

edited by Paul Bloom, Mary A. Peterson, Lynn Nadel, and Merrill F. Garrett



Language *and* Space

Language, Speech, and Communication

Statistical Language Learning, Eugene Charniak, 1994

The Development of Speech Perception, edited by Judith Goodman and Howard C. Nusbaum, 1994

Construal, Lyn Frazier and Charles Clifton, Jr., 1995

The Generative Lexicon, James Pustejovsky, 1996

The Origins of Grammar: Evidence from Early Language Comprehension, Kathy Hirsh-Pasek and Roberta Michnick Golinkoff, 1996

Language and Space, edited by Paul Bloom, Mary A. Peterson, Merrill F. Garrett, and Lynn Nadel, 1996

Corpus Processing for Lexical Acquisition, edited by Branimir Boguraev and James Pustejovsky, 1996

Methods for Assessing Children's Syntax, edited by Dana McDaniel, Cecile McKee, and Helen Smith Cairns, 1996

The Balancing Act: Combining Symbolic and Statistical Approaches to Language, edited by Judith Klavans and Philip Resnik, 1996

The Discovery of Spoken Language, Peter W. Jusczyk, 1996

[< previous page](#)

page_iii

[next page >](#)

Page iii

Language and Space

edited by Paul Bloom, Mary A. Peterson, Lynn Nadel, and Merrill F. Garrett

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

[< previous page](#)

page_iii

[next page >](#)

If you like this book, buy it!

© 1996 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Times Roman by Asco Trade Typesetting Ltd., Hong Kong and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Language and space/ edited by Paul Bloom . . . [et al.].
p. cm.(Language, speech, and communication)
Papers presented at a conference of the same name which was held
Mar. 16-19, 1994, Tucson, Ariz.
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-02403-9
1. Space and time in languageCongresses. I. Bloom, Paul, 1963- . II. Series.
p37.5.S65L36 1996
401'.9dc20

CIP

95-36427

Contents	
Preface	vii
Participants	ix
Chapter 1	
The Architecture of the Linguistic-Spatial Interface	
<i>Ray Jackendoff</i>	1
Chapter 2	
How Much Space Gets into Language?	
<i>Manfred Bierwisch</i>	31
Chapter 3	
Perspective Taking and Ellipsis in Spatial Descriptions	
<i>Willem J. M. Levelt</i>	77
Chapter 4	
Frames of Reference and Molyneux's Question: Crosslinguistic Evidence	
<i>Stephen C. Levinson</i>	109
Chapter 5	
The Confluence of Space and Language in Signed Languages	
<i>Karen Emmorey</i>	171
Chapter 6	
Fictive Motion in Language and "Ception"	
<i>Leonard Talmy</i>	211
Chapter 7	
The Spatial Prepositions in English, Vector Grammar, and the Cognitive Map Theory	
<i>John O'Keefe</i>	277

Chapter 8	
Multiple Geometric Representations of Objects in Languages and Language Learners	
<i>Barbara Landau</i>	317
Chapter 9	
Preverbal Representation and Language	
<i>Jean M. Mandler</i>	365
Chapter 10	
Learning How to Structure Space for Language: A Crosslinguistic Perspective	
<i>Melissa Bowerman</i>	385
Chapter 11	
Space to Think	
<i>Philip N. Johnson-Laird</i>	437
Chapter 12	
Spatial Perspective in Descriptions	
<i>Barbara Tversky</i>	463
Chapter 13	
A Computational Analysis of the Apprehension of Spatial Relations	
<i>Gordon D. Logan and Daniel D. Sadler</i>	493
Chapter 14	
The Language-to-Object Perception Interface: Evidence from Neuropsychology	
<i>Tim Shallice</i>	531
Chapter 15	
Space and Language	
<i>Mary A. Peterson, Lynn Nadel, Paul Bloom, and Merrill F. Garrett</i>	553
Name Index	579
Subject Index	585

Preface

The present volume consists of chapters by participants in the Language and Space conference held in Tucson, Arizona, 16–19 March 1994. In most cases the chapters have been written to reflect the numerous interactions at the conference, and for that reason we hope the book is more than just a compilation of isolated papers. The conference was truly interdisciplinary, including such domains as neurophysiology, neuropsychology, psychology, anthropology, cognitive science, and linguistics. Neural mechanisms, developmental processes, and cultural factors were all grist for the mill, as were semantics, syntax, and cognitive maps.

The conference had its beginnings in a seemingly innocent conversation in 1990 between two new colleagues at the University of Arizona (Bloom and Peterson), who wondered about the genesis of left-right confusions. One of them (M. A. P.) assumed that these confusions reflected a language problem; the other (P. B.) was quite certain that they reflected a visual perceptual problem. Curiously, it was the perception researcher who saw this issue as being mainly linguistic and the language researcher who saw it as mainly perceptual. In true academic form they decided that the best way to arrive at an answer would be to hold a seminar on the topic, which they did the very next year. Their seminar on language and space was attended by graduate students, postdoctoral fellows, and many faculty members from a variety of departments. Rather than answering the question that led to its inception, the seminar raised other questions: How do we represent space? What aspects of space can we talk about? How do we learn to talk about space? And what role does culture play in all these matters? One seminar could not explore all of these issues in any depth; an enlarged group of interested colleagues (the four coeditors) felt that perhaps several workshops might.

The Cognitive Neuroscience Program at the University of Arizona, in collaboration with the Cognitive Science Program and the Psychology Department, sponsored two one-day workshops on the relations between space and language. Although stimulating and helpful, the workshops gave rise to still other questions: How does

the brain represent space? How many kinds of spatial representations are there? What happens to spatial representations after various kinds of brain damage? Should experimental tests of the relations between space and language be restricted to closed-class linguistic elements or must the role of open-class elements be considered as well?

Given the scope of these questions, we decided to invite investigators from a variety of disciplines to a major scientific conference, and *Language and Space* took shape. The conference was judged by all to be a great success. We do not imagine that the chapters in this book provide final answers to any of the questions we first raised, but we are confident that they add much to the discussion and demonstrate the importance of the relations between space and language. We expect that increased attention will be given to this fascinating subject in the years ahead and hope that our conference, and this book, have made a significant contribution to its understanding.

Meetings cannot be held without the efforts of a considerable number of people, and the support of many funding sources. Our thanks to Pauline Smalley for all work she did in organizing the conference and making sure participants got to the right place at the right time and to Wendy Wilkins, of Arizona State University, for her gracious help both before and during the conference. We gratefully acknowledge the support of the conference's sponsors: McDonnell-Pew Cognitive Neuroscience Program, the Flinn Foundation Cognitive Neuroscience Program, and the Cognitive Science Program and Department of Psychology at the University of Arizona. We thank the participants for their intellectual energy and enthusiasm, which greatly contributed to the conference's success. Finally, we thank Amy Pierce of the MIT Press for her help with this volume. Editors Bloom and Peterson tossed a coin one evening over margaritas to determine whose name would go first.

Participants

Manfred Bierwisch
Structural Grammar Research Unit
Humboldt University, Berlin.

Paul Bloom
Departments of Psychology and
Linguistics
University of Arizona

Melissa Bowerman
Max Planck Institute for
Psycholinguistics
Nijmegen

Karen Emmorey
Salk Institute for Biological Studies
San Diego

Merrill F. Garrett
Department of Psychology and
Cognitive Science Program
University of Arizona

Ray Jackendoff
Linguistics and Cognitive Science
Program

Brandeis University
Philip N. Johnson-Laird
Department of Psychology
Princeton University

Barbara Landau
Department of Cognitive Science
University of California, Irvine

Willem J. M. Levelt
Max Planck Institute for
Psycholinguistics
Nijmegen

Stephen C. Levinson
Max Planck Institute for
Psycholinguistics
Nijmegen

Gordon Logan
Department of Psychology
University of Illinois

Jean M. Mandler
Department of Cognitive Science
University of California, San Diego

Lynn Nadel
Department of Psychology and
Cognitive Science Program
University of Arizona

John O'Keefe
Department of Anatomy and
Developmental Biology
University College, London

[< previous page](#)

[page_ix](#)

[next page >](#)

If you like this book, buy it!

Mary A. Peterson
Department of Psychology and
Cognitive Science Program
University of Arizona

Daniel D. Sadler
Department of Psychology
Indiana University of Pennsylvania

Tim Shallice
Medical Research Council, U.K.

Len Talmy
Department of Linguistics and Center
for Cognitive Science
State University of New York, Buffalo

Barbara Tversky
Department of Psychology
Stanford University

Chapter 1

The Architecture of the Linguistic-Spatial Interface

Ray Jackendoff

1.1 Introduction

How do we talk about what we see? More specifically, how does the mind/brain encode spatial information (visual or otherwise), how does it encode linguistic information, and how does it communicate between the two? This chapter lays out some of the boundary conditions for a satisfactory answer to these questions and illustrates the approach with some sample problems.

The skeleton of an answer appears in figure 1.1. At the language end, speech perception converts auditory information into linguistic information, and speech production converts linguistic information into motor instructions to the vocal tract. Linguistic information includes at least some sort of phonetic/phonological encoding of speech.¹ At the visual end, the processes of visual perception convert retinal information into visual information, which includes at least some sort of retinotopic mapping. The connection between language and vision is symbolized by the central double-headed arrow in figure 1.1. Because it is clear there cannot be a direct relation between a retinotopic map and a phonological encoding, the solution to our problem lies in elaborating the structure of this double-headed arrow.

1.2 Representational Modularity

The overall hypothesis under which I will elaborate figure 1.1 might be termed *Representational Modularity* (Jackendoff 1987, chapter 12; Jackendoff 1992, chapter 1). The general idea is that the mind/brain encodes information in many distinct formats or “languages of the mind.” There is a module of mind/brain responsible for each of these formats. For example, phonological structure and syntactic structure are distinct levels of encoding, with distinct and only partly commensurate primitives and principles of combination. Representational Modularity therefore posits that the architecture of the mind/brain devotes separate modules to these two encodings. Each

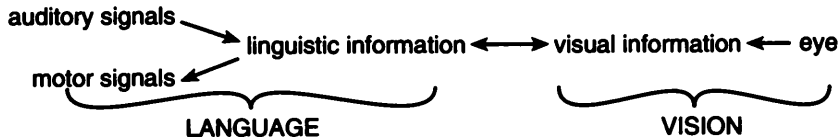


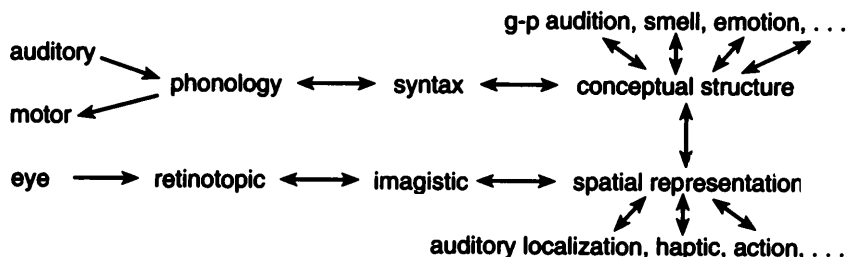
Figure 1.1

Coarse sketch of the relation between language and vision.

of these modules is domain-specific (phonology and syntax, respectively); and (with certain caveats to follow shortly) each is “informationally encapsulated” in Fodor’s (1983) sense. Representational modules differ from Fodorian modules in that they are individuated by the representations they process rather than by their function as faculties for input or output; that is, they are at the scale of individual levels of representation, rather than being entire faculties such as language perception.

A conceptual difficulty with Fodorian Modularity is that it leaves unanswered how modules communicate with each other and how they communicate with Fodor’s central, nonmodular cognitive core. In particular, Fodor’s language perception module derives “shallow representations”—some form of syntactic structure; Fodor’s central faculty of “belief fixation” operates in terms of the “language of thought,” a nonlinguistic encoding. But Fodor does not tell us how “shallow representations” are converted to the “language of thought,” as they must be if linguistic communication is to affect belief fixation. In effect, the language module is so domain-specific and informationally encapsulated that nothing can get out of it to serve cognitive purposes.² And without a theory of intermodular communication, it is impossible to approach the problem we are dealing with here, namely, how the language and vision modules manage to interact with each other.

The theory of Representational Modularity addresses this difficulty by positing, in addition to the *representation modules* proposed above, a system of *interface modules*. An interface module communicates between two levels of encoding, say L_1 and L_2 , by carrying a partial translation of information in L_1 form into information in L_2 form. An interface module, like a Fodorian module, is domain-specific: the phonology-to-syntax interface module, for instance, knows only about phonology and syntax, not about visual perception or general-purpose audition. Such a module is also informationally encapsulated: the phonology-to-syntax module dumbly takes whatever phonological inputs are available in the phonology representation module, translates the appropriate parts of them into (partial) syntactic structures, and delivers them to the syntax representation module, with no help or interference from, say, beliefs about the social context. In short, the communication among languages of the mind is mediated by modular processes as well.³

**Figure 1.2**

Slightly less coarse sketch of the relation between language and vision.

The levels of representation I will be working with here, and the interfaces among them, are sketched in figure 1.2. Each label in figure 1.2 stands for a level of representation served by a representation module. The arrows stand for interface modules. Double-headed arrows can be thought of either as interface modules that process bidirectionally or as pairs of complementary unidirectional modules (the correct choice is an empirical question). For instance, the phonology-syntax interface functions from left to right in speech perception and from right to left in speech production.

Figure 1.2 expands the “linguistic representation” of figure 1.1 into three levels involved with language: the familiar levels of phonology and syntax, plus conceptual structure, a central level of representation that interfaces with many other faculties. Similarly, “visual representation” in figure 1.1 is expanded into levels of retinotopic, imagistic, and spatial representation, corresponding roughly to Marr’s (1982) primal sketch, $2\frac{1}{2}$ -D sketch, and 3-D model, respectively; the last of these again is a central representation that interfaces with other faculties. In this picture, the effect of Fodorian faculty-sized modules emerges through the linkup of a series of representation and interface modules; communication among Fodorian faculties is accomplished by interface modules of exactly the same general character as the interface modules within faculties.

The crucial interface for our purposes here is that between the most central levels of the linguistic and visual faculties, conceptual structure and spatial representation. Before examining this interface, we have to discuss two things: (1) the general character of interfaces between representations (section 1.3); and (2) the general character of conceptual structure and spatial representation themselves (sections 1.4 and 1.5).

1.3 Character of Interface Mappings

To say that an interface module “translates” between two representations is, strictly speaking, inaccurate. In order to be more precise, let us focus for a moment on the

interface between phonology and syntax, the two best-understood levels of mental representation.

It is obvious that there cannot be a complete *translation* between phonology and syntax. Many details of phonology, most notably the segmental content of words, play no role at all in syntax. Conversely, many details of syntax, for instance the elaborate layering of specifiers and of arguments and adjuncts, are not reflected in phonology. In fact, a complete, information-preserving translation between the two representations would be pointless; it would in effect make them notational variants—which they clearly are not.

The relation between phonology and syntax is actually something more like a partial homomorphism. The two representations share the notion of *word* (and perhaps morpheme), and they share the linear order of words and morphemes.⁴ But segmental and stress information in phonology has no direct counterpart in syntax; and syntactic category (N, V, PP, etc.) and case, number, gender, and person features have no direct phonological counterparts.⁵ Moreover, syntactic and phonological constituent structures often fail to match. A classic example is given in (1).

(1) Phonological:

[This is the cat] [that ate the rat] [that ate the cheese]

Syntactic:

[This is [the cat [that ate [the rat [that ate [the cheese]]]]]]

The phonological bracketing, a flat tripartite structure, contrasts with the relentless right-embedded syntactic structure. At a smaller scale, English articles cliticize phonologically to the following word, resulting in bracketing mismatches such as (2).

(2) Phonological:

[the [big]] [house]

Syntactic:

[the [big [house]]]

Thus, in general, the phonology-syntax interface module creates only partial correspondences between these two levels.

A similar situation obtains with the interface between auditory information and phonological structure. The complex mapping between waveforms and phonetic segmentation in a sense preserves the relative order of information: a particular auditory cue may provide evidence for a number of adjacent phonetic segments, and a particular phonetic segment may be signaled by a number of adjacent auditory cues, but the overlapping “bands” of correspondence progress through the speech stream in an orderly linear fashion. On the other hand, boundaries between words, omnipresent in phonological structure, are not reliably detectable in the auditory signal; contrari-

wise, the auditory signal contains information about the formant frequencies of the speaker's voice that are invisible to phonology. So again the interface module takes only certain information from each representation into account in establishing a correspondence between them.

These examples show that each level of representation has its own proprietary information, and that an interface module communicates only certain aspects of this information to the next level up- or downstream. Representational modules, then, are not entirely informationally encapsulated: precisely to the extent that they receive information through interface modules, they are influenced by other parts of the mind.⁶

In addition to general principles of mapping, such as order preservation, an interface module can also make use of specialized learned mappings. The clearest instances of such mappings are *lexical items*. For instance, the lexical item *cat* stipulates that the phonological structure /kæt/ can be mapped simultaneously into a syntactic noun and into a conceptual structure that encodes the word's meaning. In other words, the theory of Representational Modularity leads us to regard the lexicon as a learned component of the interface modules within the language faculty (see Jackendoff forthcoming).

1.4 Conceptual Structure

Let us now turn to the crucial modules for the connection of language and spatial cognition: conceptual structure (CS) and spatial representation (SR). The idea that these two levels share the work of cognition is in a sense a more abstract version of Paivio's (1971) dual coding hypothesis. To use the terms of Mandler (chapter 9, this volume), Tversky (chapter 12, this volume), and Johnson-Laird (chapter 11, this volume), CS encodes "propositional" representations, and SR is the locus of "image schema" or "mental model" representations.

Conceptual structure, as developed in Jackendoff (1983, 1990) is an encoding of linguistic meaning that is independent of the particular language whose meaning it encodes. It is an "algebraic" representation, in the sense that conceptual structures are built up out of discrete primitive features and functions. Although CS supports formal rules of inference, it is not "propositional" in the standard logical sense, in that (1) propositional truth and falsity are not the only issue it is designed to address, and (2) unlike propositions of standard truth-conditional logic, its expressions refer not to the real world or to possible worlds, but rather to the world *as we conceptualize it*. Conceptual structure is also not entirely digital, in that some conceptual features and some interactions among features have continuous (i.e., analog) characteristics that permit stereotype and family resemblance effects to be formulated.

The theory of conceptual structure differs from most approaches to model-theoretic semantics as well as from Fodor's (1975) "Language of Thought," in that it takes for granted that lexical items have decompositions ("lexical conceptual structures," or LCSs) made up of features and functions of the primitive vocabulary. Here the approach concurs with the main traditions in lexical semantics (Miller and Johnson-Laird 1976; Lehrer and Kittay 1992; Pinker 1989; Pustejovsky 1995, to cite only a few parochial examples).

As the mental encoding of meaning, conceptual structure must include all the nonsensory distinctions of meaning made by natural language. A sample:

1. CS must contain pointers to all the sensory modalities, so that sensory encodings may be accessed and correlated (see next section).
2. CS must contain the distinction between tokens and types, so that the concept of an individual (say a particular dog) can be distinguished from the concept of the type to which that individual belongs (all dogs, or dogs of its breed, or dogs that it lives with, or all animals).
3. CS must contain the encoding of quantification and quantifier scope.
4. CS must be able to abstract actions (say running) away from the individual performing the action (say Harry or Harriet running).
5. CS must encode taxonomic relations (e.g., a bird is a kind of animal).
6. CS must encode social predicates such as "is uncle of," "is a friend of," "is fair," and "is obligated to."
7. CS must encode modal predicates, such as the distinction between "is flying," "isn't flying," "*can* fly," and "*can't* fly."

I leave it to my readers to convince themselves that none of these aspects of meaning can be represented in sensory encodings without using special annotations (such as pointers, legends, or footnotes); CS is, at the very least, the systematic form in which such annotations are couched.

For a first approximation, the interface between CS and syntax preserves embedding relations among constituents. That is, if a syntactic constituent *X* expresses the CS constituent *X'*, and if another syntactic constituent *Y* expresses the CS constituent *Y'*, and if *X* contains *Y*, then, as a rule, *X'* contains *Y'*. Moreover, a verb (or other argument-taking item) in syntax corresponds to a function in CS, and the subject and object of the verb normally correspond to CS arguments of the function. Hence much of the overall structure of syntax corresponds to CS structure. (Some instances in which relative embedding is not preserved appear in Levin and Rapoport 1988 and Jackendoff 1990, chapter 10.)

Unlike syntax, though, CS has no notion of linear order: it must be indifferent as to whether it is expressed syntactically in, say, English, where the verb precedes

the direct object, or Japanese, where the verb follows the direct object. Rather, the embedding in CS is purely relational.⁷

At the same time, there are aspects of CS to which syntax is indifferent. Most prominently, other than argument structure, much of the conceptual material bundled up inside a lexical item is invisible to syntax, just as phonological features are. As far as syntax is concerned, the meanings of *cat* and *dog* (which have no argument structure) are identical, as are the meanings of *eat* and *drink* (which have the same argument structure): the syntactic reflexes of differences in lexical meaning are extremely coarse.

In addition, some bits of material in CS are absent from syntactic realization altogether. A good example, given by Talmy (1978), is (3).

(3) The light flashed until dawn.

The interpretation of (3) contains the notion of repeated flashes. But this repetition is not coded in the verb *flash*: *The light flashed* normally denotes only a single flash. Nor is the repetition encoded in *until dawn*, because, for instance, *Bill slept until dawn* does not imply repeated acts of sleeping. Rather, the notion of repetition arises because (a) *until dawn* gives the temporal bound of an otherwise unbounded process; (b) *the light flashed* is a point event and therefore temporally bounded; and (c) to make these compatible, a principle of construal or “coercion” (Pustejovsky 1991; Jackendoff 1991) interprets the flashing as stretched out in time by repetition. This notion of repetition, then, appears in the CS of (3) but not in the LCS of any of its words.

The upshot is that the correspondence between syntax and CS is much like the correspondence between syntax and phonology. Certain parts of the two structures are in fairly regular correspondence and are communicated by the interface module, but many parts of each are invisible to the other.

Even though CS is universal, languages can differ in their overall semantic patterns, in at least three respects. First, languages can have different strategies in how they typically bundle up conceptual elements into lexical items. For example, Talmy (1980) documents how English builds verbs of motion primarily by bundling up motion with accompanying manner, while Romance languages bundle up motion primarily with path of motion, and Atsugewi bundles up motion primarily with the type of object or substance undergoing motion. Levinson (chapter 4, this volume) shows how the Guugu Yimithirr lexicon restricts the choice of spatial frames of reference to cardinal directions (see section 1.8). These strategies of lexical choice affect the overall grain of semantic notions available in a particular language. (This is of course in addition to differences in meaning among individual lexical items across languages, such as the differences among prepositions discussed by Bowerman, chapter 10, this volume.)

Second, languages can differ in what elements of conceptual structure they *require* the speaker to express in syntax. For example, French and Japanese require speakers always to differentiate their social relation to their addressee, a factor largely absent from English. Finnish and Hungarian require speakers to express the multiplicity (or repetition) of events, using iterative aspect, a factor absent from English, as seen in (3). On the other hand, English requires speakers to express the multiplicity of objects by using the plural suffix, a requirement absent in Chinese.

Third, languages can differ in the special syntactic constructions they use to express particular conceptual notions. Examples in English are the tag question (*They shoot horses, don't they?*), the "One more" construction (*One more beer and I'm leaving*) (Culicover 1972), and the "The more . . . , the more" construction (*The more you drink, the worse you feel*). These all convey special nuances that go beyond lexical meaning.

I have argued (Jackendoff 1983) that there is no language-specific "semantic" level of representation intervening between syntax and conceptual structure. Language-specific differences in semantics of the sort just listed are localized in the interface between syntactic and conceptual structures. I part company here with Bierwisch (1986), Partee (1993), and to a certain extent Pinker (1989). Within my approach, a separate semantic level is unnecessary, in part because the syntax-CS interface module has enough richness in it to capture the relevant differences; I suspect that these other theories have not considered closely enough the properties of the interface. However, the issues are at this point far from resolved. The main point, on which Bierwisch, Pinker, and I agree (I am unclear about Partee), is that there *is* a language-independent and universal level of CS, whether directly interfacing with syntax or mediated by an intervening level.

1.5 Spatial Representation

For the theory of spatial representation—the encoding of objects and their configurations in space—we are on far shakier ground. The best articulated (partial) theory of spatial representation I know of is Marr's (1982) 3-D model, with Biederman's (1987) "geonic" constructions as a particular variant. Here are some criteria that a spatial representation (SR) must satisfy.

1. SR must encode the shape of objects in a form suitable for recognizing an object at different distances and from different perspectives, that is, it must solve the classic problem of object constancy.⁸
2. SR must be capable of encoding spatial knowledge of parts of objects that cannot be seen, for instance, the hollowness of a balloon.

3. SR must be capable of encoding the degrees of freedom in objects that can change their shape, for instance, human and animal bodies.
4. SR must be capable of encoding shape variations among objects of similar visual type, for example, making explicit the range of shape variations characteristic of different cups. That is, it must support visual object *categorization* as well as visual object *identification*.
5. SR must be suitable for encoding the full spatial layout of a scene and for mediating among alternative perspectives (“What would this scene look like from over there?”), so that it can be used to support reaching, navigating, and giving instructions (Tversky, chapter 12, this volume).
6. SR must be independent of spatial modality, so that haptic information, information from auditory localization, and felt body position (proprioception) can all be brought into registration with one another. It is important to know by looking at an object where you expect to find it when you reach for it and what it should feel like when you handle it.

Strictly speaking, criteria 5 and 6 go beyond the Marr and Biederman theories of object shape. But there is nothing in principle to prevent these theories from serving as a component of a fuller theory of spatial understanding, rather than strictly as theories of high-level visual shape recognition. By the time visual information is converted into shape information, its strictly visual character is lost—it is no longer retinotopic, for example—nor, as Marr stresses, is it confined to the observer’s point of view.⁹

SR contrasts with CS in that it is geometric (or even quasi-topological) in character, rather than algebraic. But on the other hand, it is not “imagistic”—it is not to be thought of as encoding “statues in the head.” An image is restricted to a particular point of view, whereas SR is not. An image is restricted to a particular instance of a category (recall Berkeley’s objection to images as the vehicle of thought: how can an image of a particular triangle stand for all possible triangles?¹⁰), whereas SR is not. An image cannot represent the unseen parts of an object—its back and inside, and the parts of it occluded from the observer’s view by other objects—whereas SR does. An image is restricted to the visual modality, whereas SR can equally well encode information received haptically or through proprioception. Nevertheless, even though SRs are not themselves imagistic, it makes sense to think of them as encoding *image schemas*: abstract representations from which a variety of images can be generated.

Figure 1.2 postulates a separate module of imagistic (or pictorial) representation one level toward the eye from SR. This corresponds roughly to Marr’s $2\frac{1}{2}$ -D sketch. It is specifically visual; it encodes what is consciously present in the field of vision or visual imagery (Jackendoff 1987, chapter 14). The visual imagistic representation is

restricted to a particular point of view at any one time; it does not represent the backs and insides of objects explicitly. At the same time, it is not a retinotopic representation because it is normalized for eye movements and incorporates information from both eyes into a single field, including stereopsis. (There is doubtless a parallel imagistic representation for the haptic faculty, encoding the way objects feel, but I am not aware of any research on it.)

It is perhaps useful to think of the imagistic representation as “perceptual” and SR as “cognitive”; the two are related through an interface of the general sort found in the language faculty: they share certain aspects, but each has certain aspects invisible to the other. Each can drive the other through the interface: in visual perception, an imagistic representation gives rise to a spatial representation that encodes one’s understanding of the visual scene; in visual imagery, SRs give rise to imagistic representations. In other words, the relation of images to image schemas (SRs) in the present theory is much like the relation of sentences to thoughts. Image schemas are not skeletal images, but rather structures in a more abstract and more central form of representation.¹¹

This layout of the visual and spatial levels of representation is of course highly oversimplified. For instance, I have not addressed the well-known division of visual labor between the “what system” and the “where system,” which deal, roughly speaking, with object identification and object location respectively (O’Keefe and Nadel 1978; Ungerleider and Mishkin 1982; Farah et al. 1988; Jeannerod 1994; Landau and Jackendoff 1993). My assumption, perhaps unduly optimistic, is that such division of labor can be captured in the present approach by further articulation of the visual-spatial modules in figure 1.2 into smaller modules and their interfaces, much as figure 1.2 is a further articulation of figure 1.1.

1.6 Interface between CS and SR

We come at last to the mapping between CS and SR, the crucial link between the visual system and the linguistic system.¹² What do these two levels share, such that it is possible for an interface module to communicate between them?

The most basic unit they share is the notion of a physical object, which appears as a geometrical unit in SR and as a fundamental algebraic constituent type in CS.¹³ In addition, the Marr-Biederman theory of object shape proposes that object shapes are decomposed into geometric parts in SR. This relation maps straightforwardly into the part-whole relation, a basic function in CS that of course generalizes far beyond object parts.

The notions of place (or location) and path (or trajectory) play a basic role in CS (Talmy 1983; Jackendoff 1983; Langacker 1986); they are invoked, for instance, in

locational sentences such as *The book is lying on the table* (place) and *The arrow flew through the air past my head* (path). Because these sentences can be checked against visual input, and because locations and paths can be given obvious geometric counterparts, it is a good bet that these constituents are shared between CS and SR.¹⁴ (The Marr-Biederman theory does not contain places and paths because they arise only in encoding the behavior of objects in the full spatial field, an aspect of visual cognition not addressed by these theories.)

The notion of physical motion is also central to CS, and obviously it must be represented in spatial cognition so that we can track moving objects. More speculatively, the notion of force appears prominently in CS (Talmy 1985; Jackendoff 1990), and to the extent that we have the impression of directly perceiving forces in the visual field (Michotte 1954), these too might well be shared between the two representations.¹⁵

Our discussion of interfaces in previous sections leads us to expect some aspects of each representation to be invisible to the other. What might some of these aspects be? Section 1.4 noted that CS encodes the token versus type distinction (a particular dog vs. the category of dogs), quantificational relations, and taxonomic relations (a bird is a kind of animal), but that these are invisible to SR. On the other hand, SR encodes all the details of object shapes, for instance, the shape of violin or a butter knife or a German shepherd's ears. These geometric features do not lend themselves at all to the sort of algebraic coding found in CS; they are absolutely natural to (at least the spirit of) SR.

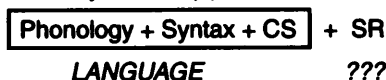
In addition to general mappings between constituent types in CS and SR, individual matchings can be learned and stored. (Learned and stored) lexical entries for physical object words can contain a spatial representation of the object in question, in addition to their phonological, syntactic, and conceptual structure. For instance, the entry for *dog* might look something like (4).

- (4)

Phono:	/dɒg/
Syntax:	+ N, - V, + count, + sing, ..
CS:	Individual, Type of Animal, Type of Carnivore
	Function: (often) Type of Pet
SR:	[3-D model w/ motion affordances]
Auditory:	[sound of barking]

In (4) the SR takes the place of what in many approaches (e.g., Rosch and Mervis 1975; Putnam 1975) has been informally called an "image of a prototypical instance of the category." The difficulty with an image of a prototype is that it is computationally nonefficacious: it does not meet the demands of object shape identification laid out as criteria 1–4 in the previous section. A more abstract spatial representation,

a. One way to view (4)



b. Another way to view (4)

**Figure 1.3**

Two ways to view the integration of spatial structures into lexical entries.

along the lines of a Marr 3-D model, meets these criteria much better; it is therefore a more satisfactory candidate for encoding one's knowledge of what the object looks like. As suggested by the inclusion of "auditory structure" in (4), a lexical entry should encode (pointers to) other sensory characteristics as well.

The idea, then, is that the "meaning" of a word goes beyond the features and functions available in CS, in particular permitting detailed shape information in a lexical SR. (A word *must* have a lexical CS; it *may* have an SR as well.) Such an approach might be seen as threatening the linguistic integrity of lexical items: as suggested by figure 1.3a, it breaks out of the purely linguistic system. But an alternative view of entries like (4) places them in a different light. Suppose one deletes the phonological and syntactic structures from (4). What is left is the nonlinguistic knowledge one has of dogs—the "concept" of a dog, much of which could be shared by a nonlinguistic organism. Phonological and syntactic structures can then be viewed as further structures tacked onto to this knowledge to make it linguistically expressible, as suggested in figure 1.3b. With or without language, the mind has to have a way to unify multimodal representations and store them as units (that is, to establish long-term memory "binding" in the neuroscience sense); (4) represents just such a unit. The structures that make this a "lexical item" rather than just a "concept" simply represent an additional modality into which this concept extends: the linguistic modality.

Having established general properties of the CS–SR interface, we must raise the question of exactly what information is on either side of it. How do we decide? The overall premise behind Representational Modularity, of course, is that each module is a specialist, and that each particular kind of information belongs in a particular module. For instance, details of shape are not duplicated in CS, and taxonomic relations are not duplicated in SR. For the general case, we can state a *criterion of economy*: all other things being equal, if a certain kind of distinction is encoded in SR,

it should not also be encoded in CS, and vice versa. I take this maximal segregation to be the default assumption.

Of course, all other things are not equal. The two modules must share enough structure that they can communicate with each other—for instance, they must share at least the notions mentioned at the beginning of this section. Thus we do not expect, as a baseline, that the information encoded by CS and SR is entirely incommensurate. Let us call this the *criterion of interfacing*.

What evidence would help decide whether a certain kind of information is in CS as well as SR? One line of argument comes from interaction with syntax. Recall that CS is by hypothesis the form of central representation that most directly interacts with syntactic structure. Therefore, if a semantic distinction is communicated to syntax, so that it makes a syntactic difference, that distinction must be present in CS and not just SR. (Note that this criterion applies only to *syntactic* and not *lexical* differences. As pointed out in section 1.4, *dog* and *cat* look exactly the same to syntax.) Let us call this the *criterion of grammatical effect*.

A second line of argument concerns nonspatial domains of CS. As is well known (Gruber 1965; Jackendoff 1976, 1983; Talmy 1978; Lakoff and Johnson 1980; Langacker 1986), the semantics of many nonspatial conceptual domains show strong parallels to the semantics of spatial concepts. Now if a particular semantic distinction appears in nonspatial domains as well as in the spatial domain, it cannot be encoded in SR alone, which by definition pertains only to spatial cognition. Rather, similarities between spatial and nonspatial domains must be captured in the algebraic structure of CS. I will call this the *criterion of nonspatial abstraction*.

1.7 A Simple Case: The Count-Mass Distinction

A familiar example will make these criteria clearer. Consider the count-mass distinction. SR obviously must make a distinction between single individuals (a cow), multiple individuals (a herd of cows), and substances (milk)—these have radically different appearances and spatial behavior over time (Marr and Biederman, of course, have little or nothing to say about what substances look like.) According to the criterion of economy, all else being equal, SR should be the *only* level that encodes these differences.

But all else is not equal. The count-mass distinction has repercussions in the marking of grammatical number and in the choice of possible determiners (count nouns use *many* and *few*, mass nouns use *much* and *little*, for example). Hence the criterion of grammatical effect suggests that the count-mass distinction is encoded in CS also.

Furthermore, the count-mass distinction appears in abstract domains. For example, *threat* is grammatically a count noun (*many threats*/**much threat*), but the semantically

very similar *advice* is a mass noun (*much advice*/**many advices*). Because the distinction between threats and advice cannot be encoded spatially—it doesn't “look like anything”—the only place to put it is in CS. That is, the criterion of nonspatial extension applies to this case.

In addition, the count-mass distinction is closely interwoven with features of temporal event structure such as the event-process distinction (Verkuyl 1972, 1993; Dowty 1979; Hinrichs 1985; Jackendoff 1991; Pustejovsky 1991). To the extent that events have a spatial appearance, it is qualitatively different from that of objects. And distinctions of temporal event structure have a multitude of grammatical reflexes. Thus the criteria of nonspatial extension and grammatical effect both apply again to argue for the count-mass distinction being encoded in CS.

A further piece of evidence comes from lexical discrepancies in the grammar of count and mass nouns. An example is the contrast between *noodles* (count) and *spaghetti* (mass)—nouns that pick out essentially the same sorts of entities in the world. A single one of these objects can be described as a singular *noodle*, but the mass noun forces one to use the phrasal form *stick* (or *strand*) of *spaghetti*. (In Italian, *spaghetti* is a plural count noun, and one can refer to a single *spaghetto*.)

Because *noodles* and *spaghetti* pick out similar entities in the world, there is no reason to believe that they have different lexical SRs. Hence there must be a mismatch somewhere between SR and syntax. A standard strategy (e.g., Bloom 1994) is to treat them as alike in CS as well and to localize the mismatch somewhere in the CS–syntax interface. Alternatively, the mismatch might be between CS and SR. In this scenario, CS has the option of encoding a collection of smallish objects (or even largish objects such as *furniture*) as either an aggregate or a substance, then syntax follows suit by treating the concepts in question as grammatically count or mass, respectively.¹⁶ Whichever solution is chosen, it is clear that SR and syntax alone cannot make sense of the discrepancy. Rather, CS is necessary as an intermediary between them.

1.8 Axes and Frames of Reference

We now turn to a more complex case with a different outcome. Three subsets of the vocabulary invoke the spatial axes of an object. I will call them collectively the “axial vocabulary.”

1. The “axial parts” of an object—its *top*, *bottom*, *front*, *back*, *sides*, and *ends*—behave grammatically like parts of the object, but, unlike standard parts such as a *handle* or a *leg*, they have no distinctive shape. Rather, they are regions of the object (or its boundary) determined by their relation to the object's axes. The up-down axis determines top and bottom, the front-back axis determines front and back, and

a complex set of criteria distinguishing horizontal axes determines sides and ends (Miller and Johnson-Laird 1976; Landau and Jackendoff 1993).

