emphasize the converse scenario, in which variation in language causes variation in cognition. The ideal candidate case study for such an account would be a domain with (a) clear variation in linguistic categories, (b) clear variation in cognition, and (c) a strong correspondence between variation in (a) and (b). The subdomain of topological relations considered in other chapters clearly varies across language and cognition, but cognition (at least in pile sorting tasks) exhibits strong universal tendencies and does not show great correspondence with varying features in language. Spatial frames of reference, however, do meet the criteria for an ideal candidate domain. The preferred spatial frame of reference (FoR) in both language and nonlinguistic tasks varies across cultures (e.g., Brown & Levinson, 1993; Pederson, 1995; Levinson, 1996; Pederson et al., 1998; Levinson, 2003), and the preferred linguistic FoR is the best predictor of nonlinguistic FoR in a broad set of demographic factors (Majid et al., 2004), presenting an ideal case for tests of linguistic relativity.

Studying spatial FoR affords an opportunity to test Kay and Kempton's (1984) two-tiered account of cognition, by which language provides a complementary but removable overlay on a universal foundation. Previous work established that nonhuman primates and toddlers have a preferred FoR that they systematically default to, suggesting a universal primate bias in FoR (Haun et al., 2006). However, adult English speakers in nonlinguistic tasks use an FoR that differs from that of young children and other primates, but aligns with the English language (Li & Gleitman, 2002). Under the two-tiered account, English-speaking adults retain the universal primate bias beneath a divergent overlay of language, but will recover this universal when language is disrupted.

Chapter 5 addresses this high-profile debate over the role of language in spatial FoR, testing Kay and Kempton's (1984) two-tiered account of cognition through verbal interference. I find no evidence that interfering with language produces a shift toward the predicted pre-linguistic mode of thought. I conclude that although language often shapes cognition through online use, this does not appear to be true in the influential case of spatial frames of reference. This finding raises the stakes of the debate around spatial frames of reference. Either language has no causal effect on cognitive FoR and previous research has widely misattributed alignment between language and thought to a causal role of language, or language learning fundamentally restructures spatial cognition in a way that is difficult to reverse.

Collectively, these studies are designed to compare structures of meaning in words and concepts in order to identify universals, variation, and specific constraints on language and cognition. The findings presented here reinforce and elaborate an emerging consensus on the relation of language and thought, by which all people share a universal conceptual foundation that may be altered by language.

Chapter 2

Universals in conceptual structure underlying language

The major premise of the universalist view of cognition is that speakers of all languages share a universal conceptual space, which is partitioned by the categories in language. Previous research on color cognition supports this view; when English speakers successively pile-sort colors, their sorting recapitulates an independently proposed hierarchy of color semantics across languages (Boster, 1986). Here we extend that finding to the semantic domain of spatial relations. Levinson et al. (2003) have proposed a hierarchy of spatial category differentiation, and we show that English speakers successively pile-sort spatial scenes in a manner that recapitulates that hierarchy. This finding provides evidence for a specific hierarchy of spatial notions as a model of universals in conceptual structure, and suggests that universal patterns observed across languages reflect general cognitive forces that are available in the minds of speakers of a single language.

2.1 Language as a mirror of the mind

A core question in cognitive science is whether the structure of language reflects the structure of the human mind. Languages vary widely, both in their formal structure and in their semantic categorization of the experienced world (Evans & Levinson, 2009). At the same time, similar structures and categories appear in unrelated languages, and many logically possible linguistic structures and categories are not attested. A natural question is whether this constrained variation in language reflects universal tendencies of human cognition.

One means of pursuing this question concerns language change. One may observe or infer general patterns in the ways languages evolve over historical time, and ask whether these patterns of change, based on observation across languages and across time, are also evident at a given moment in the minds of individuals who speak a single language.

Such a demonstration has already been made in the semantic domain of color (Boster, 1986), and here we present an analogous demonstration in the semantic domain of spatial relations. In what follows, we first describe the Boster (1986) study on color. We then describe recent work on spatial language (Levinson et al., 2003) that proposes a hierarchy for the evolution of spatial categories over historical time. We next present our study, which closely follows Boster's in design. Our central finding is that English speakers successively pile-sort spatial scenes in accordance with Levinson et al.'s (2003) proposed evolutionary hierarchy. We conclude from this finding that generalizations concerning language change may reflect cognitive forces in the mind of speakers of a single language, in the domain of space as well as in that of color.

2.2 Color categories in language and cognition

Boster (1986) asked speakers of English to successively pile-sort colors. He initially instructed participants to sort a set of eight colors into two "natural groupings" on the basis of similarity, imagining that they spoke a language with only two color terms. He then asked them to subdivide either of those two groups, making three groups total—and so on until each color was in a group by itself. Finally, he tested whether these hierarchical pile-sorts matched a linguistic hierarchy that had been proposed to represent the historical evolution of color categories across languages (Kay & McDaniel, 1978, elaborating a proposal by Berlin & Kay, 1969). That hierarchy of color term evolution is shown in Figure 2.1. The top split of this hierarchy represents the claim that a two-term color naming system will tend to group BLUE, PURPLE, GREEN, and BLACK into one category, while grouping WHITE, RED, ORANGE, and YELLOW into the other—as in the language Dani (Heider, 1972). Splits lower in the tree represent claims about finer-grained linguistic divisions, which also tend to match cross-language synchronic and diachronic data (e.g. Dougherty, 1977; Kay, 1975).

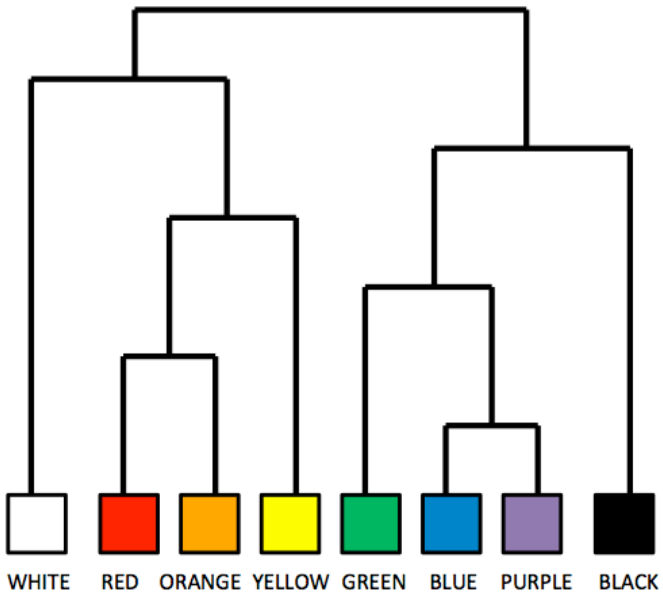

Figure 2.1: Kay and McDaniel's (1978) proposed evolutionary hierarchy of color terms.

Boster (1986) found that there was a significant tendency for successive pile-sorts by English speakers to follow the "successive differentiation" (Kay & McDaniel, 1978: 640) of this linguistic evolutionary hierarchy. This finding suggests that, at least in the semantic domain of color, the forces that produce language change over time may be present in the mind of an individual at a given moment.

2.3 An evolutionary hierarchy for spatial language

We wished to further test this claim in a different semantic domain: spatial relations. For this, we required an evolutionary hierarchy of spatial terms, to play the same role in our analysis that Kay and McDaniel's (1978) color hierarchy played in Boster's. Levinson et al. (2003) have suggested such a spatial hierarchy, based on cross-language observations of spatial systems, and drawing an explicit analogy with the above-cited work on color. They hypothesized that spatial topological categories in the world's languages evolve such that "large categories will tend...to be split into [smaller] categories over time under particular functional pressures" (Levinson et al., 2003: 512), as shown in Figure 2.2, to be interpreted as the color hierarchy in Figure 2.1 was interpreted.[3]

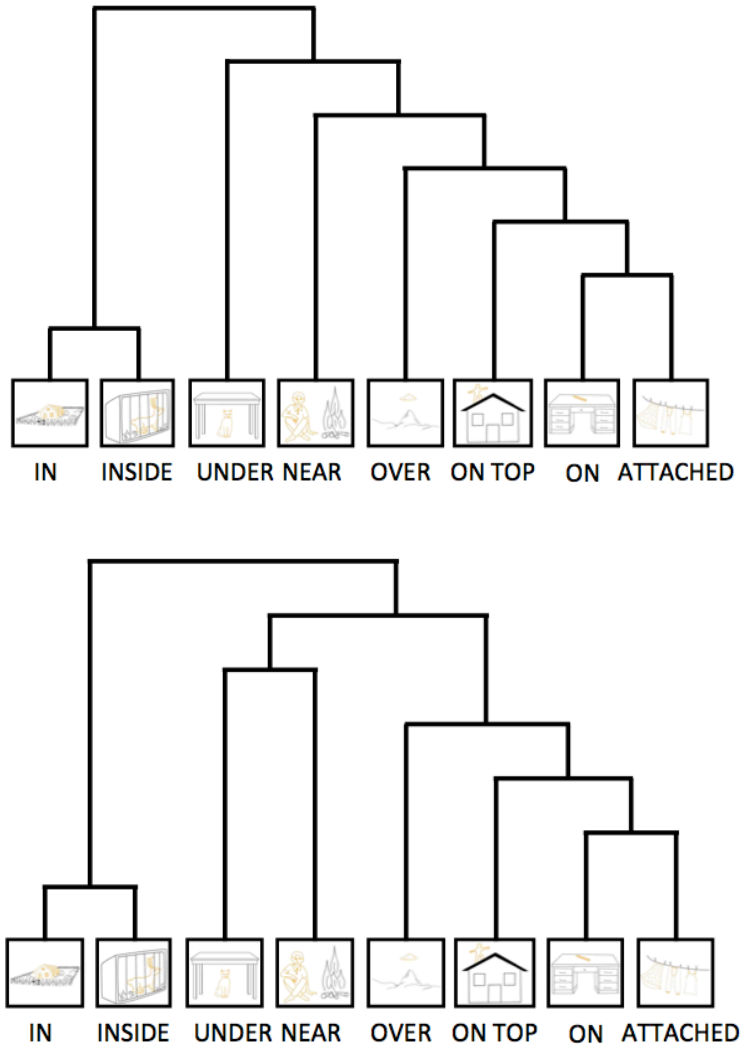

Figure 2.2: The two variants of Levinson et al.'s (2003) proposed evolutionary hierarchy of topological spatial concepts, shown with the corresponding focal scenes used in the sorting tasks.

[3] Levinson et al. (2003) actually proposed two closely related hierarchies, both of which are shown in Figure 2.2 and considered in our analyses.

2.4 Semantic evolution as gauged by pile sorting

The present study examines successive pile-sorting of spatial scenes by speakers of English, and asks whether these pile-sorts recapitulate the evolutionary spatial category hierarchy proposed by Levinson et al. (2003).

Methods

Following Boster (1986), we performed an experiment with two conditions in which participants sorted spatial stimuli. In both conditions, participants were instructed to sequentially subdivide the eight stimuli—either the line drawings of Figure 2.3 (scene sorting condition) or corresponding verbal labels (label sorting condition)—into partitions with 2, 3, 4, 5, 6, and finally 7 groups, at which point there were no further decisions to make about which group to split next.

Participants

A total of 60 members of the UC Berkeley community took part in the two conditions, with 30 participants in each. Data from 15 participants were excluded from analysis; 3 participants did not meet the study requirement that they be native English speakers, 2 reported familiarity with related research, and 10 did not follow instructions in completing the task (e.g. they failed to fully subdivide the stimuli). Accordingly, 24 participants were included in the scene sorting condition and 21 participants in the label sorting condition, all of whom had learned English by age 4 (although a number were bilingual), and were naïve to the research hypothesis and related findings.

Spatial scene sorting

Participants were presented with eight scenes from Bowerman and Pederson's Topological Relations Picture Series (TRPS; 1992). The scenes were arranged linearly on a tabletop in different randomly shuffled orders and participants were instructed to successively divide them based on the similarity of the depicted spatial relationships. Each of the eight scenes—shown in Figure 2.3—depicts an orange figure object located relative to a black background, representing the following spatial relations: NEAR (TRPS scene 37), ON (59), IN (60), ATTACHED (38), UNDER (31), INSIDE (54), ON TOP (34), and OVER (36). These particular scenes were chosen to represent focal "attractors" in spatial semantics (Levinson et al., 2003), analogous to the focal colors proposed by Berlin and Kay (1969) and used in Boster's (1986) color chip sorting task. Each focal spatial scene was selected based on (1) consistency with Levinson et al.'s (2003) characterization of focal attractors within the core spatial categories named above, and (2) the preferences of native English speakers in a pilot study, who were asked to select the best examples of each relation from the Topological Relations Picture Series.

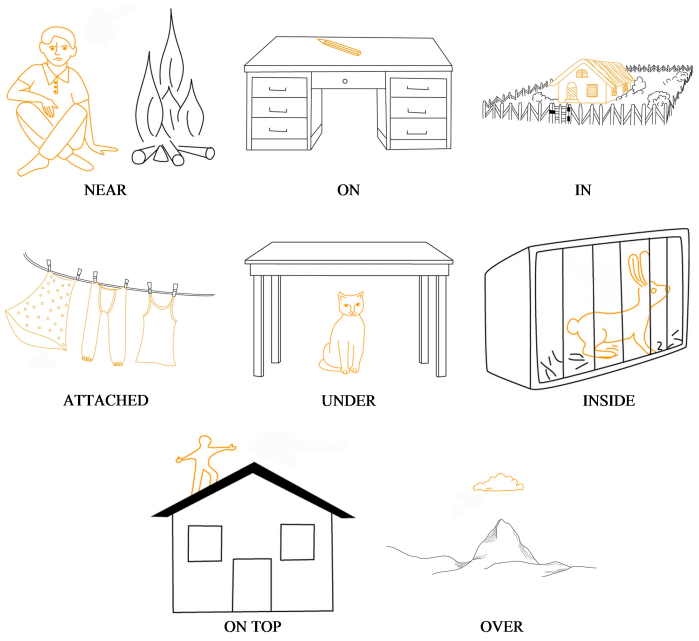

Figure 2.3: Focal scenes from the Topological Relations Picture Series used in the sorting tasks.

Instructions were adapted from Boster (1986) and asked participants to imagine they spoke a language with only two spatial words, and accordingly, to divide up the relations shown in the scenes to make two natural groupings. After participants initially split the eight scenes into two groups, they were instructed to successively subdivide their categories until all scenes were separated, and each subdivision was recorded to create an ordered hierarchy of divisions for each participant (see Figure 2.4 below for an example).

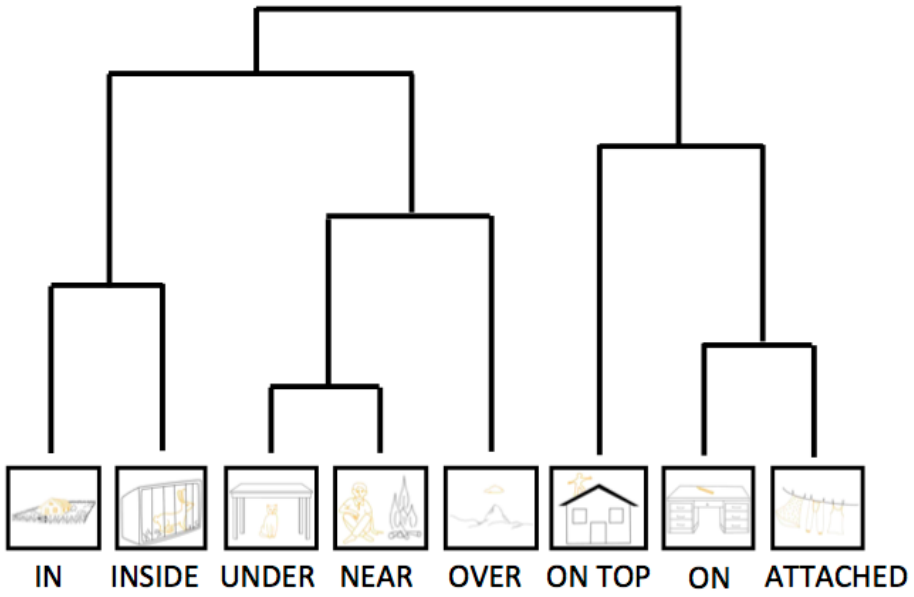

Figure 2.4: Example hierarchy from a participant in the scene sorting condition.

Spatial label sorting

The spatial label sorting task was identical to spatial scene sorting, except that in this task, participants were presented with the written English spatial expressions NEAR, ON, IN, ATTACHED, UNDER, INSIDE, ON TOP, and OVER. The labels were presented on paper in a randomly shuffled order (fixed across participants for convenience in presentation), and again, participants were instructed to successively divide the stimuli based on the similarity of the spatial relations they describe. As in Boster (1986), the images from the visual sorting task were made available to participants for reference, although they were instructed to base their partitions on the meanings of the spatial phrases themselves, rather than any specific components of the reference scenes.

Analysis

Following Boster (1986), we first measured the similarity between Levinson et al.'s (2003) hierarchy (which we refer to as the model) and the empirical data. We then compared this observed similarity to that between the data and random permutations of the model, to determine whether the observed similarity was significantly greater than chance. Finally, we asked whether there was a significant amount of residual data left unaccounted for by the model.