2. The “dimensional adjectives” *high*, *wide*, *long*, *thick*, and *deep* and their nominalizations *height*, *width*, *length*, *thickness*, and *depth* refer to dimensions of objects measured along principal, secondary, and tertiary axes, sometimes with reference to the horizontality or verticality of these axes (Bierwisch 1967; Bierwisch and Lang 1989).

3. Certain spatial prepositions, such as *above*, *below*, *next to*, *in front of*, *behind*, *alongside*, *left of*, and *right of*, pick out a region determined by extending the reference object’s axes out into the surrounding space. For instance, *in front of X* denotes a region of space in proximity to the projection of *X*’s front-back axis beyond the boundary of *X* in the frontward direction (Miller and Johnson-Laird 1976; Landau and Jackendoff 1993; Landau, chapter 8, this volume). By contrast, *inside X* makes reference only to the region subtended by *X*, not to any of its axes; *near X* denotes a region in proximity to *X* in any direction at all. Notice that many of the “axial prepositions” are morphologically related to nouns that denote axial parts.

It has been frequently noted (for instance, Miller and Johnson-Laird 1976; Olson and Bialystok 1983; and practically every chapter in this volume) that the axial vocabulary is always used in the context of an assumed frame of reference. Moreover, the choice of frame of reference is often ambiguous; and because the frame determines the axes in terms of which the axial vocabulary receives its denotation, the axial vocabulary too is ambiguous.

The literature usually invokes two frames of reference: an *intrinsic* or *object-centered* frame, and a *deictic* or *observer-centered* frame. Actually the situation is more complex. Viewing a frame of reference as a way of determining the axes of an object, it is possible to distinguish at least eight different available frames of reference (many of these appear as special cases in Miller and Johnson-Laird 1976, which in turn cites Bierwisch 1967; Teller 1969; and Fillmore 1971, among others).

A. Four *intrinsic* frames all make reference to properties of the object:

1. The *geometric* frame uses the geometry of the object itself to determine the axes. For instance, the dimension of greatest extension can determine its *length* (figure 1.4a). Symmetrical geometry often implies a top-to-bottom axis dividing the symmetrical halves and a side-to-side axis passing from one half to the other (figure 1.4b). A special case concerns animals, whose front is intrinsically marked by the position of the eyes.
2. In the *motion* frame, the *front* of a moving object is determined by the direction of motion. For instance, the front of an otherwise symmetrical double-ended tram is the end facing toward its current direction of motion (figure 1.4c).

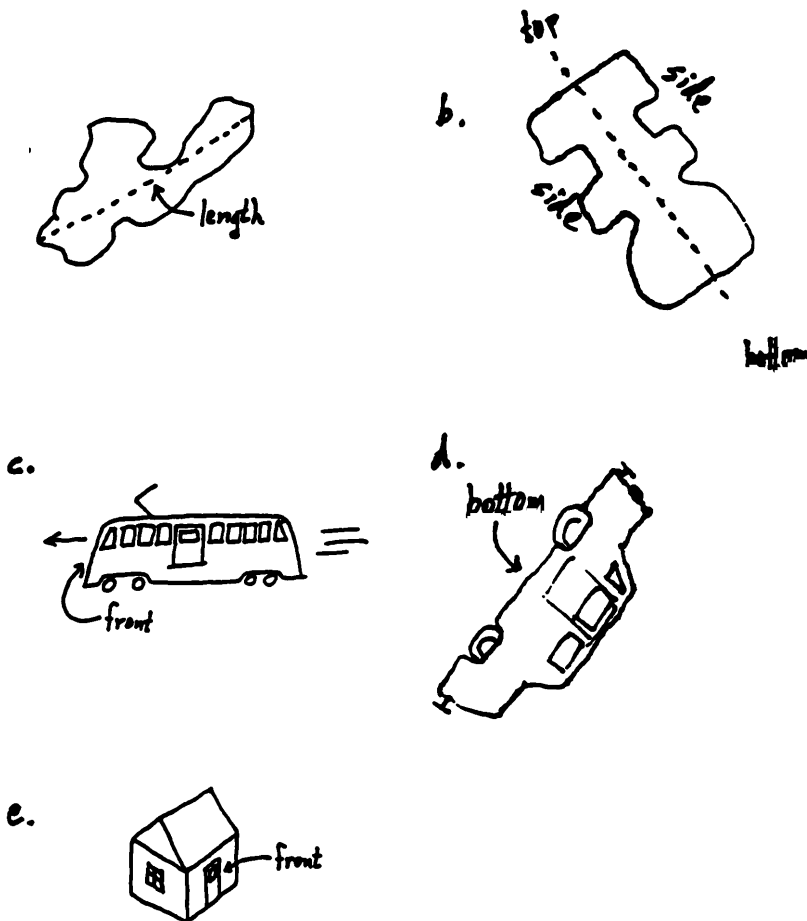


Figure 1.4
Intrinsic reference frames.

3. Two intrinsic frames depend on functional properties of the object. The *canonical orientation* frame designates as the top (or bottom) of an object the part which in the object's normal orientation is uppermost (or lowermost), even if it does not happen to be at the moment. For instance, the canonical orientation of the car in figure 1.4d has the wheels lowermost, so the part the wheels are attached to is the canonical bottom, even though it is pointing obliquely upward in this picture.
4. Intrinsic parts of an object can also be picked out according to the *canonical encounter* frame. For instance, the part of a house where the public enters is

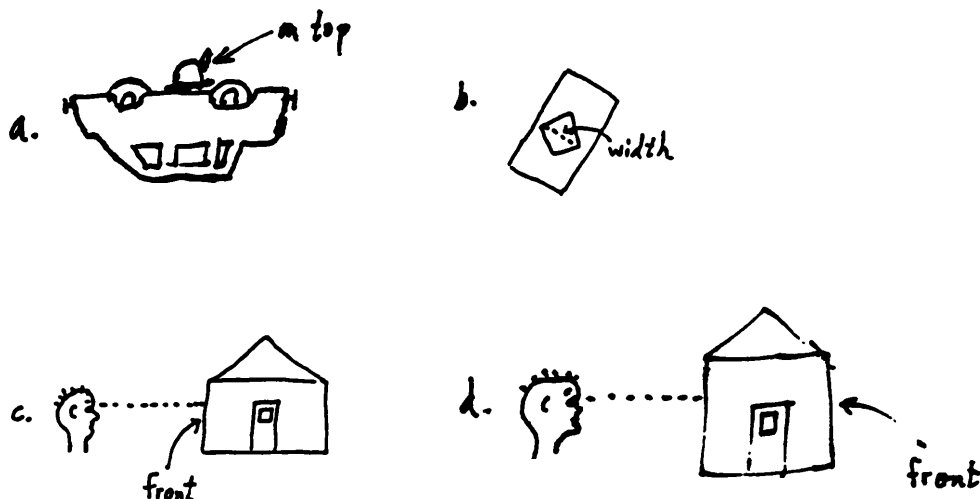


Figure 1.5
Environmental reference frames.

functionally the front (figure 1.4e). (Inside a building such as a theater, the front is the side that the public normally faces, so that the front from the inside may be a different wall of the building than the front from the outside.)

B. Four *environmental* frames project axes onto the object based on properties of the environment:

1. The *gravitational* frame is determined by the direction of gravity, regardless of the orientation of the object. In this frame, for instance, the hat in figure 1.5a is on top of the car.
2. The *geographical* frame is the horizontal counterpart of the gravitational frame, imposing axes on the object based on the cardinal directions north, south, east, and west, or a similar system (Levinson, chapter 4, this volume).
3. The *contextual* frame is available when the object is viewed in relation to another object, whose own axes are imposed on the first object. For instance, figure 1.5b pictures a page on which is drawn a geometric figure. The page has an intrinsic side-to-side axis that determines its width, regardless of orientation. The figure on the page inherits this axis, and therefore its width is measured in the same direction.
4. The *observer* frame may be projected onto the object from a real or hypothetical observer. This frame establishes the front of the object as the side facing the observer, as in figure 1.5c. We might call this the “orientation-mirroring observer frame.” Alternatively, in some languages, such as Hausa,

the front of the object is the side facing the same way as the observer's front, as in figure 1.5d. We might call this the "orientation-preserving observer frame."

It should be further noted that axes in the canonical orientation frame (figure 1.4d) are derived from gravitational axes in an imagined normal orientation of the object. Similarly, axes in the canonical encounter frame (figure 1.4e) are derived from a hypothetical observer's position in the canonical encounter. So in fact only two of the eight frames, the geometric and motion frames, are entirely free of direct or indirect environmental influence.

One of the reasons the axial vocabulary has attracted so much attention in the literature is its multiple ambiguity among frames of reference. In the preceding examples alone, for instance, three different uses of *front* appear. Only the geographical frame (in English, at least) has its own unambiguous vocabulary. Why should this be? And what does it tell us about the distribution of information between CS and SR? This will be the subject of the next section.

Before going on, though, let us take a moment to look at how frames of reference are used in giving route directions (Levelt, chapter 3, this volume; Tversky, chapter 12, this volume).

Consider a simple case of Levelt's diagrams such as figure 1.6. The route from circle 1 to circle 5 can be described in two different ways:

- (5) a. "Geographic" frame: From 1, go up/forward to 2, right to 3, right to 4, down to 5.
- b. "Observer" frame: From 1, go up/forward to 2, right to 3, straight/forward to 4, right to 5.

The problem is highlighted by the step from 3 to 4, which is described as "right" in (5a) and "straight" in (5b).

The proper way to think of this seems to be to keep track of hypothetical traveler's orientation. In the "geographic" frame, the traveler maintains a constant orientation, so that *up* always means up on the page; that is, the traveler's axes are set contextually by the page (frame B3).

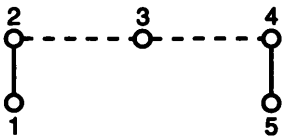


Figure 1.6
One of Levelt's "maps."

The puzzling case is the “observer” frame, where the direction from 2 to 3 is “right” and the same direction, from 3 to 4, is “straight” or “forward.” Intuitively, as Levelt and Tversky point out, one pictures oneself traveling through the diagram. From this the solution follows immediately: “forward” is determined by the observer’s last move, that is, using the motion frame (A2). The circles, which have no intrinsic orientation, play no role in determining the frame. If they are replaced by landmarks that do have intrinsic axes, as in Tversky’s examples, a third possibility emerges, that of setting the traveler’s axes contextually by the landmarks (frame B3 again). And of course geographical axes (frame B1) are available as well if the cardinal directions are known.

1.9 Lexical Encoding of Axial Vocabulary

Narasimhan (1993) reports an experiment that has revealing implications for the semantics of the axial vocabulary. Subjects were shown irregular shapes (“Narasimhan figures”) of the sort in figure 1.7, and asked to mark on them their length, width, height, or some combination of the three. Because *length*, *width*, and *height* depend on choice of axes, responses revealed subjects’ judgments about axis placement.

This experiment is unusual in its use of irregular shapes. Previous experimental research on axial vocabulary with which I am familiar (e.g., Bierwisch and Lang 1989; Levelt 1984) has dealt only with rectilinear figures or familiar objects, often only in rectilinear orientations. In Narasimhan’s experiment, the subjects have to compute axes of novel shapes on-line, based on visual input; they cannot simply call up intrinsic axes stored in long-term memory as part of the canonical representation of a familiar object.

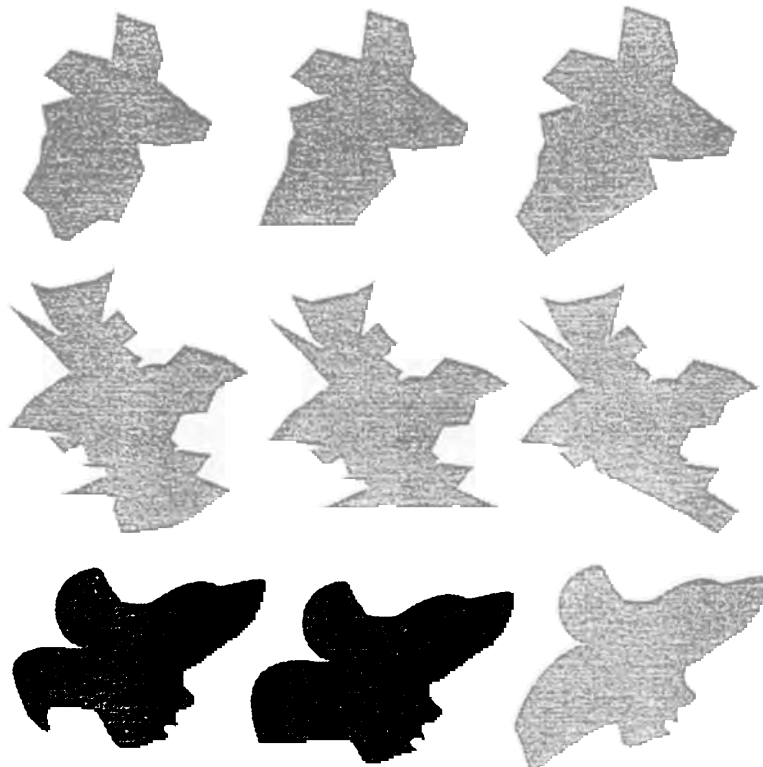
But of course linguistic information is also involved in the subjects’ responses. In particular, the dimension that the subject is asked to mark influences the choice of axis, as might be expected from the work of Bierwisch and Lang (1989). *Length* biases the subject in favor of intrinsic geometric axes (longest dimension), while *height* biases the subject toward environmental axes (gravitational or page-based contextual). Thus, confronted with a shape such as figure 1.8a, whose longest dimension is oblique to the contextual vertical, subjects tended to mark its *length* as an oblique, and its *height* as an environmental vertical. Sometimes subjects even marked these axes on the very same figure; they did not insist by any means on orthogonal axes!

The linguistic input, however, was not the only influence on the choice of axes. Details in the shape of the Narasimhan figure also exerted an influence. For example, figure 1.8b has a flattish surface near the (contextual) bottom. Some subjects (8%) apparently interpreted this surface as a base that had been rotated from its canonical orientation; they drew the height of the figure as an axis orthogonal to this base, that

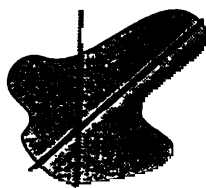
No base

Flat base

Tilted base



Up-down axis

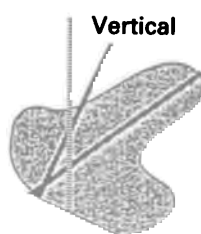


Maximum (vertical)



Observer's line of sight

Up-down axis



Vertical

Maximum

Figure 1.8

Some axis systems marked by subjects on Narasimhan figures.

is, as a “canonical vertical.” Nothing in the linguistic input created this new possibility: it had to be computed on-line from the visual input. As a result of this extra possibility, the shape presented three different choices for its axis system, as shown in the figure.

We see, then, that linguistic and visual input interact intimately in determining subjects’ responses in this experiment. However, the hypothesis of Representational Modularity does not allow us to just leave it at that. We must also ask at what level of representation (i.e., in which module) this interaction takes place. The obvious choices are CS and SR.

The fact that the subjects actually draw in axes shows that the computation of axes must involve SR. The angle and positioning of a drawn axis is continuously variable, in a way expected in the geometric SR but not expected in the algebraic feature complexes of CS.

How does the linguistic input get to SR so that it can influence the subjects’ response? That is, at what levels of representation do the words *length*, *width*, and *height* specify the axes and frames of reference they can pick out? There are two possibilities:

1. *The CS hypothesis.* The axes could be specified in the lexical entries of *length*, *width*, and *height* by features in CS such as [$\pm$ maximal], [$\pm$ vertical], [$\pm$ secondary]; the frames of reference could be specified by CS features such as [$\pm$ contextual], [$\pm$ observer]. General correspondences in the CS–SR interface would then map features into the geometry of SR. According to this story, when subjects judge the axes of Narasimhan figures, the lexical items influence SR indirectly, via these general interpretations of the dimensional features of CS. (This is, I believe, the approach advocated by Bierwisch and Lang.)
2. *The SR hypothesis.* Alternatively, we know that lexical items may contain elements of SR such as the shape of a dog. Hence it is possible that the lexical entries of *length*, *width*, and *height* also contain SR components that specify axes and frames of reference *directly* in the geometric format of SR. This would allow the axes and reference frames to be unspecified (or largely so) in the CS of these words. According to this hypothesis, when subjects judge the axes of Narasimhan figures, the SR of the lexical items interacts directly with SR from visual input.

I propose that the SR hypothesis is closer to correct. The first argument comes from the criterion of economy. Marr (1982) demonstrates, and Narasimhan’s experiment confirms, that people use SR to pick out axes and frames of reference in novel figures. In addition, people freely switch frames of reference in visuomotor tasks. For example, we normally adopt an egocentric (or observer) frame for reaching but an environmental frame for navigating; in the latter, we see ourselves moving through a

stationary environment, not an environment rushing past.¹⁷ These are SR functions, not CS functions. Consequently, axes and frames of reference cannot be eliminated from SR. This means that a CS feature system for these distinctions at best duplicates information in SR—it cannot take the place of information in SR.

Next consider the criterion of grammatical effect. If axes and frames of reference can be shown to have grammatical effects, it is necessary to encode them in CS. But in this domain, unlike the count-mass system, there seem to be few grammatical effects. The only thing special about the syntax of the English axial vocabulary is that dimensional adjectives and axial prepositions can be preceded by measure phrases, as in *three inches long*, *two miles wide* (with dimensional adjectives), and *four feet behind the wall*, *seven blocks up the street* (with axial prepositions). Other than dimensional adjectives, the only English adjective that can occur with a measure phrase is *old*; such pragmatically plausible cases as **eighty degrees hot* and **twelve pounds heavy* are ungrammatical. Similarly, many prepositions do not occur with measure phrases (**ten inches near the box*); and those that do are for the most part axial (though *away*, as in *a mile away from the house*, is not).¹⁸

Thus whether a word pertains to an axis does seem to make a grammatical difference. But that is about as far as it goes. No grammatical effects seem to depend on *which* axis a word refers to, much less which frame of reference the axis is computed in, at least in English.¹⁹ Thus the criterion of grammatical effect dictates at most that CS needs only a feature that distinguishes axes of objects from other sorts of object parts; the axial vocabulary will contain this feature. Distinguishing axes from each other and frames of reference from each other appears unnecessary on grammatical grounds.

Turning to the criterion of nonspatial extension, consider the use of axis systems and frames of reference in nonspatial domains. It is well known that analogues of spatial axes occur in other semantic fields, and that axial vocabulary generalizes to these domains (Gruber 1965; Jackendoff 1976; Talmy 1978; Langacker 1986; Lakoff 1987). But all other axis systems I know of are only one-dimensional, for example, numbers, temperatures, weights, ranks, and comparative adjectives (*more/less beautiful/salty/exciting/etc.*). A cognitive system with more than one dimension is the familiar three-dimensional color space, but language does not express differences in color using any sort of axial vocabulary. Kinship systems might be another multidimensional case, and again the axial vocabulary is not employed.

In English, when a nonspatial axis is invoked, the axis is almost always up/down (*higher number, lower rank, of higher beauty, lower temperature, my mood is up*, etc.). Is there a reference frame? One's first impulse is to say that the reference frame is gravitational—perhaps because we speak of the temperature *rising* and *falling* and of *rising in the ranks of the army*, and because *rise* and *fall* in the spatial domain

pertain most specifically to the gravitational frame. But on second thought, we really wouldn't know how to distinguish among reference frames in these spaces. What would it mean to distinguish an intrinsic upward from a gravitational upward, for example?

About the only exception to the use of the vertical axis in nonspatial domains is *time*, a one-dimensional system that goes front to back.²⁰ Time is also exceptional in that it does display reference frame distinctions. For instance, one speaks of the times *before now*, where *before* means "prior to," as though the observer (or the "front" of an event) is facing the past. But one also speaks of the *hard times before us*, where *before* means "subsequent to," as though the observer is facing the future.

A notion of frame of reference also appears in social cognition, where we speak of adopting another's point of view in evaluating their knowledge or attitudes. But compared to spatial frames of reference, this notion is quite limited: it is analogous to adopting an observer reference frame for a different (real or hypothetical) observer; there is no parallel to any of the other seven varieties of reference frames. Moreover, in the social domain there is no notion of axis that is built from these frames of reference. Thus again an apparent parallel proves to be relatively impoverished.

In short, very little of the organization of spatial axes and frames of reference is recruited for nonspatial concepts. Hence the criterion of nonspatial extension also gives us scant reason to encode in CS all the spatial distinctions among three-dimensional axes and frames of reference. All we need for most purposes is the distinction between the vertical and other axes, plus some special machinery for time and perhaps for social point of view. Certainly nothing outside the spatial domain calls for the richness of detail needed for the spatial axial vocabulary. Our tentative conclusion is that most of this detail is encoded only in the SR component of the axial vocabulary, not in the CS component; it thus parallels such lexical SR components as the shape of a dog. Let me call this the "Mostly SR hypothesis."

A skeptic committed to the CS hypothesis might raise a "functional" argument against this conclusion. Perhaps multiple axes and frames of reference are available in CS, but we do not recruit them for nonspatial concepts because we have no need for them in our nonspatial thought. Or perhaps the nature of the real world does not lend itself to such thinking outside of the spatial domain, so such concepts cannot be used sensibly.

If one insists on a "functional" view, I would urge quite a different argument. It would often be extremely useful for us to be able to think in terms of detailed variation of two or three nonspatial variables, say the relation of income to educational level to age, but in fact we find it very difficult. For a more ecologically plausible case, why do we inevitably reduce social status to a linear ranking, when it so clearly

involves many interacting factors? The best way we have of thinking multidimensionally is to translate the variables in question into a Cartesian graph, so that we can apply our multidimensional spatial intuitions to the variation in question—we can see it as a path or a region in space. This suggests that CS is actually relatively poor in its ability to encode multidimensional variation; we have to turn to SR to help us encode it. This is more or less what would be predicted by the Mostly SR hypothesis. That is, the “functional” argument can be turned around and used as evidence *for* the Mostly SR hypothesis.

The case of axes and frames of reference thus comes out differently from the case of the count-mass distinction. This time we conclude that most of the relevant distinctions are *not* encoded in CS, but only in SR, one level further removed from syntactic structure.

This conclusion is tentative in part because of the small amount of linguistic evidence adduced for it thus far—one would certainly want to check the data out crosslinguistically before making a stronger claim. But it is also tentative because we do not have enough formal theory of SR to know how it encodes axes and frames of reference. It might turn out, for instance, that the proper way to encode the relevant distinctions is in terms of a set of discrete (or digital) annotations to the geometry of SR. In such a case, it would be hard to distinguish an SR encoding of these distinctions from a CS encoding. But in the absence of a serious theory of SR, it is hard to know how to continue this line of research.

1.10 Final Thoughts

To sort out empirical issues in the relation of language to spatial cognition, it is useful to think in terms of Representational Modularity. This forces us to distinguish the levels of representation involved in language, abstract conceptual thought, and spatial cognition, and to take seriously the issue of how these levels communicate with one another. In looking at any particular phenomenon within this framework, the crucial question has proved to be at which level or levels of representation it is to be encoded. We have examined cases where the choice between CS and SR comes out in different ways. This shows that the issue is not a simple prejudged matter; it must be evaluated for each case.

For the moment, however, we are at the mercy of the limitations of theory. Compared to the richness of phonological and syntactic theory, the theory of CS is in its infancy; and SR, other than the small bit of work by Marr and Biederman, is hardly even in gestation. This makes it difficult to decide among (or even to formulate) competing hypotheses in any more than sketchy fashion. It is hoped that the present volume will spur theorists to remedy the situation.

Acknowledgments

I am grateful to Barbara Landau, Manfred Bierwisch, Paul Bloom, Lynn Nadel, Bhuvana Narasimhan, and Emile van der Zee for extensive discussion, in person and in correspondence, surrounding the ideas in this chapter. Further important suggestions came from participants in the Conference on Space and Language sponsored by the Cognitive Anthropology Research Group at the Max Planck Institute for Psycholinguistics in Nijmegen in December 1993 and of course from the participants in the Arizona workshop responsible for the present volume.

This research was supported in part by National Science Foundation grant IRI-92-13849 to Brandeis University, by a Keck Foundation grant to the Brandeis University Center for Complex Systems, and by a fellowship to the author from the John Simon Guggenheim Foundation.

Notes

1. This is an oversimplification, because of the existence of languages that make use of the visual/gestural modalities. See Emmorey (chapter 5, this volume).
2. Various colleagues have offered interpretations of Fodor in which some further vaguely specified process accomplishes the conversion. I do not find any support for these interpretations in the text.
3. Of course, Fodorian modularity can also solve the problem of communication among modules by adopting the idea of interface modules. However, because interface modules as conceived here are too small to be Fodorian modules (they are not input-output faculties), there are two possibilities: either (1) the scale of modularity has to be reduced from faculties to representations, along lines proposed here; or else (2) interfaces are simply an integrated part of larger modules and need not themselves be modular. I take the choice between these two possibilities to reflect in part a merely rhetorical difference, but also in part an empirical one.
4. Caveats are necessary concerning nonconcatenative morphology such as reduplication and Semitic inflection, where the relation between linear order in phonology and syntax is unclear, to say the least.
5. To be sure, syntactic features are frequently realized phonologically as affixes with segmental content; but the phonology itself has no knowledge of what syntactic features these affixes express.
6. Fodor's claims about informational encapsulation are largely built around evidence that semantic/pragmatic information does not immediately affect the processes of lexical retrieval and syntactic parsing in speech perception. This evidence is also consistent with Representational Modularity. The first pass of lexical retrieval has to be part of the mapping from auditory signal to phonological structure, so that word boundaries can be imposed; Fodor's discussion shows that this first pass uses no semantic information. The first pass of syntactic parsing has to be part of the mapping from phonological to syntactic structure, so that candidate semantic interpretations can subsequently be formulated and tested; this first pass uses no semantic information either. See Jackendoff 1987, chapters 6 and 12, for more detailed discussion.

7. It is surely significant that syntax shares embedding with CS and linear order with phonology. It is as though syntactic structure is a way of converting embedding structure into linear order, so that structured meanings can be expressed as a linear speech stream.

8. As a corollary, SR must support the generation of mentally rotated objects, whose perspective with respect to the viewer changes during rotation. This is particularly crucial in rotation on an axis parallel to the picture plane because different parts of the object are visible at different times during rotation—a fact noted by Kosslyn (1980).

9. Some colleagues have objected to Marr's characterizing the 3-D sketch as "object-centered," arguing that objects are always seen from some point of view or other—at the very least the observer's. However, I interpret "object-centered" as meaning that the encoding of the object is independent of point of view. This neutrality permits the appearance of the object to be computed as necessary to fit the object into the visual scene as a whole, viewed from any arbitrary vantage point. Marr, who is not concerned with spatial layout but only with identifying the object, does not deal with this further step of reinjecting the object into the scene. But I see such a step as altogether within the spirit of his approach.

10. A different sort of example, offered by Christopher Habel at the Nijmegen space conference (see acknowledgments): the "image schema" for *along*, as in *the road is along the river*, must include the possibility of the road being on either side of the river. An imagistic representation must represent the road being specifically on one side or the other.

11. It is unclear to me at the moment what relationship this notion of image schema bears to that of Mandler (1992 and chapter 9, this volume), although there is certainly a family resemblance. Mandler's formulation derives from work such as that of Lakoff (1987) and Langacker (1986), in which the notion of level of representation is not well developed, and in which no explicit connection is made to research in visual perception. I leave open for future research the question of whether the present conception can help sharpen the issues with which Mandler is concerned.

12. This section is derived in part from the discussion in Jackendoff 1987, chapter 10.

13. Although fundamental, such a type is not necessarily primitive. Jackendoff 1991 decomposes the notion of object into the more primitive feature complex [material, + bounded, –inherent structure]. The feature [material] is shared by substances and aggregates; it distinguishes them all from situations (events and states), spaces, times, and various sorts of abstract entities. The feature [+ bounded] distinguishes objects from substances, and also closed events (or accomplishments) from processes. The feature [–inherent structure] distinguishes objects from groups of individuals, but also substances from aggregates and homogeneous processes from repeated events.

14. On the other hand, it is not so obvious that places and paths are encoded in imagistic representation because we do not literally *see* them except when dotted lines are drawn in cartoons. This may be another part of SR that is invisible to imagistic representation. That is, places and paths as independent entities may be a higher-level cognitive (nonperceptual) aspect of spatial understanding, as also argued by Talmy (chapter 6, this volume).

15. Paul Bloom has asked (personal communication) why I would consider *force* but not, say *anger* to be encoded in SR because we "have the impression of directly perceiving" anger as

well. The difference is that physical force has clear geometric components—direction of force and often contact between objects—which are independently necessary to encode other spatial entities such as trajectories and orientations. Thus force seems a natural extension of the family of spatial concepts. By contrast, anger has no such geometrical characteristics; its parameters belong to the domain of emotions and interpersonal relations. Extending SR to anger, therefore, would not yield any generalizations in terms of shared components.

16. This leaves open the possibility of CS-syntax discrepancies in the more grammatically problematic cases like *scissors* and *trousers*. I leave the issue open.

17. For a recent discussion of the psychophysics and neuropsychology of the distinction between environmental motion and self-motion, see Wertheim 1994 and its commentaries. Wertheim, however, does not appear to address the issue, crucial to the present enterprise, of how this distinction is encoded so that further inferences can be drawn from it—namely, the cognitive consequences of distinguishing reference frames.

18. Measure phrases also occur in English adjective phrases as specifiers of the comparatives *more/-er than* and *as . . . as*, for instance *ten pounds heavier (than X)*, *three feet shorter (than X)*, *six times more beautiful (than X)*, *fifty times as funny (as X)*. Here they are licensed not by the adjective itself, but by the comparative morpheme.

19. Bickel 1994a, however, points out that the Nepalese language Belhare makes distinctions of grammatical case based on frame of reference. In a “personmorphic” frame for *right* and *left*, the visual field is divided into two halves, with the division line running through the observer and the reference object; this frame requires the genitive case for the reference object. In a “physiomorphic” frame for *right* and *left*, the reference object projects four quadrants whose centers are focal front, back, left, and right; this frame requires the ablative case for the reference object. I leave it for future research to ascertain how widespread such grammatical distinctions are and to what extent they might require a weakening of my hypothesis.

20. A number of people have pointed another nonvertical axis system, the political spectrum, which goes from right to left. According to the description of Bickel 1994b, the Nepalese language Belhare is a counterexample to the generalization about time going front to back: a transverse axis is used for measuring time, and an up-down axis is used for the conception of time as an opposition of past and future.

References

- Bickel, B. (1994a). Mapping operations in spatial deixis and the typology of reference frames. Working paper no. 31, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Bickel, B. (1994b). Spatial operations on deixis, cognition, and culture: Where to orient oneself in Belhare (revised version). Unpublished manuscript, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2), 115–147.
- Bierwisch, M. (1967). Some semantic universals of German adjectivals. *Foundations of Language*, 3, 1–36.

- Bierwisch, M. (1986). On the nature of semantic form in natural language. In F. Klix and H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances*, 765–784. Amsterdam: Elsevier/North-Holland.
- Bierwisch, M., and Lang, E. (Eds.) (1989). *Dimensional adjectives*. Berlin: Springer.
- Bloom, P. (1994). Possible names: The role of syntax-semantics mappings in the acquisition of nominals. *Lingua*, 92, 297–329.
- Culicover, P. (1972). OM-sentences: On the derivation of sentences with systematically unspecifiable interpretations. *Foundations of Language*, 8, 199–236.
- Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Farah, M., Hammond, K., Levine, D., and Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, 20, 439–462.
- Fillmore, C. (1971) *Santa Cruz lectures on deixis*. Bloomington: Indiana University Linguistics Club.
- Fodor, J. (1975) *The language of thought*. Cambridge, MA: Harvard University Press.
- Fodor, J. (1983) *Modularity of mind*. Cambridge, MA: MIT Press.
- Gruber, J. (1965). Studies in lexical relations. Ph.D. diss., Massachusetts Institute of Technology. Reprinted in Gruber, *Lexical structures in syntax and semantics*, Amsterdam: North-Holland, 1976.
- Hinrichs, E. (1985). A compositional semantics for *Aktionsarten* and NP reference in English. Ph.D. diss., Ohio State University.
- Jackendoff, Ray (1976). Toward an explanatory semantic representation. *Linguistic Inquiry*, 7, 89–150.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1991). Parts and boundaries. *Cognition*, 41, 9–45.
- Jackendoff, R. (1992). *Languages of the mind*. Cambridge, MA: MIT Press.
- Jackendoff, R. (forthcoming). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17, 187–201.
- Kosslyn, S. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
- Lakoff, G., and Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Landau, B., and Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–238.

- Langacker, R. (1986). *Foundations of cognitive grammar*. Vol. 1. Stanford, CA: Stanford University Press.
- Lehrer, A., and Kittay, E. (Eds.) (1992). *Frames, fields, and contrasts*. Hillsdale, NJ: Erlbaum.
- Levelt, W. (1984). Some perceptual limitations in talking about space. In A. van Doorn, W. van de Grind, and J. Koenderink (Eds.), *Limits in perception*. Utrecht: Coronet Books.
- Levin, B., and Rapoport, T. (1988). Lexical subordination. In *Papers from the twenty-fourth regional meeting of the Chicago Linguistics Society*, 275–289. Chicago: University of Chicago. Department of Linguistics.
- Mandler, J. (1992). How to build a baby: 2. Conceptual primitives. *Psychological Review*, 99, 587–604.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Michotte, A. (1954). *La perception de la causalité*. 2d ed. Louvain: Publications Universitaires de Louvain.
- Miller, G., and Johnson-Laird, P. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Narasimhan, B. (1993). Spatial frames of reference in the use of *length*, *width*, and *height*. Unpublished manuscript, Boston University.
- O’Keefe, J., and Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Oxford University Press.
- Olson, D., and Bialystok, E. (1983). *Spatial cognition*. Hillsdale, NJ: Erlbaum.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston. Reprint, Hillsdale, NJ: Erlbaum, 1979.
- Partee, B. (1993). Semantic structures and semantic properties. In E. Reuland and W. Abraham (Eds.), *Knowledge and Language*. Vol. 2, *Lexical and conceptual structure*, 7–30. Dordrecht: Kluwer.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
- Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, 41, 47–81.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Putnam, H. (1975). The meaning of “meaning.” In K. Gunderson (Ed.), *Language, mind, and knowledge*, 131–193. Minneapolis: University of Minnesota Press.
- Rosch, E., and Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- Talmy, L. (1978). The relation of grammar to cognition: A synopsis. In D. Waltz (Ed.), *Theoretical issues in natural language processing*, vol. 2, New York: Association for Computing Machinery.

- Talmy, L. (1980). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description*, vol. 3. New York: Cambridge University Press.
- Talmy, L. (1983). How language structures space. In H. Pick and L. Acredolo (Eds.), *Spatial orientation: Theory, research, and application*. New York: Plenum Press.
- Talmy, L. (1985). Force dynamics in language and thought. In *Papers from the Twenty-first Regional Meeting of the Chicago Linguistic Society*. Chicago: University of Chicago. Department of Linguistics. Also in *Cognitive Science*, 12 (1988), 49–100.
- Teller, P. (1969). Some discussion and extension of Manfred Bierwisch's work on German adjectivals. *Foundations of Language*, 5, 185–217.
- Ungerleider, L., and Mishkin, M. (1982) Two cortical visual systems. In D. Ingle, M. Goodale, and R. Mansfield (Eds.), *Analysis of visual behavior*. Cambridge MA: MIT Press.
- Verkuyl, H. (1972). *On the compositional nature of the aspects*. Dordrecht: Reidel.
- Verkuyl, H. (1993). *A theory of aspectuality*. Cambridge: Cambridge University Press.
- Wertheim, A. (1994). Motion perception during self-motion: The direct versus inferential controversy revisited. *Behavioral and Brain Sciences*, 17, 293–311.

Chapter 2

How Much Space Gets into Language?

Manfred Bierwisch

2.1 Introduction

We can talk about spatial aspects of our environment with any degree of precision we want, even though linguistic expressions—unlike pictures, maps, blueprints, and the like—do not exhibit spatial structure in any relevant way. This apparent paradox is simply due to the symbolic, rather than iconic, character of natural language. For the same reason, we can talk about color, temperature, kinship, and all the rest, even though linguistic utterances do not exhibit color, temperature, kinship, and so on. The apparent paradox nevertheless raises the by no means trivial question where and how space gets into language. The present chapter will be concerned with certain aspects of this problem, pursuing the following question:

Which components of natural language accommodate spatial information, and how?

Looking first at syntax, we observe that completely identical structures can express both spatial and clearly nonspatial situations, as in (1a) and (1b), respectively:

- (1) a. We entered Saint Peter's Cathedral.
- b. We admired Saint Peter's Cathedral.

The contrast obviously depends on the meaning of *enter* versus *admire*. Comparing (1a) with (2), we notice, furthermore, that identical or at least very similar spatial events can be expressed by means of rather different syntactic constructions:

- (2) We went into Saint Peter's Cathedral.

The conclusion that syntactic elements and relations do not accommodate spatial information seems to be confronted with certain objections, though. Thus the PP *at the end* has a temporal meaning in (3a) but a spatial one in (3b), depending on its syntactic position:

- (3) a. At the end, she signed the letter.
 b. She signed the letter at the end.

One cannot, however, assign the contrast between spatial and nonspatial interpretation to the position as such, as is evident from pairs like those in (4):

- (4) a. With this intention, she signed the letter.
 b. She signed the letter with this intention.

What we observe in (3) and (4) is rather the effect the different syntactic structure has on the compositional semantics of adjuncts (the details of which are still not really understood), determining different interpretations for the PP in (3). Pending further clarification, we will nevertheless conclude that phrase structure does not reflect spatial information *per se*. Another problem shows up in cases like (5), differing with respect to place and goal:

- (5) a. Er schwamm unter dem Steg.
 (He swam under the bridge.) *location*
 b. Er schwamm unter den Steg.
 (He swam under the bridge.) *directional*

It is, of course, not the contrast between /m/ and /n/, but rather that between dative and accusative that is relevant here. This appears to be a matter of the syntactic component. In the present case, however, the crucial distinction can be reduced to a systematic difference between a locative and a directional reading of the preposition *unter*, each associated with a specific case requirement (see Bierwisch 1988 for discussion) in languages with rich morphology. I will take up this issue in section 2.7. Whereas case can thus be shown to be related to space only as an indirect effect, this does not hold for the so-called notional or content cases.

In any case, syntax and morphology as such do not reflect spatial information. Hence the main area to be explored with respect to our central question is the semantic component, in particular the field of lexical semantics. As already mentioned with respect to (1), it is the word meaning of *enter* that carries the spatial aspect. Similarly, the contrast between place and goal in (5) is ultimately a matter of the two different readings of *unter*. Further illustrations could be multiplied at will, including all major lexical categories.

This does not mean, however, that there is a simple and clear distinction between spatial and nonspatial vocabulary. As a matter of fact, most words that undoubtedly have a spatial interpretation may alternatively carry a nonspatial reading under certain conditions. Consider (6) as a case in point:

- (6) He entered the church.

Besides the spatial interpretation corresponding to that of (1a), (6) can also have an interpretation under which it means *he became a priest*, where *church* refers to an institution and *enter* denotes a change in social relations. The verb *to enter* thus has a spatial or nonspatial interpretation depending on the reading of the object it combines with. This is an instance of what Pustejovsky (1991) calls “co-compositionality,” that is, a compositional structure where one constituent determines an interpretation of the other that is not fixed outside the combinatorial process. In other words, we must not only account for the spatial information that *enter* projects in cases like (1a) and one reading of (6), but also for the switch to the nonspatial interpretation in the second reading of (6). To conclude these preliminary considerations, in order to answer our central question, we have to investigate how lexical items relate to space and eventually project these relations by means of compositional principles.

2.2 Lexical Semantics and Conceptual Structure

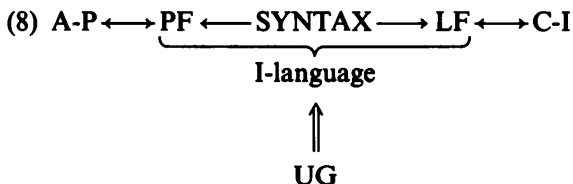
Let me begin by placing lexical and compositional semantics in the more general perspective of linguistic knowledge, that is, the internal or I-language in the sense of Chomsky (1986), which underlies the properties of external or E-language of sets of linguistic expression. Following the terminology of Chomsky (1993), *I-language* is to be construed as a computational system that determines a systematic correspondence between two different domains of mental organization:

(7) A-P $\longleftrightarrow$ I-language $\longrightarrow$ C-I

A-P comprises the systems of articulation and perception, and C-I, the systems by which experience is conceptually organized and intentionally related to the external and internal environment. I-language provides two representational systems, which Chomsky calls “phonetic form” (PF) and “logical form” (LF), that constitute the interfaces with the respective extralinguistic domains. Because there is apparently no direct relation that connects spatial information to sound structure, bypassing the correspondence established by the computational system of I-language, I will have nothing to say about PF, except where it will be useful to compare how it relates to A-P with the far more complex conceptual phenomena that concern us.

Given PF and LF as interface levels, determined by I-language and interpreted in terms of A-P and C-I, respectively, the correspondence between them is established by the syntactic and morphological operations of I-language. With this overall orientation in mind, one might consider the (species-specific) language capacity as emerging from brain structures that allow for the discrete, recursive mapping between two representational systems of different structure and origin. Assuming universal grammar (UG) to be the formal characterization of this capacity, we arrive at the

following general schema, where I-language emerges from the conditions specified by UG through the interaction with the systems of A-P and C-I:



This schema is meant as a rough orientation, leaving crucial questions to be clarified. Before I turn to details of the relation between I-language and C-I, two general remarks about UG and the organization of I-language must be made.

First, for each of the major components of I-language, universal grammar (UG) must provide two specifications:

1. A way to recruit the primitive elements by which representations and operations of the component are specified; and
2. A general format of the type of representations and operations of the component.

The most parsimonious assumption is that specification 2 is fixed across languages, emerging from the conditions of the language capacity as such. In other words, the types of representation and the operations available for I-language are given in advance.¹

As to specification 1, three types of primes are to be distinguished:

1. Primes that are recruited from and interpreted by A-P;
2. Primes that are recruited from and interpreted by C-I; and
3. Primes that function within I-language without external interpretation.

It is usually assumed that type 1, the primes of PF, namely, phonetic features and prosodic categories, are based on universally fixed options in UG. Alternatively, one might think of them as being recruited from the auditory input and articulatory patterns by means of certain constraints within UG, which provides not the repertoire of these features but rather some sort of recipe to make them up. This view would be mandatory if in fact UG were not restricted to acoustic signals but allowed also for systems like sign language. Although the details of this issue go beyond the scope of the present discussion, the notion of conditions or constraints to construct primes of I-language seems to be indispensable if we address type 2, the primes in terms of which I-language interfaces with C-I, and if semantic representations are to go beyond a highly restricted core of lexical items. I will return to these issues below. As for type 3, which must comprise the features specifying syntactic and morphological categories, these must be determined directly by the conditions on syntactic and morphological representations and operations falling under type 2, varying only to

the extent to which they can be affected by intrusion from the interface levels. This might in fact be the case for morphological categories by which syntactic conditions take up conceptual content, for example, in number, person, and so forth.

Second, the computation determined by I-language does not in general proceed in terms of primitive elements but to a large extent in terms of chunks of them fixed in lexical items. Lexical items are idiosyncratic configurations, varying from language to language, which must be learned on the basis of individual experience, but which are determined by UG with respect to their general format in accordance with specifications 1 and 2. I will call the set of lexical items, together with the general conditions to which they must conform, the “lexical system” (LS) of I-language. LS is not a separate component of I-language, alongside phonology, syntax, morphology, and semantics; rather, it cuts across all of them, combining information of all components of I-language. The general format that UG determines for lexical items is (9):

(9) [PF(*le*), GF(*le*), LF(*le*)],

where PF(*le*) determines a representation of *le* at PF;

LF(*le*) consists of primes of LF specified by *le*;

GF(*le*) represents syntactic and morphological properties of *le*.

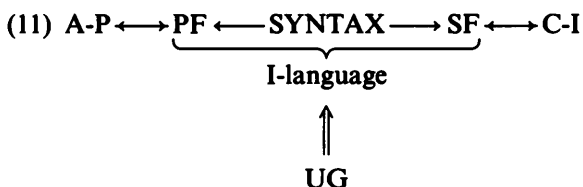
I will have more to say about the organization of lexical entries at the end of section 2.2. (9) also indicates the basic format of linguistic expressions in general, if we assume that PF(*le*), LF(*le*), and GF(*le*) can represent information of any complexity in accordance with the two requirements noted above.

With regard to the crucial question how C-I relates to I-language, there is a remarkable lack of agreement among otherwise closely related approaches. According to the conceptual framework of Chomsky (1981, 1986, 1993), LF is a level of syntactic representation whose particular status lies in its forming the interface with conceptual structure. (In Chomsky 1993, LF is in fact the only level of syntactic representation to which independent, systematic conditions apply.) The basic elements of LF are lexical items, or rather their semantic component, and the syntactic features associated with them. In other words, the primes of LF, which according to type 2 above connect I-language to C-I, are to be identified with word meanings, or more technically, with the LF part of lexical items, including complex items originating from incorporation, head movement, or other processes of “sublexical syntax” as discussed, for example, by Hale and Keyser (1993). In any case, whatever internal structure should be assigned to the semantics of lexical items is essentially a matter of C-I, not structurally reflected in I-language.

In contrast to this view, Jackendoff (1983 and subsequent work), following Katz (1972) and others, assigns lexical items a rich and systematic internal structure, which is claimed to be linguistically relevant. I will adopt this view, arguing that there are

structural phenomena directly involved in I-language that turn on the internal structure of lexical items. I call the basic elements of this structure “semantic primes,” assuming these are the elements identified in type 2 that connect I-language to C-I. Suppose now that we call the representational system based on semantic primes the “semantic form” (SF) of I-language—parallel to PF, which is based on phonetic primes. We will consequently replace schema (9) of lexical items by (10), and hence the overall schema (8) by (11):

- (10) [PF(*le*), GF(*le*), SF(*le*)] with PF(*le*) a configuration of PF, SF(*le*) a configuration of SF, and GF(*le*) a specification of morphological and syntactic properties



The system of SYNTAX is now to include the information represented at LF according to (8).² Before I take up some controversial issues that are related to these assumptions, I will briefly illustrate their empirical motivation.

The basic idea behind the organization of knowledge suggested in (11) is that I-language needs to be distinguished from the various mental systems that bear on A-P and C-I, respectively. More specifically, the conceptual interpretation *c* of a linguistic expression *e* is determined by the semantic form of *e* and the conceptual knowledge underlying C-I. As this point is crucial with respect to our central question, I will clarify the problem by means of some examples. What I want to show is twofold. On the one hand, the interpretation of an expression *e* is determined by its semantic form SF(*e*), which is based on the semantic form of lexical items exhibiting a systematic, linguistically relevant internal structure. On the other hand, the conceptual interpretation of *e*, which among other things must fix the truth and satisfaction conditions, depends in crucial respects on commonsense beliefs, world knowledge, and situational aspects, which are language-independent and must be assigned to C-I.

To begin with the second point, compare the sentences in (12):

- (12) a. He left the institute an hour ago.
 b. He left the institute a year ago.

In (12a) *leave* is (most likely) interpreted as a physical movement and *institute* as place, while the time adverbial *a year ago* of (12b) turns *leave* into a change in affiliation and *institute* into a social institution. The two interpretations of *leave the*

institute are cases of co-compositionality as already illustrated by sentence (6) above. For extensive discussion of these phenomena, where linguistic and encyclopedic knowledge interact, see, for example, Bierwisch (1983) and Dölling (1995). The most striking point of (12) is, however, that the choice between the locational and the social interpretation is determined by the contrast between *year* and *hour*. This has nothing to do, of course, with the meaning of these items as such, whether linguistic or otherwise, but with world knowledge about changes of location or institutional affiliation and their temporal frames.

In a similar vein, the physical or abstract interpretation of *lose* and *money* in (13) depends on world knowledge coming in through the different adverbial adjuncts:

- (13) a. John lost his money through a hole in his pocket.
 b. John lost his money by gambling at cards.
 c. John lost his money by speculating at the stock market.

Notice incidentally, that *his money* in (13a) refers to the coins or notes John is carrying along, while in (13c) it is likely to refer to all his wealth, again due to encyclopedic knowledge about a certain domain.

Turning now to the first point concerning the internal structure of SF(*le*), I will illustrate the issue by looking more closely at *leave*, providing at the same time an outline of the format to be assumed for SF representations. To begin with, (14) indicates the slightly simplified semantic form of *leave* as it appears in (12):

- (14) [*x* DO [BECOME [NEG [*x* AT *y*]]]]

Generally speaking, SF consists of functors and arguments that combine by functional application. The basic elements of SF in the sense mentioned in type 2 above are constants like DO, BECOME, AT, and so forth and variables like *x*, *y*, *z*. More specifically, DO is a relation between an individual *x* and a proposition *p* with the conceptual interpretation that could be paraphrased by “*x* performs *p*.” In (14), *p* is the proposition [BECOME [NEG [*x* AT *y*]]], where BECOME defines a transition into a state *s* characterized by the condition that *x* be not at *y*. In short, (14) specifies the complex condition that *x* brings about a change of state that results in *x*’s not being at *y*. For a systematic exposition of this framework in general, see Bierwisch (1988), and for the interpretation of DO and BECOME in particular, see Dowty (1979). It should be noted, at this point, that all the elements showing up in (14) are highly abstract and hence compatible with differing conceptual interpretations. Thus [*x* AT *y*] might be a spatial relation, as in (12a), or an institutional affiliation, as in (12b). Correspondingly, [*x* DO [BECOME *s*]] can be interpreted by a spatial movement or a change in social position, depending on the conceptual content of the resulting state *s*.

But why should the lexical meaning of *leave* be represented in the manner of (14), rather than simply as $[x \text{ LEAVE } y]$, if the conceptual interpretation must account for more specific details anyway? This brings us to the linguistic motivation of the internal structure stipulated for $SF(le)$. An even remotely adequate answer to this question would go far beyond the scope of this chapter, hence I can only indicate the type of motivation that bears on (14) by pointing out two phenomena. Consider first (15), which is ambiguous between a repetitive and a restitutive reading:

(15) John left the institute again.

Under the repetitive reading, (15) states that John leaves the institute for (at least) the second time, while under the restitutive reading (15) states only that John brings about of his not being at the institute, which obtained before. These two interpretations can be indicated by (16a) and (16b), respectively, where x must be interpreted by *John* and y by *the institute*, and where AGAIN is a shorthand for the SF to be assigned to *again*:

- (16) a. $[AGAIN [x \text{ DO } [BECOME [NEG [x \text{ AT } y]]]]]$
 b. $[x \text{ DO } [BECOME [AGAIN [NEG [x \text{ AT } y]]]]]$

For discussion of intricate details left out here, see von Stechow (1995). Two points are to be emphasized, however. First, the ambiguity of (15) carries over to both the physical and the institutional interpretation; it is determined by linguistic, rather than extralinguistic, conceptual conditions. Second, it could not be represented, if *leave* were to be characterized by the unanalyzed lexical meaning $[x \text{ LEAVE } y]$.

The second phenomenon to be mentioned concerns the intransitive use of *leave* as in (17):

(17) John left a year ago.

Two observations are relevant here. First, the variable y of (14) can be left without a syntactically determined value, in which case it must be interpreted by contextual conditions providing a kind of neutral origo. Second, the state $[x \text{ AT } y]$ under this condition is almost automatically restricted to the locative interpretation, which serves as a kind of default. Once again, although for different reasons, the global representation $[x \text{ LEAVE } y]$ would fail to provide the relevant structure.

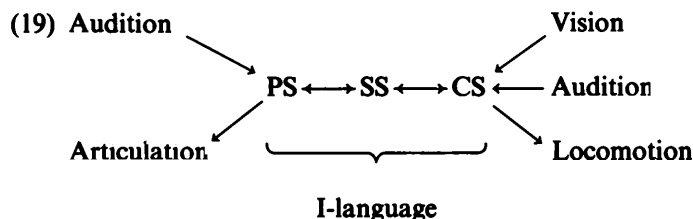
The optionality of the object of *leave* on which (17) relies brings in, furthermore, the intimate relationship between $SF(le)$ and $GF(le)$, or more specifically, the relationship between variables in $SF(le)$ and the syntactic argument structure (or subcategorization, to use earlier terminology). Suppose we include a specification of the SF variable, optionally or obligatorily interpreted by syntactic constituents, as one component into the syntactic information $GF(le)$, such that (18) would be a more complete lexical entry for *leave*:

- (18) $\underbrace{/leave/}_{PF(le)} \quad \underbrace{\text{Verb } \hat{x}(\hat{y})}_{GF(le)} \quad \underbrace{[x \text{ DO } [BECOME [NEG [x \text{ AT } y]]]]}_{SF(le)}$

Here $\hat{x}$ and $\hat{y}$ specify the obligatory subject position and the optional object position of *leave*, respectively, identifying the semantic variables to be bound by the corresponding syntactic constituents. Technically, $\hat{x}$ and $\hat{y}$ can in fact be considered as lambda operators, abstracting over the pertinent variables, such that assigning theta roles, or argument positions for that matter, amounts semantically to functional application. For details of this aspect, see, for example, Bierwisch (1988).

2.3 Remarks on Modularity of Knowledge and Representation

The main reason to distinguish SF from syntactic representations, including LF, is the linguistically relevant internal structure of lexical items connected to the conceptual interpretation of linguistic expressions. The compositional structure claimed for SF is very much in line with the proposals of Jackendoff (1983, 1987, and chapter 1, this volume) about conceptual structure (CS), with one important difference, however, which has consequences for the relation of language and space. The problem is this. Although what Jackendoff calls “lexical conceptual structure” (LCS) is—details aside—very close in spirit to the SF information $SF(le)$ of lexical items, he explicitly claims that conceptual structure (CS; and hence LCS) is an extralinguistic level of representation. In other words, CS is held to be external to I-language. Hence CS must obviously be identified with C-I (or perhaps a designated level of C-I).³ The architecture sketched in (11) is thus to be replaced by something like (19):



Jackendoff proposes a principled distinction between systems or modules of representation supporting the levels of representation indicated by the labels in (19), and interface systems or correspondence rules represented by the arrows. This proposal is connected to what he calls “representational modularity,” suggesting that autonomy of mental modules is a property of representational systems like phonological structure (PS), syntactic structure (SS), conceptual structure (CS; but also articulation, vision, etc.), rather than complex faculties like I-language. Autonomous modules of this sort are then connected to each other by interface or correspondence systems,

which—by definition—cannot be autonomous, as it is their very nature to mediate between modules.

I-language, in Jackendoff's conception, comprises PS, SS, and the correspondence rules connecting them to their adjacent levels, but not CS. The bulk of correspondence rules relating PS and SS, on the one hand, and SS and CS, on the other, are lexical items. While this is a plausible way to look at lexical items, it creates a conceptual problem. How can lexical items as part of the correspondence rules belong to I-language, if $SF(le)$, or rather LCS, does not? To put it differently, either CS (and hence LCS) is not included in I-language or lexical items belong to the system of correspondence rules included in linguistic knowledge, but not both.⁴

One might, of course, argue that the problem is not conceptual, but merely terminological, turning on the appropriate characterization of I-language, which simply cannot be schematized as in (22); the lexical system not only cuts across the subsystems within I-language, but also across language and other mental systems. I do not think this is the right solution, though, for at least three reasons.

First, there seem to be substantial generalizations that crucially depend on the linguistic nature of $SF(le)$, the principles of argument-structure being a major case in point. (This is a contention clearly shared by Jackendoff.) In this respect, then, $SF(le)$ is no less part of I-language than $PF(le)$, or even $GF(le)$.

Second, the phenomena discussed above in connection with the interpretation of *leave*, *enter*, *institute*, and so on could not reasonably be explained without accounting for their fairly abstract linguistic structure and the specific distinctions that depend on factual knowledge. In other words, there seems to be a systematic distinction between linguistic and extralinguistic factors determining conceptual and referential interpretation. If these distinctions are not captured by two levels of representation—SF and C-I in my terminology—then two aspects of CS must be distinguished in somewhat similar ways. But this would spoil the modular autonomy of CS and its extralinguistic status.

Third, the nature of correspondence rules in general remains rather elusive. To some extent, they must belong to the core of linguistic knowledge based on the principles of UG, but they appear also to depend on quite different principles of mental organization. Although one might argue that this is just a consequence of actual fact, that linguistic knowledge is not a neatly separated system of mental organization, it seems to me this conclusion can and in fact must be avoided.

Let me return, in this regard, to the initial claim schematized in (7), namely that I-language (based on UG) simply determines a systematic correspondence between the domains A-P and C-I. In this view, I-language is altogether a highly specific interface mediating two independent systems of computation and representation.

Under this perspective, PF and SF are theoretical constructs sorting out those aspects of A-P and C-I that are recruited by UG in order to compute the correspondence in question. Hence PF(*le*) and SF(*le*) represent structural conditions projected into configurations in A-P and C-I, respectively. There are no correspondence rules connecting SF(*le*) to its conceptual interpretation, or PF(*le*) to articulation for that matter. Rather, the components of PF(*le*) and SF(*le*) as such provide the interface of A-P and C-I with the language-internal computation. It is the aim of this chapter to make this view more precise with respect to the subdomain of C-I representing space.

Notice, first of all, that the difficulties concerning the status of CS are largely due to the notion of representational modularity, which is intended to overcome the inadequacies encountered by Fodor's (1983) concept of modularity. Replacing the overall language module by a number of representational systems, each of which is construed as an autonomous module, Jackendoff is forced to posit interface systems as well. Instead of speculating about the nature of these intermodular systems (are they supposed to be encapsulated and impenetrable?), I suggest we go back to the notion of modularity first proposed by Chomsky (1980), characterizing systems and subsystems of tacit knowledge, rather than levels of representation.

The notion of level of representation need by no means coincide with that of an autonomous module. To be sure, there is no system of knowledge without representations to which it applies. But neither must one module of knowledge be restricted to one level of representation, nor must a level of representation belong to only one module of knowledge. I will not go here through the intricate details of subsystems and levels of syntactic representation, where no simple correlation between levels and modules obtains. Instead, I want to indicate that, in a more general sense, different systems of rules or principles can rely on the same system of representation, determining, however, different aspects of actual representations. What I have in mind might best be illustrated by examples from different nonlinguistic domains. A simple case is the representational system consisting of sequences of digits. The same sequence, say 12121942, might happen to be your birth date, your office phone number or your bank account. Each of these interpretations belongs to a different subset of sequences, subject to different restrictions. For none of them is the fact that the number is divisible by 29 relevant; each subset defines different neighbors, different constituents, and so on. Such interpretations of the same representation are based on different rules or systems of knowledge, exploiting the same representational resources. Notice that certain operations on the representation would have the same effect for each of the interpretations, because they affect the shared properties of the representational system, while others would have different effects on alternative recruitings, as illustrated in (20a) and (20b), respectively:

- (20) a. 12121942 74612171 54915151 ...
 b. 19421212 06061942 24241942 ...

Another type of example is illustrated by the famous theme in Bach's *Art of the Fugue*:

(21)

The notes exhibit simultaneously a position within the tonal system and, because, of their “names,” within the Latin alphabet. Again, different rules apply to the two interpretations. This case is closer to what I want to elucidate than the different interpretation of digits. First, the tonal and the graphemic interpretation of the representation apply simultaneously, albeit under different interpretations. Second, the two interpretations rely on different cutouts of the shared representation. Although all notes have alphabetic names, not all letters are representable by notes.⁵ Third, the more complete interpretation (in this case the tonal one) determines the full representation, from which the additional interpretation recruits designated components, imposing its own constraints.⁶

Obviously, even though these illustrations are given in terms of external representations, it is the internal structures and the pertinent knowledge they are based on that we are interested in. In this respect, digits and notes are comparable to language, exhibiting an E- and an I-aspect. Moreover, while the examples rely on rules and elements that are more or less explicitly defined, knowledge of language is essentially based on tacit knowledge. However, the artificial character of the twofold interpretation in our examples by no means excludes the existence of the same structural relationship with respect to systems of implicit knowledge. In other words, the conceptual considerations carry over to I-language as well as other mental systems.

It might be objected that the representations considered above are not really identical under their different interpretations, especially if we try to identify the information contained in their I-representation: digits representing dates are grouped according to day, month, and year; telephone numbers, according to extensions; and so forth. In other words, the relevant elements—digits, notes, and so on—must be construed as annotated in some way with respect to the rules of different systems of knowledge. This seems to me correct, but it does not change the fact that we are dealing with annotations imposed on configurations of the same representational system. Both aspects—identity of the representational system and indication of specific affiliation—are crucial with respect to the way in which different modules of knowledge are

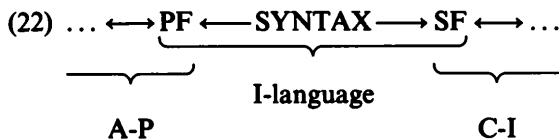
interfaced by a given representational system. These considerations lead to what might be called “modularity of knowledge,” in contrast to Jackendoff’s “representational modularity.” The moral should be obvious, but some comments seem to be indicated.

First, the notion of interface—or correspondence for that matter—is a relative concept, depending on which modules are at issue. I-language as a whole is a system that establishes an interface between A-P and C-I, with language capacity based on UG providing the requisite knowledge. Furthermore, I-language must be interfaced with A-P and C-I, respectively. This sort of interface is not based on rules that map one representation onto another, but rather on two types of knowledge that participate in one and the same representational system. In other words, PF and SF are the interfaces of I-language with A-P and C-I, respectively, which does not exclude the possibility that A-P or C-I support further levels of representation, as we will see below.

Second, if this is so, then the levels of PF and SF are each determined by (at least) two modules of knowledge, imposing conditions on, or recruiting elements of, each other, possibly adding annotations in the sense mentioned above. One might, of course, distinguish different aspects of one representation by setting up different levels of representation. While this may be helpful for descriptive purposes, it must not obscure the shared elements and properties of the representational system.

Looking more specifically at PF(*le*) under this perspective, we recognize PF(*le*) as the linguistic aspect imposed on A-P. It is based on temporal patterns determined by articulation and perception, which include various aspects such as effects of the particular speaker’s voice, emotional state, and so on. These are determined by their own subsystems but are, so to speak, ignored by I-language.

Turning finally to SF(*le*), which is of primary interest here, we will now recognize it as the designated aspect of C-I to which I-language is directly related, using configurations of C-I as elements of its own, linguistic representation. This leaves open various possibilities concerning (1) how SF components recruit elements or configurations of C-I; (2) what annotations of SF must be assumed; and (3) how rules and principles of C-I will contribute to the interface representation without being reflected in I-language. We will turn to these questions in the sections below. To conclude this section, I want to schematize the view proposed here by a slight modification of (8):



The main point is, of course, that SF is governed by conditions of I-language as well as those of C-I, although the aspect concerned need not be identical. (Parallel considerations apply to PF.) The dots in (22) indicate the (largely unknown) internal organization of C-I, to which we turn now.

2.4 The Conceptualization of Space

What interests us is the internal representation of space and the knowledge underlying it, which we might call "I-space," corresponding to I-language, and contrasting with physical, external or "E-space." I-space in this sense must be assumed to control and draw on information from a variety of sources; it is involved primarily in visual perception and kinesthetic information, but it also integrates information from the vestibular system, auditory perception, and haptic information. All these systems provide nonspatial information as well. Vision integrates color and texture; haptic and kinesthetic information distinguish, among other things, plasticity and rigidity; and so forth. I will therefore assume, following Jackendoff (chapter 1, this volume), that I-space selects information from different sources and integrates it in a particular system of spatial representation (SR). As a first approximation, SR should thus be construed as an interface representation in the sense just discussed; that is, as mediating between different perceptual and motoric modalities, on the one hand, and the conceptual system C-I, on the other, comparable to the way in which PF reconciles articulation and audition with I-language. Before looking more closely at the status of SR and its role for the relation between I-space and I-language, I will provisionally indicate the format and content to be assumed for SR.

According to general considerations, SR should meet the following conditions:

1. SR is based on a (potentially infinite) set of locations, related according to three orthogonal dimensions, with a topological and metrical structure imposed on this set.
2. Locations can be occupied by spatial entities (physical objects and their derivatives like holes, including regions, or shadows, substances, and events), such that $\text{Loc}(x)$ is a function that assigns any spatial entity x its place or location. Spatial properties of physical entities are thus related to the structure imposed on the set of locations.
3. In general, $\text{Loc}(x)$ must be taken as time-dependent, such that more completely $\text{Loc}(x, t)$ identifies the place of x at time t , presupposing standard assumptions about time intervals. (Motion can thus be identified by a sequence of places assigned to the same x by $\text{Loc}(x, t)$.)
4. In addition to dimensionality, topological structure, and metrical structure, two further conditions are determined for locations:

- a. orientation of the dimensions, marking especially a directed vertical dimension (based on gravitation);
- b. orientation with respect to a designated origo and/or observer and intrinsic conditions of objects (canonical position or motion).

Depending on how physical objects are perceived and conceptualized, the dimensionality of their locations can be reduced to two, one, or even zero dimensions. All of this would have to be made precise in a serious theory of SR. The provisional outline given by conditions 1–4 above can serve, however, as a basis for the following remarks.

Notice that although SR is transmodal in the sense already mentioned and must be considered as one of the main subsystems that contribute to the conceptual and intentional organization of experience, it should still clearly be distinguished from the level of conceptual structure (CS) for at least two interrelated reasons. First, SR is assumed to be domain-specific, representing properties and distinctions that are strictly bound to spatial experience, while conceptual structure must provide a representation for experience of all domains, including not only color, taste, smell, and auditory perception, but also emotion, social relation, goals of action, and so on, that is, information not bound to sensory domains in direct ways. Second, the type of representation at SR is depictive of or analogous to what it represents in crucial respect, while CS is abstract, propositional, algebraic, that is, nondepictive. All that is needed for a representational system to be depictive is a “functional space” in the sense explained in Kosslyn (1983), which we have in fact assumed for SR in conditions 1 and 2.

Because the distinction between the depictive nature of SR and the propositional character of CS is crucial for the further discussion, let me clarify the point by the following simplified example:

- (23) a. $\bigcirc$ b. i. $A \text{ OVER } B \ \& \ B \text{ LEFT-OF } C$
 $\square\Delta$ ii. $A \text{ OVER } B \ \& \ C \text{ RIGHT-OF } B$
 iii. $B \text{ LEFT-OF } C \ \& \ B \text{ OVER } A$
- (24) a. A corresponds to $\bigcirc$. B corresponds to $\square$. C corresponds to Δ .
 b. $x \text{ OVER } y$ corresponds $\text{Loc}(x)$



(23a) is a pictorial representation of a situation for which (23b) gives three possible propositional representations, provided the correspondences indicated in (24)—the conceptual lexicon—apply, together with the principles that relate the “functional structure” underlying (23a) to the compositional structure of the representations in

(23b). Presupposing an intuitive understanding of the correspondence in question, which could be made precise in various ways, I will point out the following essential differences between the format of (23a) and (23b):

1. Whereas there is an explicit correspondence between units representing objects in (23a) and (23b)—established by (24a)—there are no explicit units in (23a) representing the relational concepts OVER, LEFT OF, and so on in (23b), nor are there explicit elements in (23b) representing the properties of the objects in (23a), that is, the circle, the square, and so on.
2. The different distance between the objects is necessarily indicated in (23a), even though in necessarily implicit way; it is not indicated in (23b), where it could optionally be added but only in necessarily explicit manner (e.g., by adding coded units of measurement).
3. Additional properties or relations specified for an object in (23b) require a repeated representation of the object in question, while no such “anaphoric” repetition shows up in (23a); for the same reason, (23b) requires logical connectives relating the elementary propositions, while no such connectives may appear in (23a).
4. Finally, (23b) allows for various alternative representations corresponding equivalently to the unique representation in (23a), while (23a) would allow for variations that need not show up in (23b), for example, by different distances between the objects.

In general, the properties of (23a) are essentially those of mental models in the sense discussed by Johnson-Laird (1983, and chapter 11, this volume) and by Byrne and Johnson-Laird (1989), who demonstrate interesting differences between inferences based on this type of representation, as opposed to inferences based on propositional representations of type (23b). Returning to SR, it seems to be a plausible conjecture that it constitutes a pictorial representation in the sense of (23a), with objects represented in terms of 3-D models in the sense of Marr (1981), or configurations of geons as proposed by Biederman (1987). See Jackendoff (1990, and chapter 1, this volume) for further discussion. It differs from CS by formal properties like 1. to 4., allowing for essentially different operations based on its depictive character, which supports an analogical relation to conditions of E-space.

The next point to be noted is that SR as construed here is a level of representation, not necessarily an autonomous module of knowledge. Given the variety of sources it integrates, it seems in fact plausible to assume that SR draws on different systems of mental organization. According to the view proposed in the previous section, SR might rather be considered as one aspect of a representational system shared by different modalities, visual perception providing the most fundamental as well as the

most differentiated contribution. This leaves open whether, and to what extent, the SR aspect of the representational system is subject to or participates in operations like imaging or mental rotation of objects, which are argued by Kosslyn et al. (1985) to be not only depictive, but also modality-specific.

This leaves us with the question of how SR relates to the overall system C-I and the level of conceptual structure in particular. If the comments on the propositional character of CS and the depictive nature of SR are correct, then SR and CS cannot be two interlocked aspects of the same level of representation. On the other hand, SR must belong to C-I, because to the extent to which it is to be identified with Johnson-Laird's system of mental models, it supports logical operations similar in effect to those based on the propositional-level CS, albeit of a different character. The obvious conclusion is that C-I comprises at least two different levels of representation. This conclusion should not be surprising; it has in fact a straightforward parallel in I-language, where PF and SF also constitute two essentially different representational systems within the same overall mental capacity.

To carry this analogy one step further, what I have metaphorically called the "conceptual lexicon" (24) corresponds in a way to the lexical entries. Just as PF(*le*) indicates how the corresponding SF(*le*) is to be spelled out at the level of PF, the pertinent 3-D model determines the representation of a given concept on the level of SR.

More generally, and in a less metaphorical vein, the correspondence between SR and CS must provide the SR rendering of the following specifications for spatial conditions:

1. Shape of objects, that is, proportional metrical characteristics of objects and their parts with respect to their conceptually relevant axes or dimensions (3-D models);
2. Size of objects, that is, metrical characteristics of objects interacting with the relevant shape characteristics;
3. Place of objects, that is, relations of objects with respect to the location of other objects; and
4. Paths of (moving) objects, that is, changes of place in time.

Obviously, specifications 1–4 are not independent of each other. Shape, for instance, is to some extent determined by size and place of parts of an object; paths—as already mentioned—are sequences of places; and so forth. Jackendoff (chapter 1, this volume) points out further aspects and requirements to be added, which I need not repeat here. The main purpose of the outline given above is to indicate the sort of CS information that SR is to account for, without trying to actually specify the format of representations, let alone the principles or rules by which the relevant knowledge is organized.

I will conclude this sketch of the status of I-space with two comments that bear on the way spatial information is conceptually structured and eventually related to SF, and hence to I-language. First, it is worth noting that commonsense ontology, namely, the sortal and type structure of concepts, is entrenched in some way in I-space. More specifically, the informal rendering of SR in conditions 1–4 at the beginning of this section freely refers to objects, events, places, properties, relations, and so on legitimately, or in fact necessarily, I suppose, because the corresponding ontology holds also for SR. This observation, in turn, is important for two reasons: (1) in spite of its domain specificity, SR shares with general conceptual organization basic ontological structures; and (2) by virtue of this common ground, SR not only provides entities in terms of which intended reference in C-I can be established and interpreted; it also participates in a general framework that underpins the interface with general conceptual structure. I will assume, for example, that 3-D models spell out properties in SR that general conceptual knowledge combines with nonspatial knowledge about specific types of physical objects. Thus the commonsense theory about cats will include conditions about the characteristic behavior, the taxonomic classification, and so forth of cats, along with access to the shape as specified in SR. I will return to this problem in the next section.

My second comment has two parts. (1) I want to affirm that spatial representation as discussed thus far responds to properties and relations of physical objects, that is, to external conditions that constitute real, geometrical space. We are dealing with space in the literal sense, one might say, based on spatial perception of various sorts, as mentioned above. This leads to (2) the observation that spatial structures are extensively employed in many other conceptual domains. Time appears necessarily to be conceptualized as one-dimensional, oriented space with events being mapped onto intervals just like objects being mapped onto locations. Hierarchies of different sorts, such as social, evaluative, taxonomic, and so on, are construed in spatial terms; further domains—temperature, tonal scales, loudness, color—come easily to mind. More complex analogies in the expression of spatial, temporal, possessional, relations have been discussed, for example, by Gruber (1976) and by Jackendoff (1983). The conclusion from this observation is this. The basic conditions of I-space as listed at the beginning of this section seem to be available as a general framework underlying different domains of experience, which immediately raises the question of how this generalized character of basic spatial structures is to be explained. Because taxonomies, social relations, and even time do not rely on the same sources of primary experience, the transmodal aspect in question clearly must exceed I-space (in the sense assumed thus far), functioning as an organizing structure of general conceptual knowledge.

Basic structures of spatial organization must therefore either

1. constitute a general schema of conceptual knowledge imposed on different domains according to their respective conditions; or
2. originate as an intrinsic condition of I-space and are projected to other domains on demand.

According to alternative 1, actual three-dimensional space is the prevailing, dominant instantiation of an abstract structure that exists in a sense independent of this instantiation; according to alternative 2, the structure emerges as a result of experience in the primary domain. The choice between these alternatives has clear empirical impact in structural, ontogenetic, and phylogenetic respects, but it is a difficult choice to make, given the present state of understanding of conceptual structure. I tentatively assume that alternative 2 is correct for the following two reasons: (1) I-space is not only a privileged instantiation of spatial structure but is also the richest and most detailed instantiation of spatial structure, compared to other domains. Whereas I-space is basically three-dimensional, other domains are usually of reduced dimensionality, as Jackendoff (chapter 1, this volume) remarks. Orientation with respect to frame of reference is accordingly reduced to only one dimension. (2) While size and place carry over to the other domains with scalar and topological properties, shape has only very restricted analogy in other domains. I will thus consider the full structure of I-space as intrinsic to this domain due to its specific input, rather than as an abstract potential that happens to be completely instantiated in I-space only. These structural considerations might be supplemented by ontogenetic and phylogenetic considerations, which I will not pursue here.

In any case, whether imported to I-space according to alternative 1, or exported from it according to alternative 2, dimensionality and orientation require appropriate structures of other domains, or rather of conceptual structure in general, to correspond to. This is similar to what has been said earlier with respect to commonsense ontology, with its type and sortal distinctions.

It might be useful to distinguish two types of transfer of spatial structure. I will consider as *implicit transfer* the dimensionality and orientation of domains like time or social hierarchies, whose conceptualization follows these patterns automatically, that is, without explicit stipulation. In contrast, *explicit transfer* shows up in cases where dimensionality is used as a secondary organization, imposing an additional structure on primary experience. The notion of color space or property space is based on this sort of explicit transfer. The boundary between explicit and implicit transfer need not be clear in advance and might in fact vary to some extent, which would be a natural consequence of alternative 2. In what follows, I will not deal with explicit transfer but will argue that implicit transfer is a major reason for the observation

noted at the outset, namely, that there is no clear distinction between spatial and nonspatial terms. The relations expressed, for example, by *in*, *enter*, or *leave* are not restricted to space because of the implicit transfer of the framework on which they are based.