Similarity metric

In order to compare the empirical color hierarchies made by participants in his experiment to Kay and McDaniel's (1978) theoretical hierarchy representing the diachronic stages of color lexicon evolution, Boster (1986) converted each hierarchy to a similarity matrix. For each pair of colors, he determined the earliest stage in the hierarchy at which those two colors were separated into different groups, and took this to be the similarity between them. Thus, each non-identical pair had a minimal similarity of 1, meaning they were grouped together only when all eight colors were grouped together, and a maximal similarity of 7, meaning that they were the last pair to be separated, only after the other 6 colors were fully partitioned into groups of 1 each.

We applied the same analysis to the spatial hierarchies produced in this experiment, creating an 8x8 matrix representing the similarities across all pairs of spatial relations for each participant. Following Boster (1986), we then averaged across corresponding cells in the matrices from all participants in a given condition to create two group similarity matrices—one based on scene sorting and the other on label sorting. As in the color study, we used Pearson correlations to measure the similarity between matrices, where correlations were calculated based on all corresponding pairs of off-diagonal cells.

Model comparison

Given the empirical similarity matrices from each condition and Pearson correlations as a metric of similarity between such matrices, we ask whether the English speakers in our experiment created hierarchies that were systematically consistent with the cross-linguistic evolution of spatial lexicons as hypothesized by Levinson et al. (2003).

As with the empirical hierarchies, we created similarity matrices based on the Levinson et al. hierarchy. Like the Kay and McDaniel model (1978), Levinson et al.'s hierarchy includes

some variability in the relative order with which certain categories emerge. For instance, the authors leave intentional variability in whether UNDER or a cluster of ON-like relations (i.e. ON, ON TOP, ATTACHED, OVER) are split from a more general composite locative concept first. In keeping with Boster's treatment of such variability in the Kay and McDaniel model, we created two model-consistent hierarchies expressing both alternatives (shown in Figure 2.2). The similarity matrix representing the Levinson et al. model was created by averaging the similarities derived from these two model-consistent hierarchies.

We assessed the alignment of our empirical and model similarity matrices using Pearson correlations, so in order to determine whether these observed correlations were significantly greater than expected by chance, we used Monte Carlo simulations to create a distribution of comparison correlations. To do this, we randomly permuted the labels on our model similarity matrix, creating 1,000 permuted variants. Each permuted variant was comparable to the original model in that all similarity values were preserved in the matrix, but simply re-assigned to different pairings of spatial foci. We then measured the correlation between the permuted model matrix and each of the empirical matrices to determine whether the correlation between the model and the actual empirical data was greater than chance, i.e. that human data was more strongly correlated with the model than 95% of random permutations derived from it.

Following Boster's analysis, our initial comparison permuted the elements of our model, but preserved the overall structure of the hierarchy (changing only the labels of the leaves). A possible concern is that the Levinson et al. (2003) model may resemble human data more closely than other hierarchies of the same structure, but not significantly more so than alternative hierarchies with different structures. To address this concern, we performed another Monte Carlo simulation (in addition to the permutation test used by Boster) in which we created a set of 1,000 ordered, binary comparison hierarchies with random tree structures. As in the permutation analysis, we constructed a comparison distribution of similarities between the empirical data matrices and matrices corresponding to each of the random trees. If the data resembles the model to a degree greater than chance, then the empirical similarity matrices will match the actual model matrix more closely than they will match 95% of the matrices derived from random trees.

Residual analysis

Again following Boster's (1986) methods, we employed a final analysis designed to determine whether a significant portion of the observed similarity matrix data was left unexplained by the model (Hubert & Golledge, 1981). The model similarity matrix and two empirical similarity matrices were standardized by subtracting the mean of all values for each matrix from each cell in that matrix, and dividing the result by the standard deviation of the original values in that matrix. The values in each cell of the now standardized model matrix were then subtracted from corresponding cells in the standardized empirical matrices to determine the residual empirical data left unexplained by the model. We measured the Pearson correlations between these residual matrices and their corresponding empirical counterparts.

If the residual matrices no longer bear significant similarity to their full empirical counterparts, we take that to mean that the Levinson et al. (2003) model has accounted for the explainable empirical variation. In order to test the significance of the correlation between the residual and observed data, we again created a set of 1,000 simulated matrices by randomly permuting the labels on each of the residual matrices. We measured the correlations between these permuted simulations of the residual matrices and the original empirical matrix and

compared this distribution of correlations to that between the actual residual matrices and their empirical counterparts. As before, we took the observed correlation to be significant only if it was greater than that of 95% of the randomly permuted variants.

Results

Our similarity analysis found strong correlations between the Levinson et al. (2003) model matrix and the empirical matrices derived from spatial scene sorting ($r = 0.638$) and spatial term sorting ($r = 0.664$), as well as between the two empirical matrices themselves ($r = 0.861$). These correlations are presented in Table 1 below alongside the corresponding correlations from Boster (1986).

Table 2.1: Pearson correlations compared to Boster (1986).

Correlation	Present study	Boster
Image sorting vs. model	0.64	0.84
Label sorting vs. model	0.66	0.81
Image vs. label sorting	0.86	0.87

Our permutation analysis found that the scene sorting data was more strongly correlated with the actual Levinson et al. (2003) model matrix than with 995 out of 1000 permuted models, corresponding to a 1-tailed *p*-value of .005. Similarly, the spatial label sorting data was more similar to the model matrix than to 997 out of 1000 permuted versions of the model, corresponding to a 1-tailed *p*-value of .003. These results (pictured in Figures 2.5 and 2.6) confirm that the observed correlations represent a significant degree of similarity between the empirical matrices and that of the spatial hierarchy model.

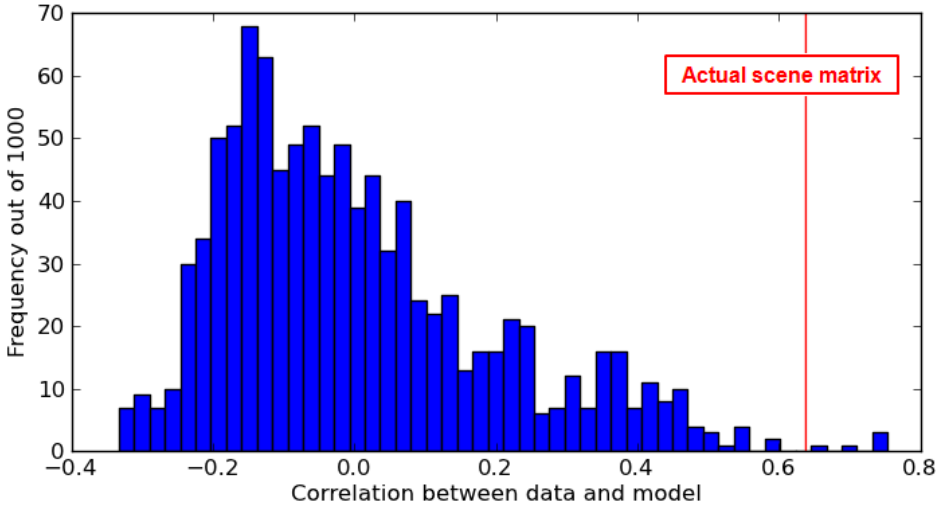

Figure 2.5: The spatial scene sorting data is more strongly correlated with the model than with 99.5% of permuted comparison models.

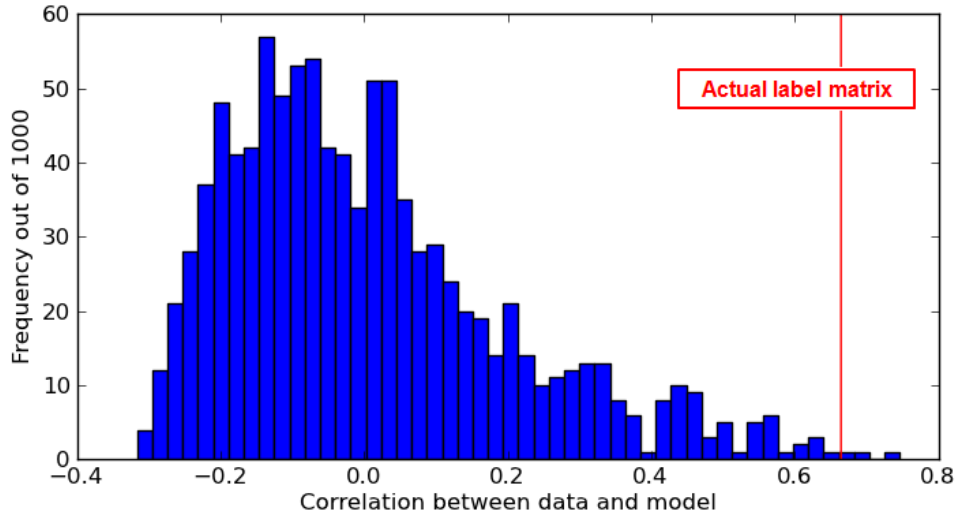

Figure 2.6: The spatial label sorting data is more strongly correlated with the model than with 99.7% of permuted comparison models.

We found comparable results in comparing the empirical data to fully random binary tree hierarchies. In both scene sorting and label sorting conditions, the empirical data was more strongly correlated with the Levinson et al. (2003) model than with 995 out of 1000 random tree hierarchies, corresponding to a 1-tailed *p*-value of .005.

The correlation between the empirical scene sorting data and the corresponding residual data after subtracting out the model-explained variation is negligible and not significant (r = -0.072; Monte Carlo 1-tailed p = 0.674). Results are comparable for tests of the correlation between empirical and residual data in the label sorting task (r = 0.073; p = 0.340), which may be interpreted analogously to Boster's (1986) study, as suggesting that the Levinson et al. (2003) model accounts for all of the explainable observed variation.

2.5 Discussion

We find substantial evidence in support of the hypothesis that English speakers synchronically recapitulate Levinson et al.'s (2003) proposed cross-linguistic patterns in the diachronic evolution of spatial lexicons. Our finding in the spatial domain directly parallels that of Boster (1986) in the color domain. Taken together, our finding and his suggest that, at least in these two semantic domains, proposed patterns of language change may be reflected in the minds of individuals at a given moment.

At the same time, there are grounds for some caution. First, as we have noted, the Levinson et al. (2003) hierarchy was intended as a tentative diachronic hypothesis, based on synchronic cross-language observation—not as a firm diachronic claim. Direct assessment of that hierarchy using historical data has to our knowledge not yet been conducted, and would be needed before our account can be considered to concern actual, rather than merely proposed, patterns of spatial language change. Second, our analyses, like Boster's (1986), were based on a comparison between model predictions and an aggregate measure of all participants' sorting. No individual participants either in Boster's (1986) study or in ours actually recapitulated the model predictions exactly—perhaps unsurprisingly given the extremely large number of hierarchical

pile-sorts that are possible, some of which are only minimally different from model predictions. Despite these cautionary notes, the present study, like Boster's (1986), has nonetheless demonstrated that such recapitulation is present as a general shared tendency—and that in this sense at least, proposed patterns of language change reflect the structure of the mind.

Chapter 3

Communicative efficiency as a source of universals

The previous chapter provides an account of universals in spatial semantics as reflections of a particular, domain-specific hierarchy of spatial notions. Other work has explained linguistic universals in a *domain-general* way (e.g., Kemp & Regier, 2012; Regier et al., 2015), suggesting that the constrained variation across languages reflects universal communicative needs. Consistent with this idea, Khetarpal et al. (2009, 2013) demonstrated that spatial category systems tend to support highly informative communication, grouping category members in a way that enables near-optimal reconstructions of a speaker's intended meaning. That finding helps to explain semantic universals and variation across languages, but does not explain how the categories in language come to assume the forms they have. This study shows that human simulation of cultural transmission in the lab produces systems of semantic categories that converge toward greater informativeness, in the domains of color and spatial relations. These findings suggest that larger-scale cultural transmission over historical time could have produced the diverse yet informative category systems found in the world's languages. This work supports the communicative efficiency account of universals in language and establishes a process through which categories in language become increasingly efficient and increasingly universal.

3.1 The origins of semantic diversity

Languages vary widely in their fundamental units of meaning—the concepts and categories they encode in single words or other basic forms. For example, some languages have a single color term spanning green and blue (Berlin & Kay, 1969), and some have a spatial term that captures the notion of being in water (Levinson et al., 2003: 496), neither of which is captured by a single word in English. Yet at the same time, similar or identical meanings often appear in unrelated languages. What explains this pattern of wide yet constrained variation?

An existing proposal suggests an explanation in terms of the functional need for *efficient communication*: that is, communication that is highly informative yet requires only minimal cognitive resources. There may be many ways for systems to be communicatively efficient, and the different category systems that we see across languages may represent different language-specific solutions to this shared communicative challenge. This idea has accounted for cross-language semantic variation in the domains of color (Regier et al., 2007; 2015), kinship (Kemp & Regier, 2012), spatial relations (Khetarpal et al., 2013), and number (Xu & Regier, 2014).

However, this prior work has also left an important question unaddressed. In a commentary on Kemp and Regier's (2012) kinship study, Levinson (2012) pointed out that although that research explains cross-language semantic variation in communicative terms, it does not tell us "where our categories come from" (p. 989); that is, it does not establish what *process* gives rise to the diverse attested systems of informative categories. Levinson suggested that a possible answer to that question may lie in a line of experimental work that explores human simulation of cultural transmission in the laboratory, and "shows how categories get honed through iterated learning across simulated generations" (p. 989). We agree that prior work explaining cross-language semantic variation in terms of informative communication has not yet addressed this central question, and we address it here.

3.2 Iterated learning and category systems

The general idea behind iterated learning studies is that of a chain or sequence of learners. The first person in the chain produces some behavior; the next person in the chain observes that behavior, learns from it, and then produces behavior of her own; that learned behavior is then observed by the next person in the chain, who learns from it, and so on. This experimental paradigm is meant to capture in miniature the transmission and alteration of cultural information across generations; the learned behavior generally changes as it is filtered through the chain of learners.

Iterated learning and related learning studies have produced a number of findings that are directly relevant to the development of informative category systems. Kirby et al. (2008) showed that iterated learning of artificial languages resulted in those languages gradually becoming more structured, suggesting that linguistic structure could emerge from the dynamics of cultural transmission. Fedzechkina et al. (2012), in a non-iterated but relevant learning study, showed that learners of an artificial language restructured their input in a way that increases the efficiency of the learned system—specifically, learners preferentially deployed case marking in contexts in which it was highly informative, although that bias was not present in the input. This finding establishes the general principle that learners may alter their input in the direction of greater efficiency. However, the study did not examine the learning of systems of semantic categories, and it is unknown whether the principle they established generalizes to the shaping of such systems. Finally, Xu et al. (2013) conducted an iterated learning study that *did* examine the learning of semantic category systems—but did not examine informativeness (see also Silvey et al., 2015). Xu et al. (2013) showed that iterated learning of color names produces systems of named color categories that are similar to those found in the world's languages. It is known that naturally-occurring color naming systems tend to support informative communication (e.g. Regier et al., 2015), so Xu et al.'s (2013) results indirectly suggest that iterated learning may lead to greater informativeness in category systems. However they did not directly test whether that is the case, and did not examine any semantic domain other than color.

Taken as a whole, the literature reviewed above leaves open two major relevant questions. (1) Does iterated learning of category systems in fact produce systems of greater informativeness? (2) If so, is this tendency toward informativeness found across different semantic domains? We pursue these questions here, to see whether they provide an answer to the challenge posed by Levinson (2012).

In what follows, we first present a computational framework for exploring semantic systems through the lens of informative communication. We then present two studies. In the

first, we reanalyze the color naming data of Xu et al. (2013), and ask whether those data reveal convergence toward informative color naming systems. In the second study, we conduct an analogous iterated learning experiment in the domain of spatial relations, and ask the same question of those data. To preview our results, we find that in both domains, systems of semantic categories become increasingly informative through the process of iterated learning. We conclude that the informative yet varied systems of categories in the world's languages may have resulted from larger-scale processes of cultural transmission.

3.3 Informative communication

We take a semantic system to be *informative* to the extent that it supports accurate mental reconstruction by a listener of a speaker's intended message (Kemp & Regier, 2012; Regier et al., 2015). Figure 3.1 illustrates this idea in the context of communicating about color in English.

In the figure, time and causality flow from left to right. The speaker has in mind a particular target color *t* drawn from the universe *U* of all colors, shown here for simplicity as a 1-dimensional spectrum. The speaker represents this target color as a probability distribution *s* over *U*, centered at *t*. In our treatment below, we will assume that the speaker is certain of the target object, so that $s(t)=1$ and $s(i)=0$ "$i \neq t$, but the framework can be generalized to accommodate speaker uncertainty about the target. The speaker wishes to communicate the target color to the listener, and so uses a word *w*: here, the English word *blue*. Having heard this word, the listener then attempts to mentally reconstruct the speaker's representation *s*, given *w*. The listener's reconstruction is also a probability distribution, *l,* and is intended to approximate the speaker's distribution *s* but is necessarily less precise, because the word *w* is semantically broad.

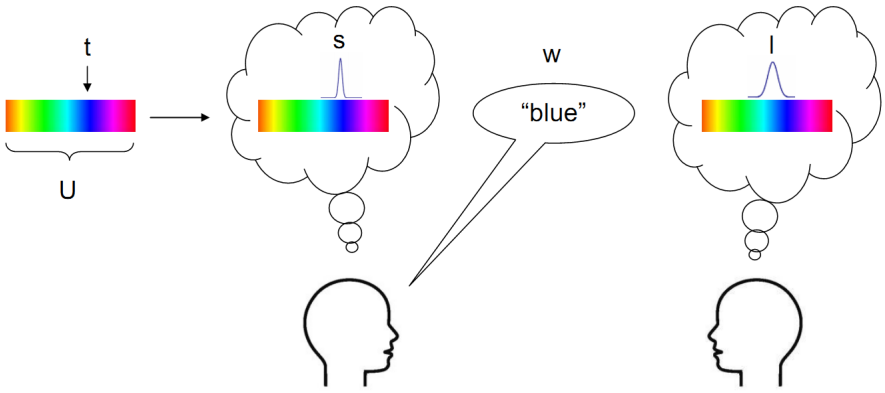

Figure 3.1: A scenario illustrating informative communication. From Regier et al. (2015).