2.5 Types of Space Relatedness in Conceptual Structure

Let us assume, to conclude the foregoing discussion, that the conceptual-intentional system (C-I) provides a level of representation (CS) by which information of different modules is integrated, looking more closely at the way in which spatial information is accommodated in CS. Notice first of all that assumptions about the properties of CS can only be justified by indirect evidence because, by definition, CS depends on various other systems relating it to domain-specific information. There seems to be general agreement, however, that CS is propositional in nature, in the sense indicated above and discussed in more detail, for example, by Fodor (1975) and by Jackendoff (1983, 1990, and chapter 1, this volume). The two main sources relied on in specifying CS are language and logic. On the one hand, CS is modeled as tightly as possible in accordance with the structure of linguistic expressions to be interpreted in CS; on the other hand, it is made to comply with requirements of logical inferences based on situations and texts.

As to the general format of CS, two very general assumptions will be sufficient in the present context. First, CS is based on functor-argument-structure, with functional application being the main (and perhaps only) type of combinatorial operation. Hence CS does not rely on sequential ordering of elements but only on nesting according to the functor-argument structure. There are various ways in which these assumptions can be made precise, a particularly explicit version being Kamp and Reyle (1993). Second, I will suppose that CS exhibits a fairly rich sortal structure provided by commonsense ontology. Both assumptions should allow CS to be interfaced with the semantic form (SF) of linguistic expressions, as discussed earlier.

I will refrain from speculations about the primitive elements of CS, with two exceptions: (1) the primes of SF must be compatible with basic or complex units of CS, if the assumptions about SF and its embedding in CS are correct; and (2) CS must accommodate information from various domains, including SR, possibly treating for example, specifications of 3-D models as basic elements that feature in CS representations. I will return to exception 2 shortly.

Note, furthermore, that CS must not be identified with encyclopedic knowledge in general. Although commonsense theories by which experience is organized and explained must have access to representations of CS, their format and organization are

to be distinguished from bare representational aspects of CS. It has been suggested (e.g., Moravcsik 1981; Pustejovsky 1991) that commonsense theories are organized by explanatory factors according to Aristotelian categories like structure, substance, function, and relation. It remains to be seen how this conjecture can be made explicit in the formal nature of commonsense knowledge. Pending further clarification, I will simply assume that C-I determines relevant aspects of CS on the basis of principles that organize experience.

Turning next to the way in which CS and commonsense knowledge integrate I-space, three observations seem to me warranted:

1. Commonsense ontology requires physical entities to exhibit spatial characteristics, including in particular shape and size of objects and portions of substance.

This observation distinguishes “aspatial” conceptual entities—mental states, informational structures (like arguments, songs, or poems), and social institutions—from those subject to spatial characterization. Although these aspatial entities are invested with spatial characteristics by the physical objects implementing them, it should be clear enough that, for example, a poem as a conceptual entity is to be distinguished from the printed letters that represent it.

2. Encyclopedic knowledge may or may not determine particular definitional or characteristic spatial properties within the limits set by (1).

This observation simply notes that spatial entities are divided into those whose typical or essential properties involve spatial characteristics, and those without specifications of this sort. *Dog, snake, child, table, or pencil* express concepts of the first type, while *animal, plant, tool, furniture* exemplify concepts of the second type, which, although inherently spatial, are not characterized by particular spatial information. Actually observation 2 does not set up a strictly binary, but rather a gradual distinction, depending on the specificity of shape, size, and positional information. Thus the concept of *vehicle* is spatially far less specific than that of *cat* or *flute*, but it still contains spatial conditions absent in the concepts of *machine* or *musical instrument*, even though these are not aspatial. Also, the specificity of spatial properties seems to vary in the course of ontogenetic development, as Landau (chapter 8, this volume) argues, showing that young children initially tend to invest concepts in general with spatial information.

3. Conceptual units may specify spatial properties or relations without involving any nonspatial properties of entities they can refer to.

While observations 1 and 2 distinguish conceptual entities with respect to their participation in spatial conceptualization, observation 3 separates conceptual units that specify purely spatial conditions for whatever entities fall within their range from conditions that inextricably involve additional conceptual information. Thus *square*,

edge, circle, top (in one reading) express strictly or exclusively spatial concepts while *dog* or *cup* include—in addition to shape and size information—further systematic conceptual knowledge.

It should be borne in mind that we are talking here about conceptual units, using linguistic expressions only as a convenient way of indication. For the time being, we ignore variability in the interpretation of lexical items, which might be of various sorts. Thus lexical items expressing strictly spatial concepts are extensively used to refer to “typical implementations” like *corner, square, margin*, and so on. Expressions for aspatial concepts, on the other hand, for example, social institutions like *parliament* or informational structures like *novel* or *sonata*, are used to refer to spatial objects where they are located or represented, as already mentioned. These are problems of conceptual shift of the sort mentioned in section 2.2, which must be analyzed in their own right.

The different spatial character of concepts discussed thus far can be schematically summarized as follows:

(25) Type of concept	Example
a. Aspatial	<i>fear, hour, duration</i>
b. Extrinsically spatial	<i>animal, robot, instrument</i>
c. Intrinsically spatial	<i>horse, man, violin</i>
d. Strictly spatial	<i>square, margin, height</i>

Observation 1 distinguishes between (25a) and (25b–d); observation 2 separates (25d) from (25a–c). “Extrinsically spatial” refers to concepts that require spatial properties but do not specify them; “intrinsically spatial” indicates the specification of (some of) these properties. It should be noted that intrinsically spatial properties might be typical or characteristic, without being definitional in the strict sense. See Keil (1987) for relevant discussion. As already mentioned, the distinction between (25b) and (25c) is hence possibly to be replaced by various steps according to the specificity of spatial information. The main point is that concepts can involve more or less specific spatial information, but need not fix it, even if they are essentially spatial.

It is worth noting that the same distinctions (with similar provisos) apply to other domains of conceptual organization, color and time being cases in point:

(26) Type of color-relatedness	Example
a. No relation	<i>live, hour, height</i>
b. Extrinsic	<i>liquid, animal, tool</i>
c. Intrinsic	<i>blood, zebra, sky</i>
d. Strict	<i>red, black, colorlessness</i>

(27) Type of time-relatedness	Example
a. No relation	<i>number, water, lion</i>
b. Extrinsic	<i>fear, commettee, travel</i>
c. Intrinsic	<i>death, inauguration, beat</i>
d. Strict	<i>hour, beginning, duration</i>

There are numerous problems in detail, which would have to be clarified with respect to the particular domains in question. The point at issue is merely that the observations 1–3 noted above are not an isolated phenomenon of space.

Thus far I have illustrated the distinctions in question with respect to objects of different sorts. The observations apply, however, in much the same way to other ontological types, such as properties, relations, and functions; (28) gives a sample illustration:

(28)	Property	Relation
a. Aspatial	<i>clever, sober, famous</i>	<i>acknowledge, during</i>
b. Extrinsic	<i>colored, wet, solid</i>	<i>kill, show, write</i>
c. Intrinsic	<i>striped, broken, open</i>	<i>close, pierce, squeeze</i>
d. Strict	<i>upright, long, slanting</i>	<i>under, near, place</i>

Notice, once again, that we are talking about concepts, not about the nouns, verbs, adjectives, prepositions expressing them. In addition to distinctions blurred by this practice, further difficulties must be observed. Thus *long*, as shown in the appendix below, expresses actually a three-place relation, rather than a property. The main point should be clear, however. Concepts of different types are subject to the distinctions related to observations 1–3.

The distinctions discussed thus far are directly related to two additional observations important in the present context. First, there are, on the one hand, concepts with a fairly rich array of different conditions—Pustejovsky's (1991) "qualia structure," for example—integrated into theories of commonsense explanation. Concepts of natural kinds like *dog* or *raven*, but also artifacts like *car* or *elevator*, combine more or less specific shape and size information with knowledge about function, behavior, substance, and so on that might be gradually extended on the basis of additional experience. On the other hand, there are relatively spare concepts such as *near*, *square*, *stand*, based on highly restricted conditions of only one or two domains. Let me call these two kinds "rich concepts" and "spare concepts," for the sake of discussion. There is, of course, no sharp boundary here, but the difference is relevant in two respects: (1) spare concepts might in fact enter into conditions of rich concepts, with rich concepts being subject to further elaboration, while spare concepts are just what they are; and (2) it is essentially rich concepts that constitute commonsense theories:

although spare concepts like *in* or *long* can feature in explanations, they do not explain anything. Contrasting, for example, *record* and *circle*, we notice that *circle* is part of the shape information in *record*, which relies, however, on knowledge explaining sound storage (in varying degrees of detail), while nothing (beyond mere geometry) is explained by *circle*. For almost trivial reasons, the distinction of rich and spare concepts relates to (but is not identical with) the distinction between extrinsic and intrinsic spatial concepts, as opposed to strictly spatial concepts. Strictly spatial concepts can be integrated into intrinsically spatial ones, but not vice versa.

Related to this is the second observation. Specifications represented in SR can be relied on in CS in two ways, which I will call “explicit” and “implicit.” Detailed shape information, for instance, represented in SR by 3-D models, enters the pertinent concepts implicitly, which means that neither the internal structure of 3-D models nor the properties reconstructing them like “four-legged” or “long-necked” enter CS representations, but rather the shape information as a whole. In contrast, strictly spatial concepts like *behind*, *far*, *tall*, and so on must explicitly represent the relevant spatial conditions in terms of conceptual primitives. One might take this as a corollary of the classification illustrated in (25) in the following sense:

Strictly spatial concepts represent spatial information explicitly in terms of conceptual primes; intrinsically spatial concepts represent spatial information implicitly, that is, encapsulated in configurations of SR.

The moral of all of this with respect to our initial question would thus be something like the following. CS extracts information from SR in two ways: (1) encapsulated in SR configurations that are only treated holistically, defining, so to speak, an open set of primes in terms of conditions in SR, and (2) explicitly represented by means of conceptual primes that directly recruit elements of SR. Because we have further assumed that CS is the interface of C-I with I-language, it follows that SF has two types of access to SR. I will return to this point below. Although I take this moral to be basically correct as a kind of guideline, there are essential provisos to be made, even if the notion of explicit and implicit representation can be made formally precise, and even if the usual problems with borderline cases can be overcome.

A major problem to be faced in this connection is the fact that in CS strictly spatial (i.e., explicit) concepts must appropriately combine with implicit spatial information. Thus, for the complex concepts expressed by *short man*, *long table*, or *steep roof*, the strictly spatial information of *short*, *long*, or *steep* must be able to extract the relevant dimensional and orientational information from the encapsulated shape representation of *man*, *table*, or *roof*. A useful proposal to overcome this problem is the notion of object schemata developed in Lang (1989). An object schema specifies the conditions that explicit representations could extract from encapsulated shape informa-

tion, in particular, dimensionality, canonical orientation and subordination of axes relative to each other. Even though an object schema is less specific than a 3-D model, it is not just a simplification of the model, but rather its rendering in terms of primes of the strictly spatial sort. An object schema makes 3-D models respond to explicitly spatial concepts, so to speak. Notice that there are default schemata also for extrinsically spatial concepts that do not provide a specified 3-D model, as combinations like *long instrument* show. For details see Bierwisch and Lang (1989) and Lang (1989).

A final distinction emerging from the observations about I-space and C-I should be noted. As a consequence of the implicit transfer imposing basic structures of I-space on other domains, which we noted above, it seems plausible to assume that explicitly spatial concepts like *in*, *length*, and *around* do in fact relate to I-space and other domains to which the pertinent structures are transferred. In other words, we are led to a distinction between elements of CS that are exclusively interpreted in SR and elements that are neutral in this respect, being interpreted by structures of SR that transfer to other domains. The latter would include only explicit concepts, which are strictly spatial only if interpreted in I-space.

Not surprisingly, we found a fairly rich typology of different elements and configurations thereof in CS, depending only on the way in which SR as a representational system relates to I-space as well as other cognitive domains. I would like to stress that the observations from which this typology derives, are not stipulated conditions but simply consequences of basic assumptions about the architecture of subsystems of C-I and their internal organization.

2.6 Basic Spatial Terms: Outline of a Program

Assuming that the relation of spatial cognition and conceptual structure is to be construed along the lines sketched thus far, the central question we posed at the outset boils down to two related questions:

1. How is I-space reflected in CS?
2. How are spatial aspects of CS taken up in SF?

We have already dealt with question 1. A partial answer to question 2 is implied by the assumption that SF and CS, although determined by distinct and autonomous systems of knowledge, need not be construed as disjoint representational systems, but rather as ways to recruit pertinent configurations according to different modules of knowledge. Pursuing now question 2 in more detail, I will stick to the assumption made earlier, that SF can be thought of as embedded in CS, such that the conditions on the format of SF representations outlined in section 2.2 would carry over to the format of CS, unless specific additional requirements are motivated by independent

evidence concerning the nature of CS. Such additional requirements might relate, for example, to commonsense ontology and the sortal system it induces.

With these prerequisites, the main issue raised by question 2 is which elements of CS are recruited for lexicalization in I-language. An additional point concerning further grammaticalization in terms of morphological categories will be taken up in section 2.7. I will restrict the issue of lexicalization to strictly spatial concepts for two reasons: (1) to go beyond obvious, or even trivial, statements with respect to encapsulated information of intrinsically spatial concepts, including the intervening effects of object schemata, would by far exceed the limits of this chapter; and (2) understanding the lexicalization of strictly spatial concepts would be a necessary precondition in any case.

Given these considerations, the following research strategy seems to be promising, and has in fact been followed implicitly by a great deal of research in this area. First we define the system of basic spatial terms (BST, for short) of a given language, and then we look at the properties they exhibit with respect to question 2. The notion of basic spatial terms has been borrowed from Berlin and Kay's (1969) basic color terms and is similar in spirit, though different in certain respects. Because space is a far more complex than color, BSTs cannot, for example, be restricted to adjectives, as basic color terms can.

Basic spatial terms can be characterized by the following criteria:

1. BSTs are lexical items [PF(*le*), GF(*le*), SF(*le*)] that belong to the basic (i.e., morphologically simple), native, core of the lexical system of a given language;
2. In their semantic form [SF(*le*)], BSTs identify strictly spatial units in the sense discussed above.

Thus *short*, *under*, *side*, *lie* are BSTs, while *hexagonal* and *squeeze* are not, violating criterion 1 and criterion 2, respectively. It should be emphasized that BST is a purely heuristic notion with no systematic impact beyond its role in setting up a research strategy. Hence one might relax or change the criteria should this be indicated in order to arrive at relevant generalizations or insights. Thus my aim in assuming these criteria is not to justify the delimitation they define, but rather to rely on them for practical reasons.

It is immediately obvious that the two criteria, even in their rather provisional form, lead to various systematically related subsystems of BSTs:

1. Linguistically, BSTs belong to different syntactic and morphological categories (verbs, nouns, prepositions, adjectives, and perhaps classifiers and inflections for Case);
2. Conceptually, BSTs are interpreted by different aspects of space (size, shape, place, change of size, motion, etc.).

Of particular interest is, of course, the relation between linguistic (1) and conceptual (2) subsystems, whether systematic or incidental. Ultimately, a research strategy taking BSTs as a starting point is oriented toward (at least) three aims, all of which are related to our central question:

- Identification of the conceptual repertoire available to BSTs. This includes in particular the question whether universal grammar provides an a priori system of potential conceptual distinctions that can be relied on in the SF of BSTs—parallel to what is generally assumed for PF primes—or whether the distinctions made in SF are abstracted from actual experience and its conceptualization.
- Identification of basic patterns, either strict or preferential, by which UG organizes BSTs with respect to their SF, as well as their syntactic and morphological properties.
- Identification of systematic options that distinguish languages with respect to the repertoire and the patterns they rely on. This problem might be couched in terms of parameters allowing for a restricted number of options, or simply as different ways to idiosyncratically exploit the range of possibilities provided by principles of C-I and UG.

As a preliminary illustration, I will have a look at the reasonably well understood structure of dimensional adjectives (DAs, for short) like *long*, *high*, *tall*, *short*, and *low*, the interpretation of which combines conditions on shape and size. Generally speaking, a DA picks out a particular, possibly complex, dimensional aspect of the entity it applies to and assigns it a quantitative value. Characteristically, DAs come in antonymous pairs like *long* and *short*, specifying somehow opposite quantitative values with respect to the same dimension. Thus the sentences in (29) state that the maximal dimension of the boat is above or below a certain norm or average, respectively:

- (29) a. The boat is long.
b. The boat is short.

The opposite direction of quantification specified by antonymous DAs creates rather intriguing consequences, however, as can be seen in (30):

- (30) a. The boat is twenty feet long and five feet wide.
b. *The boat is ten feet short and three feet narrow.
c. The boat is ten feet longer than the truck.
d. The boat is ten feet shorter than the truck.

In other words, a measure phrase like *ten feet* can naturally be combined only with the “positive” DA—hence the deviancy of (30b)—except for the comparative, where it combines with the positive as well as the negative DA. These and a wide range of

other phenomena discussed in Bierwisch (1989) can be accounted for, if DAs are assumed to involve three elements: (1) an object x evaluated with respect to a specified dimension; (2) a value v to be compared with; and (3) a difference y by which x either exceeds or falls short of v . While x and y are bound to argument positions to be filled in by syntactic constituents the DA combines with, v is left unspecified in the positive and made available for a syntactically explicit phrase by the comparative morpheme. Using the notational conventions illustrated in (18), the following entries for *long* and *short* can be given:

(31) /long/ Adj $\hat{x}(\hat{y}) [[\text{QUANT} [\text{MAX } x]] = [v + y]]$

|
Deg

(32) /short/ Adj $\hat{x}(\hat{y}) [[\text{QUANT} [\text{MAX } x]] = [v - y]]$

|
Deg

As in (18), the entry for *leave*, $\hat{x}$ and $\hat{y}$ are operators binding semantic variables to syntactic arguments, where the optional degree complement is morphologically marked by the grammatical feature Deg that selects measure phrases and other degree complements. Semantically, *long* and *short* are identical except for the different functor $+$ as opposed to $-$. The common functor MAX picks up the maximal dimension of the argument x , which then is mapped onto an appropriate scale by the operator QUANT. The scalar value thus determined must amount to the sum or difference of v and y , where the choice of the value for v is subject to rather general semantic conditions responsible for the phenomena illustrated by (29) and (30). One option for the choice of the variable v is N_C , indicating the norm or average of the class C which x belongs to. It accounts for the so-called contrastive reading that shows up in (29), while in (30) v must be specified as the initial point 0 of the scale selected by QUANT.

Three points can be made on the basis of this fairly incomplete illustration. First, the semantic form of dimensional adjectives, providing one type of BSTs, has a nontrivial compositional structure in the sense introduced in section 2.2, from which crucial aspects of the linguistic behavior of these items can be derived. Second, the elements making up the SF of these items have an obvious interpretation in terms of the structural conditions provided by SR, even though this interpretation is anything but trivial. Especially the way in which MAX and other dimensional operators like VERT or SEC for the vertical or secondary dimension of x are to be interpreted follows intricate conditions spelled out in detail in Lang (1989). Third, the entries (31) and (32) immediately account for the fact that *long* and *short* apply not only to spatial

entities in the narrower sense but to all elements for which a maximal dimension is defined, such as *a long trip*, *a short visit*, *a long interval*, and so on, due to the projection of spatial conditions to other domains in the sense discussed above. Note that the choice of the scale and its units determined by QUANT must be appropriately specified as a consequence of the interpretation of MAX. I will place this initial illustration of BSTs in a wider perspective in the appendix, looking at further conditions for basic patterns and their variation.

2.7 Grammaticalization of Space

The elements and configurations considered thus far are supposed to be part of the semantic form of I-language. As part of the interface, they determine directly the conceptual interpretation of linguistic expressions; their impact on the computational structure of I-language, for example, via argument positions, is only indirect and does not depend on their spatial interpretation as such.

The problem to be considered briefly in this section concerns the relation between elements of the morphosyntactic structure of I-language and spatial interpretation. As rationale for this question, there are categories of I-language that clearly enter strictly morphological and syntactic relations and operations such as agreement, concord, and categorial selection, but that are obviously related to conditions of conceptual interpretation. Person, number, gender, and tense are obvious cases in point. Before taking up this problem with respect to spatial properties, I will briefly consider the status of grammatical categories with semantic impact more generally.

The problem to be clarified is the need to reconcile two apparently incompatible claims. On the one hand, morphological and syntactic primes, type 3 as indicated in section 2.2, differ from phonetic features and semantic components by the lack of any extralinguistic interpretation, their content being restricted to their role within the computational system of I-language. On the other hand, there cannot be any doubt that, for example, tense or person do have semantic purport in some way.

The way out of this apparent dilemma can be seen by looking more closely at number as a paradigm case. [$\pm$ Plural] is clearly a feature that enters the morphosyntactic computation of English and many other languages. The details of inflection, concord, and agreement that depend on this feature need not concern us here; it is clear enough that these are strictly formal conditions or operations. It is equally clear there must be some kind of an operator in SF related to [$+$ Plural] that imposes a condition on individual variables turning their interpretation into a multiplicity of individuals, although the details once again need not concern us. The relation between these two aspects becomes clear in cases of conflict, such as the *pluralia tantum*

of (33), where “glasses” refers to a set of objects in (33a), but to a single object in (33b):

- (33) a. Their glasses were collected by the waiter.
 b. His glasses were sitting on his nose.

Obviously, the feature [+ Plural] of “glasses” cannot be responsible for the set reference in (33a), as it must be lacking in (33b). Another type of conflict is illustrated by (34), where “who” must allow for set interpretation, as shown by (34a), but does not provide the plural antecedent required by “each other”:

- (34) a. Who was invited? (Eve, Paul, and Max were invited.)
 b. *Who does not talk to each other? (Eve and Paul.)

Further types of dissociation between morphological number and semantic individual/set interpretation could easily be added. The conclusion to be drawn from these observations is obvious. The feature [$\pm$ Plural] is related to, but not identical to, the presence or absence of the semantic set operator. More specifically, [+ Plural] in the default case is related to the operator SET; [– Plural] to the lack of this operator. How this relation is to be captured is a nontrivial problem, which resembles in some respects the phonological realization of [$\pm$ Plural] and other morphological categories. Thus the suffix /–s/ is the default realization of [+ Plural] for English Nouns, but is, of course, just as different from [+ Plural] as SET is. Notice, however, that both the phonological realization and the semantic interpretation of the default case might be instrumental in fixing the morphological category in acquisition as well as in language change. Similar, albeit more complex, accounts might be given for categories like gender and its relation to sex and animateness, or tense and its relation to temporal reference.

More generally, for morphological categories, the following terminological convention seems to be useful:

A semantic condition—that is, a configuration of primes of SF—is grammaticalized, if there is a morphological category *M* to which *C* is related by certain rules or conditions *R*.

The conditions *R* should be considered as the semantic counterpart to inflectional morphology, which relates morphological categories to configurations in PF. I am not going to make serious proposals as to the formal nature of *R* at the moment. The simplest assumption would be to associate a morphological category, such as [+ Plural], with some element in SF, such as SET, in a way that will be suspended in specifically marked cases. The potential suppression of the association would then be a consequence of the autonomous character of the morphological category, whereas

its actual realization indicates the conceptual purport of the formal category in question. Instead of pursuing these speculations, I will briefly look at the grammaticalization of spatial components in the sense specified in the above convention.

Two candidates are of primary interest in this respect: (1) case systems including sufficiently rich distinctions of so-called notional cases; and (2) classifier systems, corresponding to location and shape, respectively. We must expect in general not a straight and simple realization of spatial information by these categories, but rather a more or less systematic mapping, whose transparency will vary, depending on how entrenched the morphological categories are in autonomous computational relations like concord and agreement.

That notional cases are related to spatial information about location is uncontroversial and has been the motivation for the localistic theory of case mentioned earlier. In agglutinative languages like Hungarian, there is no clear boundary separating postpositions from cases. The semantic information related to locational and directional cases largely matches the schema of the corresponding prepositions discussed in the appendix, as shown in simple cases like (35):

- (35) a. a haz- ban 'in the house'
 the house in
 b. Budapest-ben
 'in Budapest'
 c. Budapest-re
 'to Budapest'

Even though things are far less transparent in more elaborate systems, it is sufficiently clear that place information can be grammaticalized by inflectional categories. For an extensive study of complex case systems (including Lak and Tabassarian) that is relevant under this perspective, even though it is committed to a different theoretical framework, see Hjelmslev (1935–37, part 1).

Classifier systems are subject to similar variations with respect to differentiation and grammatical systematization. A characteristic example is Chinese, where classifiers are obligatory with numerals for syntactic reasons, and related to shape in cases like (36):

- (36) a. *tiao* (longish, thin objects)
 yi tiao jie
 one CL street
 'one street'
 liang tiao he
 two CL river
 'two rivers'

- b. *zhang* (planar objects)
 - liang zhang xiangpian
 - two CL photograph
 - 'two photographs'
 - san zhang zhuozi
 - three CL table
 - 'three tables'
- c. *kuai* (three-dimensional objects)
 - yi kuai zhuan
 - one CL brick
 - 'one brick'
 - san kuai feizao
 - three CL soap
 - 'three cakes of soap'

The SF conditions to which these classifiers are related are not particular 3-D models but rather abstract object schemata of the sort mentioned above, which must be available, among others, for dimensional adjectives of English or German, for Tzeltal positional adjectives discussed in the appendix, but also for positional verbs like *lie*, *sit*, or *stand*, albeit in different modes of specification. Even though the details need clarification, it should be obvious that shape information can correspond to grammatical categories.

I will conclude these sketchy remarks on the grammaticalization of space with two more general considerations concerning the range and limits of these phenomena. There are, in fact, two opposite positions in this respect. The first position takes spatial structure as immediately supporting the computational structure of I-language and the categories of syntax and morphology. A tradition directly relevant is the locationist theory of case, according to which not only notional but also structural cases are to be explained in terms of spatial concepts like distance, contact, coherence, and orientation. The most ambitious account along these lines is given in Hjelmslev (1935–37), a slightly less rigorous proposal is developed in Jakobson (1936). While these theories are concerned with case only, more recent proposals of so-called cognitive grammar as put forward, for example, in Langacker (1987) extend spatial considerations to syntax in general. I will restrict myself to the locationist case theory. To cover the range of phenomena related to the varying structural properties of case, an extremely abstract construal of space must be assumed that has little, if any, connection to spatial cognition as sketched in section 2.4. Spatial structure is thereby turned into a completely general system of formal distinctions that makes the explanation either vacuous or circular. Even more crucially, the way in which case is

related to spatial conditions is notoriously opaque and indirect. In many languages case is involved in the distinction between place and direction, as mentioned above (see appendix for illustration). On the other hand, the dative/accusative contrast of German for example, *in der Schule* (in the school) versus *in die Schule* (into the school), is a purely formal condition connected to the semantic form of locative and directional *in*, respectively; it does not by itself express location or direction. This is borne out by the fact that “zur Schule” (to the school) requires the dative, even though it is directional. The conclusion to be drawn here has already been stated. Cases, like number, gender, tense, and person, and morphological categories in general are elements of the computational structure that may correspond to conceptual distinctions, but that do not in general represent those distinctions directly. In other words, spatial distinctions as represented in SF can correspond to elements of grammatical form, as should be expected, but are clearly to be distinguished from them.

The second position, which is in a way the opposite of the first one, is advocated by Jackendoff (chapter 1, this volume). Comparing two options with regard to the encoding of space, Jackendoff argues that axial systems and the pertinent frames of reference are represented in spatial representation but generally not in conceptual structure. The claim, presumably, applies to spatial structure in general. It is based on the following consideration. A clear indication for the conceptual encoding of a given distinction is the effect it has on grammatical structure. As a case in point, Jackendoff notes the count-mass distinction, which has obvious consequences for morphosyntactic categories in English. That comparable effects are missing for practically all spatial distinctions, at least in English, is then taken as an indication that they are not represented in conceptual structure, but only in spatial representation. I agree with Jackendoff in assuming that grammatical effects indicate the presence of the pertinent distinctions in conceptual structure. But it seems to me that the conclusion is the opposite because the major spatial patterns are no less accessible for grammatical effects than conceptual distinctions related to person, number, gender, tense, definiteness, or the count-mass distinction. Given the provisos just discussed, shape may correspond to classifiers; location may correspond to notional case; and size may correspond to degree and constructions like comparative, equative, and so on. Whether and which spatial distinctions are taken up explicitly by elements of semantic form and whether these correspond, furthermore, to effects in computational aspects of I-language, is a matter of language particular variation. English keeps most of them within the limits of lexical semantics. But this does not mean that they are excluded from grammatical effects in other languages, nor that they are excluded from conceptual and semantic representations of English expressions.

2.8 Conclusion

The overall view of how language accommodates space that emerges from these considerations might be summarized as follows:

1. Spatial cognition or I-space can be considered a representational domain within the overall system of C-I of conceptual and intentional structure integrating various perceptual and motoric modalities.
2. Representations of I-space must be integrated into propositional representations of conceptual structure, where in particular shape, size, and location of objects and the situations in which they are involved will be combined with other aspects of commonsense knowledge. Conceptual representation of spatial structure provides, among other things, more abstract schemata specifying the dimensionality of objects and situations, the axes and frames of reference of their location, and metrical scales with respect to which size is determined.
3. Linguistic knowledge or I-language interfaces with conceptual structure, recruiting configurations of it by basic components of semantic form, where strictly spatial concepts are to be identified as configurations that interpret elements of SF by exclusively spatial conditions on objects and situations.
4. Spatial information "visible" in I-language is thus restricted to strictly spatial concepts and their combinatorial effects, all other spatial information being supplied by representations of C-I and the commonsense knowledge on which they are based.
5. The computational categories of I-language, which map semantic form onto phonetic form, seem to fall into two types: syntactic categories, which serve the exclusively computational conditions of I-language, and morphological categories, which may correspond in more or less transparent ways to configurations in SF (or PF for that matter). The distinction between these two types of categories varies for obvious reasons, depending on the systematicity of the correspondence in question. Thus tense, person, and number are usually more transparent than (abstract) case or infinite categories of verbs. Categories of the combinatorial system, however transparent their correspondence might be to elements of the interfaces of I-language with other mental systems, are nevertheless components of the formal structure of I-language.

With all the provisos required by the wide range of unsolved or even untouched problems, the question raised initially might be answered as follows:

I-space is accommodated by semantic form in terms of primitives interpreted by strictly spatial concepts.

In what follows, I will illustrate the types of questions that arise with respect to the program sketched in section 2.6 by looking somewhat more closely at locative prepositions and dimensional adjectives, relating to place and shape, respectively.

To begin with, I will consider a general schema that covers a wide range of phenomena showing up within the system of locative propositions. By means of the notational conventions introduced in (18) and (31) above, the lexical entry for the preposition *in* can be stated as follows:

According to this analysis, based on Bierwisch (1988) and Wunderlich (1991), the semantic form of *in* is composed of a number of elements, including the relation LOC and the functor INT, which specifies the interior of its argument. In other words, instead of a simple relation IN, we assume a compositional structure, which I will now motivate by a number of comments.

(38) /the garden/ [DP, +Obj, ...] DEF u_i [GARDEN] u_i

(39) /in the garden/ [PP, ...] $\hat{x}$ [DEF u_i]: [x [LOC [INT u_i]]]]

(40) a. the man in the garden
b. The man is waiting in the garden.
c. The man is in the garden.

The main point to be noted here is the way in which the saturation of argument positions imposes conditions on the variables provided by the lexical SF(*le*) of *in*. I will take up the consequences of this point shortly.

A final remark on the argument positions of *in* concerns the optionality of its object, indicated by bracketing $\hat{y}$ in (37). It accounts for the intransitive use in cases like (41), where y is left as a free variable in SF(*le*) and will be specified by default conditions applying in C-I without conditions from SF.

(41) He is not in today.

Semantic Primes The variables x and y in (37) are related by the constants LOC and INT. Both are explicitly spatial in the sense that they identify conceptual components that represent simple (possibly primitive) spatial conditions. The interpretation of *in* can thus be stated more precisely as follows:

- (42) a. x LOC p identifies the condition that the location of x be (improperly) included in p
 b. INT y identifies a location determined by the boundaries of y , that is, the interior of y

Three comments are to be made with respect to this analysis.

First, additional conditions applying to x and y will affect how LOC and INT are interpreted in C-I. Relevant conditions include in particular the dimensionality of the object schema conceptually imposed on x and y , alongside with further conceptual knowledge. Thus the actual location of the theme in (43b) would rather be expressed by *under* if it were identical to that in (43a):

- (43) a. The fish is in the water.
 b. The boat is in the water.

A similar case in point is the following contrast:

- (44) a. He has a strawberry in his mouth.
 b. He has a pipe in his mouth.

Both “water” and “mouth” are associated with a three-dimensional object schema in (43a) and (44a) but conceptualized as belonging to a two-dimensional surface in (43b) and (44b). Knowledge about fishes, boats, fruits, and pipes supports the different construal of both INT and LOC. Somewhat different factors apply to the following cases:

- (45) a. There are some coins in the purse.
 b. There is a hole in the purse.

In (45a) *purse* relies on the object schema of a container; in (45b) the conditions coming from *hole* enforce the substance schema. Notice that in (45) it is only the interpretation of INT that varies, while in (43) and (44) the inclusion determined by LOC differs accordingly. The differences resulting from theme or relatum may enter into inferences. Thus from (45a) and (46) the conclusion (47a) derives, but (47b) does not follow from (45b) and (46):

(46) The purse is in my briefcase.

- (47) a. There are some coins in my briefcase.
b. There is a hole in my briefcase.

I do not think that *water*, *mouth*, *purse* are lexically ambiguous; although the way in which conceptual knowledge creates the differences in question is by no means a trivial issue, it must be left aside here. In any case, there is no reason to assume that *in* is ambiguous between (37) and some other lexical SF(*le*). The different interpretations illustrated by (42)–(47), to which further variants could easily be added, are due to conditions of I-space and conceptual knowledge not reflected in the lexical SF(*le*) of *in*.

Second, the conditions identified by LOC and INT are subject to implicit transfer to domains other than I-space:

- (48) a. He came in November.
b. several steps in the calculation
c. The argument applies only in this case.
d. readings in linguistics
e. He lost his position in the bank.

Again, the specification of the theme and/or the relatum provides the conditions on which LOC and INT are interpreted. Examples like those in (48) indicate, however, that the notion of BST crucially depends on how implicit transfer of spatial structures is construed. In one possible interpretation, *in* is a BST only if it relates to I-space, but not if it relates (in equally literal fashion) to time or institutions. It seems to me an important observation that *in* under this construal of BST is not an exclusively spatial term, but I do not think that this terminological issue creates serious problems. I will thus continue to use BST without additional comment.

And third, the range of I-space conditions identified by INT depends on the distinctions a given language happens to represent explicitly in SF by distinct primes. Thus English and German, for example, contrast INT with a prime ON with roughly the following property:

ON *y* identifies a location that has direct contact with (the designated side of), but does not intersect with, *y*.

This yields the different interpretations of, for example, *the nail in the table* and *the nail on the table*—assuming that SF(*le*) of *on* is [x LOC [ON y]]—whereas in Spanish *el clavo en la mesa* would apply to both cases because there is no *in/on* contrast in Spanish, such that the surface of the table could provide the location identified by INT.

The Pattern of Locative Prepositions I have assumed throughout that the categorization inherent in the primes of SF determines the compositional structure of SF according to general principles of I-language. Hence the variation in patterns of lexical representations I will briefly look at are fully determined by the basic elements involved. What is nevertheless of interest is the systematicity of variation these lexical representations exhibit.

The first point to be noted is the obvious generalization about locative prepositions, all of which instantiate schema (49), where F is a variable ranging over functors that specify locations determined by y :

(49) [x LOC [Fy]]

Not only do *in* and *on* fit into (49), specifying F by INT and ON, respectively, but also *near*, *under*, *at*, *over*, and several other prepositions, using pertinent constants to replace F . It is not obvious, however, whether schema (49) covers the full range of conditions that locative prepositions can impose. Thus Wunderlich (1991) claims that, for example, *along*, *across*, and *around* are more complex, introducing an additional condition, as illustrated in (62):

(50) /along/ [$-V, -N, \dots$] ($\hat{y}$) $\hat{x}$ [[x LOC[PROX y]]: [x PARALLEL
[MAX y]]]

PROX y and MAX y determine the proximal environment and the maximal extension of y , respectively. If this is correct, the general schema of locative prepositions is (51) instead of (49):

(51) [[x LOC[Fy]]: [xCy]] where C is a condition on x and y

C might be a configuration of basic elements, as exemplified in (50), all of which must have a direct, explicit spatial interpretation, in order to keep to the limits of BST.

Another systematic aspect of locative prepositions concerns their relation to directional counterparts, as shown for English and German examples in (52):

- (52) a. They were in the school. They went into the school.
 Sie waren in der Schule. Sie gingen in die Schule.
 b. The ball was under the table. The ball rolled under the table.
 Der Ball war unter dem Tisch. Der Ball rollte unter den Tisch.

Semantically, the directional preposition identifies a path whose end is specified by the corresponding locative preposition. Let $\text{CHANGE } p$ be an operator that turns the proposition p into the terminal state of change or path. The general schema of a standard directional preposition would then be (53):

- (53) $\text{CHANGE } [[x \text{ LOC}[Fy]]: [xCy]]$ where $\text{CHANGE } [\dots]$ identifies a transition whose final state is specified by $[\dots]$

The relevant observation in the present context is the systematic status of CHANGE in lexical structure. Besides mere optionality in cases like *under*, *over*, *behind*, which can be used as locative or directional prepositions, the occurrence of CHANGE is connected to *-to* in *onto*, *into*. In languages like Russian, German, and Latin with appropriate morphological case, CHANGE is largely related to accusative, to be checked by the object of the preposition. Using notational devices introduced in phonology, the relation in question can be expressed as in (54) for German *in*:

- (54) /in/ $[-V, -N, \alpha\text{Dir}]$ $\begin{array}{c} \hat{y} \hat{x} [(\alpha \text{CHANGE})[x \text{ LOC } [\text{INT } y]]] \\ | \\ [-\alpha\text{Obl}] \end{array}$

This means that *in* is either directional, assigns $-$ oblique case and contains the CHANGE component, or it is locative, assigns $+$ oblique case and does not contain CHANGE .