The listener distribution is determined in different ways for different semantic domains, depending on the character of the domain. In the color and space analyses below, as in earlier work in these domains (Regier et al., 2007; 2015; Khetarpal et al., 2013), we assume a similarity-based listener distribution: the listener reconstructs the speaker's intended meaning by assigning mass to each object *i* in the domain (here, each color *i*) as a function of how similar *i* is to the objects in the category named by *w*:

$$l(i) \propto \sum_{j \in cat(w)} sim(i,j) \qquad (1)$$

This captures the intuition that category-central referents (those with high similarities to other members) are the most expected targets when that category is used. The similarity $sim(i,j)$ between objects i and j is determined separately for different domains, as described in our studies below, and provides a necessarily domain-specific treatment of the domain-general concept of similarity.

Given the speaker s and listener l distributions, we define the communicative cost $c(t)$ of communicating object t under a given semantic system to be the information lost in communication: that is, the information lost when l is taken as an approximation to s. We formalize this as the Kullback-Leibler divergence between s and l. In the case of speaker certainty as assumed here, this quantity reduces to surprisal:

$$c(t) = D_{KL}(s \,\|\, l) = \sum_{i \in U} s(i) \log_2 \frac{s(i)}{l(i)} = \log_2 \frac{1}{l(t)} \quad (2)$$

Finally, we define the communicative cost for the domain as a whole to be the expected communicative cost over all objects in the domain universe U:

$$E[c] = \sum_{i \in U} n(i)c(i) \quad (3)$$

Here $n(i)$ is the probability that the speaker will wish to talk about object i. In the analyses below, as in earlier work in color and space (Regier et al., 2007; 2015; Khetarpal et al., 2013), we assume for simplicity that $n(i)$ is uniform. We take a semantic system to be informative to the extent that it exhibits low $E[c]$. A system could increase its informativeness through the addition of more categories; in our analyses we control for this possibility by comparing (groups of) systems with the same number of categories.

3.4 Study 1: Color

Xu et al. (2013) showed that iterated learning of color naming yields categorical partitions of color space that are similar to color naming systems found in the world's languages. They measured the distance between color categories produced in their experiment and those in the World Color Survey (WCS: Cook et al., 2005), the largest existing publicly available database of color naming data, containing color naming data from speakers of 110 languages of non-industrialized societies. Xu et al. (2013) found that as color naming systems in their iterated learning task were transmitted across generations of learners, the systems became more similar to those in WCS languages. In a separate study, Regier et al. (2015) assessed the communicative cost of color naming systems in the languages of the WCS, using the formal framework described above, and showed that the majority of these systems are highly informative, despite their diversity.

Taken together, these earlier findings suggest that color naming systems produced under iterated learning may come to resemble those found in languages through gradual increases in informativeness over generations. However, that proposal of increasing informativeness under iterated learning has not been directly tested. We test it here, by reanalyzing the color naming data from Xu et al. (2013)'s iterated learning experiment in terms of the framework described above.

Methods

Iterated learning of color

Xu et al. (2013) trained an initial generation of 20 participants on random partitions of color space into 3-6 categories, and then asked them to recall those categories by labeling a set of color chips accordingly. The next set of 20 participants each studied the assignment of labels to color chips of a single first generation learner, and created their own labelings in turn, which were then used to train the subsequent generation. This procedure was iterated over 20 chains of learners with 13 generations of learners each. In each generation of each chain, participants created a full color naming system by assigning a category label to each of the 330 color chips in the color naming array used in the WCS. Xu et al. then measured the dissimilarity between these transmitted category systems, at each generation, and the color naming systems of the WCS. They measured dissimilarity using variation of information (VI: Meilă, 2007), a distance measure between different groupings of the same set of items.

The data in Figure 3.2 (red line, left y-axis) are from Xu et al. (2013). These data show that as color naming systems are filtered through generations of learners, they become more similar to the natural systems of the WCS, as Xu et al. reported. We wish to ascertain whether this change also reflects a gradual increase in informativeness, brought about through transmission.

Communicative cost

In order to assess the informativeness of a given color naming system, we need to specify how similarity is determined in that domain (recall Equation 1). As in earlier work in this domain (Regier et al., 2007; 2015), we take the similarity of two colors *i* and *j* to be a Gaussian function of the perceptual distance between them:

$$sim(i,j) = \exp(-c \times dist(i,j)^2) \qquad (4)$$

Following Regier et al. (2007; 2015), the scaling factor *c* is set to .001 for all analyses reported here, and *dist*(*i*,*j*) is the distance between colors *i* and *j* in the CIELAB color space. Given this, we can now assess the informativeness of a given color naming system following Equations 1-4.

Results

Figure 3.2 (blue line, right y-axis) shows the average communicative cost $E[c]$ of the 20 color naming systems in Xu et al's (2013) study, over the 13 generations of that study. Generation 0 corresponds to the random initial partitions supplied to the first generation of participants in training.

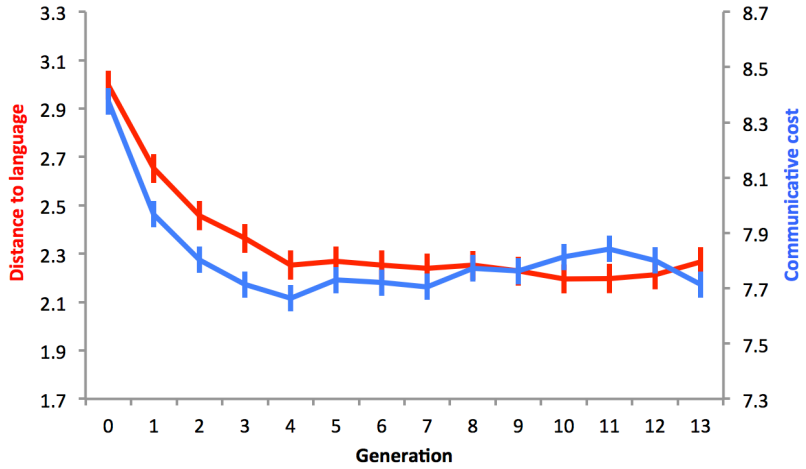

Figure 3.2: Average distance to WCS languages (red; left y-axis), and communicative cost (blue; right y-axis) of artificial systems of color categories, over generations of iterated learning. Bars indicate standard error of the mean.

It can be seen that these color naming systems exhibit decreasing communicative cost (increasing informativeness) over the first four generations of learners, after which no further systematic change is seen. This pattern of change over time closely parallels that seen in the similarity of lab-generated color naming systems to those of actual languages (red line). This finding suggests that artificial color naming systems come to resemble those found in languages through a transmission process that favors systems of greater informativeness.

3.5 Study 2: Spatial relations

Does iterated learning lead to increasing informativeness across multiple domains, or only in the domain of color? To answer this question, we conducted an analogous study in a different semantic domain, that of spatial relations.

Languages categorize the spatial domain in a wide variety of ways that nonetheless show certain recurring tendencies (e.g. Levinson et al., 2003). Figure 3.3 gives a quick sense for this variation.

Additionally, spatial systems across languages tend to support informative communication (Khetarpal et al., 2013). In both of these respects, space is like color. However it is unlike color in that it is more complex. Perceptual color space is defined with respect to just three dimensions: hue, saturation, and lightness. In contrast, the mental representations underlying the kinds of spatial relations shown in Figure 3.3 appear to rely on a much wider range of spatial features (Levinson et al., 2003; Xu & Kemp, 2010).

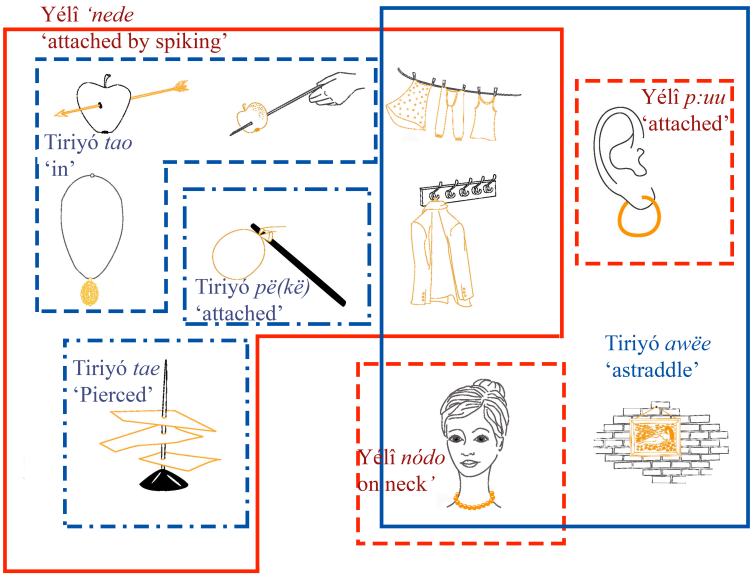

Figure 3.3: Ten spatial relations, as categorized in two languages: Tiriyó and Yélî-Dnye. Adapted from Levinson et al. (2003).

We considered spatial naming data, collected both in the field and in the lab, relative to a standard stimulus set: the Topological Relations Picture Series (TRPS: Bowerman & Pederson, 1992). The spatial scenes in Figure 3.3 above are from the TRPS. The full TRPS is a set of 71 such line drawings depicting different spatial relations. Each image shows an orange figure object located relative to a black background object. We wished to discover whether iterated learning of category systems over these stimuli would converge toward the spatial systems of natural languages, and toward greater informativeness, in a parallel to the color findings reported above.

Methods

Iterated learning of spatial relations

Fifty undergraduates at UC Berkeley took part in the study in return for class credit, forming 5 transmission chains of 10 generations each. Each participant completed an iterated learning task in which they studied and then attempted to recall category assignments for 4-category partitions of the 71 TRPS scenes.

Participants were instructed to learn spatial categories from an "alien language" by observing a series of scenes paired with visual sentences. In each training trial, a scene from the TRPS was presented for 5 seconds along with a visual sentence describing that scene in a hypothetical alien language. The visual sentence consisted of three smaller images beneath the main scene, as shown in Figure 3.4. The visual sentences showed the figure and ground objects from the main scene separately, and a colored patch indicating the alien spatial category to which the spatial relationship between figure and ground belongs. For example, in Figure 3.4, the participant is labeling the spatial relation apple-in-bowl as belonging to the category marked by red. Other scenes would be labeled by other colors, for a total of four color-coded categories.

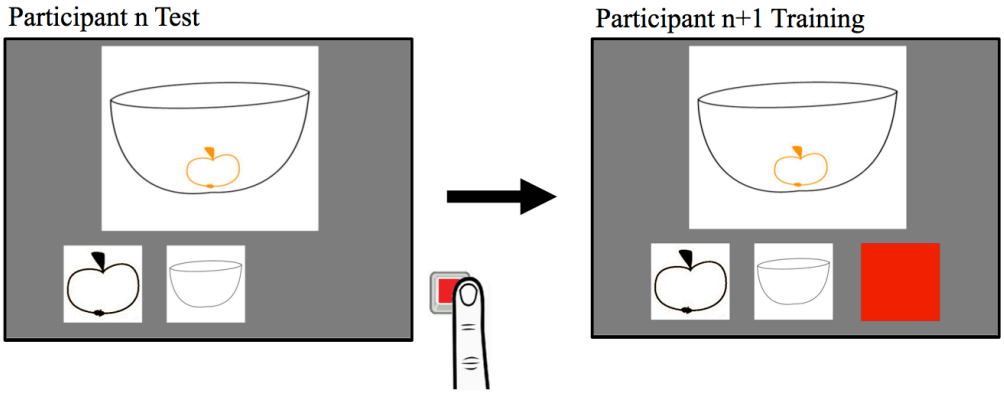

Figure 3.4: Example test and training trials from two consecutive generations of a transmission chain.

Participants completed two training sessions in which each of the 71 TRPS scenes was presented one at a time in random order paired with a color representing the spatial category to which that scene belongs. After two rounds of training, participants were shown the scenes and visual sentences a final time, but without the color label, and categorized each spatial relationship according to the alien language by pressing colored keys to indicate category assignments. Color labels and their locations on the keyboard were counterbalanced across participants within each iterated learning chain.

As in Xu et al.'s (2013) study, each of the 5 chains was initialized as a random partition of the 71 TRPS scenes into four roughly equally-sized categories, which the first participants in each chain studied during training and attempted to reproduce in the following test session. All subsequent participants in each chain were trained on the responses of the previous participant and were instructed to reproduce them as closely as possible, but were not aware that any of the data had any connection to other participants.

We excluded any participants whose categorization accuracy was at or below chance or who reported that they relied principally on non-spatial information (e.g. the objects involved) to learn the spatial categories.

Distance to languages

Analogous to Xu et al. (2013), we measured the dissimilarity between these transmitted spatial category systems at each generation, and the spatial systems of languages. Our target languages were a convenience sample: Arabic, Basque, Chichewa, Dutch, English, Japanese, Máíhɨki, Mandarin Chinese, and Spanish. The naming data were collected by our group in Arabic (unpublished), Chichewa (Carstensen, 2011), Japanese (unpublished), Mandarin Chinese (Tseng et al., 2016), and Spanish (ibid). Spatial labels in the remaining languages was collected by collaborators who kindly shared their data with us; we thank Asifa Majid for contributing the Basque (Levinson et al., 2003) and Dutch data (Khetarpal et al., 2010), Naveen Khetarpal for the English data (ibid), and Grace Neveu and Lev Michael for the Máíhɨki data (Khetarpal et al., 2013). All data were collected relative to the TRPS scenes. For each language, we assigned to each TRPS scene the spatial term that was applied to that scene by the plurality of native speakers interviewed. This procedure yielded labels for all TRPS scenes, in each language. Following Xu et al. (2013), we used variation of information (VI) to measure the distance between category systems obtained through iterated learning, and those found in these languages.

Communicative cost

In order to assess informativeness for spatial relations, as for color, we needed an independent measure of similarity. We took the similarity between any two spatial relations stimuli to be determined by pile-sorting of those stimuli in a separate study. Khetarpal et al. (2010) asked native English speakers to sort the TRPS scenes into piles based on the similarity of the spatial relationships they depict (this study is described in greater detail in the next chapter). We took the similarity of any two scenes to be the proportion of participants who sorted those two scenes into the same pile in Khetarpal et al.'s (2010) data.[4] Given this specification of similarity, we assessed the informativeness of spatial naming systems following Equations 1-3.

Results

Figure 3.5 (red line, left y-axis) shows the average distance (VI) between the spatial naming systems generated through iterated learning, and those of our language sample. This distance gradually decreases, as the systems are shaped by transmission from generation to generation. Thus, as in the case of color, iterated learning leads to spatial naming systems that become increasingly similar to those of natural languages.

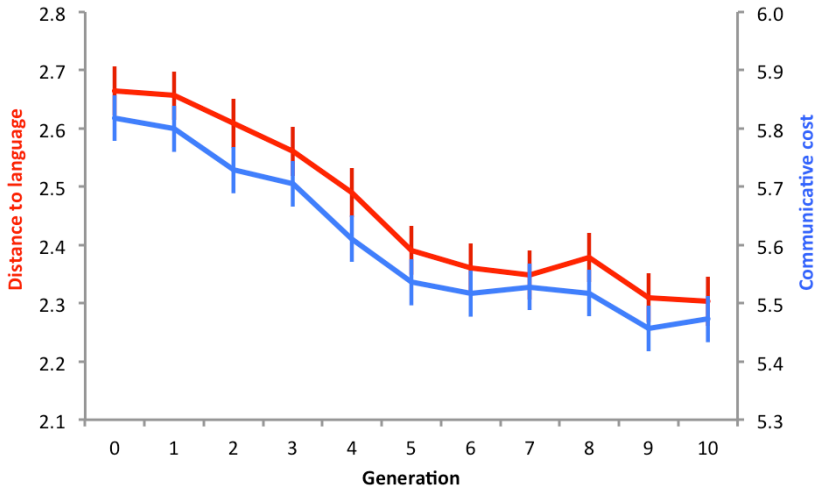

Figure 3.5: Average distance to languages (red; left y-axis), and communicative cost (blue; right y-axis) of artificial systems of spatial categories, over generations of iterated learning. Bars indicate standard error of the mean.

For comparison, Figure 3.5 (blue line, right y-axis) shows the average communicative cost of category systems across generations in our experiment. As in the case of color, this quantity also decreases as systems are transmitted from generation to generation, showing that transmitted spatial systems become more informative as they are transmitted. Moreover, again as in the case

[4] While this measure of similarity is derived from pile sorting by speakers of English, analyses of this pile sort data find very small amounts of language-aligned pile sorting (Tseng et al., 2016). In fact, sorting is more similar across speakers of different languages than across halves of one language group in split-half reliability (Khetarpal et al., 2010), suggesting that this measure of similarity is generally comparable across languages.

of color, this decrease closely tracks the decrease in distance to language, suggesting that iterated learning produces spatial systems that resemble those of languages through a transmission process that favors informative categories.

A natural concern is that the participants in our experiment may have been influenced by their knowledge of English, and that the increasing proximity of the learned systems to those of actual languages may have been driven by English semantic structuring. We feel this concern should be lessened by three observations (not shown in the figure): (1) the learned category systems get progressively closer to all languages considered, including those with categories that cross-cut English spatial terms; (2) the learned category systems are closer to some other languages (e.g. Arabic, Chichewa, and Mandarin Chinese) than they are to English; and (3) the same qualitative results obtain when English is excluded from the set of languages to which the learned category systems are compared. Given this, it seems plausible that the increasing proximity to languages may have been driven in large part by universal semantic tendencies and cognitive forces, rather than by the English language itself.

Increases in both informativeness and language-like semantic structuring are illustrated below in Figure 3.6. The figure shows scenes from a single category at the beginning (left panel) and end (right panel) of our experiment. After transmission through 10 generations of learners, the meaning of the category has been altered through the loss of many initial members depicting a wide variety of spatial relations, down to a set of scenes exemplifying a novel relational category that expresses the notion "tightly around", or encirclement and tight fit. This spatial notion is intuitively clear, yet does not correspond to a single spatial term in English, the primary language of our participants.

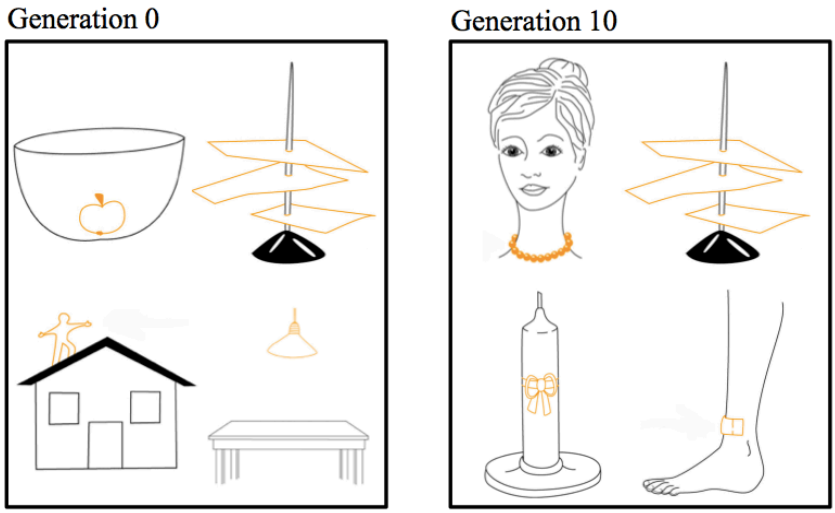

Figure 3.6: Representative scenes showing the semantic reorganization of a single category over transmission.

3.6 Discussion

We have shown that iterated learning produces semantic systems that tend toward informative category structure, and also toward similarity with human languages. We find this pattern in two domains—color and spatial relations—suggesting that it may hold more generally across

domains. To the extent that these findings do generalize, they suggest an answer to Levinson's (2012) question of how diverse category systems across languages assume their highly informative character.

Our findings reported here suggest support for a specific account of the origins of the semantic diversity seen in the world's languages, as a natural result of shared communicative principles, operating across communities of language learners, and across time.

Chapter 4

Efficiency as a source of universals in cognition

The preceding chapter accounts for language evolution in terms of general principles of efficient communication but does not explore the relevance of these principles for cognition more broadly. The study in this chapter asks whether the same principles that govern efficient semantic systems also characterize nonlinguistic cognition. Inspired by an earlier treatment of color memory (Lantz and Stefflre, 1964), this question adopts a view of thought as self-directed communication over time. To evaluate this view, we follow a prior study by Khetarpal et al. (2010), which assessed language and cognition in speakers of English and Dutch via pile sorting of spatial stimuli. We replicate previous universalist findings in two new languages, Chichewa and Máíhɨki, finding that speakers of all four languages, despite large differences in the granularity of their linguistic spatial systems, made pile sorts with similarly fine granularity. We propose that this universal tendency reflects a drive for efficiency in cognition, analogous to that found in language in the previous chapter. Here, we provide an account of spatial cognition in which conceptual categories optimize the trade-off between informativeness (making for fine-grained and intuitively organized spatial categories) and simplicity (limiting the number of categories). We find that pile sorts made by speakers of Máíhɨki and Chichewa match this universal account more closely than they match the semantics of the sorters' native languages. These results suggest that across languages, spatial cognition reflects universal pressures for efficient categorization, and the observed universals in category structure and granularity result from these pressures.

4.1 Universals and variation in spatial cognition

Speakers of Chichewa, a Bantu language in East Africa, use the spatial term 'mu' to refer to what speakers of English commonly describe as a goldfish *in* a bowl, a hat *on* a head, and a belt *around* a waist. Do speakers of different languages think about the world in the same way and speak about it differently, or do they think about it in systematically different ways?

These two possibilities correspond to the universalist and relativist views of language and thought outlined in the introduction to this dissertation. The universalist view is supported in the spatial domain by the observation that diverse languages share recurring tendencies in spatial categorization, suggesting a universal conceptual repertoire for space (Levinson et al., 2003; Khetarpal et al., 2009; Khetarpal et al., 2013). The relativist view is also supported by findings in the spatial domain, which show that differing categories in language correspond to divergence in

basic spatial cognition, and suggest that language has an important role in thought (Hermer & Spelke, 1996; Levinson, 1996; Levinson et al., 2002; Majid et al., 2004; Haun et al., 2006).

As noted at the outset of this dissertation, there is an existing proposal that posits a reconciliation of these seemingly dissonant views, suggesting that there is a universal conceptual basis for the categorical distinctions that languages make, but that this conceptual foundation may be altered by language (Kay & Kempton, 1984; Hespos & Spelke, 2004; Regier & Kay, 2009; Tseng et al., 2016). This account is supported by the findings of Khetarpal et al. (2010), who assessed cognition through pile sorting and found evidence for two forces in spatial thought: a weak language-specific force and strong universal force. Many studies have addressed the semantics of spatial language, which provide a characterization of the language-specific force (e.g., Bowerman & Pederson, 1992; Lucy, 1992; Munnich et al., 2001; Bowerman & Choi, 2003), but less is known about the nature of universals in spatial cognition. How do speakers of diverse languages come to produce pile-sorts with highly similar spatial categories? What drives these similarities?

The previous chapter showed that general principles of efficient communication can account for universals in linguistic structure; here, we ask whether the same principles that govern semantic systems also characterize nonlinguistic cognition. To do so, we apply the efficient communication framework (Regier et al., 2015) to cognition, making the assumption that cognition operates in a way that is broadly similar to a communicative system. This view of thought parallels previous work by Lantz and Stefflre (1964) in the domain of color, which treats memory "as though it were a situation in which an individual communicates to himself through time using the brain as a channel." Following this treatment generally, and the efficient communication framework in particular, we propose an account of universals in spatial cognition based on efficiency. Following Khetarpal et al., we assess spatial cognition via pile sorting across languages, repeat their analyses on this new dataset, and finally compare this data to the proposed account of efficient spatial cognition. To preview the results, we find that the patterns seen in spatial pile sorting across languages closely match my predictions. These findings establish a domain-general account of universals in spatial cognition, and demonstrate that cognition, like language, follows general principles of efficiency.

In what follows, we review the previous findings by Khetarpal et al. (2010), assess the generality of these findings with data from two additional languages, and then evaluate the account of efficient cognition against the spatial category systems made in pile sorting by speakers of diverse languages.

Khetarpal et al. (2010)

In their earlier study, Khetarpal et al. (2010) assessed categories in thought and language by asking speakers of English and Dutch to sort a set of cards depicting spatial scenes into piles based on the similarity of the spatial relation portrayed, and to name the relation on each card. Although the naming systems of the two languages differed, the sorting systems observed were quite similar, revealing universal tendencies in spatial categorization. Specifically, both Dutch- and English-speaking participants tended to sort the cards into piles that were significantly more similar to the Dutch linguistic system than that of English. Khetarpal et al. suggest that this apparent privilege of Dutch names is explained by a further finding: pile-sorts across speakers of both languages tended to be fine-grained, and the Dutch language partitions spatial relationships in a similarly fine-grained way.

In addition to this universal tendency, Khetarpal et al. found that participants' sorting systems nonetheless diverged as a function of the speaker's native language: while both English and Dutch participants sorted more like the Dutch language overall, Dutch sorters showed this Dutch-aligned tendency more strongly than English sorters did. The results of this study simultaneously reflect universal and language-specific influences on spatial cognition. Taken as a whole, they provide support for the idea that humans share a universal conceptual framework which can be further modulated by language.[5]

While promising, this evidence in support of a unifying perspective has two major limitations that make its conclusions difficult to generalize. First, the study finds strong support for universal tendencies in data from related languages and cultures, which makes the generality of these findings difficult to interpret. Second, the authors suggest that universals in cognition are driven by a tendency to be fine-grained, but this suggestion is post hoc and provides no account of *why* this tendency exists or *what* granularity is preferable. These shortcomings are discussed in detail below. The issue of limited diversity will be addressed in our extension of Khetarpal et al. (section 4.2), and the underspecified account of universals in granularity will be taken up together with our account of cognitive universals in general (section 4.3).

Diversity in the language sample

The first limitation of Khetarpal et al. (2010) is the result of homogeneity in the dataset: the data is drawn from speakers of two languages that share similar cultures and similar cultural biases to those inherent in the stimuli. Their study compared linguistic and non-linguistic categorizations of spatial relations using the Topological Relations Picture Series (TRPS; Bowerman & Pederson, 1992). A weakness of this approach is that these stimuli reflect largely Western cultural biases, depicting objects, and sometimes spatial relationships, which are commonplace in Western industrial life but often quite rare in many other environments. Thus, the criticism can be raised that the shared Indo-European origins and parallel cultural contexts of these two languages, rather than a universal conceptual repertoire, may explain the observed similarities in non-linguistic categorization. Do these findings reflect true universal similarities, or simply commonalities of Western culture? It is unclear whether the previous universalist findings by Khetarpal et al. generalize to more diverse languages. Vague

In our extension of this work, we seek to address this shortcoming. We test an assumption of the previous conclusions: that the cross-language tendency toward fine-grained nonlinguistic distinctions is due to universals of cognition, rather than shared linguistic origins or culture. In order to discriminate between these two possible explanations, we would ideally want to consider linguistic and behavioral data derived from speakers of languages that are (1) non-Indo-European, (2) spoken in a non-European culture, and (3) diverse in their semantic granularity—ideally, one language that is coarser in spatial naming than the languages previously examined, and one finer. In this study we consider two such languages: Chichewa is a Bantu language spoken in East Africa, and has a spatial naming system that is considerably

[5] The role of language is described here as modulating cognition to reflect agnosticism about the process(es) that may produce language-specific patterns in pile sorting, which are not addressed in this study. These data are consistent with relatively strong views of linguistic relativity, in which language fundamentally restructures nonlinguistic conceptual representations and also with much weaker variants in which underlying cognition is unaffected and language alters pile sorting behavior through arguably more superficial mechanisms like attention or linguistic priming. See Tseng et al. (2016) for a reanalysis of universal and language-specific patterns in this data, and a possible process by which these influences interact.

coarser than those of Dutch and English, and Máíhɨki, a Western Tucanoan language spoken in the northeastern Peruvian Amazon, has a considerably finer spatial naming system.

We obtained spatial language and categorization data from speakers of Chichewa (Carstensen, 2011) and Máíhɨki (which was kindly collected and shared by collaborator Grace Neveu under the supervision of Lev Michael) to compare with corresponding data from Dutch and English speakers. These languages were selected for their accessibility, the cultural diversity of their speakers (the majority of whom are raised and live in largely agrarian, non-industrialized towns and villages), and the diverse origins and typology of the languages themselves. These are the first non-Indo-European languages for which pile sorting data were collected and while Levinson et al. (2003) examined spatial naming systems from a diverse set of language families, these are the first languages to be examined from the Bantu and Tucanoan families.

Spatial systems in the world's languages vary substantially in how they carve everyday spatial relations into categories like 'in' and 'on'. The four languages considered here represent a broad sample of this semantic variability, shown in Figure 4.1 as an overlay to a subset of spatial relational scenes from the TRPS (Bowerman & Pederson, 1992). As indicated by the dashed gold line, all six of these scenes fall into a single expansive category, 'pa', in Chichewa. English and Dutch categorize these scenes using systems with both differing granularity and substantial semantic cross-cutting, depicted by the solid orange (English 'through', 'on', and 'in', clockwise from the top left) and dotted red (Dutch 'in', 'door', 'aan', and 'op') lines. Finally, Máíhɨki distinguishes each scene with a unique lexical category (clockwise from the top left: 'de', 'jui', 'sɨi', 'ɨmɨjai', 'tai', and 'bi', roughly meaning hanging, though, attached, on, floating, and held in the mouth).

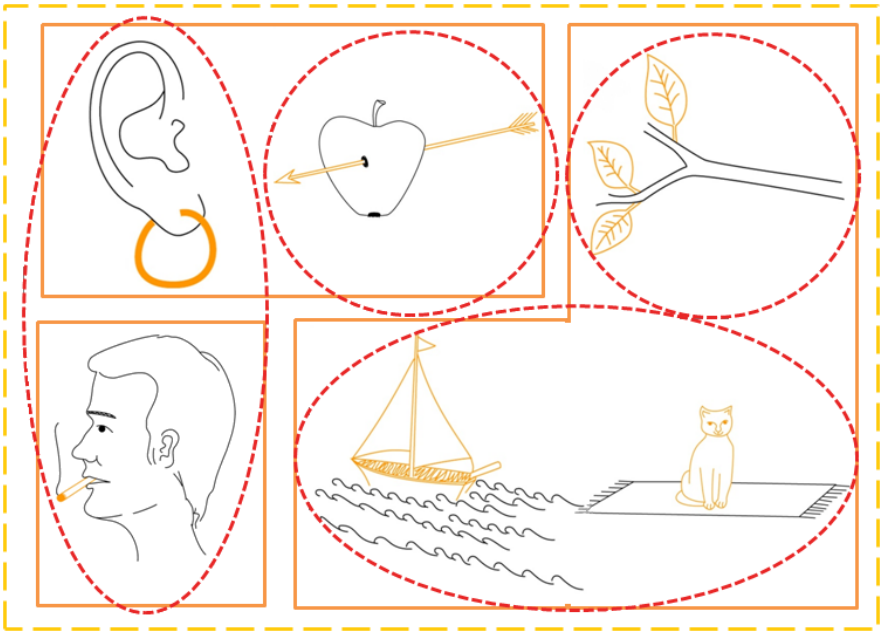

Figure 4.1: Cross-language variation in spatial semantics. The spatial categories of Chichewa, English, and Dutch are indicated by dashed gold, solid orange, and dotted red lines, respectively. In Máíhɨki, each of these scenes falls into a unique spatial category.

Granularity in language and thought

The second issue limiting the generality of conclusions from Khetarpal et al. (2010) is their post hoc and underspecified account of universals in cognition. The authors observed a cross-language tendency toward Dutch-like sorting and proposed an explanation for this universal tendency, which hinges on the observation that some languages, like Dutch, are finer-grained than others in the spatial domain. For example, Chichewa speakers use a single term, 'mu', to refer to scenes depicting spatial relations that English speakers distinguish as 'around' and 'in', as shown in Figure 4.2. Khetarpal et al. suggested that spatial cognition tends to be universally fine-grained, and that speakers of all languages may therefore sort finely and thus more like finer-grained languages—in their case, Dutch. They support this idea by analyzing height (a measure of granularity) across pile sorts, and showing that speakers of both English and Dutch tend to sort in fine-grained ways. Thus, Khetarpal et al. show that a cross-language preference for fine granularity exists, but they do not explain why this tendency exists, or what degree of fine granularity is preferable.

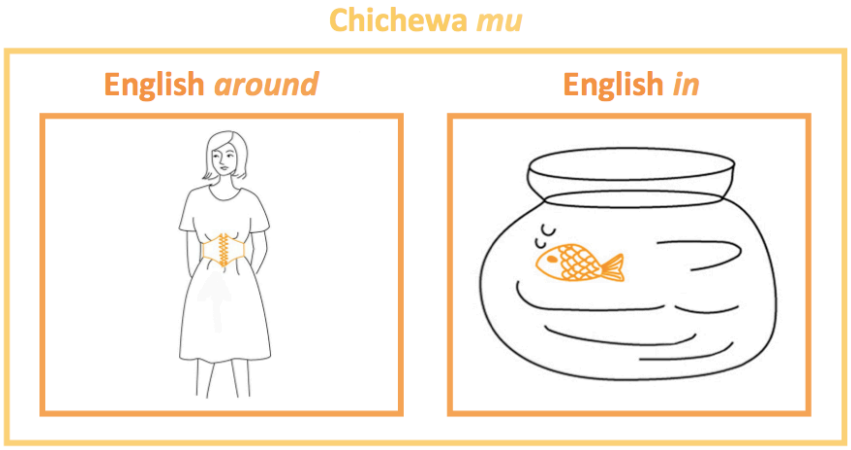

Figure 4.2: An example of varying granularity in semantic systems. The English terms 'around' and 'in' are used to distinguish spatial senses that fall under a single broad spatial category, 'mu' in Chichewa.

This previous account leaves open several important questions. How general is this granularity preference? Do we find support for this proposed universal tendency in granularity in how Máíhɨki and Chichewa speakers pile sort? If so, why? Do speakers know something about how best to partition spatial relations, regardless of language? Do they follow universal principles in constructing these categories?