Typological Variation Thus far, the general patterns of prepositions have been considered as the frame by which lexical knowledge of a given language is organized. Crosslinguistic comparison reveals variations of a different sort, one of which concerns what might be called “lexical packaging,” that is, the way components of basic schema (49) are realized by separate formatives. A straightforward alternative is found, for example, in Korean, as can be seen in (55), taken from Wunderlich (1991):

- (55) Ch'aeksang- (ui)- ui- e kkotpyong i iss- ta
 desk Gen top Loc vase Nom be there Pres
 'There is a vase on the desk.'

The relatum *ch'aeksang* (optionally marked for genitive) functions as complement of the noun *ui*, which identifies the top or surface of its argument and provides the complement of the locative element *e*. In other words, LOC and F of (49) are realized by separate items with roughly the entries in (56), yielding (57):

Dimensional Adjectives

(61) /long/ Adj $\hat{x}(\hat{y})$ [[QUANT [MAX x]] = [$v + y$]]

Deg

Variables and Argument Structure As already mentioned, (61) expresses the fact that dimensional adjectives in English are syntactically two-place predicates, relating an object (or event) x to an optional complement of the DA that specifies a degree y , realized by appropriate measure phrases, as in (62), or more complex expressions as in (63):

- A particular point that distinguishes DAs from locative Ps concerns the variable *v* and the particular conditions that apply to it, as mentioned earlier. Due to this variable, DAs are semantically three-place relations, rather than two-place relations like prepositions. This becomes in fact visible when comparative morphology or the *too* construction make the variable accessible to syntactic specification:

- In a way, *than Bill* and *for this garage* are complements that explicitly specify the variable *v* under particular syntactic conditions.

Semantic Primes The variables x , y , and v are related in (61) by means of the four constants QUANT, MAX, =, and +, of which only MAX has a specifically spatial interpretation, identifying the maximal dimension with respect to the shape of y , while QUANT, =, and + identify quasi-arithmetical operations underlying quantitative, scalar evaluations quite generally. More specifically, [QUANT Y] is a function that maps arbitrary dimensions Y on an appropriate abstract scale, and = and + have the usual arithmetical interpretation with respect to scalar values. In other words, *long* is a spatial term only insofar as MAX determines dimensional conditions that rely on shape and size of objects or events; the shape and the size information contained in *long* and *short* are defined by MAX, on the one hand, and by QUANT, =, and + or −, on the other. Hence semantically, shape and size are interlocked in ways that differ remarkably from their interpretation in SR. Also, the quantitative conditions may carry over to various other domains: *old* and *young* are strictly temporal; *heavy* and *light* are gravitational; and so forth.

The Pattern of Dimensional Adjectives The characteristic properties of DAs show up more clearly if we look at the general schema of their SF, which automatically accounts for the fact that they usually come in antonymous pairs as already noted:

$$(65) \text{ [[QUANT [DIM } y]] = [v \pm x]]$$

The second point of variability in (65) besides the $\pm$ alternation is indicated by DIM, which marks the position for different dimensional components. Where *long/short* pick out the maximal dimension, *high/low* pick out the actually vertical axis by means of VERT, and *tall* combines both MAX and VERT. As a matter of fact, the constants replacing the variable DIM in (65) turn an adjective into a spatial term like *tall* or *thin*, a temporal term like *young* or *late*, a term qualifying movement, like *fast* and *slow*, and so forth.

It might be noted that the interpretation of the different dimensional constants requires the projection of an appropriate object schema on the term providing the value for x : *a tall sculpture* induces a schema whose maximal dimension is vertical for *sculpture*, which does not provide this condition by itself. As *ball* would not allow for a schema of this sort, *a tall ball* is deviant. For details of this mechanism see Lang (1989).

Typological Variation Thus far, we have considered variation within schema (65). I will now indicate some of the possibilities to modify the schema itself in various ways. An apparently simple modification is shown by languages like Russian, which do not allow measure phrases with DAs. *10 m long* could not come out *10 m dlinnij*; measure phrases can only be combined with the respective nouns, that is, by constructions like

dlinna 10 metrov, corresponding to *length of 10 meters*. This suggests that Russian DAs do not have a syntactic argument position for degree complements, preserving otherwise schema 84. Things seem to be a bit more complicated, though: measure phrases with comparatives are possible, although only in terms of prepositional phrases with *na*. *2 m longer*, for example, translates into the adjectival construction *na 2 m dlinnej*. I cannot go into the details of this matter.

We have already seen a much more radical variation of schema (65), exemplified by Tzeltal positional adjectives. Here, not only the degree argument position is dropped, but the whole quantificational component, retaining only [DIM *x*], but supplying it with a much more detailed system of specifications, as indicated provisionally in (59a). This is not merely a matter of quantity; rather, it attests a different strategy to recruit conditions on shape and position of objects. Where the twenty-odd DAs of most Indo-European languages rely on object schemata in a rather abstract and indirect way, the positional adjectives of Tzeltal include fairly specific, strictly spatial specifications of objects to which they apply.

Although organizing principles and actual details of Tzeltal positional adjectives remain to be explored, rather subtle, but clear distinctions determining alternatives in DAs of German, Russian, Chinese, and Korean have been isolated in Lang (1995). Object schemata in Chinese seem to be based on proportion of dimensions, while Korean takes observer orientation as prominent; a similar preference distinguishes German and Russian.

Let me summarize the main points of this rather provisional sketch of basic spatial terms. First, among the entries of the core lexical system of I-language, there is a subsystem of items that are strictly spatial in the sense illustrated in section 2.5. Their semantic form [SF(*le*)] consists exclusively of primes that are explicitly interpreted in terms of conditions of I-space. Even though the delimitation of this subsystem is subject to intervening factors, such as implicit or explicit transfer of interpretation, its elements play a theoretically relevant role for the linguistic representation of space. Second, there are characteristic consequences with respect to the linguistic properties of these items, as shown by the appearance of degree phrases, and argument structure more generally. Hence the compositional structure of the SF of these terms must be assumed to belong to I-language, their basic elements being components of a representational aspect determined by UG. Finally, there is remarkably systematic variation among different languages with respect to both the choice of basic distinctions recruited for lexicalization and the different types of packaging according to more general patterns. In general, then, the analysis of basic spatial terms, even though it could be illustrated only by two types of cases, promises to give us a more detailed understanding of how (much) space gets into language.

Acknowledgments

The present chapter benefits from discussions at various occasions. Besides the members of the Max Planck Research Group on Structural Grammar, I am indebted to the participants of the project on Spatial and Temporal Reference at the Max Planck Institute for Psycholinguistics; further discussions included Dieter Gasde, Paul Kiparsky, Ewald Lang, Stephen Levinson, and Dieter Wunderlich. Particular debts are due to Ray Jackendoff, whose stimulating proposals are visible throughout the paper, even if I do not agree with him in certain respects.

Notes

1. This view is in line with fundamental developments in recent linguistic theory, including the minimalist program proposed in Chomsky (1993). Although it is still compatible with the possibility of parametric variation regarding the way options provided by specification 2 are exploited in individual languages, this sort of parametric variation should be considered as bound to lexical information, and thus ultimately to the choice of primitives in the sense of specification 1. I will examine more concrete possibilities along these lines in section 2.6.
2. This does not necessarily imply a proliferation of levels of representations, stipulating LF in addition to SF. One might in fact consider LF a systematic categorization imposed on SF, just as PF must be subject to certain aspects of syntactic structure.
3. Even though Chomsky (1993) refers to A-P and C-I occasionally as “performance systems,” it should be clear that they must be construed as computational systems with their own specific representational properties.
4. It should be noted that Jackendoff considers the phonological structure (i.e., PF) as belonging properly to I-language, although he recognizes the need for correspondence rules connecting it to articulation and perception.
5. Thus, in order to honor Schönberg, Alban Berg in his “*Lyrische Suite*” introduces a theme that consists of the notes es (=e-flat)-c-h (=b)-e-g, representing all and only the letters in Schönberg corresponding to the German rendering of notes.
6. A very special “interface representation” in the intended sense is the system of numbering used in Gödel’s famous proof of the incompleteness of arithmetic, where numbers are given two mutually exclusive systematic interpretations, one stating properties of the other.

References

- Berlin, B., and Kay, P. (1969). *Basic color terms*. Berkeley: University of California Press.
- Biedermann, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Bierwisch, M. (1983). Semantische und konzeptuelle Repräsentation lexikalischer Einheiten. In R. Ruzicka and W. Motsch (eds.), *Untersuchungen zur Semantik: Studia Grammatica XXII*, 61–100, Berlin, Akademie-Verlag.

- Bierwisch, M. (1988). On the grammar of local prepositions. In M. Bierwisch, W. Motsch, and I. Zimmermann (Eds.), *Syntax, Semantik, und Lexikon: Rudolf Ruzicka zum 65. Geburtstag*, 1–65. Berlin: Akademie-Verlag.
- Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch and E. Lang (Eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation*, 71–261. Heidelberg, New York: Springer.
- Bierwisch, M., and Lang, E. (1989). Somewhat longer—much deeper—further and further. In Bierwisch and Lang (Eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation*, 471–514. Heidelberg, New York: Springer.
- Brown, P. (1994). The INS and ONS of Tzeltal locative expressions: The semantics of static descriptions of location. *Linguistics*, 32, 743–790.
- Byrne, R. M. J., and Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language*, 28, 564–575.
- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Chomsky, N. (1993). A minimalist program for linguistic theory. In K. Hale and S. J. Keyser (Eds.), *Essays in linguistics in honor of Sylvian Bromberger: The view from Building 20*, 1–52. Cambridge, MA: MIT Press.
- Dölling, J. (1995). Ontological domains, semantic sorts, and systematic ambiguity. *International Journal of Human-Computer Studies*, 43, 785–807.
- Dowty, D. R. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Fodor, J. A. (1975). *The language of thought*. New York: Cromwell.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Gruber, J. S. (1976). *Studies in lexical relations*. Amsterdam: North-Holland.
- Hale, K., and Keyser, S. J. (1993). On argument structure and the lexical expression of syntactic relations. In Hale and Keyser (Eds.), *Essays in linguistics in honor of Sylvian Bromberger: The view from Building 20*, 53–109. Cambridge, MA: MIT Press.
- Hjelmslev, L. (1935–37). *La catégorie des cas*. Århus: Universitetsforlaget.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Jakobson, R. (1936). Contribution to the general theory of case: General meanings of the Russian cases. In R. Jakobson, *Russian and Slavic grammar studies: 1931–1981*, 59–103. Berlin, New York: Mouton. (Original version: Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutungen der russischen Kasus. Selected Writings, Vol. II, 23–71.)

- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press; Cambridge, MA: Harvard University Press.
- Kamp, H., and Reyle, U. (1993). *From discourse to logic*. Dordrecht: Kluwer.
- Katz, J. J. (1972). *Semantic theory*. New York: Harper and Row.
- Keil, F. C. (1987). Conceptual development and category structure. In U. Neisser (Ed.), *Concepts and conceptual development*, 175–200. Cambridge: Cambridge University Press.
- Kosslyn, S. M. (1983). *Ghosts in the mind's machine*. New York: Norton.
- Kosslyn, S. M., Holtzmann, J. D., Farah, M. J., and Gazzaniga, M. S. (1985). A computational analysis of mental image generation: Evidence from functional dissociations in split-brain patients. *Journal of Experimental Psychology: General*, 114, 311–341.
- Lang, E. (1989). The semantics of dimensional designation of spatial objects. In M. Bierwisch and E. Lang (Eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation*, 263–417. Heidelberg, New York: Springer.
- Lang, E. (1995). Basic dimension terms: A first look at universal features and typological variation. *FAS-Papers in Linguistics*, 1, 66–100.
- Langacker, R. W. (1987). Nouns and verbs. *Language*, 63, 53–94.
- Levinson, S. C. (1990). Figure and ground in Mayan spatial description. Paper delivered to the conference Time, Space, and the Lexicon. Nijmegen: Max Planck Institute for Psycholinguistics, November.
- Marr, D. (1981). *Vision*. San Francisco: Freeman.
- Moravcsik, J. M. E. (1981). How do words get their meanings? *Journal of Philosophy*, 78, 5–24.
- Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, 17, 409–441.
- von Stechow, A. (1995). Lexical decomposition in syntax. In E. Urs et al. (Eds.), *The lexicon in the organization of language: Selected papers from 1991 Konstanz Conference*, 81–117. Amsterdam: Benjamins.
- Wunderlich, D. (1991). How do prepositional phrases fit into compositional syntax and semantics? *Linguistics*, 29, 591–621.

Chapter 3

Perspective Taking and Ellipsis in Spatial Descriptions

Willem J. M. Levelt

3.1 Thinking for Speaking

There exists happy agreement among students of language production that speaking normally involves a stage of conceptual preparation. Depending on the communicative situation, we decide in some way or another on *what* to express. Ideally, this choice of content will eventually make our communicative intention recognizable to our audience or interlocutor. The result of conceptual preparation is technically termed a *message* (or a string of messages); it is the conceptual entity the speaker will eventually express in language, that is, formulate.

But there is more to conceptual preparation than considering what to say, or *macroplanning*. There is also *microplanning*. The message has to be of a particular kind; it has to be tuned to the target language and to the momentary informational needs of the addressee. This chapter is about an aspect of microplanning that is of paramount importance for spatial discourse, namely *perspective taking*.

In an effort to cope with the alarming complexities of conceptual preparation, I presented a figure in my book *Speaking* (1989) that is reproduced here as figure 3.1. It is intended to express the claim that messages must be in some kind of propositional or “algebraic” format (cf. Jackendoff, chapter 1, this volume) to be suitable for formulation. In particular, they must be composed out of *lexical concepts*, that is, concepts for which there are words or morphemes in the speaker’s language. An immediate corollary of this notion is that conceptual preparation will, to some extent, be specific to the target language. Lexical concepts differ from language to language. A lexical concept in one language may be nonlexical in another and will therefore need a slightly different message to be expressed. To give one spatial example (from Levelt 1989), there are languages such as Spanish or Japanese that treat deictic proximity in a tripartite way: proximal-medial-distal. Other languages, such as English or Dutch, have a bipartite system, proximal-distal. Spanish use of *aquí-ahí-alli* requires to construe distance from speaker in a different way than English use of *here-there*.

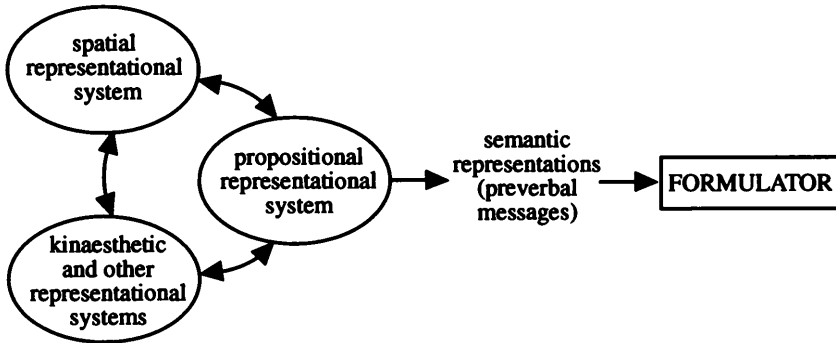


Figure 3.1

The mind harbors multiple representational systems that can mutually interact. But to formulate any representation linguistically requires its translation into a semantic, “propositional” code (reproduced from Levelt 1989).

Slobin (1987) has usefully called this “thinking for speaking,” which is an elegant synonym for microplanning.

Thinking for speaking is always involved when we express nonpropositional, in particular spatial, information. Figure 3.1 depicts the notion that when we talk about our spatial, kinesthetic, musical, and so on experiences, we cast them in propositional form. This necessarily requires an act of abstraction. When talking about a visual scene, for instance, we attend to entities that are relevant to the communicative task at hand, and generate predications about these entities that accurately capture their spatial relations within the scene. This process of abstracting from the visual scene for speaking I will call “perspective taking.” Although this term will in the present chapter be restricted to its original spatial domain, it is easily and fruitfully generalized to other domains of discourse (cf. Levelt 1989).

3.2 Perspective Taking

Perspective taking as a process of abstracting spatial relations for expression in language typically involves the following operations:

1. Focusing on some portion of the scene whose spatial disposition (place, path, orientation) is to be expressed (Talmy 1983). I will call this portion the “referent.”
2. Focusing on some portion of the field with respect to which the referent’s spatial disposition is to be expressed. I will call this portion the “relatum.”
3. Spatially relating the referent to the relatum (or expressing the referent’s path or orientation) in terms of what I will call a “perspective system.”

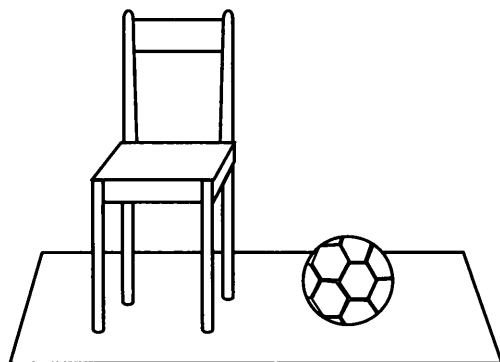


Figure 3.2

This spatial array can be described in myriad ways, depending on the choice of referent, relatum, and perspective.

Let me exemplify this by means of figure 3.2. One way of describing this scene is

(1) I see a chair and a ball to the right of it.

Here the speaker introduces the chair as the relatum and then expresses the spatial disposition of the ball (to the right of the chair). Hence, the ball is the referent. The perspective system in terms of which the relating is done is the *deictic* system, that is, a speaker-centered *relative* system.¹ When you focus on the relatum (the chair), your gaze must turn to your right in order to focus on the referent (the ball). That is why the ball is to the right of the chair in this system.

Two things are worth noticing now. First, you can swap relatum and referent, as in (2):

(2) I see a ball and a chair to the left of it.

This is an equally valid description of the scene; it is only a less preferred one. Speakers tend to select smaller and more foregrounded objects as referents and larger or more backgrounded entities as relata. Here they tend to follow the Gestalt organization of the scene (Levelt 1989). Second, you can take another perspective system. You can also describe the scene as (3):

(3) I see a chair and a ball to its left.

This description is valid in the *intrinsic* perspective system. Here the referent's location is expressed in terms the relatum's intrinsic axes. A chair has a front and a back, a left and a right side. The ball in figure 3.2 is at the chair's left side, no matter from which viewpoint the speaker is observing the scene. Still another perspective system allows for the description in (4):

(4) I see a chair and a ball north of it.

This description is valid if indeed ball and chair are aligned on a north-south dimension. This is termed an *absolute* system; it is neither relative to the speaker's nor to the relatum's coordinate system, but rather to a fixed bearing.

The implication of these two observations is that perspective is linguistically free. There is no unique way of perspective taking. There is no biologically determined one-to-one mapping of spatial relations in a visual scene to semantic relations in a linguistic description of that scene. And cultures have taken different options here, as Levinson and Brown have demonstrated (Levinson 1992a,b; Brown and Levinson 1993). Speakers of Guugu Yimithirr are exclusive users of an absolute perspective system, Mopan speakers are exclusive users of an intrinsic system, Tzeltal uses a mix of absolute and intrinsic perspectives, and English uses all three systems. Similarly, there are personal style differences between speakers of the same language. Levelt (1982b) found that, on the same task, some speakers consistently use a deictic system whereas others consistently use an intrinsic perspective system. Finally, the same speaker may prefer one system for one purpose and another system for another purpose as Tversky (1991) and Herrmann and Grabowski (1994) have shown.

This freedom of perspective taking does not mean, however, that the choice of a perspective system is arbitrary. Each perspective system has its specific advantages and disadvantages in language use, and these will affect a culture's or a speaker's choice. In other words, there is a *pragmatics* of perspective systems.

In the rest of this chapter I will address two issues. The first one is pragmatics. I will compare some advantages and disadvantages in using the three systems introduced above; the deictic, the intrinsic, and the absolute systems. In particular, I will ask how suitable these systems are for spatial reasoning, how hard or easy they are to align between interlocutors, and to what extent the systems are mutually interactive.

The second issue goes back to figure 3.1 and to "thinking for speaking." I defined perspective taking as a speaker's mapping of a spatial representation onto a propositional (or *semantic*) representation for the purpose of expressing it in language. A crucially important question now is whether the spatial representations themselves are already "tuned to language." For instance, a speaker of Guugu Yimithirr, who exclusively uses absolute perspective, may well have developed the habit of representing *any* spatial state of affairs in an oriented way, whether for language or not. After all, *any* spatial scene may become the topic of discourse at a different place and time. The speaker should then have remembered the scene's absolute orientation. Levinson (1992b) presents experimental evidence that this is indeed the case. On the other hand, I argued above that perspective is free. A speaker is not "at the mercy" of a spatial representation in thinking for speaking. In the strongest non-Whorfian

case, spatial representations will be language-independent, and it is perspective taking that maps them onto language-specific semantic representations. One way of sorting this out is to study how speakers operate when they produce *spatial ellipsis* (such as in *go right to blue and then Ø to purple*, here Ø marks the position where a second occurrence of *right* is elided). I will specifically ask whether ellipsis is generated from a perspectivized or from a perspective-free representation. If the latter turned out to be the case, that would plead for the existence of perspective-free spatial representations.

3.3 Some Properties of Deictic, Intrinsic, and Absolute Perspective

Of many aspects that may be relevant for the use of perspective systems I will discuss the following three: (1) their inferential potential, (2) their ease of coordination between interlocutors, and (3) their mutual support or interference.

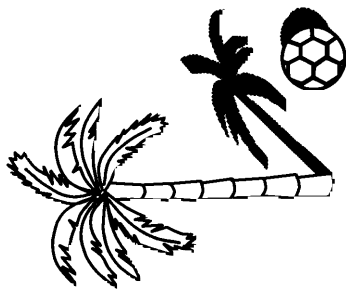
3.3.1 Inferential Potential

Spatial reasoning abounds in daily life (cf. Byrne and Johnson-Laird 1989; Tversky 1991). Following road directions, equipment assembly instructions, spatial search instructions, or being involved in spatial planning discourse all require the ability to infer spatial layouts from linguistic description. And the potential for spatial inference is crucially dependent on the perspective system being used. In Levelt (1984) I analyzed some essential logical properties of the deictic and intrinsic systems; I will summarize them here and extend the analysis to the absolute system.

Converseness An attractive logical property is converseness. Perspective systems usually (though not always) involve directional opposites, such as front-back, above-below, north-south. If the two-place relation expressed by one pole is called R and the one by the other pole by R^{-1} , then converseness holds if $R(A, B) \Rightarrow R^{-1}(B, A)$. For instance, if object A is above object B , B will be below A .

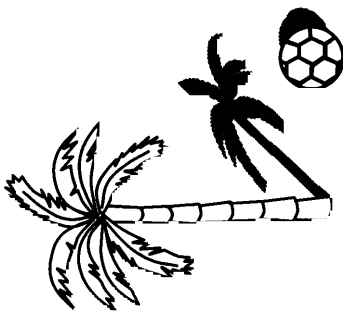
Converseness holds for the deictic system and for most cases² of the absolute system, but not for the intrinsic system. This is demonstrated in figure 3.3. Assuming that it is about noon somewhere in the Northern Hemisphere with the sun shining, the shadows of the tree and ball indicate that the ball is east of the tree. Using this absolute bearing, the tree must be west of the ball, where west is the converse of east. Converseness also holds for the (three-place) deictic relation. From the speaker's point of view, the ball (referent) is to the right of the tree (relatum), which necessarily implies that the tree (referent) is to the left of the ball (relatum). But it is easy to violate converseness for the intrinsic system. The ape can be on the right side ("to the right") of the bear at the same time the bear is on the right side ("to the right") of the

Absolute



The ball is East of the tree
*
The tree is West of the ball

Deictic



The ball is to the right of the tree
*
The tree is to the left of the ball

Intrinsic



The ape is to the right of the bear
*
The bear is to the right of the ape

Figure 3.3

Converseness holds for the absolute and deictic systems, but not for the intrinsic system.

ape. It is therefore impossible to infer the relation between relatum and referent from the relation between referent and relatum in the intrinsic system, which is a major drawback for spatial reasoning.

Transitivity Transitivity holds if from $R(A, B)$ and $R(B, C)$, it follows that $R(A, C)$. This is the case for the absolute and deictic systems, but not for the intrinsic system. This state of affairs is demonstrated in figure 3.4. The flag, tree, and ball scene depicts the transitivity of “east of” in the absolute system and of “to the right of” in the deictic system. For the intrinsic system it is easy to construct a case that violates transitivity. This is the case for the bear, cow, and ape scene. The user of an intrinsic system cannot rely on transitivity. From A is to the right of B , and B is to the right of C , one cannot reliably conclude that A is to the right of C , and so forth. Hence one cannot create a chain of inference, using the previous referent as a relatum for the next one.

These are serious drawbacks of the intrinsic system. Converseness and transitivity are very desirable properties if you want to make inferences from spatial premises. And spatial reasoning abounds in everyday discourse, for instance, in following route directions, in jointly planning furniture arrangements or equipment assembly, and so on. I will shortly discuss further drawbacks of the intrinsic system for spatial reasoning.

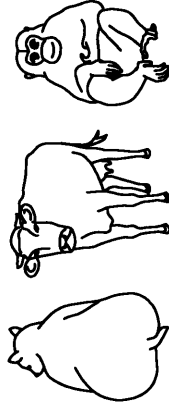
3.3.2 Coordination between Interlocutors

It is more the exception than the rule that interlocutors make explicit reference to the perspective system they employ in spatial discourse (for references and discussion, see Levelt 1989, 51). Usually there is tacit agreement about the system used, but not always. An example of nonagreement turned up in an experiment where I asked subjects to describe colored dot patterns in such a way that other subjects would be able to draw them from the tape-recorded descriptions. An example of such a pattern is presented in figure 3.5. Subjects were instructed to start at the arrow. It turned out that most subjects used deictic perspective. A typical deictic description of this pattern is the following:

Begin with a yellow dot. Then one step up is a green dot and further up is a brown dot. Then right to a blue dot and from there further right to a purple dot. Then one step down there is a red dot. And left of it is a black one.

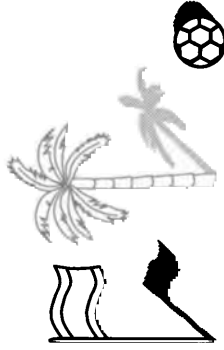
Although the dot pattern was always flat on the table in front of the subject, moves toward and away from the subject were typically expressed by vertical dimension terms (*up*, *down*). This is characteristic for deictic perspective, because it is viewer-centered. It essentially tells you where the gaze moves (see Levelt 1982b; Shepard and Hurwitz 1984). For the pattern in figure 3.5, the gaze moves up, up, right, right,

Intrinsic

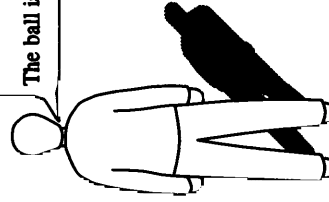


The ape is to the right of the bear
 The bear is to the right of the cow
 *
 The ape is to the left of the cow

Deictic



The ball is to the right of the tree
 The tree is to the right of the flag
 *
 The ball is to the right of the flag



The ball is East of the tree
 The tree is East of the flag
 *
 The ball is East of the flag

Absolute

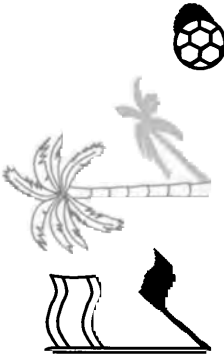


Figure 3.4
 Transitivity holds for the absolute and deictic systems, but not for the intrinsic system.

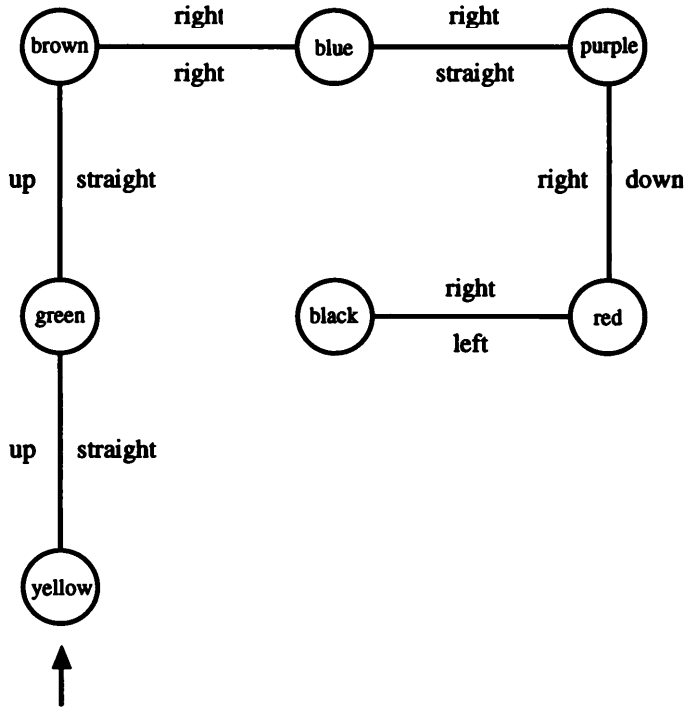


Figure 3.5

Pattern used in a spatial description task. The nodes were colored (here replaced by color names). On the outside of the arcs are the dominant directional terms used in deictic descriptions; on the inside, the ones used dominantly in intrinsic descriptions.

down, and left. These directional terms in the description are depicted at the exterior side of the pattern. Notice that all terms would have been different if the pattern had been turned by 90 degrees.

But other subjects used the intrinsic system. They described the scene as if they were moving through it or leading you through it. This is a typical intrinsic description.³

You start at a yellow point. Then go straight to a green dot and straight again to brown. Now turn right to a blue dot and from there straight to a purple dot. From there turn right to red and again right to a black dot.

There are no vertical dimension terms here. The description is not viewer-centered, but derives from the intrinsic directions of the pattern itself; the directional terms

would still be valid if the pattern were turned by 90 degrees. The interior of figure 3.5 depicts the directional terms used in this intrinsic description.

When I gave the deictic descriptions to subjects for drawing, they usually reproduced the pattern correctly. But when I presented the intrinsic description, subjects' drawings tended to be incorrect, and systematically so. Most reproductions are like the one in figure 3.6, which is a typical example. What has happened here is obvious. The listener tacitly assumes a deictic perspective and forces the intrinsic description into this deictic Procrustean bed. The incongruent term *straight* is interpreted as "up." This, then, is a case of failing speaker/hearer coordination.

Coordination failures can be of different kinds. In this example the listener tacitly assumes one perspective system where the speaker has in fact used a different one. Our deictic and intrinsic systems are subject to this confusion because many of the

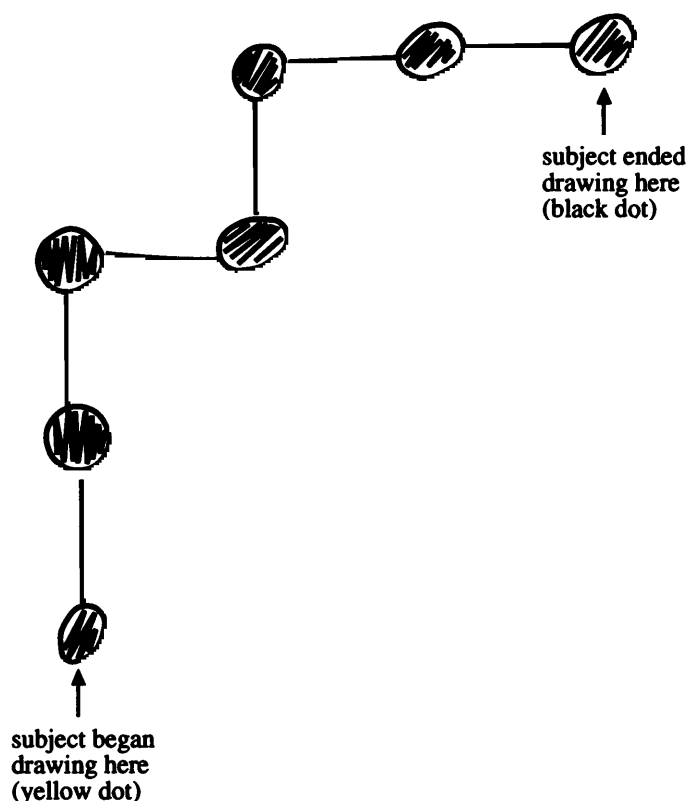


Figure 3.6

A subject's reconstruction of the pattern in figure 3.5 from its intrinsic description.

dimensional terms are the same or similar in the two systems. But also within the same perspective system coordination failure can arise.

For the deictic system, a major problem in coordination is that the system derives from the speaker's viewpoint, that is, the speaker's position and orientation in the scene. And because the viewpoints are never fully shared, there is continuous switching back and forth in conversation between the coordinate systems of the interlocutors. The interlocutors must keep track of their partners' viewpoints throughout spatial discourse.

This contrasts with the intrinsic and absolute systems, which are speaker-independent. The intrinsic system, however, requires that the interlocutor is aware of the relatum's orientation. The utterance *the ball is to the right of the chair* can only effectively localize the ball for the interlocutor if not only the chair's position is known, but also its orientation. In a perceptual scene, therefore, the intrinsic system requires *recognition* of the relatum on the part of the listener, not only awareness of its localization.

The felicity of speaker/hearer coordination in the intrinsic system is, therefore, crucially dependent on the shared image of the relatum. First, coordination in the intrinsic system is only possible if the relatum is oriented. Any object that does not have an intrinsic front is excluded as a base for the front/back and left/right dimensions (Miller and Johnson-Laird 1976). Second, frontness is an *interpretative* category, not a strictly visual one. There is no visual feature that characterizes both the front of a chair and the front of a desk (see figure 3.7a–b). These properties are *functional* ones, derived from our characteristic uses of these objects, and these uses

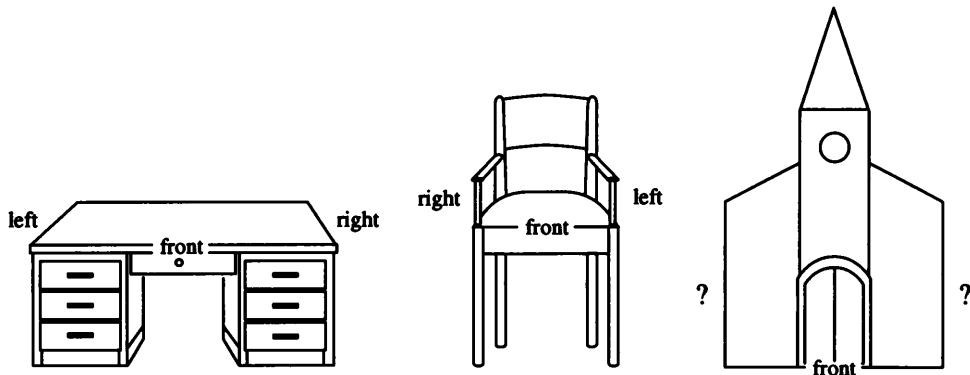


Figure 3.7

The alignment of an object's left, front, and right side does not depend on its spatial, but on its functional, properties.

can be complex. What we experience as the front side of a church from the outside (figure 3.7c) is its rear or back from the inside. Still worse, the alignment of an object's front, left, and right is not fixed, but dependent on its characteristic use (compare the alignments for chair and desk in figures 3.7a and 3.7b); it may even be undetermined or ambiguous (as is the case for the church in figure 3.7c).

Not all intrinsic systems share all of these problems. Levinson (1992a) was able to show that speakers of Tzeltal are much more vision-bound in deriving the intrinsic, orientation-determining parts of objects than English or Dutch, which tend to use a more functional approach. Still, the use of intrinsic perspective always requires detailed interpretation of the relatum's shape, and this has to be shared between interlocutors. These problems do not arise for the deictic and absolute systems.

So far we discussed some of the coordination problems in utilizing the deictic or the intrinsic system. What about speaker/hearer coordination in terms of an absolute system? Here, the interlocutors must agree on absolute orientation, for instance on what is north. Even if such a main direction is indicated in the landscape as a tilt or a coastline, dead reckoning will be required if successful spatial communication is to take place in the dark, in the fog, farther away from one's village, or inside unfamiliar dwellings (Levinson 1992b). The only absolute dimension that is entirely unproblematic is verticality, for which we have a designated sensory system (and even this one can nowadays be tampered with; see Friederici and Levelt 1990 for some experimental results in outer space). So even an absolute system is not without its drawbacks in spatial communication.

3.3.3 Interaction between Perspective Systems

When language users have access to more than a single perspective system, additional problems arise. A first problem already appeared in the previous section. Interlocutors must agree on a system, or must at least be aware of the system used by their partners in speech. This mechanism failed in the network description task in figure 3.6. Various factors can contribute to the establishment of agreement. One important factor is the choice of a default solution. Depending on the communicative task at hand, interlocutors tend to opt for the same solution (Taylor and Tversky 1996; Herrmann and Grabowski 1994). In addition, a speaker's choice of perspective is often given away by the terminology typical for that perspective. When a speaker uses terms such as *north* or *east*, the chosen perspective cannot be deictic or intrinsic. And there are more subtle differences. I have mentioned the presence of vertical dimension terms in deictic directions in a horizontal plane and their total absence in intrinsic directions (the relevant data are to be found in Levelt 1982b). Hence, for these descriptions, presence or absence of vertical dimension terms gives away which perspective system is being used. Surprisingly, the subjects in my experiment completely

ignored this distinctive information when they drew patterns such as in figure 3.6. There are still other linguistic cues. When you say *The chair is on Peter's left*, you are definitely using the intrinsic system, and so is the Frenchman who says *la chaise est à la gauche de ma soeur* (Hill 1982), or the German who utters *Der Stuhl ist zu ihrer Linken* (Ehrich 1982). I am not familiar with any empirical study about the effectiveness of such linguistic cues in transmitting the speaker's perspective to the listener.