We address the generality of these findings in the next section, an extension of Khetarpal et al., and in section 4.3, we provide a more specific account of the granularity universal as the result of pressures for efficiency in cognition, as well as in communication.

4.2 Extending an account of categories in language and thought

Following Khetarpal et al. (2010), we instructed native speakers of Chichewa (Carstensen, 2011) and Máíhɨki (data kindly collected by collaborator Grace Neveu) to partition a set of spatial

scenes through pile sorting and by labeling the scenes in their native languages. We then measured the granularity of both sorting and naming partitions across all four languages to determine whether Chichewa and Máíhɨki speakers create spatial categories that respect the previously observed universal granularity preference in spatial cognition.

Participants

A total of 24 native English speakers (Khetarpal et al., 2010), 24 native Dutch speakers (Khetarpal et al., 2009), 38 native Chichewa speakers (Carstensen, 2011), and 7 native Máíhɨki speakers (unpublished data contributed by Grave Neveu; see Khetarpal et al., 2013 for a related treatment of the Máíhɨki naming data) took part in both the nonlinguistic and linguistic tasks, administered in their native languages and home countries of the United States, the Netherlands, Malawi, and Peru, respectively.

Nonlinguistic categorization

Participants sorted the 71 scenes in the TRPS into piles based on the spatial relation depicted in each scene. Each scene showed an orange figure object positioned relative to a black ground object and participants were instructed to group the scenes into piles based on the similarity of these spatial relations. Participants were given the stimulus scenes shuffled in varying orders and informed that they could make as few or as many piles as they chose, rearrange their piles as they felt necessary, and could take as much time as they wanted.

Labeling

After completing the sorting task, participants were asked to name the spatial relation depicted on each card. Labels picking out the target and ground objects were supplied in the participant's native language and the participant filled in the blank between these labels to complete a sentence specifying the figure's location in relation to the ground, except for Máíhɨki, whose speakers received labels in Spanish (if clarification was necessary) and produced their own labels in Máíhɨki. Data was sanitized to collapse over responses that differed only in components without spatial meaning (e.g., variations in verb tense were standardized and verbs lacking spatial content, such as 'living,' but not 'piercing,' were removed). The sanitized responses were taken to be the participant's linguistic categories for the depicted spatial relations, in which scenes named with the same spatial construction compose a single category. The most frequent label supplied across participants for each scene (referred to here as modal terms, though they also include multi-word phrases) were taken to be the language's categories for the spatial scenes.

Height analysis

To quantify the relative coarse-grainedness of the modal terms for each language and the pile sorts of each speaker, we used Coxon's (1999) measure of height, which is the sum of the number of possible pairs in each category. This metric can be written as follows, where g_i is the number of items in group i:

$$height = \sum_i \binom{g_i}{2} = \sum_i g_i(g_i - 1)/2 \qquad (1)$$

Results

As anticipated, we found the Chichewa spatial naming system to be considerably coarser in its partitioning of spatial scenes than either Dutch or English, naming 61 of the 71 total scenes with just 3 modal terms. Conversely, Máíhɨki speakers used a naming system that was substantially finer-grained than the other three languages in our sample, employing 22 unique modal labels for the 71 spatial scenes,[6] whereas Dutch speakers used 13 modal terms, English speakers used 11, and Chichewa used only 9 unique modal terms. These generalizations about granularity are confirmed quantitatively by our height analysis, which shows that Chichewa is more coarse-grained in its spatial naming than both Dutch and English, and Máíhɨki more fine-grained.

To compare granularity across language and cognition, we compute the height of each pile sort and of the modal terms for each language. Figure 4.3 below presents the heights of Máíhɨki, Dutch, English, and Chichewa, shown relative to the pile sorts created by speakers of each language, the large majority of which prize a small region of the granularity range that falls between Máíhɨki and Dutch.

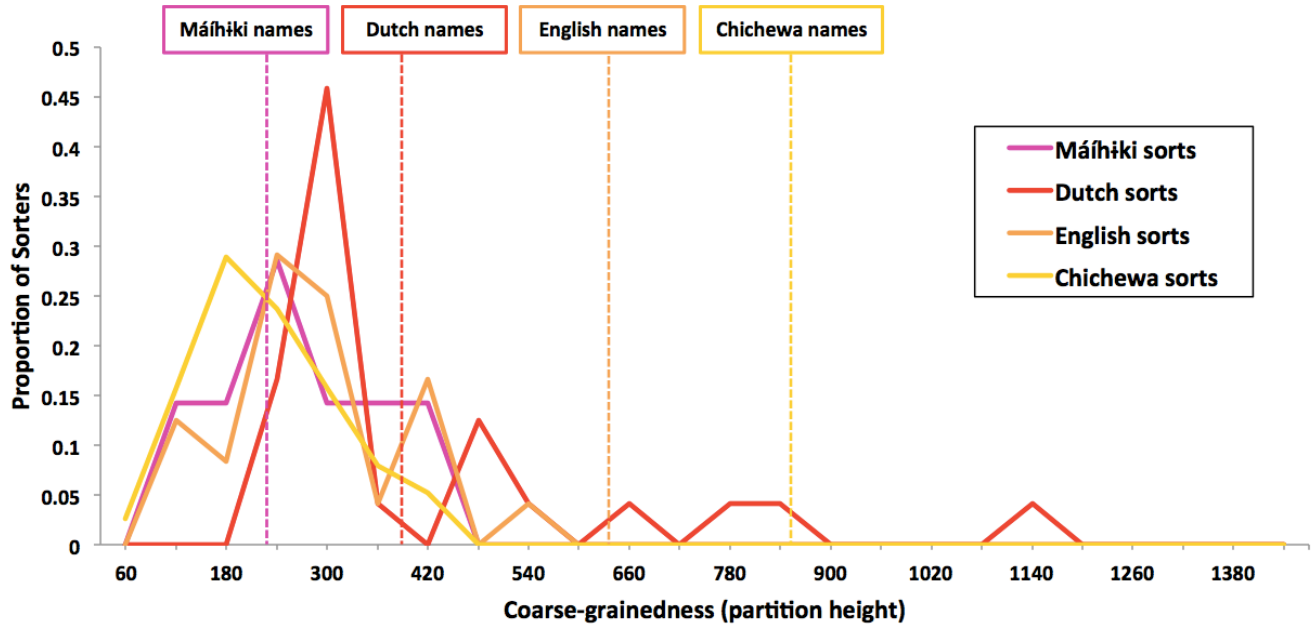

Figure 4.3: Granularity of spatial categories in language compared to pile sorting. Dotted lines indicate the granularity of each language's modal names. Solid lines represent the distribution of granularity in pile sorts, color coded by the sorter's native language. The majority of pile sorts are relatively fine-grained.

[6] While Máíhɨki speakers produced labels for all 71 spatial scenes of the TRPS, 3 of these scenes (TRPS scene numbers 4, 26, and 55) failed to elicit labels that described a static spatial relationship. Further consultation with several participants suggests that these scenes depict relations (e.g., negative space, as in scene 26 showing a crack in a cup) that are outside the domain of spatial naming in Máíhɨki and would instead be characterized as possessives or other nonspatial relations. For consistency in our cross-language comparison, these 3 scenes are excluded from all analyses; we obtain similar results regardless of their exclusion.

As illustrated in Figure 4.3, height analysis of the pile sorts shows that participants tend to categorize spatial relationships with similar degrees of granularity and in ways that are more fine-grained than most of the linguistic systems analyzed in this study. This finding replicates that of Khetarpal et al. (2010) and extends their account of spatial cognition as fine-grained. However, the nature of this universal preference in granularity is still unclear. What causes this cross-linguistic preference, and is it limited to the granularity of category systems or does it also reflect universals in how people group spatial notions? We now turn to these questions.

4.3 Efficiency as a source of universals in cognition

Previous work by Khetarpal et al. (2010) suggests that nonlinguistic categories are constrained by a universal preference, of unknown origin, for fine-grained sorting. What explains this preference, and patterns of pile-sorting more generally? To pursue this question, we use the efficient communication framework described in section 3.3, which has been used to demonstrate that categories across languages near-optimally partition referents within the domains of color, kinship, and natural kinds for efficient communication (Kemp & Regier, 2012; Regier et al., 2015). This previous work shows that categories in language are universally informative within a framework that nonetheless accommodates wide variation in semantic systems, and specifies a principle (simplicity) that constrains variation. Here, we evaluate whether this framework can also explain the universal tendencies and variation in pile sorting, making the parallel assumption that categories in thought are universally informative with constrained variation. The assumption that cognition respects the same efficiency pressures as language is supported by a previous account of color cognition, in which the authors "view memory as though it were a situation in which an individual communicates to himself through time using the brain as a channel" (Lantz & Stefflre, 1964). Following this treatment generally, and the efficient communication framework in particular, we create partitions of spatial scenes that are theoretically optimal in clustering together similar spatial relations and distinguishing different ones. We then compare the best such partition—our account of optimal categorization—to pile sorts by speakers of diverse languages to determine whether the universals in their sorting reflect our account of categorization as efficient groupings of spatial meaning.

Theoretically optimal partitions

Principles of efficient categorization, as explored in the previous chapter, define a theoretically optimal partition of the spatial stimuli into any specified number of categories, that is, the partition with the minimum communication cost for that number of categories. To create a single theoretically optimal partition as our comparison partition, we began by finding the optimal configuration for partitions varying in size from 2 to 16 categories. We considered partitions with up to 16 categories in order to exceed the median number of piles produced in sorting by Máíhɨki and Chichewa speakers, which was 14 piles.[7] For each number of categories, we

[7] Additionally, we observed that improvements in granularity (measured by height) as a function of partition size dropped off for partitions larger than about 6 categories.

generated five random partitions of the 71 TRPS scenes and computed their communicative cost as specified in the spatial relations study from section 3.5, with one difference, explained below.

Communicative cost requires a measure of similarity between category members and in section 3.5, we took the similarity between any two spatial relations stimuli to be determined by pile-sorting of those stimuli in Khetarpal et al.'s previous study (2010). In analyzing the categories of native English speakers for section 3.5, we defined the similarity of any two items in a category as the proportion of English-speaking participants who sorted the corresponding scenes into the same pile in Khetarpal et al.'s data. Here we seek an optimal partition of spatial scenes to assess pile sorts created by Máíhɨki and Chichewa speakers, so we instead combine pile sort data from English and Dutch as a rough proxy for cross-linguistic similarities. Accordingly, we take the similarity of any two items in a category to be the proportion of English- *and* Dutch-speaking participants who sorted the corresponding scenes into the same pile in Khetarpal et al.'s study (there was no normalization of data by language as this study included the same number of speakers for each).

Given the random partitions with 2 to 16 categories and a definition of the communicaitve cost associated with each, we then optimize each system of categories with respect to its cost. Communicative cost is a measure of expected error in using a category to pick out a referent, summed across all possible referents. Hence, more informative and precise systems of categories are assigned lower communicative costs. We created 75 optimized partitions (15 sizes x 5 random initializations), minimizing communicative cost by steepest descent. In doing so, we iteratively considered the category membership of each scene, reassigning scenes to the category that would produce the largest decrease in overall communicative cost for the partition. We considered all scenes in random order and repeated this steepest descent category reassignment procedure until no improvements to communicative cost were possible through the reassignment of any scene. Given the set of 75 optimized partitions, we then chose the partition with the lowest communicative cost for each number of categories, and took these 15 partitions to be the best spatial category systems for each number of terms.

From the set of 15 best category systems, we select a single system as our optimal partition of spatial relations. The efficient communication framework defines an "optimal frontier" across all systems with minimal communicative cost for their number of categories, so we select a single best system based on the system's silhouette (Rousseeuw, 1987). Silhouette is a metric for evaluating the validity of clustering in a partition, and assesses the fit of a partition as the average fit for each item to its assigned category, relative to the nearest alternative category. The (lack of) fit of item *i* to its assigned category *a* is defined as below:

$$a(i) = \sum_{j \in cat(a)} dissim(i, j) \qquad (2)$$

As in our measure of communicative cost, we take the similarity between spatial scenes *i* and *j* to be the frequency with which English- and Dutch-speaking participants sorted the corresponding scenes into the same pile in Khetarpal et al.'s (2010) experiment, such that the maximum similarity is 1 (where scenes *i* and *j* were always grouped together) and minimum is 0 (scenes *i* and *j* were never grouped together). Here, we take the dissimilarity between scenes *i* and *j* to be their co-sorting similarity subtracted from 1. Given $a(i)$ as the measure of dissimilarity between scene *i* and its assigned category *a*, we define $b(i)$ as the dissimilarity between scene *i* and the nearest alternative category, i.e., the category other than *a* for which the average dissimilarity $b(i)$ is lowest. With this, we compute the silhouette of item *i* as below:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \qquad (3)$$

The average silhouette $s(i)$ across all members of category a is a measure of how tightly clustered the category members are within the category, compared to the nearest alternative categories to which each member could have been assigned. Accordingly, the average $s(i)$ across all spatial scenes indicates how appropriately those scenes have been clustered by the partition under consideration. The silhouette for each of the 15 best n-category paritions is shown in Figure 4.4 below, plotted against the number of categories and granularity of the partition.

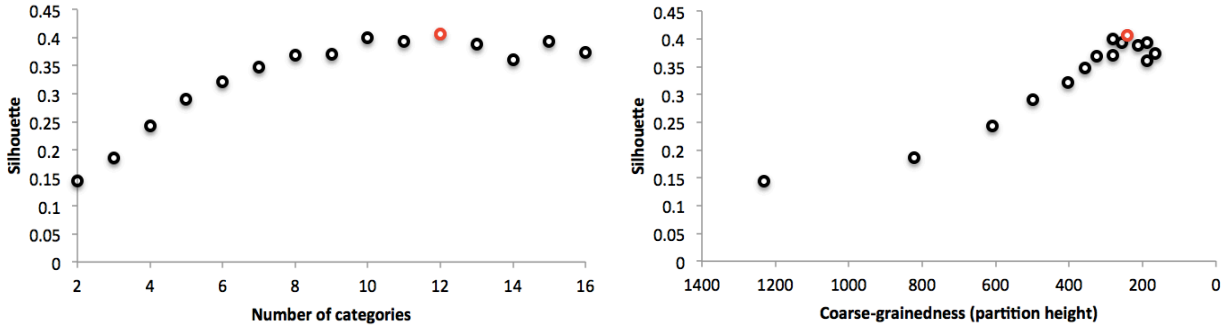

Figure 4.4: Comparison of clustering (silhouette) and complexity for optimized spatial category systems. Complexity is gauged as the number of categories in the left panel and as the granularity of the category system (i.e., height, the number of potential categorical distinctions that are not made) in the right panel. The partition with the best overall silhouette is plotted in red.

Comparing across the 15 best n-category partitions, we take the system with the most tightly clustered categories (i.e., the largest silhouette) to be the optimal system overall. Comparing silhouette scores and granularity, we find that this optimal 12-term spatial system and its neighboring systems cluster in the same range of granularity that speakers across languagues tend to produce in sorting spatial scenes. This observation is consistent with the idea that cross-language preferences for fine-grained sorting reflect the specific universal tendencies represented by our optimal partition. However, these results do not directly assess whether speakers' categories reflect the category structure of our optimal partition in addition to its granularity. To do so, we will measure the similarity between pile sort paritions and the optimal partition in terms of category membership using edit distance.

Edit distance dissimilarity between partitions

A pile sort of the full set of spatial stimuli is a partition of those stimuli into groups; the names a language applies to those stimuli are also a partition of those stimuli into groups. In order to determine the degree of dissimilarity between partitions, we used edit distance. The edit distance between two partitions, A and B, is defined as the minimum number of changes needed to transform partition A into partition B, where each change involves moving an element from one

group in partition A to another. Following Khetarpal et al. (2010), we used the Hungarian algorithm for bipartite graph matching (Deibel et al., 2005) to calculate the edit distance.

Efficient categorization as a cognitive universal

Does the granularity preference observed across languages stem from universals in nonlinguistic conceptual structure? If this is the case, and if our optimal partition captures these universals of conceptual structure, then we would expect that to the extent speakers sort unlike their native language, they sort similarly to our optimal partition.

To assess whether universals in nonlinguistic categorization resemble our optimal partition, we compare each pile sort's distance from the optimal partition against its distance from the categories of the sorter's native language. For this, we considered each of the pile sorts made by Máíhɨki and Chichewa speakers,[8] and computed the edit distance (as specified above) from the pile sort to (a) the modal terms of the sorter's native language and (b) our optimal partition. We then calculated a difference score for each pile sort by taking the distance to the sorter's native language and subtracting the distance to our optimal partition. Thus, the difference score is a measure of tendency toward our optimal partition over native language, for which a score of zero indicates that a pile sort was equidistant from the sorter's native language and the theoretically optimal partition. Accordingly, positive scores indicate greater similarity to the theoretically optimal partition identified above and negative indicate a bias toward the sorter's native language. These difference scores are plotted in Figure 4.5 below.

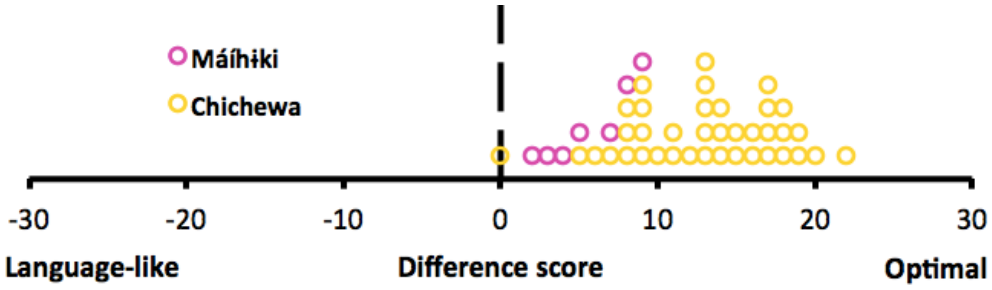

Figure 4.5: Difference scores indicating whether pile sort categorization is closer to our optimal partition or the categories of the sorter's native language.

With the exception of a single Chichewa speaker whose pile sort was equidistant to Chichewa and our optimal partition, all of the Máíhɨki and Chichewa speakers sorted in a way that was

[8] We test our account against the pile sorts made by speakers of Máíhɨki and Chichewa for several reasons. Our account of efficient categorization incorporates similarity that is derived from pile sorting, and we wish to avoid testing our optimal partition against the pile sorts (or sorts by speakers of the same languages) that were used to define similarity within it. We also seek to avoid lingusitic and cultural similarities between the sorters whose data is used to create and test our optimal partition, and because English and Dutch are related languages spoken in two very similar cultures, a clean comparison should avoid accounting for either on the basis of the other. Given these constraints, accounting for Máíhɨki and Chichewa pile sorting on the basis of English- and Dutch-derived similarity provides the most conservative test because (1) the semantic systems of Máíhɨki and Chichewa fall at the extremes of the granularity spectrum covered by our language sample, and (2) as the only language with finer granularity than our partition, Máíhɨki provides the only test case in which speakers can sort more finely than our optimal partition by reproducing the semantic distinctions of their native language.

more like our optimal partition than like the semantics of their native language. This finding demonstrates that the strong cross-linguistic tendencies we observe in both sorting granularity and language-like structure reflect our proposed optimal partition, suggesting that the category systems made in pile sorting follow efficient universal patterns of spatial meaning.

4.4 Discussion

Despite the great range in spatial naming represented across the four languages considered here—both in the granularity and the shape of their spatial categories—we found that speakers of all four languages sorted finely, in line with the predictions of Khetarpal et al. (2010). These results extend the previous account and further support their suggestion that spatial cognition, unlike spatial language, is universally fine-grained.

Furthermore, the success of our optimal partition demonstrates that these universal tendencies extend beyond the granularity of categories to conceptual structure, and that spatial cognition lies closer to these conceptual universals than to the semantics of one's native language. This universalist finding may have been expected for Chichewa, as the granularity analysis showed that Chichewa is a coarse-grained language whose speakers sort finely, and thus necessarily unlike their native language. Máíhɨki, however, provides a stronger test of our account; it is the only language in our sample that enables its speakers to fully indulge their granularity preferences and sort even more finely than our optimal partition simply by reproducing the categories of their language. The finding that Máíhɨki speakers sorted finely, and did so in a manner that more closely reflects the structure of our optimal partition than the semantics of their similarly fine-grained native language suggests that our account of efficient partitioning well characterizes the universal tendencies in cognition.

More broadly, the success of this optimal partition in characterizing the universal tendencies of pile sorts suggests that these universals reflect (and may result from) a process that parcels meanings by optimizing the trade-off between informativeness and simplicity (Kemp & Regier, 2012; Regier et al., 2015). These findings provide additional support for the view that nonlinguistic cognition, even in deviating from the structure of language, follows the same general principles of efficient communication that constrain meanings across languages (see Khetarpal et al., 2009 and 2013 for constraints on spatial semantics; c.f. Lantz & Stefflre (1964) on memory as communication directed to oneself).

Searching for linguistic relativity under the streetlights?

The cognitive universals explored here and in the previous chapters are responsible for the preponderance of behavior across languages in pile sorting spatial relations. In spite of their smaller explanatory power for spatial categorization, the role of language-specific forces in thought represents a consequential open question, which the next chapter will examine in more depth.

A natural next step in considering the role of language in spatial cognition would be to address the contribution of language to the spatial categorization task in this chapter, and a recent study does so. Tseng, Carstensen, Regier, & Xu (2016) reanalyze these data and produce a computational model of the trade-off between universal and language-aligned forces in categorization. They find a small but significant effect of native language in predicting how speakers categorize spatial scenes when nonlinguistic information is uncertain, i.e., when the

scene being sorted is similar to multiple categories. This model is grounded in Kay and Kempton's (1984) two-tiered account of cognition, and its results support the view that language is invoked as a strategy to resolve ambiguity in perceptual information on a universal, nonlinguistic tier of thought.

This recent work helps to characterize the role of language in spatial cognition and contributes to a growing literature on experimental paradigms and facets of cognition that show little to no effect of language. However, it's possible that such studies, and Tseng et al. in particular, underestimate the effects of linguistic relativity more generally. In keeping with Kay and Kempton's proposal, the model in Tseng et al. incorporates language in its predictions only when nonlinguistic information is ambiguous, providing no opportunity for language to influence many of the category assignment predictions. More broadly, the pile sorting tasks used in Tseng et al. and the preceding chapters produced small amounts of language-aligned pile sorting relative to universal sorting, so these datasets (and possibly the experimental paradigm of pile sorting) are limited in their potential to provide information about the role of language in spatial thought. To avoid the linguistic relativity equivalent of searching for one's keys where the streetlights are, the next chapter departs from topological spatial relations and will instead assess the role of language within spatial frames of reference, an area of spatial cognition that forms the basis for an ongoing debate about the role of language in spatial cognition.

Chapter 5

Characterizing the role of language in thought: Spatial frames of reference

The present and final study in this dissertation evaluates the role of language in spatial frames of reference. This represents a departure from the previous three chapters, which presented universalist accounts for cross-language phenomena concerning topological spatial relations. In contrast, spatial frames of reference are an area of cognition in which language strongly predicts nonlinguistic cognition (e.g., Brown & Levinson, 1993; Pederson, 1995; Levinson, 1996; Pederson et al., 1998; Levinson, 2003; Majid et al., 2004). Accordingly, frames of reference make for an ideal testbed with significant potential for identifying effects of linguistic relativity. In departing from the preceding themes, this work engages a high-profile debate and affords an opportunity to assess Kay and Kempton's (1984) two-tiered account of cognition. This chapter probes the role of language in online spatial reasoning, using linguistic interference to prevent participants from relying on language in solving a spatial task. In previous work, adult English speakers have been shown to use a spatial frame of reference that differs from that of nonhuman primates and toddlers (Haun et al., 2006), suggesting that learning the spatial frame of reference used in English may motivate a switch away from universal modes of spatial thought. We find that under linguistic interference, despite a sharp increase in error, adult English speakers fail to readopt the spatial frame of reference used by nonhuman primates and toddlers. This finding rules out the possibility that language affects spatial frames of reference online and accordingly argues against Kay and Kempton's account, which predicts a removable online role of language. This result raises the stakes of the debate over the role of language in nonlinguistic spatial frames of reference—either something other than language causes alignment between linguistic and nonlinguistic frames of reference, or language learning fundamentally restructures nonlinguistic spatial cognition in a way that is difficult to reverse.

5.1 Language and spatial frames of reference

Intuitions about spatial frames of reference differ profoundly across languages. Speakers of Dutch and Tzeltal (among others) systematically select opposite solutions to spatial questions, whether matching visual patterns, reconstructing an array, or tracing a route (Brown & Levinson, 1993; Pederson, 1995; Pederson et al., 1998; Levinson, 2003). These differences in nonlinguistic cognition closely resemble differences in spatial language—specifically, speakers' responses

tend to align with the spatial coordinate system preferred in their native language (ibid; for reviews of this work see Levinson, 1996 and Majid et al., 2004).

These coordinate systems, known as frames of reference (FoRs), vary in how they code the location or direction of objects in space, and fall into three general categories: self-based, object-based, and environment-based. The self-based, or egocentric, FoR depends on the viewer's perspective, and objects are located accordingly as being in front, behind, left, or right of other objects from this perspective, for instance "the fork is to the left of the plate". The object-based, or intrinsic, FoR codes object locations with respect to the axes or position of other landmark objects, for descriptions like "the fork is alongside the plate". The environment-based, or absolute, FoR references a fixed coordinate system, which is independent of the viewer's perspective or the axes of another object, such as "the fork is west of the plate". The egocentric FoR can be broadly contrasted with both intrinsic and absolute FoRs, which pattern similarly in that they emphasize either fixed coordinates or landmarks external to the viewer's perspective, and are correspondingly independent of the viewer; here we refer to both of these viewer-independent FoRs collectively as the allocentric FoR (Levinson, 1996).

Languages vary in which FoRs they use to encode spatial relations among small-scale objects in "table-top space." English speakers, for example, tend to use egocentric descriptions of locations at this scale, rarely producing sentences like "the fork is west of the plate". In contrast, Tzeltal speakers routinely describe even small-scale spatial relations in absolute terms (Levinson, 1994; Pederson et al., 1998).

Previous work has tested whether this difference in preferred linguistic FoRs is reflected in spatial thought outside of language by exploiting the differences between the egocentric and allocentric FoRs through rotation. Under a 180° rotation, the coordinates of an allocentric FoR are preserved, but the coordinates of an egocentric FoR are reversed.[9] In one such paradigm, the animals-in-a-row task (Levinson & Schmitt, 1993), participants are asked to study an array of ordered animal figures arranged on a table in a line facing the participant's left or right side. The participant is then rotated 180° and instructed to recreate the array they studied on a second table. In reproducing the array, speakers of languages like English and Dutch, who tend to describe objects at this scale in egocentric terms, usually rotate the array, reconstructing the row of animals relative to their body. This approach preserves the left-right mapping of animal order and orientation, such that the left- and rightmost animals are the same and the row continues to face left or right relative to the participant. In contrast, speakers of Tzeltal and other languages that prefer allocentric spatial descriptions characteristically translate the array between tables, preserving the order and orientation of the animals with respect to cardinal directions or other external coordinates or landmarks (Pederson et al., 1998).

Findings from this paradigm, and similar table turning tasks, have been widely replicated across languages (for a review of these studies, see Majid et al., 2004), and the alignment between nonlinguistic behavior and the dominant FoR of participants' languages is often taken as evidence for linguistic relativity (e.g., Levinson, 1996; Pederson et al., 1998; Levinson et al., 2002; but cf. Li & Gleitman, 2002). However, this relativist view is nuanced by findings from developmental and cross-species research, which reveal strong universals in spatial FoR across

[9] Here we contrast egocentric and allocentric FoRs, as behavior after a 180° rotation will appear identical under an absolute FoR with fixed coordinates and most intrinsic FoRs based on e.g., near or far objects, furniture, or the geometry of a room. Local landmarks, such as an identical statue that appears on both the study and reconstruction tables, can be used to discriminate between absolute, intrinsic, and relative FoRs after a 90° rotation (see Levinson et al., 2002), but are unnecessary in contrasting an egocentric FoR with allocentric strategies more broadly.

young children and nonhuman primates. Specifically, child speakers of egocentric languages tend to perform table turning tasks allocentrically, as do other primates, including orangutans, gorillas, bonobos, and chimpanzees (Haun et al., 2006). Similarly, young children learning egocentric languages tend to infer allocentric meanings for novel spatial words (Shusterman & Li, 2016). These findings, in combination with the cross-linguistic studies on adults, suggest that humans have a pre-linguistic cognitive bias toward allocentric FoRs, but that this component of spatial cognition is restructured or overridden for speakers of languages that privilege an egocentric FoR (Majid et al., 2004).[10] This implies that adult speakers of egocentric languages like English have transitioned from allocentric reasoning in childhood to egocentric reasoning as adults. Here, we seek to test whether this switch to the egocentric FoR in adult speakers of egocentric languages is caused by language in an online fashion.

If this shift in FoRs is driven by online use of language, then removing language should prevent or attenuate the use of the egocentric FoR in adult English speakers. In this study, we will test a possible mechanism for linguistic relativity by disrupting access to language during a table turning task.

5.2 Manifestations of linguistic relativity

In this experiment, we seek to test one of several accounts of the role of language in determining nonlinguistic frames of reference, specifically, that language influences cognition in an online way, prompting use of a specific FoR. This possibility lies in the middle of a spectrum of possible relations between language and cognition. On one end of this spectrum is the possibility that FoRs in language do not influence thought. By this account, even when language and cognition align, language is not the causal factor in cognition and some other influence explains any correlation between them. At the other end of the spectrum is the possibility that language has direct, enduring effects on cognition, to the point of overwriting alternative modes of thought to establish new ones. This view represents a strong version of linguistic relativity, or the Sapir-Whorf hypothesis, by which the language we speak determines our conceptualization of the world.

A more nuanced view of linguistic relativity falls in the middle of the spectrum, and holds that language influences thought, but that there are non-linguistic factors that shape it as well. Kay and Kempton (1984) advanced a classic version of this account, which reformulates the Sapir-Whorf hypothesis as a two-tiered view of cognition. On this view, language has two tiers:

> "one, a kind of rock-bottom, inescapable seeing-things-as-they are (or at least as human beings cannot help but see them) and a second, in which the metaphors implicit in the grammatical and lexical structures of language cause us to classify things in ways that could be otherwise (and are otherwise for speakers of different languages)" (p. 76).

[10] We note here that the pre-linguistic cognitive bias and cross-linguistically varying preferences in FoR that we describe reflect facility in learning and tendencies in nonlinguistic tasks, rather than hard limitations on cognition. Children and adults learn and use frames of reference in ways that deviate from the tendencies described, and do so in response to a range of contextual and communicative cues that are not discussed here (see e.g., Shusterman & Li, 2016 on learning and use of spatial terms and FoRs and Li & Gleitman, 2002 on flexible use of FoRs in adults).

By this account, universal cognitive tendencies on the first tier underlie language-driven and relativistic reasoning on the second tier, and language affects cognition when it is engaged online for reasoning. However, this online effect of linguistic relativity is not permanent—underlying nonlinguistic representations are still available, and can be revealed by temporarily disabling the second tier. Accordingly, this view holds that disruption of linguistic reasoning will lead people to fall back on the first tier, and respond according to a universal, cross-culturally shared mode of cognition. With respect to spatial frames of reference, this account predicts that if language shapes cognition in an online fashion, adult speakers of egocentric languages will fall back on the allocentric frame of reference shared across cultures by young children and across species by nonhuman primates.

A standard test of the two-tiered account of cognition removes this second tier by taking language offline through verbal interference tasks. In these experiments, a language-centric task is used to occupy the participant's verbal resources and thereby temporarily disable the use of language to reason about concurrent or interleaved tasks. If interfering with language changes the way a cognitive task is performed, then this suggests that language affects the typical performance through online recruitment. Verbal interference tasks have demonstrated an online role of language in mediating language-specific categorical perception of color (e.g., Kay & Kempton, 1984; Roberson & Davidoff, 2000; Gilbert et al., 2006; Winawer et al., 2007) and exact numerosity (Frank et al., 2008; Frank et al., 2012). Additionally, verbal interference has been shown to disrupt spatial reorientation, suggesting that language plays an online role in at least some types of spatial cognition (Hermer-Vazquez et al., 1999, but cf. Ratliff & Newcombe, 2008).

Here, we test whether reasoning with spatial frames of reference in a table turning task also depends on online access to language. The accounts we have reviewed provide three contrasting hypotheses but only two discriminable outcomes:

1. The non-language hypothesis: Language does not influence spatial FoR. Participants under verbal interference will perform the table turning task in line with their native language and comparably to those in the control condition. That is, English speakers will preferentially use the egocentric FoR, with no increase in allocentric FoR use under verbal interference.
2. The offline language hypothesis: Language influences spatial FoR in an enduring way which does not require online recruitment of linguistic resources. In adult English speakers, the linguistically preferred egocentric FoR has permanently replaced the pre-linguistic tendency toward the allocentric FoR. As above, participants under verbal interference will perform the table turning task in line with their native language and comparably to those in the control condition. That is, English speakers will preferentially use the egocentric FoR, with no increase in allocentric FoR use under verbal interference.
3. The online language hypothesis: Language influences spatial FoR in a temporary and removable way which requires online recruitment of linguistic resources. In adult English speakers, the linguistically preferred egocentric FoR dominates cognition that is mediated by language. However, verbal interference will remove this effect, leading English speakers to perform the table turning task in accordance with their pre-linguistic tendencies. Thus, English speakers under verbal interference will show an increase in use of the allocentric FoR relative to the control condition.

In the previous studies on color, number, and spatial reorientation, verbal interference disrupted typical language-specific behavior and produced a shift toward universal behavior, ruling out the non-language and offline hypotheses in support of an online effect of language on cognition. If we observe a similar shift from egocentric to allocentric responding in the verbal interference condition, this will demonstrate an online effect of language on cognition. Alternatively, if verbal interference fails to produce a shift from egocentric to allocentric responding, this result will rule out the online account, leaving the offline and non-language hypotheses, but (as in other domains) we would not be able to discriminate between them. However, this finding *would* indicate that the role of language in spatial frames of reference is unlike the cases of color, number, and spatial reorientation.

5.3 Linguistic interference and spatial frames of reference

This study tests whether egocentric responding in nonlinguistic tasks is mediated by online recruitment of linguistic resources. If English speakers' shift from allocentric spatial strategies in childhood to egocentric strategies in adulthood reflects the acquisition of a complementary, removable, language-driven mode of cognition, then we would expect linguistic interference to attenuate use of the egocentric FoR, prompting adults to respond following the allocentric FoR preferred by children and nonhuman primates.