Two problems that arise with multiple perspectives are *alignment* and *preemption*. Different perspectives may or may not be aligned in a particular situation, and if they are not aligned, one perspective may gain (almost) full dominance, more or less preempting the other perspectives. This is most easily demonstrated from the use of vertical dimension terms, such as in *A is above/below B*. The basis for verticality is different in the three systems under consideration. In the absolute system verticality is determined by the direction of gravity. In the intrinsic system it is determined by the top/bottom dimension of the relatum. In the deictic system it is probably determined by the direction of your retinal meridian (Friederici and Levelt 1990). In any perceptual situation these three bases of verticality may or may not coincide. Let us consider situations where there is a ball as referent and a chair as relatum and there is an observer/speaker.⁴ The ball can now be *above* the chair with respect to one, two, or all three of these bases. The eight possibilities that arise are depicted in figure 3.8.⁵

The appropriateness of saying *the ball is above the chair* varies dramatically for the depicted speaker in the eight scenes. This we know from the work by Carlson-Radvansky and Irwin (1993), who put subjects in the positions depicted in figure 3.8 and asked them to name the spatial relation between the referent and the relatum. Although the scenes were formally the ones in figure 3.8, they varied widely in the objects depicted and in backgrounds.⁶ Figure 3.8 shows the percentage of "above" responses for each configuration. Clearly, absolute perspective is quite dominant here (scenes *a–d* are "above" cases in absolute perspective). But in the absence of absolute *above*, intrinsic *above* keeps having some force, whether or not it is aligned with deictic *above* (scenes *e* and *g*, respectively). Deictic *above* alone, however, (scene *f*) is insufficient to release "above" responses. More generally, the deictic dimension does not seem to contribute much in any combination. But further work by the same authors (Carlson-Radvansky and Irwin 1994), in which reaction times of judgments were measured for the same kind of scenes, showed that all three relevant systems contribute to the reaction times. The three systems mutually facilitate or interfere, depending on their alignment. In addition, the reaction times roughly follow the judgment data in figure 3.8. The fastest responses are for *above* in absolute perspective, followed by intrinsic and then deictic *above* responses.

These findings throw a new light on a discussion of my "principle of canonical orientation" (Levelt 1984) by Garnham (1989). I had introduced that principle to

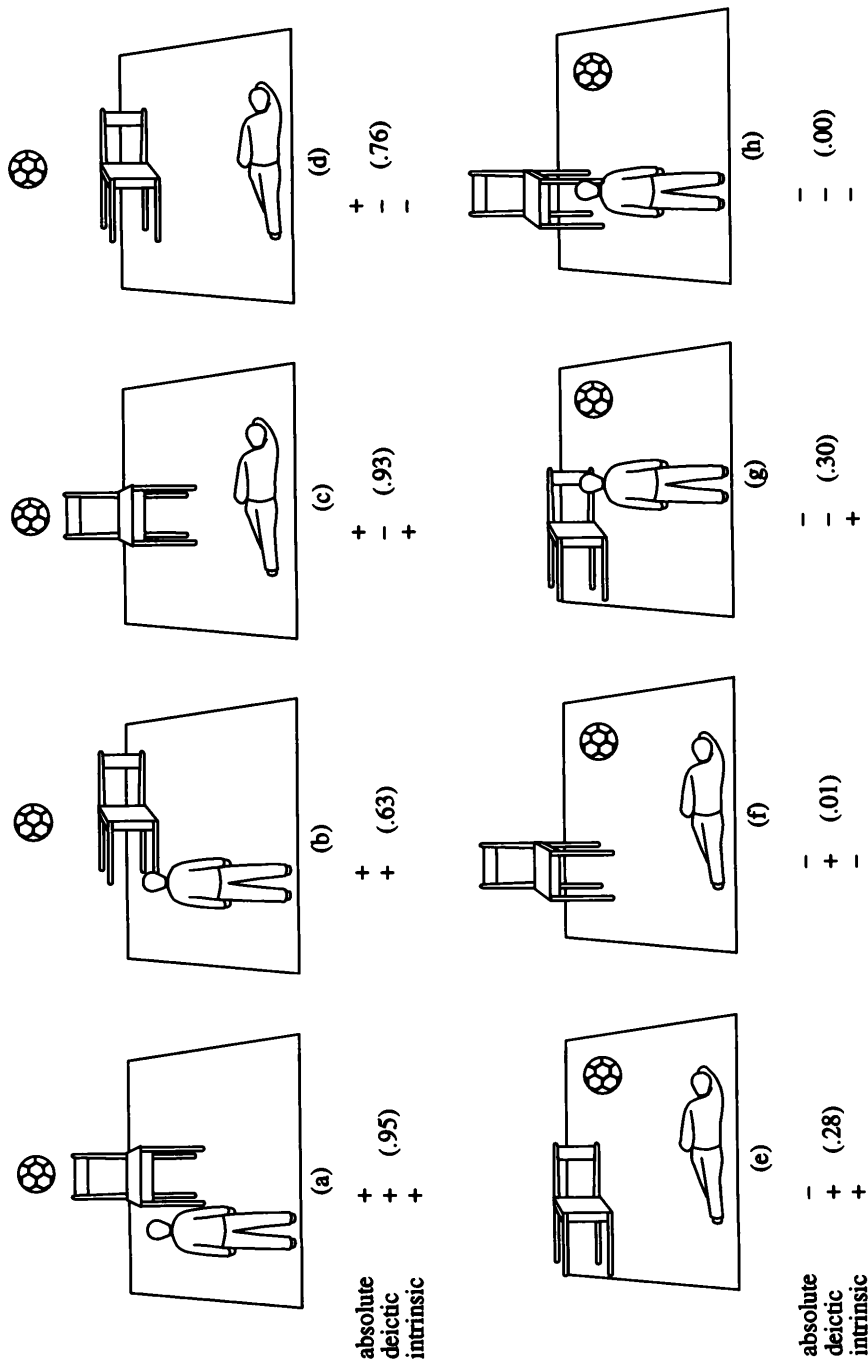
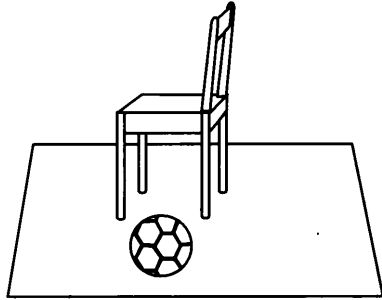


Figure 3.8

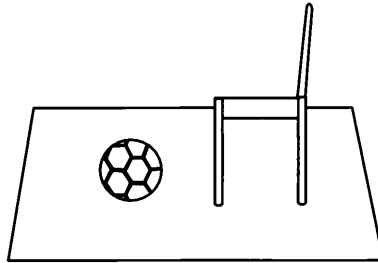
The ball is *above* the chair from all three perspectives (a), from two perspectives (b, c, e), from just one perspective (d, f, g), or from no perspective (h). The numbers in brackets are the percentage of "above" responses obtained in a study by Carlson-Radvansky and Irwin (1993) for scenes with the same formal properties as these.

"The ball is to the left of the chair."

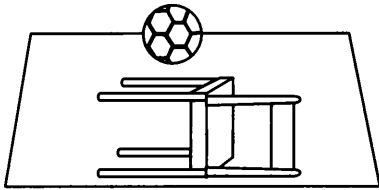
"The ball is in front of the chair."



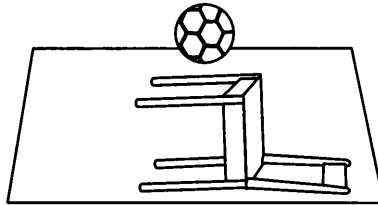
a.



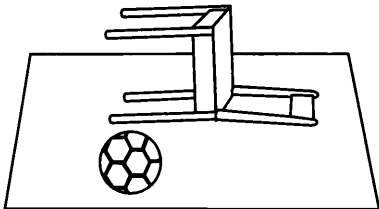
d.



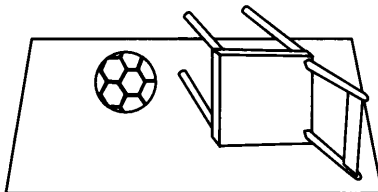
b.



e.



c.



f.

Figure 3.9

According to the principle of canonical orientation, the ball can be intrinsically *to the left of* the chair in (a) and (c), but not in (b). It can be intrinsically *in front of* the chair in (d) and (f), but not in (e).

account for certain cases where the intrinsic system is “immobilized” when it conflicts with the deictic system. Because the principle is directly relevant to the present discussion of alignment and preemption, I cite it here from the original paper:

Principle of canonical orientation For the intrinsic system to refer to a relatum’s intrinsic dimension, that dimension must be in canonical position with respect to the perceptual frame of orientation of the referent.

The principle of canonical orientation is easily demonstrated from figure 3.9. Cases *a*, *b*, and *c*, in the left-hand side of the figure, refer to the intrinsic description *the ball is to the left of the chair*.

According to the principle of canonical orientation this is a possible description in (a). The description refers to the relatum’s intrinsic left/right dimension. That dimension is in canonical orientation to the relatum’s perceptual frame. The perceptual frame for the chair’s orientation is in this case the normal gravitational field. The chair is in canonical position with respect to this perceptual frame. In particular, the chair’s left/right dimension has a canonical direction, that is, it lays in a plane that is horizontal in the perceptual frame. However, the description is virtually impossible in (b). Here the left/right dimension of the chair (the relatum) is not in canonical position; it is not in a horizontal plane, given the perceptual frame. Finally and surprisingly, it is for many native speakers of English acceptable to say *the ball is to the left of the chair* in case of (c). Here the chair is not in canonical position either, but the chair’s left/right dimension is; it is in a horizontal plane of the perceptual frame. Hence the principle of canonical orientation is satisfied in this case.

The state of affairs is similar for the intrinsic description *the ball is in front of the chair*. This description is fine for (d). It is, however, virtually unacceptable for (e), and this is because the front/back dimension of the relatum (the chair) is not in a canonical, horizontal plane with respect to the perceptual frame. Although in (f) the chair is not in canonical position, its front/back dimension is. Hence the description is again possible according to the principle, which agrees with intuitions of many native speakers of English to whom I showed the scene (the formal experiment has never been done, though).

Why does the principle refer to “the perceptual frame of orientation of the referent,” and not just to “the perceptual frame of orientation”? In figure 3.9 it is indeed impossible to distinguish between these two. The perceptual frame of the ball is the visual scene as a whole. Its orientation, and in particular its vertical direction, determines whether some dimension of the relatum (the chair) is in canonical position. More generally, a referent’s perceptual frame of orientation will normally be the experienced vertical, as it derives from vestibular and visual environmental cues, and

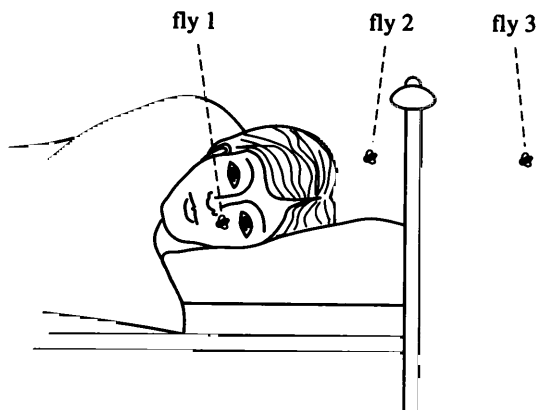


Figure 3.10

According to the principle of canonical orientation, fly 1 can be intrinsically *to the left of* John's nose, and fly 2, but not fly 3, can be *above* John's head (reproduced from Levelt 1984).

will be the same for referent and relatum. But there are exceptions in which a dominant visual Gestalt adopts the function of perceptual frame for the referent. This can happen in the scene of figure 3.10, which is reprinted here from Levelt (1984).

In that paper I argued that it is not impossible in this case to say about fly 2 in the picture: *there is a fly above John's head* even though the top/bottom dimension of John's head is not in canonical orientation. And this is in agreement with the principle. To show this, let us consider the figure in some more detail, beginning at the location of fly 1. Here John's face is a quite dominant background pattern which may become the perceptual frame of orientation for the fly. In that case, the principle of canonical orientation predicts that it is appropriate to say, *there is a fly to the left of John's nose*. This is because the intrinsic left/right dimension in which the fly is spatially related to John's nose is canonically oriented with respect to the perceptual frame. It is in a plane perpendicular to the top/bottom dimension of the face. And fly 2 may similarly take John's face as its perceptual frame, because it is so close to it. If this is a subject's experience, then it is appropriate to say *there is a fly above John's head*, according to the principle. The experimental findings by Carlson-Radvansky and Irwin (1993; cf. figure 3.8g) now confirm that this can indeed be the case.⁷ Fly 3 is further away from John's head and does not naturally take John's head as its perceptual frame of reference. Hence it is less appropriate here to say it is "above" John's head. Notice that in these three cases John's head itself has the bed and its normal gravitational orientation as its perceptual frame. Hence the perceptual frame of the referent can be different from the larger perceptual frame in which the relatum

is embedded. In other words, there can be a hierarchy of frames, and it is not necessarily the case that the referent and the relatum share a frame.

Garnham (1989) challenged the principle of canonical orientation. Although he agreed with the intuitions concerning the scenes in figure 3.9, he rejected those with respect to figure 3.10. That allowed him to ignore the distinction between the referent's and the relatum's perceptual frame and to formulate a really simple principle, the "framework vertical constraint," which says that "no spatial description may conflict with the meanings of *above* and *below* defined by the framework in which the related objects are located." But the results by Carlson-Radvansky and Irwin (1993) for scenes *e* and *g* in figure 3.8 contradict this because, according to Garnham, *above/below* derives in this case from the normal gravitational framework. Hence there is a conflict between the meaning of *above* in this framework and the description *the ball is above the chair*, which should make this description impossible according to his constraint, but it does not. The findings are, however, in agreement with the principle of canonical orientation because the experiments involved cases such as the one just discussed for fly 2 in figure 3.10.

Garnham's critique of my 1984 formulation of the principle can, in part, be traced back to a vagueness of the term *canonical position*. It does not positively exclude the following strict interpretation: the dimension on which the intrinsic location is made should coincide with the same dimension in the perceptual frame. This is obviously false, as Garnham (1989) correctly pointed out. For instance, "if a vehicle is parked across a street, a bollard [traffic post] to the intrinsic right of the vehicle can still be described as to its right" (p. 59), even if the perceptual frame for the bollard is given by the street (whose right side is opposite to the vehicle's right side). The only tenable interpretation of "canonical position" is a weaker one:

Canonical position The top-down dimension of the relatum is in canonical position if it coincides with the vertical dimension of the referent's perceptual frame. The left-right and front-back dimensions of the relatum are in canonical orientation if and only if they are in a plane that is perpendicular to the vertical in the referent's perceptual frame.

With this further specification, then, the principle of canonical orientation seems to be in agreement with intuition and with experimental data. If in a scene canonical orientation does not hold, the intrinsic system is evaded by the standard average European (SAE) language user; it is preempted by the deictic or by the absolute system.⁸

In this section I have discussed various properties of perspective systems that are of pragmatic significance. We have seen that systems differ in inferential potential and

in their demands on coordination between interlocutors. We also have seen that if one system is dominant, concurring systems are not totally dormant in the speaker's mind. Their rivalry appears from the kind and speed of a subject's spatial judgments, and the outcome depends on quite abstract properties of the rivaling systems, as is the implication of the principle of canonical orientation.

3.4 Ellipsis in Spatial Expressions

Perspective taking is one aspect of our thinking for speaking. When we talk about spatial configurations, we create predications about spatial properties of entities or referents in the scene. These predications usually relate the entity to some relatum in terms of some perspective system. In short, the process of perspective taking maps a spatial representation onto a propositional or semantic one. The latter is the speaker's *message*, which consists of lexical concepts, that is, concepts for which there are words in the speaker's target language.

This state of affairs is well exemplified in figure 3.5. The same pattern is expressed in two systematically different ways, dependent on the speakers' perspectives. Figure 3.11 represents one critical detail (circled) of this example. Depending on the perspective taken, the same referent/relatum relation is expressed as *left* or as *right*. Figure 3.11 expresses that the choice of lexical concept (and ultimately of lexical item) depends on the perspective system being used, that is, on thinking for speaking. It is important to be clear on the underlying assumption here. It is that the spatial representation is itself perspective-free; it is neither intrinsic nor deictic. This assumption may or may not be correct, and I will return to it below.

The issue in this section is whether spatial ellipsis originates *before* or *after* perspective taking. In other words, does the speaker decide not to mention a particular feature of the spatial representation, or rather, does the speaker decide not to express a particular lexical concept? In the first case we will speak of "deep ellipsis"; in the latter case, of "surface ellipsis" (roughly following Hankamer and Sag 1976 on "deep" and "surface anaphora").

Compare the following two descriptions from our data. Both relate to the encircled trajectory in the left pattern of figure 3.12, plus the move that precedes it. The first description is nonelliptic with respect to the directional expression, the second one is elliptic in that respect.

Full deictic: "Right to yellow. Right to blue. Finished."

Elliptic deictic: "From pink we go right one unit and place a yellow dot. One, er, one unit from the yellow dot we place a blue dot."

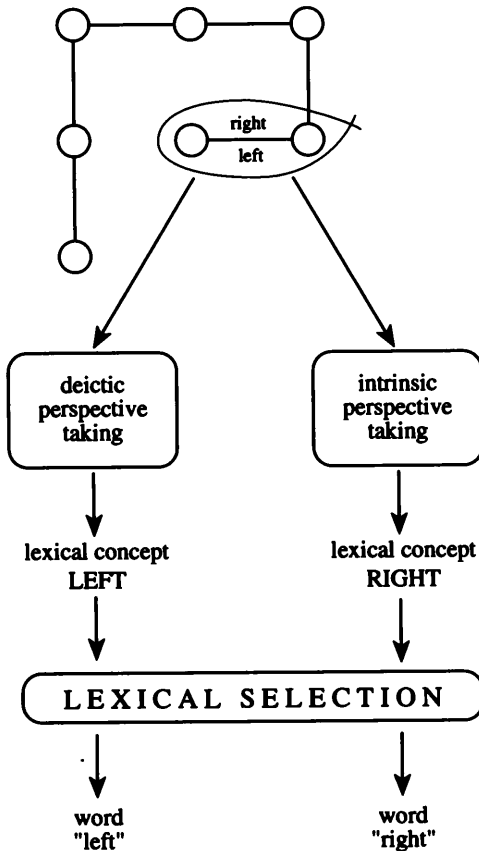


Figure 3.11

Perspective taking is involved in deriving a semantic or “propositional” representation from a spatial representation.

The crucial feature of the latter, elliptic expression is that it contains no spatial term that relates the blue dot to the (previous) yellow one. How does the speaker create this ellipsis? There are, essentially, two possibilities. The first one is that the speaker in scanning the spatial configuration recognizes that the new visual direction is the same as the previous one. Before getting into perspective taking, the speaker decides not to prepare that direction for expression again. This is deep ellipsis. The second possibility is that the speaker does apply deictic perspective to the second move, thus activating the lexical concept RIGHT a second time. This repeated activation of the concept then leads to the decision not to formulate the lexical concept a second time,

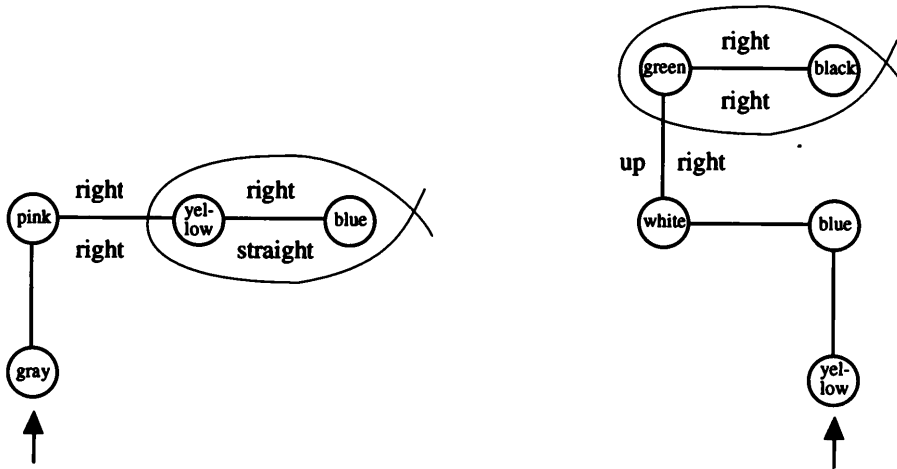


Figure 3.12

Deictic and intrinsic descriptions for two patterns. Can the last spatial term (*right, straight*) be deleted?

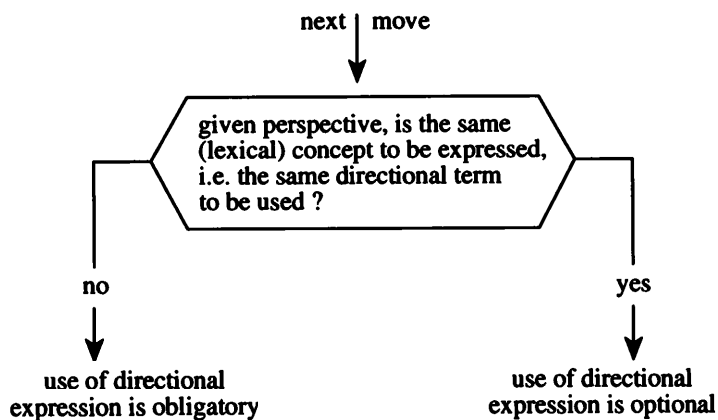
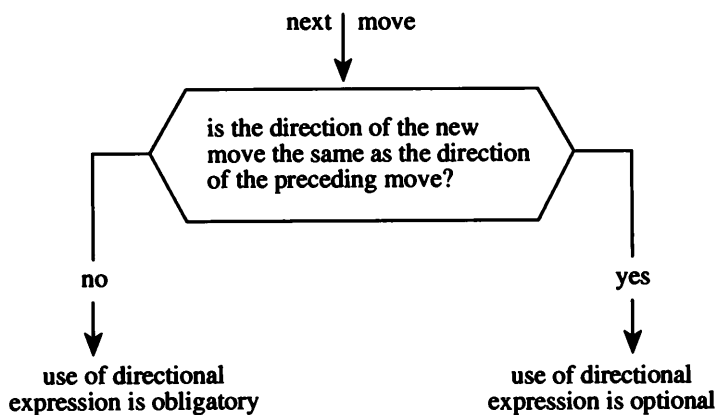
that is, not to repeat the word *right*. This is surface ellipsis. These two alternatives are depicted in figure 3.13.

The alternatives can now be distinguished by observing what happens in descriptions from an intrinsic perspective. Here is an instance of a full intrinsic description of the same trajectory:

Full intrinsic: “Then to the right to a yellow node and straight to a blue node.”

Can the same state of affairs be described elliptically? This should produce something like: *Then to the right to a yellow node and to a blue node*. The answer is not obvious; intuitions waver here. In case of deep ellipsis this should be possible. Just as the previous deictic speaker, the present intrinsic one will scan the spatial scene and recognize that the new direction is the same as the previous one and the speaker may decide not to prepare it again for expression; it is *optional* to mention the direction. But in case of surface ellipsis the intrinsic speaker has a problem. In the intrinsic system the direction of the first move is mapped onto the lexical concept *RIGHT*, whereas the direction of the second move is mapped onto *STRAIGHT*. Because the latter is not a repetition of the former, it has to be formulated in speech. In other words, the condition for surface ellipsis is not met for the intrinsic speaker; it is *obligatory* to use a directional expression.

This state of affairs can now be exploited to test empirically whether spatial ellipsis is deep or surface ellipsis. Does ellipsis occur in intrinsic descriptions of this kind? If

MODEL 1 "Surface ellipsis" (ellipsis is perspective- dependent)**MODEL 2 "Deep ellipsis" (ellipsis is perspective- independent)****Figure 3.13**

Surface ellipsis versus deep ellipsis. Is it reiterating a lexical concept or a spatial direction that matters?

so, we have an argument for deep ellipsis. And we can create an alternative case where surface ellipsis is possible for intrinsic descriptions, but not deep ellipsis. An example concerns the encircled trajectory in the right pattern of figure 3.12. A normal full intrinsic description of this trajectory (plus the previous one) is

Full intrinsic: Then right to green. And then right to black.

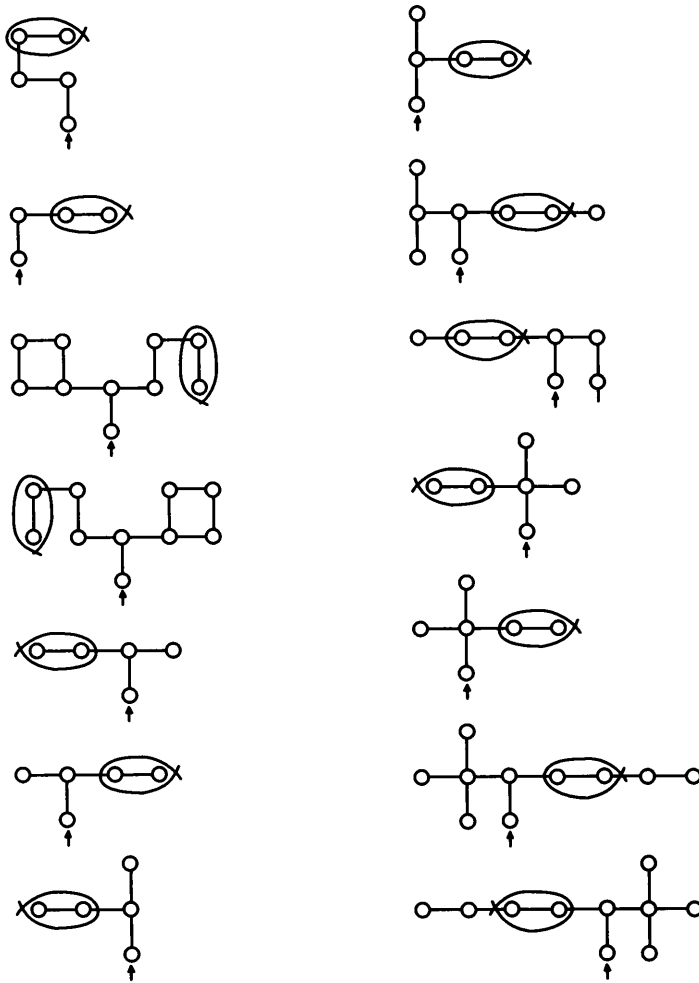
Is surface ellipsis possible here, producing “Then right to green. And to black” or some similar expression? That is an empirical issue. It should be clear that neither deep nor surface ellipsis is possible in a deictic description of this pattern. Take this full deictic description from our data:

Full deictic: From white we go up to a green circle. And from the green circle we go right to a black circle.

Surface ellipsis is impossible here because “right” is not a repetition of the previous directional term (“up”). Deep ellipsis is impossible because the trajectory direction is different from the previous one. Hence, if we find ellipsis in such cases, we will have to reject both models.

In an experiment reported in Levelt (1982a,b) we had asked 53 subjects to describe 53 colored dot patterns, among them those in figure 3.12. I will call the circled moves in these patterns “critical moves” because the surface and deep models make predictions about them that differ critically for deictic and intrinsic descriptions in the way just described. Among the test patterns there were 14 that contained such critical moves; they are given in figure 3.14. I checked all 53 subjects to determine whether they made elliptic descriptions for any of these 14 critical trajectories. I removed all subjects who did not have a consistent perspective over these 14 critical patterns; a subject’s 14 pattern descriptions should either be all deictic or all intrinsic. This left me with 31 consistent deictic subjects and 13 consistent intrinsic ones,⁹ and hence with $44 \times 14 = 616$ pattern descriptions to be checked. In this set I found a total of 43 cases of ellipsis.¹⁰ These are presented in table 3.1.

The table presents predictions and results under both models of ellipsis. For each critical move I determined whether a directional term would be obligatory or optional (i.e., elidable) under the model in deictic and in intrinsic descriptions (such as I did above for the critical moves of the patterns in figure 3.12). Hence there are four cases per model. The table presents the actual occurrence of ellipsis for these four cases within each model. It should be noticed that the two models make the same predictions with respect to deictic descriptions; if use of a directional term is obligatory under the surface model, it is also obligatory under the deep model and vice versa. But this is not so for the intrinsic descriptions.

**Figure 3.14**

Fourteen test patterns containing “critical moves,” including the two example patterns of figure 3.12. Each test pattern includes either the one or the other example pattern as a substructure (though rotated in two cases). The critical moves are circled.

Table 3.1
Distribution of Elliptical Descriptions under Surface and Deep models of Ellipsis

Model	Surface ellipsis			Deep ellipsis		
	deictic	intrinsic	Total	deictic	intrinsic	Total
Directional term is						
obligatory	1	18	19	1	0	1
optional	24	0	24	24	18	42
Total	25	18	43	25	18	43

If a model says “obligatory,” but ellipsis does nevertheless occur, that model is in trouble. How do the two models fare? It is immediately obvious from the table that the surface model is out. Where it prescribes obligatory use of a directional term, there are no less than 18 violations among the intrinsic descriptions (i.e., cases of ellipsis) and one among the deictic descriptions, for a total of 19. That is almost half our sample. In contrast, the deep model is in good shape; there is only one deictic description that violates it.¹¹ All other deictic and all intrinsic descriptions respect the deep model.

These findings show that the decision to skip mentioning a direction is really an early step in thinking for speaking. It precedes the speaker’s application of a perspective; the speaker’s linguistic perspective system is irrelevant here. The decision is based on a visual or imagistic representation, not on a semantic (lexical-conceptual) representation (see figure 3.11). This is, probably, the same level of representation where linearization decisions are taken. When we describe 2-D or 3-D spatial patterns (such as the patterns in figure 3.14 or the layout of our living quarters), we must decide on some order of description because speech is a linear medium of expression. The principles governing these linearization strategies (Levelt 1981, 1989) are non-linguistic (and in fact nonsemantic) in character; they relate exclusively to the image itself.

But these very clear results on ellipsis create a paradox. If ellipsis runs on a perspective-free spatial representation, spatial representations are apparently *not* perspectivized. But this contradicts the convincing experimental findings reported by Brown and Levinson (1993) and by Levinson (chapter 4, this volume), which show that when a language uses absolute perspective, its speakers use *oriented* (i.e., perspective-dependent) spatial representations in nonlinguistic spatial matching tasks. For instance, the subject is shown an array of two objects A and B on a table, where A is (deictically) left of B (hence A-B). Then the subject is turned around 180° to another table with two arrays of the same objects, namely, A-B and B-A, and then asked to

indicate which of the two arrays is identical to the one the subject saw before. The “absolute” subject invariably chooses the B-A array, where A is deictically to the *right* of B. What the subject apparently preserves is the absolute direction of the vector $\overrightarrow{AB}$. A native English or Dutch subject, however, typically produces the deictic response (A-B). Hence spatial representations *are* perspectivized already, in the sense that they follow the dominant perspective of the language even in nonlinguistic tasks, that is, where there is no “thinking for speaking” taking place.¹²

How to solve this paradox? One point to note is that the above ellipsis data and Brown and Levinson’s (1993) data on oriented spatial representations involve different perspectives, and the ellipsis predictions are different for different perspectives. As can be seen from table 3.1, columns 1 and 4, the same predictions result from the deep and the surface model under deictic perspective. The two models can only be distinguished when the speaker’s perspective is intrinsic (cf. columns 2 and 5); violations under deictic perspective could only show that *neither* model is correct. In this respect, absolute perspective behaves like deictic perspective. If a speaker’s perspective is absolute, the deep and surface models of ellipsis make the same predictions; if two arcs have the same spatial direction or orientation, the corresponding lexical concepts will be the same as well (e.g., both north, or both east).

In other words, ellipsis data of the kind analyzed here can only distinguish between the deep and surface models if the speaker’s perspective is intrinsic. One could then argue that Brown and Levinson’s findings show that absolute and deictic perspective are “Whorfian,” that is, a property of the spatial representation itself. If, in addition, the intrinsic system is not Whorfian in the same sense, the above ellipsis data would be explained as well.

The problem is, of course, why intrinsic perspective should be non-Whorfian. After all, speakers of Mopan, exclusive users of intrinsic perspective, will profit from registering the position of foregrounded objects relative to background objects that have intrinsic orientation. If at some later time the scene is talked about from memory, that information about intrinsic position will be crucial for an intrinsic spatial description. But if we discard the option of excluding intrinsic perspective from “Whorfianness,” the paradox remains.

More important, it seems to me, is the fact noted in the introduction that perspective is linguistically free. There is no “hard-wired” mapping from spatial to semantic representations. What we pick out from a scene in terms of entities and spatial relations to be expressed in language is not subject to fixed laws. There are preferences, for sure, following Gestalt properties of the scene, human interest, and so on, but they are no more than preferences. Similarly, we can go for one perspective or another if our culture leaves us the choice, and this chapter has discussed various reasons for choosing one perspective rather than another, depending on communica-

tive intention and situation. It is correct to say that Guugu Yimithirr speakers can choose from only one, absolute perspective, but that does not obliterate their freedom in expressing spatial configurations in language. The choice of referents, relata, spatial relations to be expressed, the pattern of linearization chosen when the scene is complex, and even the decision to express absolute perspective at all (e.g., *A* is north of *B*, rather than *A* is in *B*'s neighborhood) are prerogatives of the speaker that are not thwarted by the limited choice of perspective. As all other speakers, the Guugu Yimithirr can attend to various aspects of their spatial representations; they can express in language what they deem relevant and in ways that are communicatively effective. This would be impossible if the spatial representation dictated its own semantics. Hence, Brown and Levinson's (1993) important Whorfian findings cannot mean that spatial and semantic representations have a "hard-wired" isomorphism. A more likely state of affairs is this. A culture's dominant perspective makes a speaker attend to spatial properties that are relevant to that perspective because it will facilitate (later) discourse about the scene. In particular, these attentional biases make the speaker register in memory spatial features that are perspective-specific, such as the absolute orientation of the scene. This does not mean, however, that an ellipsis decision must make reference to such features. That one arc in figure 3.12 is a continuation of another arc is a spatial feature in its own right that is available to a speaker of any culture. Any speaker can attend to it and make it the ground for ellipsis. In other words, the *addition* of perspective-relevant spatial features does not preempt or suppress the registration of other spatial properties that can be referred to or used in discourse.

3.5 Conclusion

This chapter opened by recalling, from Levelt (1989), the distinction between macroplanning and microplanning. In macroplanning we elaborate our communicative intention, selecting information whose expression can be effective in revealing our intentions to a partner in speech. We decide on what to say. And we linearize the information to be expressed, that is, we decide on what to say first, what to say next, and so forth. In microplanning, or "thinking for speaking," we translate the information to be expressed in some kind of "propositional" format, creating a semantic representation, or message, that can be formulated. In particular, this message must consist of lexical concepts, that is, concepts for which there are words in the target language. When we apply these notions to spatial discourse, we can say that macroplanning involves selecting referents, relata, and their spatial relations for expression. Microplanning involves, among other things, applying some perspective system that will map spatial directions/relations onto lexical concepts.

The chapter has been largely about microplanning, in particular about the pragmatics of different perspective systems. It has considered the advantages and disadvantages of deictic, intrinsic, and absolute systems for spatial reasoning and for speaker/hearer coordination in spatial discourse. It has also considered how a speaker deals with situations in which perspective systems are not aligned.

“Thinking for speaking” led, as a matter of course, to the question whether this perspectival thinking is just for speaking or more generally permeates our spatial thinking, that is, in some Whorfian way. The discussed recent findings by Levinson and Brown strongly suggest that such is indeed the case. I then presented experimental data on spatial ellipsis showing that perspective is irrelevant for a speaker’s decision to elide a spatial direction term. Having speculated that the underlying spatial representation might be perspective-free, contrary to the Whorfian findings, I argued that this is paradoxical only if the mapping from spatial representations onto semantic representations is “hard-wired.” But this is not so; speakers have great freedom in both macro- and microplanning. There are no strict laws that govern the choice of relatum and referent, that dictate how to linearize information, and so forth. In particular, there is no law that the speaker must acknowledge orientedness of a spatial representation (if it exists) when deciding on what to express explicitly and what implicitly. There are only (often strong) preferences here that derive from Gestalt factors, cultural agreement on perspective systems, ease of coordination between interlocutors, requirements of the communicative task at hand, and so on.

Still, it is not my intention to imply that anything goes in thinking for speaking. Perspective systems are interfaces between our spatial and semantic “modules” (in Jackendoff’s sense, chapter 1, this volume), performing well-defined restricted mapping operations. The interfacing requirements are too specific for these perspective systems to be totally arbitrary. But much more challenging is the dawning insight from anthropological work that there are only a few such systems around. What is it in our biological roots that makes the choice so limited?

Notes

1. I am in full agreement with Levinson’s taxonomy of frames of reference (here called “perspective systems”) in chapter 4 of this volume. The main distinction is between relative, intrinsic, and absolute systems, and each has an egocentric and an allocentric variant. The three perspective systems discussed here are relative egocentric (=deictic), intrinsic allocentric, and absolute allocentric. The relative systems are three-place relations between referent, relatum, and base entity (“me” in the deictic system); the intrinsic and absolute systems are two-place relations between referent and relatum.

2. Brown and Levinson (1993) present the case of Tenejapan, where the traverse direction in the absolute system is not polarized, that is, spanned by two converse terms; there is just one

term meaning “traverse.” Obviously, the notion of converseness is not applicable. The notion of transitivity, however, *is* applicable and holds for this system (see below in text).

3. Barbara Tversky (personal communication) has correctly pointed out that Bühler (1934) would treat this case as a derived form of deixis, “Deixis am Phantasma,” where the speaker *imagines* being somewhere (for instance in the network). There would be two speakers then, a real one and imaginary one, each forming a base for a (different) deictic system. This is unobjectionable as long as we do not confound the two systems. But Bühler’s case is not strong for this network. It is not essential in the route-type description that “I” (the speaker in his imagination) make the moves and turns. If there were a ball rolling through the pattern, the directional terms would be just the same. But a ball doesn’t have deictic perspective. What the speaker in fact does in this description is to use the last directed *path* as the relatum for the subsequent path. The new path is straight, right, or left from the current one. Hence it is the *intrinsic* orientation of the current path that is taken as the relatum.

4. I am ignoring a further variable, the listener’s viewpoint/orientation. Speakers can and often do express spatial relations from the *interlocutor’s* perspective, as in *for you, the ball is to the left of the chair*. Conditions for this usage have been studied by Herrmann and his colleagues (cf. Herrmann and Grabowski 1994).

5. Here I am considering only one case of nonalignment, namely, a 90° angle between the relevant bases. Another case studied by Carlson-Radvansky and Irwin (1993) is 180° nonalignment.

6. Carlson-Radvansky and Irwin do not discuss item-specific effects, although it is likely that the type of relatum used is not irrelevant. It is the case, though, that their statistical findings always agree between subject and item analyses. Another point to keep in mind is that the experimental procedure may invite the development of “perspective strategies” on the part of subjects, and occasionally the employment of an “unusual” perspective.

7. Carlson-Radvansky and Irwin included several scenes that were formally of the same type as scene (g) in figure 3.8, among them the one in figure 3.9 with fly 2.

8. There is, however, no reason why this should also hold in other cultures. Stephen Levinson (personal communication), for instance, has presented evidence that the principle does not hold for speakers of Tzeltal, who can use their intrinsic system when the relatum’s critical dimension is not in canonical orientation. But the Tzeltal intrinsic system differs substantially from the standard average European (SAE) intrinsic system (see Levinson 1992a). What is intrinsic *top/bottom* in SAE is “longest dimension” or the “modal axis” of an object in Tzeltal; the former, but not the latter, has a connotation of verticality.

9. These numbers differ from those reported in Levelt (1982b) because the present selection criterion is a different one.

10. My criterion for ellipsis was a strict one. There should, of course, be no directional term, but there also should be no coordination that can be interpreted as one directional term having scope over two constituents, as in *From pink right successively yellow and blue* or *A road turns right from pink and meets first yellow and then blue*. I have excluded all cases where subjects mention a line on which the nodes are located.

11. The case occurs in a deictic description of the fourth pattern down the first column in figure 3.14. It goes as follows. *From there left to a pink node. And from there to a green node.* This obviously violates both models of ellipsis. I prefer to see it as a mistake or omission.

12. The discussion that follows in the text is much inspired by discussions with Stephen Levinson.

References

- Brown, P., and Levinson, S. C. (1993). Linguistic and nonlinguistic coding of spatial arrays: Explorations in Mayan cognition. Working paper no. 24, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer. A major part on deixis from this work appeared in translation in R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action: Studies in deixis and related topics*, 9–30. Chichester: Wiley, 1982.
- Byrne, R. M. J., and Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language*, 28, 564–575.
- Carlson-Radvansky, L. A., and Irwin, D. E. (1993). Frames of reference in vision and language: Where is above? *Cognition*, 46, 223–244.
- Carlson-Radvansky, L. A., and Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *Journal of Memory and Language*, 33, 646–671.
- Ehrich, V. (1982). The structure of living space descriptions. In R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action: Studies in deixis and related topics*, 219–249. Chichester: Wiley.
- Friederici, A. D., and Levelt, W. J. M. (1990). Spatial reference in weightlessness: Perceptual factors and mental representations. *Perception and Psychophysics*, 47, 253–266.
- Garnham, A. (1989). A unified theory of the meaning of some spatial relational terms. *Cognition*, 31, 45–60.
- Hankamer, J., and Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry*, 7, 391–426.
- Herrmann, T., and Grabowski, J. (1994). *Speechen: Psychologie der Sprachproduktion*. Heidelberg: Spektrum.
- Hill, A. (1982). Up/down, front/back, left/right: A contrastive study of Hausa and English. In J. Weissenborn and W. Klein (Eds.), *Here and there: Crosslinguistic studies on deixis and demonstration*, 13–42. Amsterdam: Benjamins.
- Levelt, W. J. M. (1981). The speaker's linearization problem. *Philosophical Transactions of the Royal Society, London*, B95, 305–315.
- Levelt, W. J. M. (1982a). Linearization in describing spatial networks. In S. Peters and E. Saarinen (Eds.), *Processes, beliefs, and questions*, 199–220. Dordrecht: Reidel.
- Levelt, W. J. M. (1982b). Cognitive styles in the use of spatial direction terms. In R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action: Studies in deixis and related topics*, 251–268. Chichester: Wiley.

- Levelt, W. J. M. (1984). Some perceptual limitations on talking about space. In A. van Doorn, W. van de Grind, and J. Koenderink (Eds.), *Limits of perception: Essays in honour of Maarten A. Bouman*, 323–358. Utrecht: VNU Science Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA.: MIT Press.
- Levinson, S. C. (1992a). Vision, shape, and linguistic description: Tzeltal body-part terminology and object description. Working paper no. 12, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Levinson, S. C. (1992b). Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. Working paper no. 13, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Shepard, R. R., and Hurwitz, S. (1984). Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition*, 18, 161–193.
- Slobin, D. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, and H. Filip (Eds.), *Berkeley Linguistics Society: Proceedings of the Thirteenth Annual Meeting*, 435–444. Berkeley: Berkeley Linguistics Society.
- Talmy, L. (1983). How language structures space. In H. Pick and L. Acredolo (Eds.), *Spatial orientation: Theory, research, and application*. New York: Plenum Press.
- Taylor, H. A., and Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language* (in press).
- Tversky, B. (1991). Spatial mental models. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, vol. 27, 109–146. New York: Academic Press.

Chapter 4

Frames of Reference and Molyneux's Question: Crosslinguistic Evidence

Stephen C. Levinson

4.1 What This is All About

The title of this chapter invokes a vast intellectual panorama; yet instead of vistas, I will offer only a twisting trail. The trail begins with some quite surprising cross-cultural and crosslinguistic data, which leads inevitably on into intellectual swamps and minefields—issues about how our “inner languages” converse with one another, exchanging spatial information.

To preview the data, first, languages make use of different frames of reference for spatial description. This is not merely a matter of different use of the same set of frames of reference (although that also occurs); it is also a question of *which* frames of reference they employ. For example, some languages do not employ our apparently fundamental spatial notions of *left/right/front/back* at all; instead they may, for example, employ a cardinal direction system, specifying locations in terms of *north/south/east/west* or the like.

There is a second surprising finding. The choice of a frame of reference in linguistic coding (as required by the language) correlates with preferences for the same frame of reference in nonlinguistic coding over a whole range of nonverbal tasks. In short, there is a cross-modal tendency for the same frame of reference to be employed in language tasks, recall and recognition memory tasks, inference tasks, imagistic reasoning tasks, and even unconscious gesture. This suggests that the underlying representation systems that drive all these capacities and modalities have adopted the same frame of reference.

These findings, described in section 4.2, prompt a series of theoretical ruminations in section 4.3. First, we must ask whether it even makes sense to talk of the “same” frame of reference across modalities or inner representation systems.¹ Second, we must clarify the notion “frame of reference” in language, and suggest a slight reformation of the existing distinctions. Then we can, it seems, bring some of the distinctions made in other modalities into line with the distinctions made in the study of

language, so that some sense can be made of the idea of “same frame of reference” across language, nonverbal memory, mental imagery, and so on. Finally, we turn to the question Why does the same frame of reference tend to get employed across modalities or at least across distinct inner representation systems? It turns out that information in one frame of reference cannot easily be converted into another, distinct frame of reference. This has interesting implications for what is known as “Molyneux’s question,” the question about how and to what extent there is cross-modal transfer of spatial information.

4.2 Cross-Modal Transfer of Frame of Reference: Evidence from Tenejapan Tzeltal

To describe where something (let us dub it the “figure”) is with respect to something else (let us call it the “ground”) we need some way of specifying angles on the horizontal. In English we achieve this either by utilizing features or axes of the ground (as in “the boy is at the front of the truck”) or by utilizing angles derived from the viewer’s body coordinates (as in “the boy is to the left of the tree”). The first solution I shall call an “intrinsic frame of reference”; the second, a “relative frame of reference” (because the description is relative to the viewpoint—from the other side of the tree the boy will be seen to be to the right of the tree). The notion “frame of reference” will be explicated in section 4.3 but can be thought of as labeling distinct kinds of coordinate systems.

At first sight, and indeed on close consideration (see, for example, Clark 1973; Miller and Johnson-Laird 1976), these solutions seem inevitable, the only natural solutions for a bipedal creature with particular bodily asymmetries on our planet. But they are not. Some languages use just the first solution. Some languages use neither of these solutions; instead, they solve the problem of finding angles on the horizontal plane by utilizing fixed bearings, something like our cardinal directions north, south, east, and west. Spatial descriptions utilizing such a solution can be said to be in an “absolute” frame of reference (because the angles are not relative to a point of view, i.e., are not relative, and are also independent of properties of the ground object, i.e., are not intrinsic). A tentative typology of the three major frames of reference in language, with some indication of the range of subtypes, will be found in section 4.3. Here I wish to introduce one such absolute system, as found in a Mayan language.

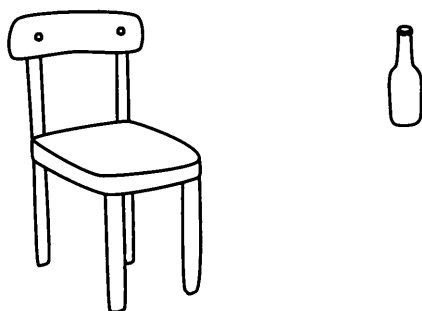
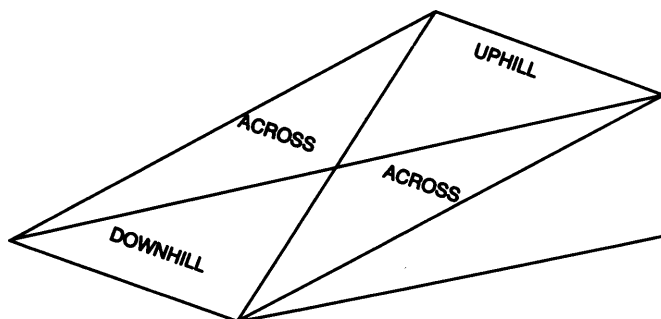
Tzeltal is a Mayan language widely spoken in Chiapas, Mexico, but the particular dialect described is spoken by at least 15,000 people in the Indian community of Tenejapa; I will therefore refer to the relevant population as Tenejapans. The results reported here are a part of an ongoing project, conducted with Penelope Brown (Brown and Levinson 1993a,b; Levinson and Brown 1994).

4.2.1 Tzeltal Absolute Linguistic Frame of Reference

Tzeltal has an elaborate intrinsic system (see Brown 1991; Levinson 1994), but it is of limited utility for spatial description because it is usually only employed to describe objects in strict contiguity. Thus for objects separated in space, another system of spatial description is required. This is in essence a cardinal direction system, although it has certain peculiarities. First, it is transparently derived from a topographic feature: Tenejapa is a large mountainous tract, with many ridges and crosscutting valleys, which nevertheless exhibits an overall tendency to fall in altitude toward the north-northwest. Hence *downhill* has come to mean (approximately) north, and *uphill* designates south. Second, the coordinate system is deficient, in that the orthogonal *across* is labeled identically in both directions (east and west); the particular direction can be specified periphrastically, by referring to landmarks. Third, there are therefore certain ambiguities in the interpretation of the relevant words. Despite this, however, the system is a true fixed-bearing system. It applies to objects on the horizontal as well as on slopes. And speakers of the language point to a specific direction for *down*, and they will continue to point to the same compass bearing when transported outside their territory. Figure 4.1 may help to make the system clear.

The three-way semantic distinction between *up*, *down*, and *across* recurs in a number of distinct lexical systems in the language. Thus there are relevant abstract nominals that describe directions, specialized concrete nominals of different roots that describe, for example, edges along the relevant directions, and motion verbs that designate ascending (i.e., going south), descending (going north), and traversing (going east or west). This linguistic ramification, together with its insistent use in spatial description, make the three-way distinction an important feature of language use.

There are many other interesting features of this system (Brown and Levinson 1993a), but the essential points to grasp are the following. First, this is the basic way to describe the relative locations of all objects separated in space on whatever scale. Thus if one wanted to pick out one of two cups on a table, one might ask for, say, the uphill one; if one wanted to describe where a boy was hiding behind a tree, one might designate, say, the north (downhill) side of the tree; if one wanted to ask where someone was going, the answer might be "ascending" (going south); and so forth. Second, linguistic specifications like our *to the left*, *to the right*, *in front*, *behind* are not available in the language; thus there is no way to encode English locutions like "pass the cup to the left," "the boy is in front of the tree," or "take the first right turn."² Third, the use of the system presupposes a good sense of direction; tests of this ability to keep track of directions (in effect, to dead reckon), show that Tenejapans, even



"The bottle is uphill of the chair."

waxal ta y-ajik'ol xila te limits
standing at its-uphill chair the bottle

Figure 4.1

Tenejapan Tzeltal uphill/downhill system.

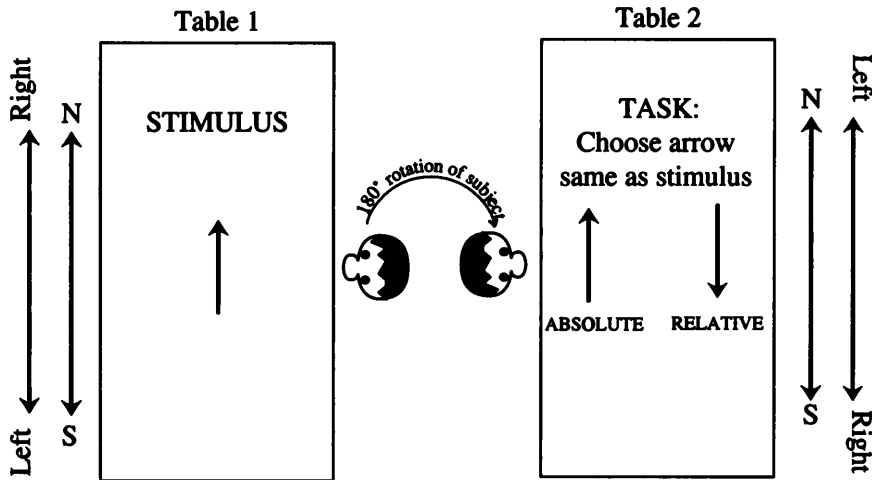


Figure 4.2
Underlying design of the experiments.

without visual access to the environment, do indeed maintain the correct bearings of various locations as they move in the environment.

In short, the Tzeltal linguistic system does not provide familiar viewer-centered locations like "turn to the left" or "in front of the tree." All such directions and locations can be adequately coded in terms of antecedently fixed, absolute bearings. Following work on an Australian language (Haviland 1993; Levinson 1992b) where such a linguistic system demonstrably has far-reaching cognitive consequences, a series of experiments were run in Tenejapa to ascertain whether nonlinguistic coding might follow the pattern of the linguistic coding of spatial arrays.

4.2.2 Use of an Absolute Frame of Reference in Nonverbal Tasks

4.2.2.1 Memory and Inference As part of a larger comparative project, my colleagues and I have devised experimental means for revealing the underlying nonlinguistic coding of spatial arrays for memory (see Baayen and Danziger 1994). The aim is to find tasks where subjects' responses will reveal which frame of reference, intrinsic, absolute, or relative, has been employed during the task. Here we concentrate on the absolute versus relative coding of arrays. The simple underlying design behind all the experiments reported here can be illustrated as follows. A male subject, say, sees an array on a table (table 1): an arrow pointing to his right, or objectively to the north (see figure 4.2). The array is then removed, and after a delay, the subject

is rotated 180 degrees to face another table (table 2). Here there are, say, two arrows, one pointing to his right and one to his left—that is, one to the north and one to the south. He is then asked to identify the arrow like the one he saw before. If he chooses the one pointing to his right (and incidentally to the south), it is clear that he coded the first arrow in terms of his own bodily coordinates, which have rotated with him. If he chooses the other arrow, pointing north (and to his left), then it is clear that he coded the original array without regard to his bodily coordinates, but with respect to some fixed bearing or environmental feature. Using the same method, we can explore a range of different psychological faculties: recognition memory (as just sketched), recall memory (by, for example, asking the subject to place an arrow so that it is the same as the one on table 1) and various kinds of inference (as sketched below).

We will describe here just three such experiments in outline form (see Brown and Levinson 1993b for further details and further experiments). They were run on at least twenty-five Tenejapan subjects (depending on the experiment) of mixed age and sex, and a Dutch comparison group of at least thirty-nine subjects of similar age/sex composition. As far as the distinction between absolute and relative linguistic coding goes, Dutch like English relies heavily of course on a *right/left/front/back* system of speaker-centered coordinates for the description of most spatial arrays. So the hypothesis entertained in all the experiments is the following simple Whorfian conjecture: the coding of spatial arrays—that is, the conceptual representations involved—in a range of nonverbal tasks should employ the same frame of reference that is dominant in the language used in verbal tasks for the same sort of arrays. Because Dutch, like English, provides a dominant relative frame of reference, we expect Dutch subjects to solve all the nonlinguistic tasks utilizing a relative frame of reference. On the other hand, because Tzeltal offers only an absolute frame of reference for the relevant arrays, we expect Tenejapan subjects to solve the nonlinguistic tasks utilizing an absolute frame of reference. Clearly it is crucial that the instructions for the experiments, or the wording used in training sessions, do not suggest one or another of the frames of reference. Instructions (in Dutch or Tzeltal) were of the kind “Point to the pattern you saw before,” “Remake the array just as it was,” “Remember just how it is,” that is, as much devoid of spatial information as possible, and as closely matched in content as could be achieved across languages.

Recall Memory

Method The design was intended to deflect attention from memorizing direction towards memorizing order of objects in an array, although the prime motive was to tap recall memory for direction.³ The stimuli consisted of two identical sets of four model animals (pig, cow, horse, sheep) familiar in both cultures. From the set of four,

three were aligned in random order, all heading in (a randomly assigned) lateral direction on table 1. Subjects were trained to memorize the array before it was removed, then after a three-quarters of a minute delay to rebuild it "exactly as it was," first with correction for misorders on table 1, then without correction under rotation on table 2. Five main trials then proceeded, with the stimulus always presented on table 1, and the response required under rotation, and with delay, on table 2. Responses were coded as "absolute" if the direction of the recalled line of animals preserved the fixed bearings of the stimulus array, and as "relative" if the recalled line preserved egocentric *left* or *right* direction.

Results Ninety-five percent of Dutch subjects were consistent relative coders on at least four out of five trials, while 75% of Tzeltal subjects were consistent absolute coders by the same measure. The remainder failed to recall direction so consistently. For the purposes of comparison across tasks, the data have been analyzed in the following way. Each subject's performance was assigned an index on a scale from 0 to 100, where 0 represents a consistent relative response pattern and 100 a consistent absolute pattern; inconsistencies between codings over trials were represented by indices in the interval. The data are displayed in the graph of figure 4.3, where subjects from each population have been grouped by 20-point intervals on the index.

As the graph makes clear, the curves for the two populations are approximately mirror images, except that Tenejapan subjects are less consistent than Dutch ones. This may be due to various factors: the unfamiliarity of the situation and the tasks, the "school"-like nature of task performed by largely unschooled subjects, or to interference from an egocentric frame of reference that is available but less dominant. Only two Tenejapan subjects were consistent relative coders (on 4 out of 5 trials). This pattern is essentially repeated across the experiments. The result appears to confirm the hypothesis that the frame of reference dominant in the language is the frame of reference most available to solve nonlinguistic tasks, like this simple recall task.

Recognition Memory

Method Five identical cards were prepared; on each there was a small green circle and a large yellow circle.⁴ The trials were conducted as follows. One card was used as a stimulus in a particular orientation; the subject saw this card on table 1. The other four were arrayed on table 2 in a number of patterns so that each card was distinct by orientation (see figure 4.4). The subject saw the stimulus on table 1, which was then removed, and after a delay the subject was rotated and led over to table 2. The subject was asked to identify the card most similar to the stimulus. The eight trials

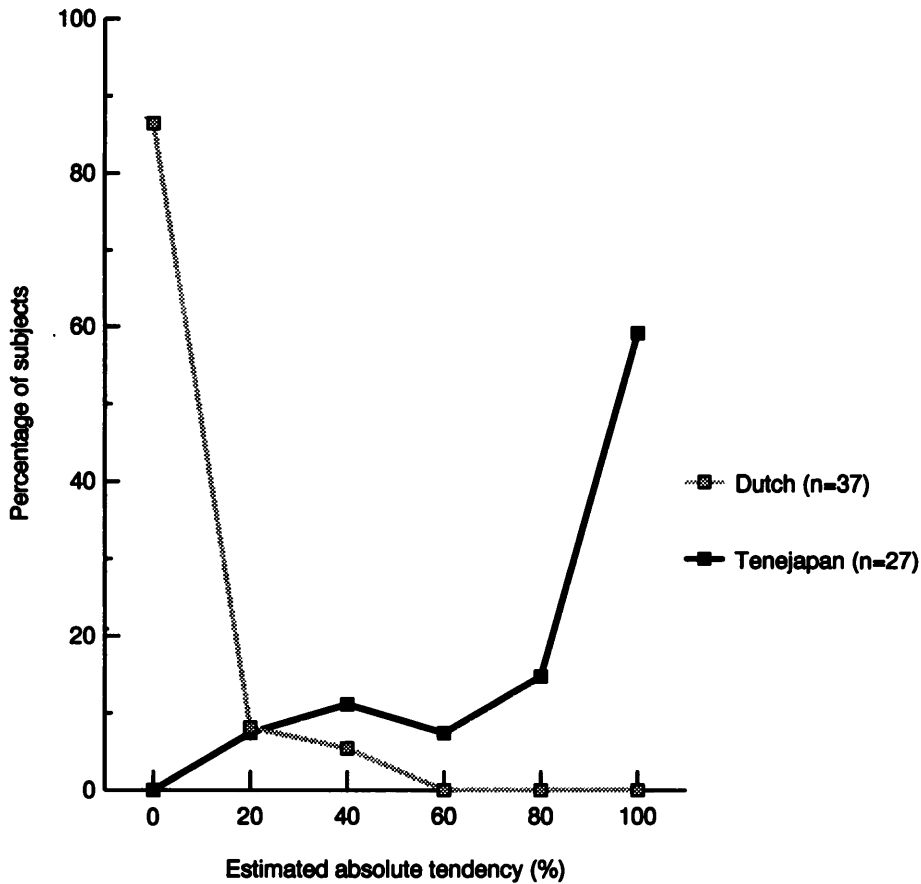


Figure 4.3
Animals recall task: direction.

were coded as indicated in figure 4.3: if the card which maintained orientation from an egocentric point of view (e.g., “small circle toward me”) was selected, the response was coded as a relative response, while the card which maintained the fixed bearings of the circles (“small circle north”) was coded as an absolute response. The other two cards served as controls, to indicate a basic comprehension of the task. Training was conducted first on table 1, where it was made clear that sameness of type rather than token identity was being requested.

Results We find the same basic pattern of results as in the previous task, as shown in figure 4.5. Once again, the Dutch subjects are consistently relative coders, while the

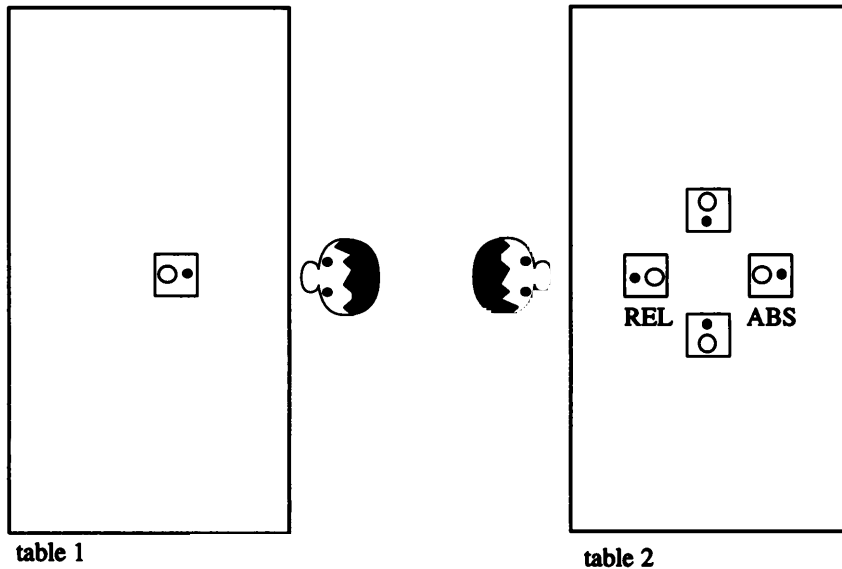


Figure 4.4

Chips recognition task: “absolute” versus “relative” solutions.

Tenejapans are less consistent. Nevertheless, of the Tenejapan subjects who performed consistently over 6 or more of 8 trials, over 80% were absolute coders. The greater inconsistency of Tenejapan subjects may be due to the same factors mentioned above, but there is also here an additional factor because this experiment tested for memory on both the transverse and sagittal (or north-south and east-west) axes. As mentioned above, the linguistic absolute axes are asymmetric: one axis has distinct labels for the two half lines north and south, while the other codes both east and west identically (“across”). If there was some effect of this linguistic coding on the conceptual coding for this nonlinguistic task, one might expect more errors or inconsistency on the east-west axis. This was indeed the case.

Transitive Inference Levelt (1984) observed that relative, as opposed to intrinsic, spatial relations support transitive and converse inferences; Levinson (1992a) noted that absolute spatial relations also support transitive and converse inferences (see also Levelt, chapter 3, this volume). This makes it possible to devise a task where, from two spatial arrays or nonverbal “premises,” a third spatial array, or nonverbal “conclusion” can be drawn by transitive inference utilizing either an absolute or a relative frame of reference. The following task was designed by Eric Pederson and Bernadette Schmitt, and piloted in Tamilnadu by Pederson (1994).

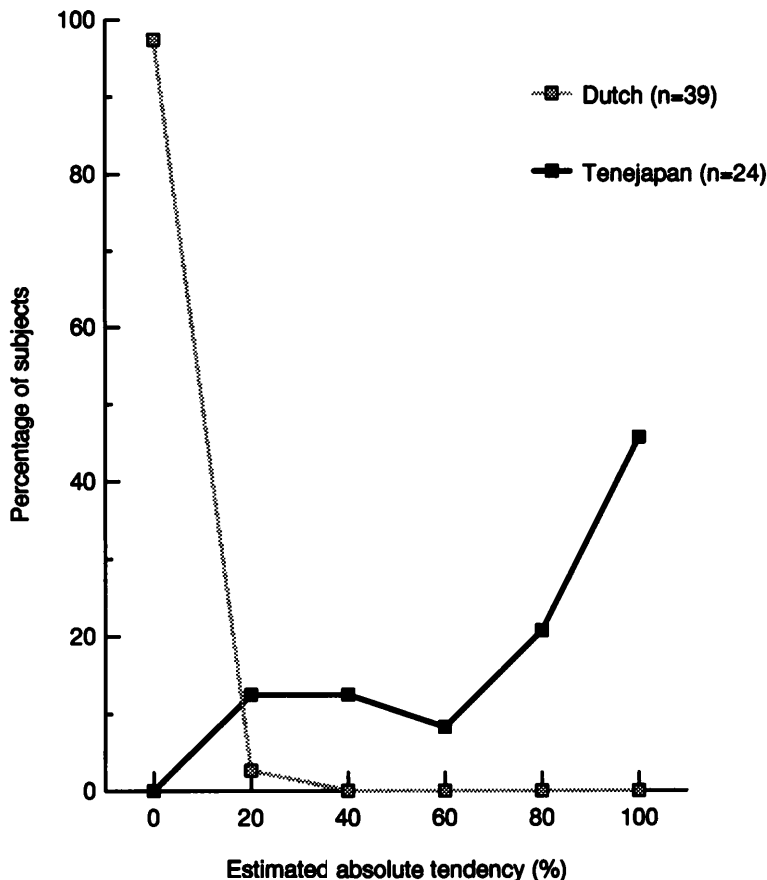
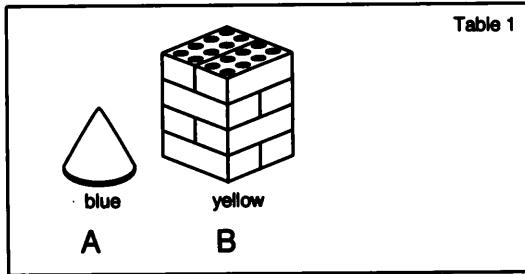
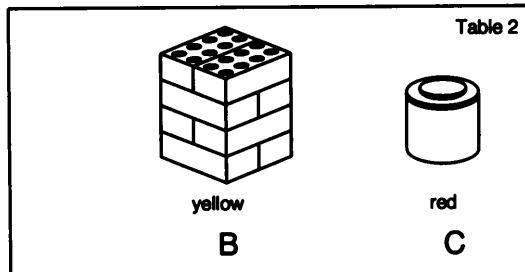


Figure 4.5
Chips recognition task.

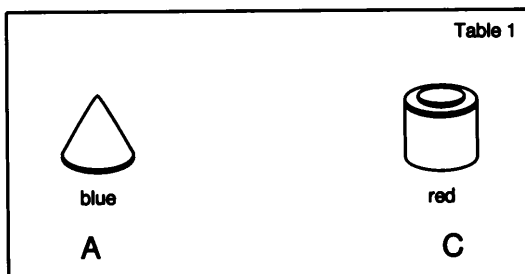
Design Subjects see the first nonverbal “premise” on table 1, for example, a blue cone A and a yellow cube B arranged in a predetermined orientation. The top diagram in figure 4.6 illustrates one such array from the perspective of the viewer. Then subjects are rotated and see the second “premise,” a red cylinder C and the yellow cube B in a predetermined orientation on table 2 (the array appearing from an egocentric point of view as, for example, in the second diagram in figure 4.6). Finally, subjects are rotated again and led back to table 1, where they are given just the blue cone A and asked to place the red cylinder C in a location consistent with the previous nonverbal “premises.” For example, if a female subject, say, sees (“premise 1”)



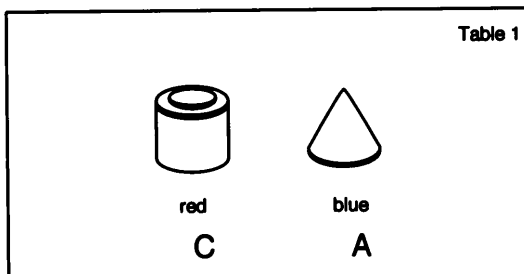
First 'premise'



Second 'premise'



Relative Solution



Absolute Solution

Figure 4.6
Transitive inference—the visual arrays.

the yellow cube to the right of the blue cone, then ("premise 2") the red cylinder to the right of the yellow cube, when given the blue cone, she may be expected to place the red cylinder C to the right of the blue cone A. It should be self-evident from the top two diagrams in figure 4.6, representing the arrays seen sequentially, why the third array (labeled the "relative solution") is one natural nonverbal "conclusion" from the first two visual arrays.

However, this result can only be expected if the subject codes the arrays in terms of egocentric or relative coordinates which rotate with her. If instead the subject utilizes fixed bearings or absolute coordinates, we can expect a different "conclusion"—in fact the reverse arrangement, with the red cylinder to the left of the blue cone (see the last diagram labeled "absolute solution" in figure 4.6)! To see why this is the case, consider figure 4.7, which gives a bird's-eye view of the experimental situation. If the subject does not use bodily coordinates that rotate with her, the blue cone will be, say, south of the yellow cube on table 1, and the red cylinder farther south of the yellow cube on table 2; thus the conclusion must be that the red cylinder is south of the blue cone. As the diagram makes clear, this amounts to the reverse arrangement from that produced under a coding using relative coordinates. In this case, and in half the trials, the absolute inference is somewhat more complex than a simple transitive inference (involving notions of relative distance), but in the other half of the trials the relative solution was more complex than the absolute one in just the same way.

Method Three objects distinct in shape and color were employed. Training was conducted on table 1, where it was made clear that the positions of each object relative to the other object—rather than exact locations on a particular table—was the relevant thing to remember. When transitive inferences were achieved on table 1, subjects were introduced to the rotation between the first and second premises; no correction was given unless the placement of the conclusion was on the orthogonal axis to the stimulus arrays. There were then ten trials, randomized across the transverse and sagittal axes (i.e., the arrays were either in a line across or along the line of vision).

Results The results are given in the graph in figure 4.8. Essentially, we have the same pattern of results as in the prior memory experiments: Dutch subjects are consistently relative coders, and Tenejapan subjects strongly tend to absolute coding, but more inconsistently. Of the Tenejapans who produced consistent results on at least 7 out of 10 trials, 90% were absolute coders (just two out of 25 subjects being relative coders). The reasons for the greater inconsistency of Tenejapan performance are presumably the same as in the previous experiment: unfamiliarity with any such procedure or test situation and the possible effects of the weak Absolute axis (the east-west axis lacking

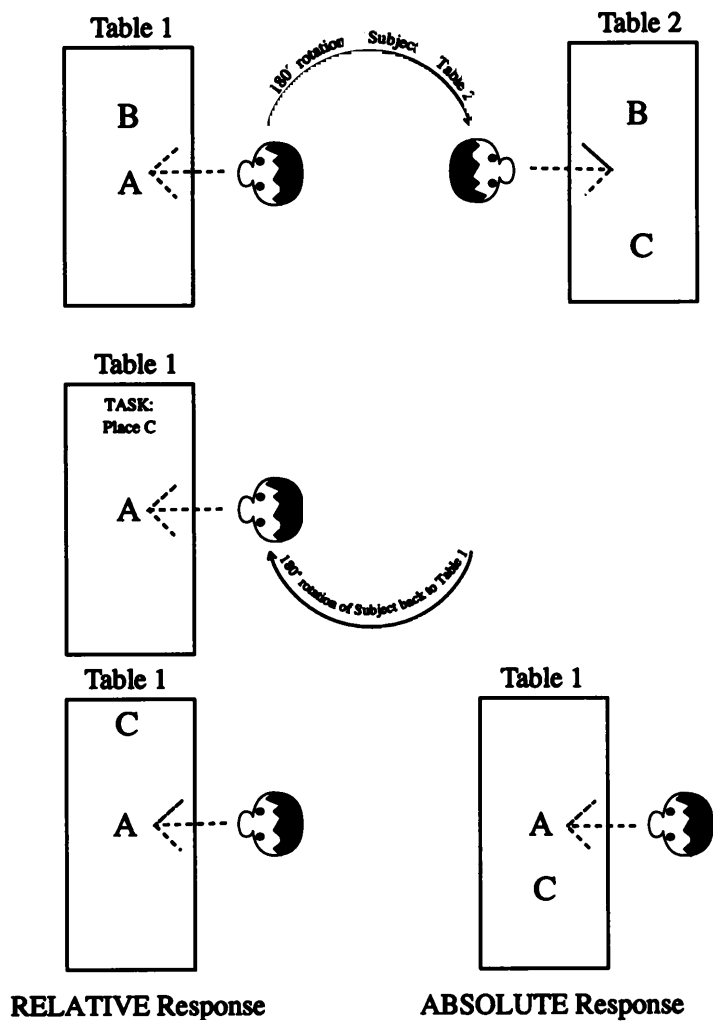


Figure 4.7
Transitive inference—bird's-eye view of experimental situation.

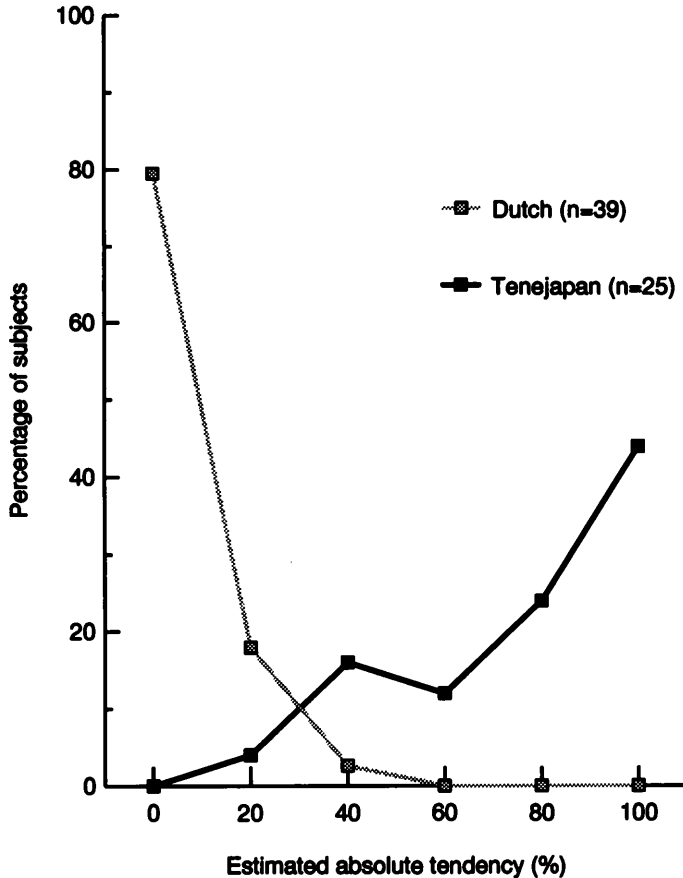


Figure 4.8
Transitive inference task

distinct linguistic labels for the half lines). Once again, Tenejapans made most errors or performed most inconsistently, on the east-west axis.