Methods

To gauge nonlinguistic FoR, we used Levinson and Schmitt's (1993) animals-in-a-row task, which has been used extensively across languages. The relationship between language and nonlinguistic FoR is a topic of active debate, about which there are many conflicting views (compare e.g., Li & Gleitman (2002) to Levinson et al. in the same issue). For this reason, we sought a well-controlled design for verbal interference in order to equate the difficulty of our interference task across participants and thereby maximize the interpretability of our results. Accordingly, we decided to use a tailored verbal interference task to tax each participant according to their capacity. This paradigm, established by Frank et al. (2012) uses an adaptive staircase procedure to determine the appropriate interference difficulty for each participant and customize their task accordingly. We take it as an unofficial gold standard for verbal interference paradigms.

Participants

Forty-one undergraduate students at UC Berkeley were recruited on campus and took part in the study in exchange for course credit or $10. All participants were native speakers of English who had learned English by age 4 (although a number were bilingual), and were naïve to the research hypothesis and related findings. One participant was excluded from the verbal interference condition due to experimenter error, leaving 20 participants in that condition, and 20 in the control condition with no verbal interference.

Testing room

The experiment was conducted in a small rectangular room, with a door on the west wall, window (with blinds closed) on the east wall, bookshelf on the north wall, and whiteboard on the south wall. A desk and computer in the southeast corner of the room were used to run the verbal interference staircase task before the FoR task began. Two identical tables were placed parallel to each other, 4.5 feet apart against the north and south walls, and stimuli were arranged on these tables along the east-west axis. Participants studied the stimulus array on one table and then turned 180° and walked to the second table to reconstruct the array, with the study and recall tables counterbalanced between subjects.

Materials

The stimuli were four rubber animals: a cow, a duck, a fish, and a sheep (pictured in Figure 5.1). The animals were roughly 2 inches tall, 4 inches long, and 1.5 inches wide. Each animal was a different color and shape, but all were symmetrical along their head-to-tail axis.

Procedures

Participants in both conditions were tested individually. The FoR task was a slightly modified version of the Levinson and Schmitt animals-in-a-row task (1993), with 5 training trials and 5 test trials. This task was adapted from Levinson et al. (2002), Pederson et al. (1998), and Li and Gleitman (2002). The verbal interference (VI) task was adapted from Frank et al. (2012) and required participants to remember a different sequence of consonants while performing each trial of the animals-in-a-row task.

Verbal interference staircase

In both the VI and control conditions, participants began by completing 60 trials of an adaptive staircase to match the difficulty of the verbal interference based on each participant's ability. For participants in the VI condition, the goal of the staircase task was to determine how many consonants they could retain in memory while performing at a constant level on both the consonant recall and a concurrent task; control participants completed the staircase task to equate fatigue across conditions. This staircase procedure was identical, except as noted below, to that of the VI staircase in Frank et al. (2012; see Experiment 3 for a complete description of the procedures). Participants were seated at a desk and presented with a string of consonants, which flashed on the screen one at a time for 200ms with 100ms between consonants. The first string contained 2 consonants, and participants were instructed to remember the letters they had seen (and constantly repeat them aloud, unlike Frank et al.) while they completed a visual search task. The visual search task consisted of an array with 24 letters scattered around the screen. In half of the trials, all letters were capital T, and in the other half of the trials, there was a single capital L amongst the Ts. Participants pressed a button to indicate whether an L was present and on the next screen they typed the consonants they had been instructed to keep in memory. Consonant strings were scored as correct as long as they contained the correct set of consonants, regardless of order. Participants were given feedback on their performance in both the consonant recall and visual search. A unique consonant sequence was used on every trial and the length of the

consonant string was incremented by 1 after every two trials with correct performance for both the consonant recall and visual search. The number of consonants was decreased by 1 after every trial in which an error was made on one or both of the tasks. The structure of the staircase task was described to participants before they began and they were encouraged to do as well as possible. The number of consonants in the interference task was set to the average number of consonants for the last 25 trials of the staircase, rounded up.

Animals-in-a-row task

After completing the verbal interference staircase, participants moved on to the animals-in-a-row task; the only difference between the control and VI conditions was that participants in the VI condition continued to do the consonant recall task concurrently with animals-in-a-row, while those in the control condition completed the animals-in-a-row trials without the consonant recall task.

For the animals-in-a-row task, participants were led to one of the identical tables along the north and south walls, and instructed to "stand like this", facing one of the tables.[11] Two experimenters were present, and always stood 1-2 feet past the ends of the tables on either side of the participant (to the east and west), and moved symmetrically, standing behind the participant and out of their line of sight when they faced either table.[12] The first experimenter provided instructions for the table turning task and arranged the animals while the second timed task intervals and recorded responses. The first experimenter began the first training trial by placing 3 of the 4 animals in the center of the table in front of the participant, creating a line in which all animals faced either to the right or left from the participant's perspective (as shown in Figure 5.1), and asked the participant to study the animals. After a study interval of 5 seconds, the experimenter removed the animals and held them for a delay interval of 15 seconds while standing to the side of the table. After the delay, the first experimenter placed all four animals in a pile in the center of the same table and instructed the participant to make the array again "just the same." When the participant finished, the experimenter corrected any mistakes in which animals were used, the order they appeared in, and the direction they faced by physically re-arranging the animals and saying "the correct is…like this." The experimenter then removed all of the animals, saying "we're going to do that all again," and repeated this procedure for an additional training trial with a different array of animals. In the control condition, training proceeded like this for a total of 5 trials, but in the VI condition, the experimenter reintroduced the consonant recall task on the third training trial. For this, the second experimenter stood in front of the stimulus presentation table and held a tablet to present the consonant strings with the same timing as in the staircase task. The participant was again instructed to remember the string of letters while continuously repeating them aloud, and after two practice trials repeating the letters alone, the consonant recall task was given concurrently with each trial of the animals-in-a-row task. Accordingly, each trial in the VI condition consisted of the presentation of a consonant string, a 5-second study interval, a 15-second delay, reconstruction of the animal array, and a

[11] Throughout the task, the experimenters avoided using spatial language by demonstrating positions and orientations, using deictic terms (like "here"), and pointing to indicate spatial meanings as needed.

[12] The experimenters were always out of view when the participant faced a table, but still may have provided salient landmarks. If so, then they would have patterned along with the other features of the testing room (including the door, window with blinds closed, furniture, and rectangular geometry of the room), as absolute or intrinsic landmarks, as they maintained their positions in absolute coordinates throughout the table turning task.

final repetition of the consonant string. Participants repeated the consonant string continuously during the trial and their VI accuracy was scored based on their final repetition after reconstructing the animal array.

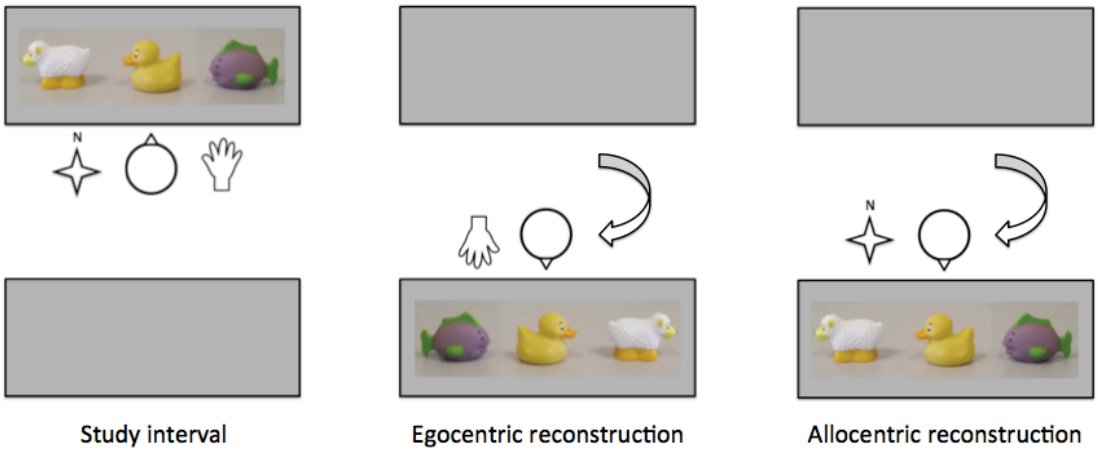

Figure 5.1: Animals-in-a-row task setup. The left panel shows the study interval at the stimulus presentation table. The middle and right panels depict egocentric and allocentric reconstructions of the animal array after the 180° rotation to the recall table.

After the 5 training trials conducted at the stimulus presentation table, participants in both conditions completed 5 test trials at the second table after a rotation. The transition to test trials was minimally marked by the experimenter explaining that they would now do something "very similar" and affirming that the participant should "do everything else just the same." Participants in the VI condition were then presented with consonant strings, and in both conditions, the animals were again arranged on the stimulus presentation table for a 5-second study period and removed for a 15-second delay. In the last 5 seconds of this delay, the experimenter gently turned the participant by their shoulders 180° to face the second table, where they placed all four animals in a pile at the center of the table. Participants reconstructed the row of animals on the second table (while continuing to repeat the consonant string in the VI condition), and when finished, were instructed to turn back to the stimulus presentation table. This procedure was repeated 4 more times for a total of 5 test trials during which the experimenter did not provide feedback on the participant's responses. The order, direction, and identity of the animals were randomized across the training and test trials; the sequence of randomized arrays was the same for all participants.

Response coding

Reconstructions of the animal array in test trials were coded as egocentric, allocentric, or error. For a response to be coded as egocentric or allocentric, the direction, order, and identity of the animals had to be correct according to that frame of reference (i.e., rotationally correct for the egocentric FoR and translationally correct for the allocentric FoR). Responses were coded as errors if there was an error in the order or identity of the animals used, or if animal order and facing direction were consistent with differing FoRs (e.g., the animals furthest to the west and

east remained the same but were arranged facing west when before they had faced east). In the VI condition, only test trials in which the consonant string was reproduced correctly were included in the analysis, resulting in the exclusion of 33 test trials out of 100 total test trials across the 20 participants.

Results

In this experiment, we use verbal interference to test whether language plays an online role in English speakers' preferential use of the egocentric FoR. If language has an *online* effect on spatial cognition, in which adult English speakers retain an underlying cross-cultural and cross-species preference for the allocentric FoR, then verbal interference should: 1) reduce participants' reliance on the egocentric FoR, and 2) reveal the underlying bias for the allocentric FoR. Consequently, participants in the VI condition should produce fewer egocentric and more allocentric responses in the animals-in-a-row task.

However, if language permanently alters the preferred cognitive FoR (the offline language hypothesis), or if language is not the cause of varying preferences in cognitive FoRs across cultures (the non-language hypothesis), then we should see preferential use of the egocentric FoR across conditions, with no increase in allocentric FoR use under verbal interference.

Figure 5.2 below compares the proportions of egocentric, allocentric, and error responses across the control and VI conditions. Participants in the VI condition produced significantly fewer egocentric responses (Mann-Whitney $U = 45.5$, $p < .0001$ two-tailed), but significantly more errors (Mann-Whitney $U = 75.5$, $p = .0008$ two-tailed) than those in the control condition. There was no significant difference in allocentric responses across conditions (Mann-Whitney $U = 178.5$, $p = .5687$ two-tailed).[13]

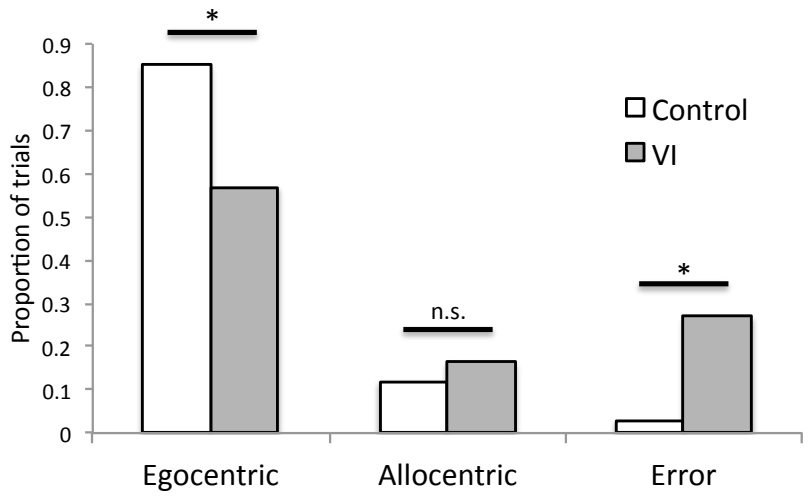

Figure 5.2: Frequency of each FoR with and without verbal interference.

[13] The *p*-values reported here are not corrected for multiple comparisons. However, both significant *p*-values remain significant at $\alpha = .01$ with Bonferroni correction.

In summary, we found that verbal interference *did* affect participants' behavior, but not as predicted by the online language hypothesis. Under verbal interference, participants produced fewer egocentric responses (as predicted by the online hypothesis), but did not make more allocentric responses (against prediction), and instead, there was an increase in error. This pattern of results is inconsistent with an online effect of language on nonlinguistic FoR, but consistent with either of the alternative accounts, in which language has either an offline effect or no effect on spatial FoR in nonlinguistic tasks.

Manipulation check and verbal shadowing replication

A possible concern is that our verbal interference task failed to disrupt participants' use of language in the VI condition. If our VI taxed participants' cognitive resources generally but not their verbal resources in particular, then we would expect, as we found, to see an increase in error (due to the task demands) but continued use of the egocentric FoR (because participants could still use language). A straightforward way to address this possibility would be to include an additional interference condition that is matched in difficulty to our verbal interference, but nonverbal in nature. Frank et al. (2012) developed such a task, but their nonverbal interference taxes spatial memory, and so would not provide a clean comparison given the spatial nature of our FoR task. However, the results of Frank et al. (2012) in the domain of number can be seen as a manipulation check demonstrating that their VI task (which we reproduced) does interfere with linguistic processes to a greater extent than (their) nonverbal interference of matched difficulty.

As an additional test, we replicated our study using a verbal shadowing paradigm in which participants continuously listened to and repeated speech while performing the table turning task, and obtained comparable results. As a manipulation check for this verbal shadowing task, we tested participants' ability to recall a 9-digit number presented while they engaged in shadowing and found that the number sequences produced by shadowing participants were significantly less similar (in edit distance) to the target sequence than responses produced by control participants who were not shadowing.[14]

5.4 Discussion

This study evaluates a possible role of language in nonlinguistic cognition involving spatial frames of reference. Cross-linguistic research has previously shown that people who speak languages with differing spatial FoRs tend to solve nonlinguistic spatial tasks differently, according to the preferred spatial FoR of their native language (e.g., Majid et al., 2004). At the same time, other work has demonstrated strong universal tendencies toward the allocentric FoR in young children and nonhuman primates (e.g., Haun et al., 2006; Shusterman & Li, 2016). This study tests a possible resolution of these findings, in the form of a proposal by Kay and Kempton (1984), which characterizes human cognition as having two tiers, the first of which is pre-linguistic, driven by perception, and universally shared across human cultures and languages, whereas the second tier is shaped by the grammar and lexicon of the specific language that we

[14] We further replicated our main finding using the staircase VI task in two additional studies: a pilot study with slightly different parameters in the staircase task, and a second study using a 3-animal variant of the animals-in-a-row task, in which participants are not required to remember the identity of the animals (see Levinson et al., 2002 and Li & Gleitman, 2002 for other uses and discussion of this task variant). Both studies included similar numbers of participants and produced comparable results to those reported in the main text.

speak. Previous support for this account has shown that verbal interference produces a shift from culturally variable, language-aligned responding to universal patterns shared across languages (e.g., Winawer et al., 2007; Frank et al., 2012).

Following this tradition, we tested whether English speakers under verbal interference shift away from their language's preferred egocentric FoR toward the allocentric FoR preferred by young children and nonhuman primates. Despite a sharp increase in error under linguistic interference, we found no evidence of a switch toward allocentric strategies. This finding in the spatial domain deviates from analogous studies in color and number, where linguistic interference produced a tendency toward universal response patterns. In contrast, our results suggest that spatial FoR does not follow the two-tiered account of cognition, and English speakers' use of the egocentric FoR does not represent an online effect of language.

Accordingly, our study suggests that the primate bias toward the allocentric FoR is overridden by something other than online influences from language, but does not discriminate between the two remaining possibilities. This finding raises the stakes of the debate over the role of language in nonlinguistic spatial frames of reference—either something other than language causes alignment between linguistic and nonlinguistic FoRs, or language restructures nonlinguistic cognition in a way that is difficult to reverse.

Chapter 6

Conclusions

6.1 Findings and implications

Why do languages parcel human experience into categories in the ways they do, and to what extent do these categories in language shape our view of the world? This dissertation examines these questions in the domain of spatial cognition, drawing on approaches from the areas of semantic typology, language evolution, and linguistic relativity. The evidence considered here demonstrates that universal cognitive and communicative pressures shape the constrained variation observed across languages, that the semantic structures in language and general principles of communication provide insight in characterizing the nature of thought, and finally, that our conceptions of the world can closely parallel the varied structures in language even when language does not play an active role in thought.

The first three studies in this dissertation characterize strong cognitive universals observed across languages in the spatial domain. The first of these tests a typological account of conceptual structure derived from cross-linguistic studies of spatial semantics. I find support for this hierarchical model of spatial relational notions in two pile sorting experiments where participants successively subdivide spatial stimuli. These findings (1) demonstrate that individuals respect a hierarchy of spatial notions that has been proposed to guide language change over time, (2) provide evidence for a structured model of spatial thought, and (3) establish an explicit account of universal conceptual space in this domain.

The second study characterizing universals, in Chapter 3, examines the evolutionary pressures that shape linguistic universals in spatial semantics, taking up a challenge from Stephen Levinson (2012) to explain how the highly informative semantic systems in language arise. Through simulated language learning and transmission in the lab, I show (1) that human reproductions of random partitions of color and spatial relations reflect learning biases consistent with domain-general principles of efficient communication, and (2) that these biases are accompanied by a convergence toward the semantic structure of language in these domains. These findings provide convergent support for universals in spatial cognition, which are gauged explicitly in Chapter 2 via sorting and implicitly in this chapter through a measure of bias in learning. Further, the close relationship between communicative efficiency and language-like structure reinforce the findings of Khetarpal et al. (2009; 2013), which show that general principles of communication account for the constrained diversity observed in spatial semantic systems across languages. They also suggest that this framework may explain universals and variation in stages of language evolution (such as those along the proposed diachronic hierarchy

in Chapter 2) and in spatial cognition, which was argued in Chapter 2 to underlie this linguistic diversity.

Chapter 4 builds on the conclusions of the preceding chapter and related work on cross-language semantics in color (Regier et al., 2007; 2015), kinship (Kemp & Regier, 2012), and number (Xu & Regier, 2014), extending the framework of efficient communication beyond semantic structure to explain nonlinguistic cognition. In previous work, Khetarpal et al. (2010) identified strong universal tendencies in pile sorts of spatial scenes by Dutch and English speakers, which I reproduce in two unrelated languages, Máíhiki and Chichewa, showing that speakers of all four languages sort with the same fine granularity despite large variation in the granularity of their semantic systems. I propose an account for these universal tendencies in terms of general principles from the efficient communication framework, characterizing pile sorts as optimizing a trade-off between informativeness (making for fine-grained and intuitively organized spatial categories) and simplicity (limiting the number of categories). I show that this account explains universal tendencies in the granularity of pile sorting and captures universals in how spatial meanings are categorized. In doing so, the account provides a domain-general characterization of universals and variation in spatial cognition. Moreover, these results suggest that cognitive universals reflect a trade-off between informativeness and simplicity, parallel to the account of semantics proposed by Kemp and Regier (2012). More broadly, the finding that cognition follows the same principles that constrain meanings across languages supports the suggestion of Lantz and Stefflre (1964), by which memory can be seen as self-directed communication. While the pile sorting task in this study does not directly address memory, it may engage a similarly categorical process by requiring participants to track the range of spatial relations in a large set of stimuli, and the meanings of each group they make as they organize and update their piles. On this view, by which pile sorting engages several subtasks that tax memory, the tendency toward fine-grained—and thus more complex—pile sorting is particularly surprising. This preference for fine granularity may indicate that in the case of cognition, the trade-off between informativeness and simplicity favors informativeness to a greater extent; that is, cognition may be biased toward informativeness where language is biased toward simplicity.

Taken together, the studies in Chapters 2-4 provide evidence for a universal conceptual space underlying semantic universals; demonstrate a domain-general process by which principles of efficiency may shape informative semantic systems; and extend this account of efficient semantics to explain universals in cognition. Within the domain of space, they build on earlier work to more comprehensively describe universals in language and thought, validating a universal conceptual hierarchy of spatial notions, revealing biases in spatial category learning, and identifying universals in the granularity and structure of pile sort categories across languages. These findings characterize and expand upon the nature of cognitive universals in the spatial domain and others.

Much of this dissertation considers cognition through language, which serves as a window onto thought. In the final study, Chapter 5, I ask whether language also determines thought. For this, I consider spatial frames of reference in cognition, a case in which the relation between language and thought is contested (e.g., Levinson et al., 2002; Li & Gleitman, 2002). Work in this domain has suggested that there is a cross-cultural and cross-species bias toward the allocentric FoR in primate cognition (Haun et al., 2006), but English-speaking adults deviate from this cognitive universal, preferring the egocentric FoR used pervasively in their language (Li & Gleitman, 2002). Under analogous circumstances in the cases of color (e.g., Kay & Kempton, 1984), number (Frank et al., 2008), and spatial reorientation (Hermer-Vazquez et al.,

1999), linguistic interference has been shown to induce a switch from language-specific to universal modes of thought. Chapter 5 tests for a parallel online effect of language in spatial FoRs, and finds instead that linguistic interference disrupts egocentric responding without producing a shift toward the hypothetically underlying allocentric FoR. This result demonstrates that language does not affect spatial FoR in an online way, challenging the generality of Kay and Kempton's (1984) Whorfian proposal, by which language constitutes a complementary and removable tier of cognition. Moreover, this finding contrasts with the accounts of color, number, and spatial reorientation, and correspondingly demonstrates that the relation between language and thought can vary across areas of cognition and subdomains of spatial cognition. This finding raises the stakes of the debate: either something other than languages motivates the use of language-specific FoRs in cognition, or spatial language fundamentally restructures spatial cognition in a way that is difficult to reverse experimentally.

As a whole, the body of work in this dissertation extends an emerging consensus by which all people share a universal conceptual foundation that may be altered by language. The research here further elaborates this account in the domain of spatial cognition, suggesting the following view of the relation between spatial language and thought: speakers of all languages share universals in conceptual structure, which underlie the universal tendencies in language (Chapter 2). These cognitive universals correspond to biases in language learning that act as selective pressures in language evolution. In particular, these pressures favor efficient categories, producing semantic systems that are near-optimally informative (Chapter 3). Moreover, efficiency acts as an organizing principle in both language and cognition, accounting for universals and variation in systems of categories across languages (Khetarpal et al., 2010; Carstensen et al., under revision) and in nonlinguistic systems of meaning more broadly (Chapter 4). However, even in cases where language and cognition align closely across cultures, language-like reasoning can occur without online access to language, challenging the view that language actively influences cognition during thought.

6.2 Concluding remarks

Much remains to be determined about the nature of both language-specific and universal forces in cognition, and their interaction. However, in recent years, the oscillation between universalist and relativist accounts of cognition has become less dramatic and the debate has instead gained interesting complexity and nuance. The work presented here represents additional steps in moving this debate toward an equilibrium, in which fascinating questions receive grounded and broadly informed answers.

The universals observed in this dissertation are characterized by accounts that emphasize broad functional considerations in explaining the structure of language and thought. These findings underscore the importance of pressures external to the mind (from communication), constraints from basic cognitive processes (biases and limitations in learning and memory), and general principles of efficiency (simplicity and informativeness) in shaping both language and cognition across cultures.

The findings of this dissertation in the domain of space, taken together with parallels in color, number, and other domains, reinforce an emerging consensus on the relation of language and thought, by which all people share a universal conceptual foundation that may be altered by language. The research here further elaborates this account, suggesting that universals and variation in both language and thought may derive to some extent from general principles of

efficiency. At the same time, it challenges the generality of a classic (Kay & Kempton, 1984) formulation of this view, motivating future research. In both complementing and challenging an emerging consensus on language and thought, this dissertation informs our view of language, a defining feature of human cognition, and contributes to a more complete understanding of the nature of thought.

References

Abbott, J. T., Griffiths, T. L., & Regier, T. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences*, *113*(40), 11178-11183.

Baddeley, R., and Attewell, D. (2009). The relationship between language and the environment: Information theory shows why we have only three lightness terms. *Psychological Science*, *20*(9), 1100-1107.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327-18332.

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

Boroditsky, L. & Gaby, A. (2010). Remembrances of times East: Absolute spatial representations of time in an Australian aboriginal community. *Psychological Science, 21*(11), 1635-1639.

Boster, J. (1986). Can individuals recapitulate the evolutionary development of color lexicons? *Ethnology*, 25(1), 61-74.

Bowerman, M., & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 387-427). Cambridge, MA: MIT Press.

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

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology*, *49*(3), 454.

Brown, P. & Levinson, S. (1993). Linguistic and nonlinguistic coding of spatial arrays: Explorations in Mayan cognition. Working paper 24.

Butterworth, B., Reeve, R., Reynolds, F., & Lloyd, D. (2008). Numerical thought with and without words: Evidence from indigenous Australian children. *Proceedings of the National Academy of Sciences, 105,* 13179-13184.

Carstensen, A. (2011). Universals and variation in spatial language and cognition: Evidence from Chichewa. Undergraduate thesis, University of California, Berkeley.

Carstensen, A., Khetarpal, N., Majid, A., Malt, B., Sloman, S., & Regier, T. (under revision). Cross-language universals and variation in cognition: The cases of space and artifacts.

Carstensen, A., Neveu, G., Michael, L., & Regier, T. (2013). Thinking in ways we don't speak: Evidence for a universal preference in semantic granularity [Abstract]. In M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 3893). Austin, TX: Cognitive Science Society.

Carstensen, A., & Regier, T. (2013). Individuals recapitulate the proposed evolutionary development of spatial lexicons. In M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 293-298). Austin, TX: Cognitive Science Society.

Carstensen, A., Xu, J., Smith, C.T., & Regier, T. (2015). Language evolution in the lab tends toward informative communication. In D. Noelle et al. (Eds.), *Proceedings of the 37th*

Annual Meeting of the Cognitive Science Society (pp. 303-308). Austin, TX: Cognitive Science Society.

Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf hypothesis and probabilistic inference: Evidence from the domain of color. *PLOS ONE 11(7):* e0158725.

Cook, R., Kay, P., & Regier, T. (2005). The World Color Survey database: History and use. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 223-242). Elsevier.

Coxon, A. (1999). *Sorting data: Collection and analysis*. Thousand Oaks, CA: Sage Publications.

Croft, W. (2003). *Typology and universals: Second edition*. Cambridge, UK: Cambridge University Press.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature, 398,* 203-204.

Deibel, K., Anderson, R., & Anderson, R. (2005). Using edit distance to analyze card sorts. *Expert Systems, 22*, 129-138.

Dougherty, J. (1977). Color categorization in West Futunese: Variability and change. In M. Sanchez (Ed.), *Sociocultural dimensions of language change*. New York: Academic.

Dowman, M. (2007). Explaining color term typology with an evolutionary model. *Cognitive Science*, 31(1), 99-132.

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.

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.

Frank, M., Fedorenko, E., & Gibson, E. (2008). Language as a cognitive technology: English-speakers match like Pirahã when you don't let them count. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 439-444). Austin, TX: Cognitive Science Society.

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

Frank, M., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology*, 64(1), 74-92.

Gilbert, A., Regier, T., Kay, P., & Ivry, R. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*(2), 489-494.

Gilbert, A., Regier, T., Kay, P., & Ivry, R. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and language*, *105*(2), 91-98.

Haspelmath, M., Dryer, M.S., Gil, D., & Comrie, B (Eds). (2005). *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Haun, D.B., Rapold, C.J., Call, J., Janzen, G., & Levinson, S.C. (2006). Cognitive cladistics and cultural override in Hominid spatial cognition. *Proceedings of the National Academy of Sciences*, *103*(46), 17568-17573.

Heider, E. (1972). Probabilities, sampling, and ethnographic method: The case of Dani colour names. *Man*, New Series, 7, 448-466.

Heider, E. R., & Olivier, D. C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, *3*(2), 337-354.

Hermer, L., & Spelke, E.S. (1994). A geometric process for spatial reorientation in young children. *Nature, 370*, 57-59.

Hermer-Vazquez, L., & Spelke, E.S. (1996). Modularity and development: The case of spatial reorientation. *Cognition*, *61*, 195-232.

Hermer-Vazquez, L., Spelke, E.S., & Katsnelson, A. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, 39(1), 3-36.

Hespos, S.J. & Spelke, E.S. (2004). Conceptual precursors to language. *Nature, 430*, 453-456.

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

Holmes, K.J., & Regier, T. (2016). Categorical perception beyond the basic level: The case of warm and cool colors. *Cognitive Science*.

Hubert, L. & Golledge, R. (1981). A heuristic method for the comparison of related structures. *Journal of Mathematical Psychology*, 23(3), 214-226.

f (pp. 295–319). Cambridge, UK: Cambridge University Press.

Kalish, M.L., Griffiths, T.L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review 14:* 288-294.

Kay, P. (1975). Synchronic variability and diachronic change in basic color terms. *Language in Society*, 4, 257-270.

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

Kay, P. & McDaniel, C. (1978). The linguistic significance of the meanings of basic color terms. *Language*, 54, 610-646.

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

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 Meeting of the Cognitive Science Society* (pp. 358-363). Austin, TX: Cognitive Science Society.

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

Khetarpal, N., Neveu, G., Majid, A., Michael, L., & Regier, T. (2013). Spatial terms across languages support near-optimal communication: Evidence from Peruvian Amazonia, and computational analyses. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 764-769). Austin, TX: Cognitive Science Society.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681-10686.

Kirby, S. (2002). Learning, bottlenecks, and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge: Cambridge University Press.

Kuschel, R., & Monberg, T. (1974). 'We don't talk much about colour here': A study of colour semantics on Bellona Island. *Man*, *9*(2), 213-242.

Lantz, D. & Stefflre, V. (1964). Language and cognition revisited. *The Journal of Abnormal and Social Psychology*, 69(5), 472-481.

Levinson, S. (1994). Vision, shape, and linguistic description: Tzeltal body-part terminology and object description. *Linguistics*, 32, 791-856.

Levinson, S. C. (1996). Frames of reference and Molyneux's question: Cross-linguistic evidence. In P. Bloom, M. Peterson, L. Nadel & M. Garrett (Eds.), *Language and space* (pp. 109-169). Cambridge, MA: MIT Press.

Levinson, S. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge University Press.

Levinson, S. (2012). Kinship and human thought. *Science*, 336(6084), 988-989.

Levinson, S., Kita, S., Haun, D., & Rasch, B. (2002). Returning the tables: Language affects spatial reasoning. *Cognition*, 84(2), 155-188.

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

Levinson, C., & Schmitt, B. (1993). Animals in a row. In *Cognition and Space Kit Version 1.0* (pp. 65-69).

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*(3), 265-294.

Lindsey, D.T., & Brown, A.M. (2006). Universality of color names. *Proceedings of the National Academy of Sciences*, *103*(44), 16608-16613.

Lindsey, D.T., & Brown, A.M. (2009). World Color Survey color naming reveals universal motifs and their within-language diversity. *Proceedings of the National Academy of Sciences*, *106*(47), 19785-19790.

Lucy, J. (1992). *Grammatical categories and cognition*. Cambridge: Cambridge University Press.

Meilă, M. (2007). Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, 98(5), 873-895.

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

Newcombe, N. S. & Ratliff, K. R. (2007). Explaining the development of spatial reorientation: Modularity-plus-language versus the emergence of adaptive combination. In J. M. Plumert & J. P. Spencer (Eds.), *The emerging spatial mind* (pp. 53-76). New York, NY: Oxford University Press.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language*, 557-589.

Piantadosi, S., Tily, H., and Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526-3529.

Pick, H & Acredolo, L. (1983). *Spatial orientation: Theory, research, and application*. New York, NY: Plenum Press.

Ratliff, K. & Newcombe, N. (2008). Is language necessary for human spatial reorientation? Reconsidering evidence from dual task paradigms. *Cognitive Psychology*, 56(2), 142-163.

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLOS ONE*, *11*(4), e0151138.

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., Kay, P., & Khetarpal, N., (2009). Color naming and the shape of color space. *Language*, 85, 884-892.

Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence* (pp. 237-263). Wiley-Blackwell.

Roberson, D. & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, *28*(6), 977-986.

Roberson, D., Davidoff, J., Davies, I.R.L., & Shapiro, L.R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50*, 378-411.

Roberson, D., Davies, I., & 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., 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*(2), 752-762.

Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

Russell, B. (1948). Human knowledge: Its scope and limits (pp. 60). New York: Simon and Schuster.

Shepard, R.N. & Cooper, L.A. (1986). *Mental images and their transformations*. The MIT Press, Cambridge.

Shepard, R.N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.

Shusterman, A., & Li, P. (2016). Frames of reference in spatial language acquisition. *Cognitive Psychology*, *88*, 115-161.

Silvey, C., Kirby, S., & Smith, K. (2015). Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive Science*, 39(1), 212-226.

Talmy, L. (1983). How language structures space. In Herbert L. Pick, Jr. & Linda P. Acredolo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225-282). New York: Plenum Press.

Taylor, C., Clifford, A., & Franklin, A. (2013). Color preferences are not universal. Journal of Experimental Psychology: General. 142(4), 1015-1027.

Tolman, E. C., & Honzik, C. H. (1930). "Insight" in rats. University of California, Publications in Psychology, 4, 215-232.

Tseng, C., Carstensen, A., Regier, T., & Xu, Y. (2016). A computational investigation of the Sapir-Whorf hypothesis: The case of spatial relations. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 2231-2236). Austin, TX: Cognitive Science Society.

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

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*(19), 7780-7785.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Routledge & Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations / Philosophische Untersuchungen*. John Wiley & Sons.

Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences*, 14, 138-146.

Wyszecki, G. & Stiles, W. S. (1967). *Color Science*. New York, NY: Wiley.

Xu, J., Dowman, M., & Griffiths, T. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1758), 20123073.

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.

Xu, Y. & Regier, T. (2014). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1802-1807). Austin, TX: Cognitive Science Society.

Zwaan, R. (2003). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of learning and motivation*, 44, 35-62.