Discussion The results from these three experiments, together with others unreported here (see Brown and Levinson 1993b), all tend in the same direction. While Dutch subjects utilize a relative conceptual coding (presumably in terms of notions like *left*, *right*, *in front*, *behind*) to solve these nonverbal tasks, Tenejapan subjects predominantly use an absolute coding system. This is of course in line with the coding built into the semantics of spatial description in the two languages. The same pattern holds across different psychological faculties: the ability to recall spatial arrays, to

recognize those one has seen before, and to make inferences from spatial arrays. Further experiments of different kinds, exploring recall over different arrays and inferences of different kinds, all seem to show that this is a robust pattern of results.

The relative inconsistency of Tenejapan performance might simply be due to unfamiliar materials and procedures in this largely illiterate, peasant community. But as suggested above, errors or inconsistencies accumulated on one absolute axis in particular. However, because the experiments were all run on one set of fixed bearings, the error pattern could have been due equally to a strong versus weak egocentric axis (and in fact it is known that the left-right axis—here coinciding with the east-west axis—is less robust conceptually than the front-back axis). Therefore half the subjects were recalled and the experiments rerun on the orthogonal absolute bearings. The results showed unequivocally that errors and inconsistencies do indeed accumulate on the east-west absolute axis (although there also appears to be some interference from egocentric axes). This is interesting because it shows that Tenejapan subjects are not simply using an ad hoc system of local landmarks, or some fixed-bearing system totally independent of the language; rather, the conceptual primitives used to code the nonverbal arrays seem to inherit the particular properties of the semantics of the relevant linguistic distinctions.

This raises the skeptical thought that perhaps subjects are simply using linguistic mnemonics to solve the nonverbal tasks. However, an effective delay of at least three-quarters of a minute between losing sight of the stimulus and responding on table 2 would have required constant subvocal rehearsal for the mnemonic to remain available in short-term memory. Moreover, there is no particular reason why subjects should converge on a linguistic rather than a nonlinguistic mnemonic (like crossing the fingers on the relevant hand, or using a kinesthetic memory of a gesture—which would yield uniform relative results). But above all, two other experimental results suggest the inadequacy of an account in terms of a conscious strategy of direct linguistic coding.

4.2.2.2 Visual Recall and Gesture The first of these further experiments concerns the recall of complex arrays. Subjects saw an array of between two and five objects on table 1, and had to rebuild the array under rotation on table 2. Up to five of these objects had complex asymmetries, for example, a model of a chair, a truck, a tree, a horse leaning to one side, or a shoe. The majority of Tenejapan subjects rebuilt the arrays preserving the absolute bearings of the axes of the objects. This amounts to mental rotation of the visual array (or of the viewer) on table 1 so that it is reconstructed on table 2 as it would look like from the other side. Tenejapans prove to be exceptionally good at this, preserving the metric distances and precise angles between objects. It is far from clear that this could be achieved even in principle by a linguistic

coding: the precise angular orientation of each object and the metric distances between objects must surely be coded visually and must be rebuilt under visual control of the hands. This ability argues for a complex interaction between visual memory and a conceptual coding in terms of fixed bearings: an array that is visually distinct may be conceptually identical, and an array visually identical may be conceptually distinct (unlike with a system of relative coding, where what is to the left side of the visual field can be described as to the left). Thus being able to “see” that an array is conceptually identical to another in absolute terms may routinely involve mental rotation of the visual image. That a particular conceptual or linguistic system may exercise and thus enhance abilities of mental rotation has already been demonstrated for American Sign Language (ASL) by Emmorey (chapter 5, this volume). Tenejapans appear to be able to memorize a visual image of an array tagged, as it were, with the relevant fixed bearings.

There is another line of evidence that suggests that the Tenejapan absolute coding of spatial arrays is not achieved by conscious, artificial use of linguistic mnemonics. To show this, one would wish for some repetitive, unconscious nonverbal spatial behavior that can be inspected for the underlying frame of reference that drives it. There is indeed just such a form of behavior, namely, unreflective spontaneous gesture accompanying speech. Natural Tenejapan conversation can be inspected to see whether, when places or directions are referred to, gestures preserve the egocentric coordinates appropriate to the protagonist whose actions are being described, or whether the fixed bearings of those locations are preserved in the gestures. Preliminary work by Penelope Brown shows that such fixed bearings are indeed preserved in spontaneous Tenejapan gesture.⁵ A pilot experiment seems to confirm this. In the experiment, a male subject, say, facing north, sees a cartoon on a small portable monitor with lateral action from east to west. The subject is then moved to another room where he retells the story as best he can to another native speaker who has not seen the cartoon. In one condition, the subject retells the story facing north; in another condition the subject retells the story facing south. Preliminary results show that at least some subjects under rotation systematically preserve the fixed bearing of the observed action (from east to west) in their gestures, rather than the direction coded in terms of left or right. (Incidentally, the reverse finding has been established for American English by McCullough 1993). Because subjects had no idea that the experimenter was interested in gesture, we can be sure that the gestures record unreflective conceptualization of the directions. Although the gestures of course accompany speech, gestures preserving the fixed bearings of the stimulus often occur without explicit mention of the cardinal directions, suggesting that the gestures reflect an underlying spatial model, at least partially independent of language.

4.2.3 Conclusion from the Tenejapan Studies

Putting all these results together, we are led to the conclusion that the frame of reference dominant in the language, whether relative or absolute, comes to bias the choice of frame of reference in various kinds of nonlinguistic conceptual representations. This correlation holds across a number of "modalities" or distinct mental representations: over codings for recall and recognition memory, over representations for spatial inference, over recall apparently involving manipulations of visual images, and over whatever kind of kinesthetic representation system drives gesture. These findings look robust and general; similar observations have previously been made for an Aboriginal Australian community that uses absolute linguistic spatial description (Haviland 1993; Levinson 1992b), and a cross-cultural survey over a dozen non-Western communities shows a strong correlation of the dominant frame of reference in the linguistic system and frames of reference utilized in nonlinguistic tasks (see Baayen and Danziger 1994).

4.3 Frames of Reference across Modalities

Thus far, we have seen that (1) not all languages use the same predominant frame of reference and (2) there is a tendency for the frame of reference predominant in a particular language to remain the predominant frame of reference across modalities, as displayed by its use in nonverbal tasks of various kinds, unconscious gesture, and so on. The results seem firm; they appear to be replicable across speech communities, but the more one thinks about the implications of these findings, the more peculiar they seem to be. First, the trend of current theory hardly prepares us for such Whorfian results: the general assumption is rather of a universal set of semantic primes (conceptual primitives involved in language), on the one hand, and the identity or homomorphism of universal conceptual structure and semantic structure, on the other. Second, ideas about modularity of mind make it seem unlikely that such cross-modal effects could occur. Third, the very idea of the same frame of reference across different modalities, or different internal representation systems specialized to different sensory modalities, seems incoherent.

In order to make sense of the results, I shall in this section attempt to show that the notion "same frame of reference across modalities" is, after all, perfectly coherent, and indeed already adumbrated across the disciplines that study the various modalities. This requires a lightning review of the notion "frame of reference" across the relevant disciplines (section 4.3.1 and 4.3.2); it also requires a reformation of the linguistic distinctions normally made (section 4.3.3). With that under our belts, we can then face up to the peculiarity, from the point of view of ideas about the

modularity of mind, of this cross-modal adoption of the same frame of reference (section 4.4). Here some intrinsic properties of the different frames of reference may offer the decisive clue: if there is to be any cross-modal transfer of spatial information, we may have no choice but to fixate predominantly on just one frame of reference.

4.3.1 “Spatial Frames of Reference”

The notion of “frames of reference” is crucial to the study of spatial cognition across all the modalities and all the disciplines that study them. The idea is as old as the hills: medieval theories of space, for example, were deeply preoccupied by the puzzle raised by Aristotle, the case of the boat moored in the river. If we think about the location of an object as the place that it occupies, and the place as containing the object, then the puzzle is that if we adopt the river as frame of reference, the boat is moving, but if we adopt the bank as frame, then it is stationary (see Sorabji 1988, 187–201 for a discussion of this problem, which dominated medieval discussions of space).

But the phrase “frame of reference” and its modern interpretation originate, like so much else worthwhile, from Gestalt theories of perception in the 1920s. How, for example, do we account for illusions of motion, as when the moon skims across the clouds, except by invoking a notion of a constant perceptual window against which motion (or the perceived vertical, say) is to be judged? The Gestalt notion can be summarized as “a unit or organization of units that collectively serve to identify *a coordinate system* with respect to which certain properties of objects, including the phenomenal self, are gauged” (Rock 1992, 404; emphasis mine).⁶

In what follows, I will emphasize that distinctions between frames of reference are essentially distinctions between underlying coordinate systems and not, for example, between the objects that may invoke them. Not all will agree.⁷ In a recent review, philosophers Brewer and Pears (1993) ranging over the philosophical and psychological literature, conclude that frames of reference come down to the selection of reference objects. Take the glasses on my nose—when I go from one room to another, do they change their location or not? It depends on the “frame of reference”—nose or room.⁸ This emphasis on the ground or relatum or reference object⁹ severely underplays the importance of coordinate systems in distinguishing frames of reference, as I shall show below.¹⁰ Humans use multiple frames of reference: I can happily say of the same assemblage (ego looking at car from side, car’s front to ego’s left): “the ball is in front of the car” and “the ball is to the left of the car,” without thinking that the ball has changed its place. In fact, much of the psychological literature is concerned with ambiguities of this kind. I will therefore insist on the emphasis on coordinate systems rather than on the objects or “units” on which such coordinates may have their origin.

4.3.2 "Frames of Reference" across Modalities and the Disciplines that Study Them

If we are to make sense of the notion "same frame of reference" across different modalities, or inner representation systems, it will be essential to see how the various distinctions between the frames of reference proposed by different disciplines can be ultimately brought into line. This is no trivial undertaking, because there are a host of such distinctions, and each of them has been variously construed, both within and across the many disciplines (such as philosophy, the brain sciences, psychology, and linguistics) that explicitly employ the notion "frames of reference." A serious review of these different conceptions would take us very far afield. On the other hand, some sketch is essential, and I will briefly survey the various distinctions in table 4.1, with some different construals distinguished by the letters *a*, *b*, *c*.¹¹

First, then, "relative" versus "absolute" space. Newton's distinction between absolute and relative space has played an important role in ideas about frames of refer-

Table 4.1

Spatial Frames of Reference: Some Distinctions in the Literature

"Relative" versus "absolute":

(*philosophy, brain sciences, linguistics*)

- a. Space as relations between objects versus abstract void
- b. Egocentric versus allocentric
- c. Directions: Relations between objects versus fixed bearings

"Egocentric" versus "allocentric"

(*developmental and behavioral psychology, brain sciences*)

- a. Body-centered versus environment-centered (Note many ego centers: retina, shoulder, etc.)
- b. Subjective (subject-centered) versus objective

"Viewer-centered" versus "object-centered" or "2½-D sketch" versus "3-D models"

(*vision theory, imagery debate in psychology*)

"Orientation-bound" versus "orientation-free"

(*visual perception, imagery debate in psychology*)

"Deictic" versus "intrinsic"

(*linguistics*)

- a. Speaker-centric versus non-speaker-centric
- b. Centered on speaker or addressee versus thing
- c. Ternary versus binary spatial relations

"Viewer-centered" versus "object-centered" versus "environment-centered"

(*psycholinguistics*)

- = "gaze tour" versus "body tour" perspectives
- = ?"survey perspective" versus "route perspective"

ence, in part through the celebrated correspondence between his champion Clarke and Leibniz, who held a strictly relative view.¹² For Newton, absolute space is an abstract, infinite, immovable, three-dimensional box with origin at the center of the universe, while relative space is conceived of as specified by relations between objects. Psychologically, Newton claimed, we are inclined to relative notions: “Relative space is some moveable dimension or measure of the absolute spaces, which our senses determine by its position to bodies . . . and so instead of absolute places and motions, we use relative ones” (quoted in Jammer 1954, 97–98). Despite fundamental differences in philosophical position, most succeeding thinkers in philosophy and psychology have assumed the psychological primacy of relative space—space anchored to the places occupied by physical objects and their relations to one another—in our mental life. A notable exception is Kant, who came to believe that notions of absolute space are a fundamental intuition, although grounded in our physical experience, that is, in the use of our body to define the egocentric coordinates through which we deal with space (Kant 1768; see also Van Cleve and Frederick 1991). O’Keefe and Nadel (1978; see also O’Keefe 1993 and chapter 7, this volume) have tried to preserve this Kantian view as essential to the understanding of the neural implementation of our spatial capacities, but by and large psychologists have considered notions of “absolute” space irrelevant to theories of the naive spatial reasoning underlying language (see Clark 1973; Miller and Johnson-Laird 1976, 380). (Absolute notions of space may, however, be related to cognitive maps of the environment—discussed under the rubric of “allocentric” frames of reference below.)

Early on, the distinction between relative and absolute space acquired certain additional associations; for example, relative space became associated with egocentric coordinate systems, and absolute space with non-egocentric ones (despite Kant 1768),¹³ so that this distinction is often confused with the egocentric versus allocentric distinction (discussed below). Another interpretation of the relative versus absolute distinction, in relating relativistic space to egocentric space, goes on to emphasize the different ways coordinate systems are constructed in relative versus absolute spatial conceptions: “Ordinary languages are designed to deal with relativistic space; with space relative to the objects that occupy it. Relativistic space provides three orthogonal coordinates, just as Newtonian space does, but *no fixed units of angle or distance are involved, nor is there any need for coordinates to extend without limit in any direction*” (Miller and Johnson-Laird 1976, 380; emphasis mine). Thus a system of fixed bearings, or cardinal directions, is opposed to the relativistic “space concept,” whether egocentric or object-centered, which Miller and Johnson-Laird (1976, 395) and many other authors, like Clark (1973), Herskovits (1986) and Svorou (1994, 213), have assumed to constitute the conceptual core of human spatial thinking. But because, as we have seen, some languages use as a conceptual basis coordi-

nate systems with fixed angles (and coordinates of indefinite extent), we need to recognize that these systems may be appropriately called "absolute" coordinate systems. Hence I have opposed relative and absolute frames of reference in language (see section 4.3.3).

Let us turn to the next distinction in table 4.1, namely, "egocentric" versus "allocentric." The distinction is of course between coordinate systems with origins within the subjective body frame of the organism, versus coordinate systems centered elsewhere (often unspecified). The distinction is often invoked in the brain sciences, where there is a large literature concerning frames of reference (see, for example, the compendium in Paillard 1991). This emphasizes the plethora of different egocentric coordinate systems required to drive all the different motor systems from saccades to arm movements (see, for example, Stein 1992), or the control of the head as a platform for our inertial guidance and visual systems (again see papers in Paillard 1991). In addition, there is a general acceptance (Paillard 1991, 471) of the need for a distinction (following Tolman 1948; O'Keefe and Nadel 1978) between egocentric and allocentric systems. O'Keefe and Nadel's demonstration that something like Tolman's mental maps are to be found in the hippocampal cells is well known.¹⁴ O'Keefe's recent (1993) work is an attempt to relate a particular mapping system to the neuronal structures and processes. The claim is that the rat can use egocentric measurements of distance and direction toward a set of landmarks to compute a non-egocentric abstract central origo (the "centroid") and a fixed angle or "slope." Then it can keep track of its position in terms of distance from centroid and direction from slope. This is a "mental map" constructed through the rat's exploration of the environment, which gives it fixed bearings (the slope), but just for this environment. Whether this strictly meets the criteria for an objective, "absolute," allocentric system has been questioned (Campbell 1993, 76–82).¹⁵ We certainly need to be able to distinguish mental maps of different sorts: egocentric "strip maps" (Tolman 1948), allocentric landmark-based maps with relative angles and distances between landmarks (more Leibnizian), and allocentric maps based on fixed bearings (more Newtonian).¹⁶ But in any case, this is the sort of thing neurophysiologists have in mind when they oppose "egocentric" and "allocentric" frames of reference.¹⁷

Another area of work where the opposition has been used is in the study of human conceptual development. For example, Acredolo (1988) shows that, as Piaget argued, infants have indeed only egocentric frames of reference in which to record spatial memories; but contrary to Piaget (Piaget and Inhelder 1956), this phase lasts only for perhaps the first six months. Thereafter, they acquire the ability to compensate for their own rotation, so that by sixteen months they can identify, say, a window in one wall as the relevant stimulus even when entering the room (with two identical windows) from the other side. This can be thought of as the acquisition of a

non-egocentric, “absolute” or “geographic” orientation or frame of reference.¹⁸ Pick (1993, 35) points out, however, that such apparently allocentric behavior can be mimicked by egocentric mental operations, and indeed this is suggested by Acredolo’s (1988, 165) observation that children learn to do such tasks by adopting the visual strategy “if you want to find it, keep your eyes on it (as you move).”

These lines of work identify the egocentric versus allocentric distinction with the opposition between body-centered and environment-centered frames of reference. But as philosophers point out (see, for example, Campbell 1993), ego is not just any old body, and there is indeed another way to construe the distinction as one between subjective and objective frames of reference. The egocentric frame of reference would then bind together various body-centered coordinate systems with an agentive subjective being, complete with body schema, distinct zones of spatial interaction (reach, peripheral vs. central vision, etc.). For example, phenomena like “phantom limbs” or proprioceptive illusions argue for the essentially subjective nature of egocentric coordinate systems.

The next distinction on our list, “viewer-centered” versus “object-centered,” comes from the theory of vision, as reconstructed by Marr (1982). In Marr’s well-known conceptualization, a theory of vision should take us from retinal image to visual object recognition, and that, he claimed, entails a transfer from a viewer-centered frame of reference, with incremental processing up to what he called the “ $2\frac{1}{2}$ -D sketch,” to an object-centered frame of reference, a true 3-D model or structural description.¹⁹ Because we can recognize an object even when foreshortened or viewed in differing lighting conditions, we must extract some abstract representation of it in terms of its volumetric properties to match this token to our mental inventory of such types. Although recent developments have challenged the role of the 3-D model within a modular theory of vision,²⁰ there can be little doubt that at some conceptual level such an object-centered frame of reference exists. This is further demonstrated by work on visual imagery, which seems to show that, presented with a viewer-centered perspective view of a novel object, we can mentally rotate it to obtain different perspectival “views” of it, for example, to compare it to a prototype (Shepard and Metzler 1971; Kosslyn 1980; Tye 1991, 83–86). Thus at some level, the visual or ancillary systems seem to employ two distinct reference frames, viewer-centered and object-centered.

This distinction between viewer-centered and object-centered frames of reference relates rather clearly to the linguistic distinction between deictic and intrinsic perspectives discussed below. The deictic perspective is viewer-centered, whereas the intrinsic perspective seems to use (at least in part) the same axial extraction that would be needed to compute the volumetric properties of objects for visual recognition (see Landau and Jackendoff 1993; Jackendoff, chapter 1, this volume; Landau, chapter 8,

this volume; Levinson 1994). This parallel will be further reinforced by the reformation of the linguistic distinctions suggested in section 4.3.3.

This brings us to the "orientation-bound" versus "orientation-free" frames of reference.²¹ The visual imagery and mental rotation literature might be thought to have little to say about frames of reference. After all, visual imagery would seem to be necessarily at most $2\frac{1}{2}$ -D and thus necessarily in a viewer-centered frame of reference (even if mental rotations indicate access to a 3-D description). But recently there have been attempts to understand the relation between two kinds of shape recognition: one where shapes are recognized without regard to orientation (thus with no response curve latency associated with degrees of orientation from a familiar related stimulus), and another where shapes are recognized by apparent analog rotation to the familiar related stimulus. The Shepard and Metzler (1971) paradigm suggested that only where handedness information is present (as where enantiomorphs have to be discriminated) would mental rotation be involved, which implicitly amounts to some distinction between object-centered and viewer-centered frames of reference; that is discrimination of enantiomorphs depends on an orientation-bound perspective, while the recognition of simpler shapes may be orientation-free.²² But some recent controversies seem to show that things are not as simple as this (Tarr and Pinker 1989; Cohen and Kubovy 1993). Just and Carpenter (1985) argue that rotation tasks in fact can be solved using four different strategies, some orientation-bound and some orientation-free.²³ Similarly, Takano (1989) suggests that there are four types of spatial information involved, classifiable by crossing *elementary* (simple) versus *conjunctive* (partitionable) forms with the distinction between *orientation-bound* and *orientation-free*. He insists that only orientation-bound forms should require mental rotation for recognition. However, Cohen and Kubovy (1993) claim that such a view makes the wrong predictions because handedness identification can be achieved without the mental rotation latency curves in special cases. In fact, I believe that despite these recent controversies, the original assumption—that only objects lacking handedness can be recognized without mental rotation—must be basically correct for logical reasons that have been clear for centuries.²⁴ In any case, it is clear from this literature that the study of visual recognition and mental rotation utilizes distinctions in frames of reference that can be put into correspondence with those that emerge from, for example, the study of language. Absolute and relative frames of reference in language (to be firmed up below) are both orientation-bound, while the intrinsic frame is orientation-free (Danziger 1994).

Linguists have long distinguished "deictic" versus "intrinsic" frames of reference, because of the rather obvious ambiguities of a sentence like "the boy is in front of the house" (see, for example, Leech 1969, 168; Fillmore 1971; Clark 1973). It has also been known for a while that linguistic acquisition of these two readings of terms like

in front, behind, to the side of is in the reverse direction from the developmental sequence egocentric to allocentric (Pick 1993): *intrinsic* notions come resolutely earlier than *deictic* ones (Johnston and Slobin 1978). Sometimes a third term, *extrinsic*, is opposed, to denote, for example, the contribution of gravity to the interpretation of words like *above* or *on*. But unfortunately the term *deictic* breeds confusions. In fact there have been at least three distinct interpretations of the deictic versus intrinsic contrast, as listed in table 4.1: (1) speaker-centric versus non-speaker-centric (Levelt 1989); (2) centered on any of the speech participants versus not so centered (Levinson 1983); (3) ternary versus binary spatial relations (implicit in Levelt 1984 and chapter 3, this volume; to be adopted here). These issues will be taken up in section 4.3.3, where we will ask what distinctions in frames of reference are grammaticalized or lexicalized in different languages.

Let us turn now to the various distinctions suggested in the psychology of language. Miller and Johnson-Laird (1976), drawing on earlier linguistic work, explored the opposition between deictic and intrinsic interpretations of such utterances as “the cat is in front of the truck”; the logical properties of these two frames of reference, and their interaction, have been further clarified by Levelt (1984, 1989, and chapter 3, this volume). Carlson-Radvansky and Irwin (1993, 224) summarize the general assumption in psycholinguistics as follows:

Three distinct classes of reference frames exist for representing the spatial relationships among objects in the world ... *viewer-centered frames*, *object-centered frames*, and *environment centered frames* of reference. In a viewer-centered frame, objects are represented in a retinocentric, head-centric or body-centric coordinate system based on the perceiver’s perspective of the world. In an object-centered frame, objects are coded with respect to their intrinsic axes. In an environment-centered frame, objects are represented with respect to salient features of the environment, such as gravity or prominent visual landmarks. In order to talk about space, vertical and horizontal coordinate axes must be oriented with respect to one of these reference frames so that linguistic spatial terms such as “above” and “to the left of” can be assigned. (Emphasis added)

Notice that in this formulation frames of reference inhere in spatial perception and cognition rather than in language: *above* may simply be semantically general over the different frames of reference, not ambiguous (Carlson-Radvansky and Irwin (1993, 242).²⁵ Thus deictic, intrinsic, and extrinsic are merely alternative labels for the linguistic interpretations corresponding, respectively, to viewer-centered, object-centered, and environment-centered frames of reference.

There are other oppositions that psycholinguists employ, although in most cases they map onto the same triadic distinction. One particular set of distinctions, between different kinds of survey or route description, is worth unraveling because it has caused confusion. Levelt (1989, 139–144) points out that when a subject describes a

complex visual pattern, the linearization of speech requires that we “chunk” the pattern into units that can be described in a linear sequence. Typically, we seem to represent 2-D or 3-D configurations through a small window, as it were, traversing the array; that is, the description of complex static arrays is converted into a description of motion through units or “chunks” of the array. Levelt (chapter 3, this volume) has examined the description of 2-D arrays, and found two strategies (1): a *gaze tour* perspective, effectively the adoption of a fixed deictic or viewer-centered perspective; and (2) a body or *driving tour*, effectively an intrinsic perspective, where a pathway is found through the array, and the direction of the path used to assign *front*, *left*, and so on from any one point (or location of the window in describing time). Because both perspectives can be thought of as egocentric, Tversky (1991; see also Taylor and Tversky in press and Tversky, chapter 12, this volume) opts to call Levelt's intrinsic perspective a “*deictic* frame of reference” or “*route* description” and his deictic perspective a “*survey* perspective.”²⁶ Thus Tversky's “deictic” is Levelt's “intrinsic” or nondeictic perspective! This confusion is, I believe, not merely terminological but results from the failure in the literature to distinguish coordinate systems from their origins or centers (see section 4.3.3).

Finally, in psycholinguistic discussions about frames of reference, there seems to be some unclarity, or sometimes overt disagreement, at *which level*—perceptual, conceptual or linguistic—such frames of reference apply. Thus Carlson-Radvansky and Irwin (1993, 224) make the assumption that a frame of reference must be adopted within some spatial representation system, as a precondition for coordinating perception and language, whereas Levelt (1989; but see Levelt, chapter 3, this volume) has argued that a frame of reference is freely chosen in the very process of mapping from perception or spatial representation to language (see also Logan and Sadler, chapter 13, this volume). On the latter conception, frames of reference in language are peculiar to the nature of the linear, propositional representation system that underlies linguistic semantics, that is, they are different ways of conceiving the same percept in order to talk about it.²⁷

The view that frames of reference in linguistic descriptions are adopted in the mapping from spatial representation or perception to language seems to suggest that the perceptions or spatial representations themselves make no use of frames of reference. But this of course is not the case: there has to be some coordinate system involved in any spatial representation of any intricacy, whether at a peripheral (sensory) level or at a central (conceptual) level. What Levelt's results (chapter 3, this volume) or Friederici and Levelt's (1990) seem to establish, is that frames of reference at the perceptual or spatial conceptual level do not necessarily determine frames of reference at the linguistic level. This is exactly what one might expect. Language is flexible and it is an instrument of communication—thus it naturally allows us, for

example, to take the other person's perspective. Further, the ability to cast a description in one frame or another implies an underlying conceptual ability to handle multiple frames, and within strict limits (see below) to convert between them. In any case, we need to distinguish in discussions of frames of reference between at least three levels: (1) perceptual, (2) conceptual, and (3) linguistic; and we need to consider the possibility that we may utilize distinct frames of reference at each level (but see section 4.4).

There is much further pertinent literature in all the branches of psychology and brain science, but we must leave off here. It should already be clear that there are many, confusingly different classifications, and different construals of the same terms, not to mention many unclarities and many deep confusions in the thinking behind them. Nevertheless, there are some obvious common bases to the distinctions we have reviewed. It is clear for example, that on the appropriate construals, "egocentric" corresponds to "viewer-centered" and "2½-D sketch" to "deictic" frame, while "intrinsic" maps onto "object-centered" or "3-D model" frames of reference; "absolute" is related to "environment-centered"; and so forth. We should seize on these commonalities, especially because in this chapter we are concerned with making sense of the "same frame of reference" across modalities and representational systems. However, before proposing an alignment of these distinctions across the board, it is essential to deal with linguistic frames of reference, whose troubling flexibility has led to various confusions.

4.3.3 Linguistic Frames of Reference in Crosslinguistic Perspective

Cursory inspection of the linguistic literature will give the impression that the linguists have their house in order. They talk happily of topological vs. projective spatial relators (e.g., prepositions like *in* vs. *behind*), deictic versus intrinsic usages of projective prepositions, and so on (see, for example, Bierwisch 1967; Lyons 1977; Herskovits 1986; Vandeloise 1991; and psycholinguists Clark 1973; Miller and Johnson-Laird 1976). But the truth is less comforting. The analysis of spatial terms in familiar European languages remains deeply confused,²⁸ and those in other languages almost entirely unexplored. Thus the various alleged universals should be taken with a great pinch of salt (in fact, many of them can be directly jettisoned). One major upset is the recent finding that many languages use an "absolute" frame of reference (as illustrated in the case of Tzeltal) where European languages would use a "relative" or viewpoint-centered one (see, for example, Levinson 1992a,b; Haviland 1993). Another is that some languages, like many Australian ones, use such frames of reference to replace so-called topological notions like *in*, *on*, or *under*. A third is that familiar spatial notions like *left* and *right* and even sometimes *front* and *back* are missing from many, perhaps a third of all languages. Confident predictions

and assumptions can be found in the literature that no such languages could occur (see, for example, Clark 1973; Miller and Johnson-Laird 1976; Lyons 1977, 690).

These developments call for some preliminary typology of the frames of reference that are systematically distinguished in the grammar or lexicon of different languages (with the caveat that we still know little about only a few of them). In particular, we shall focus on what we seem to need in the way of coordinate systems and associated reference points to set up a crosslinguistic typology of the relevant frames of reference. In what follows I shall confine myself to linguistic descriptions of static arrays, and I shall exclude the so-called topological notions, for which a new partial typology concerning the coding of concepts related to *in* and *on* is available (Bowerman and Pederson in prep.).²⁹ Moreover, I shall focus on distinctions on the horizontal plane. This is not whimsy: the perceptual cues for the vertical may not always coincide, but they overwhelmingly converge, giving us a good universal solution to one axis. But the two horizontal coordinates are up for grabs: there simply is no corresponding force like gravity on the horizontal.³⁰ Consequently there is no simple solution to the description of horizontal spatial patterns, and languages diverge widely in their solutions to the basic problem of how to specify angles or directions on the horizontal.

Essentially, three main frames of reference emerge from these new findings as solutions to the problem of description of horizontal spatial oppositions. They are appropriately named "intrinsic," "relative" and "absolute," even though these terms may have a somewhat different interpretation from some of the construals reviewed in the section above. Indeed, the linguistic frames of reference potentially *crosscut* many of the distinctions in the philosophical, neurophysiological, linguistic, and psychological literatures, for one very good reason. Linguistic frames of reference cannot be defined with respect to the origin of the coordinate system (in contrast to, for example, egocentric vs. allocentric). It will follow that the traditional distinction deictic versus intrinsic collapses—these are not opposed terms. All this requires some explanation.

We may start by noting the difficulties we get into by trying to make the distinction between deictic and intrinsic. Levelt (1989, 48–55) organizes and summarizes the standard assumptions in a useful way: we can cross-classify linguistic uses according to (a) whether they presume that the coordinates are centered on the speaker (*deictic*) or not (*intrinsic*); and (b) whether the *relatum* (ground) is the speaker or not. Suppose then we call the usage "deictic" just in case the coordinates are centered on the speaker, "intrinsic" otherwise. This yields, for example, the following classification of examples:

- (1) The ball is in front of me.

Coordinates: **Deictic** (i.e., origin on speaker)

Relatum: Speaker

- (2) The ball is in front of the tree.
Coordinates: **Deictic** (i.e., origin on speaker)
Relatum: Tree
- (3) The ball is in front of the chair (at the chair's front).
Coordinates: **Intrinsic** (i.e., origin not on speaker)
Relatum: Chair

Clearly, it is the locus of the origin of the coordinates that is relevant to the traditional opposition deictic versus intrinsic, otherwise we would group (2) and (3) as both sharing a nondeictic relatum. The problem comes when we pursue this classification further:

- (4) The ball is in front of you.
Coordinates: **Intrinsic** (origin on addressee, not speaker)
Relatum: Addressee
- (5) The ball is to the right of the lamp, from your point of view.
Coordinates: **Intrinsic** (origin on addressee)
Relatum: Lamp

Here the distinction deictic versus intrinsic is self-evidently not the right classification, as far as frames of reference are concerned. Clearly, (1) and (4) belong together: the interpretation of the expressions is the same, with the same coordinate systems; there are just different origins—speaker and addressee, respectively (moreover, in a normal construal of “deictic,” inclusive of first and second persons, both are “deictic” origins). Similarly, in another grouping, (2) and (5) should be classed together: they have the same conceptual structure, with a viewpoint (acting as the origin of the coordinate system), a relatum distinct from the viewpoint, and a referent—again the origin alternates over speaker or addressee.

We might therefore be tempted simply to alter the designations, and label (1), (2), (4), and (5) all “deictic” as opposed to (3) “intrinsic.” But this would produce a further confusion.

First, it would conflate the distinct conceptual structures of our groupings (1) and (4) versus (2) and (5). Second, the conceptual structure of the coordinate systems in (1) and (4) is in fact shared with (3). “The ball is in front of the chair” presumes (on the relevant reading) an intrinsic front and uses that facet to define a search domain for the ball; but just the same holds for “the ball is in front of me/you.”³¹ Thus the logical structure of (1), (3), and (4) is the same: the notion “in front of” is here a *binary* spatial relation, with arguments constituted by the figure (referent) and the ground (relatum), where the projected angle is found by reference to an intrinsic or inherent facet of the ground object. In contrast, (2) and (5) have a different logical

structure: "in front of" is here a ternary relation, presuming a viewpoint V (the origin of the coordinate system), a figure, and ground, all distinct.³² In fact, these two kinds of spatial relation have quite different logical properties, as demonstrated elsewhere by Levelt (1984, and chapter 3, this volume), but only when distinguished and grouped in this way. Let us dub the binary relations "intrinsic," but the ternary relations "relative" (because the descriptions are always relative to a viewpoint, in contradistinction to "absolute" and "intrinsic" descriptions).

To summarize then, the proposed classification is

- (1') The ball is in front of me
Coordinates: **Intrinsic**
Origin: Speaker
Relatum: Speaker
- (3') The ball is in front of the chair (at the chair's front)
Coordinates: **Intrinsic**
Origin: Chair
Relatum: Chair
- (4') The ball is in front of you
Coordinates: **Intrinsic**
Origin: Addressee
Relatum: Addressee
- (2') The ball is in front of the tree
Coordinates: **Relative**
Origin: Speaker
Relatum: Tree
- (5') The ball is to the right of the lamp, from your point of view
Coordinates: **Relative**
Origin: Addressee
Relatum: Lamp
- (6') John noticed the ball to the right of the lamp
For John, the ball is in front of the tree.
Coordinates: **Relative**
Origin: Third person (John)
Relatum: Lamp (or Tree)

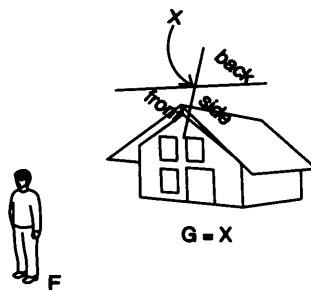
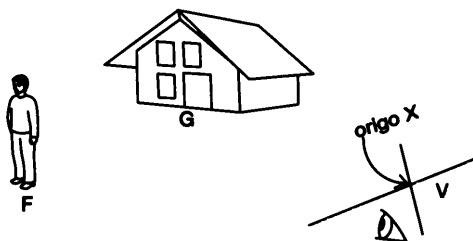
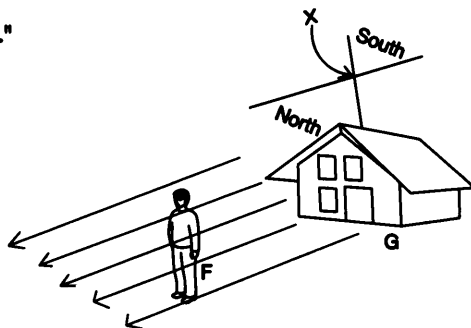
Note that use of the intrinsic system of coordinates entails that relatum (ground) and origin are constituted by the same object (the spatial relation is binary, between F and G), while use of the relative system entails that they are distinct (the relation is

ternary, between F , G , and viewpoint V). Note, too, that whether the center is deictic, that is, whether the origin is speaker (or addressee), is simply irrelevant to this classification. This is obvious in the case of the grouping of (1'), (3'), and (4') together. It is also clear that although the viewpoint in relative uses is normally speaker-centric, it may easily be addressee-centric or even centered on a third party as illustrated in (6'). Hence deictic and intrinsic are *not* opposed; instead, we need to oppose coordinate systems as intrinsic versus relative, on the one hand, and origins as deictic and non-deictic (or, alternatively, egocentric vs. allocentric), on the other. Because frames of reference are coordinate systems, it follows that in language, frames of reference cannot be distinguished according to their characteristic, but variable, origins.

I expect a measure of resistance to this reformation of the distinctions, if only because the malapropism "deictic frame of reference" has become a well-worn phrase. How, the critic will argue, can you define the frames of reference if you no longer employ the feature of deicticity to distinguish them? I will expend considerable effort in that direction in section 4.3.3.2. But first we must compare these two systems with the third system of coordinates in natural language, namely, absolute frames of reference. Let us review them together.

4.3.3.1 The Three Linguistic Frames of Reference As far as we know, and according to a suitably catholic construal, there are exactly three frames of reference grammaticalized or lexicalized in language (often, lexemes are ambiguous over two of these frames of reference, sometimes expressions will combine two frames,³³ but often each frame will have distinct lexemes associated with it).³⁴ Each of these three frames of reference encompasses a whole family of related but distinct semantic systems.³⁵ It is probably true to say that even the most closely related languages (and even dialects within them) will differ in the details of the underlying coordinate systems and their geometry, the preferential interpretation of ambiguous lexemes, the presumptive origins of the coordinates, and so on. Thus the student of language can expect that expressions glossed as, say, intrinsic *side* in two languages will differ considerably in the way in which *side* is in fact determined, how wide and how distant a search domain it specifies, and so on. With that caveat, let us proceed.

Let us first define a set of primitives necessary for the description of all systems.³⁶ The application of some of the primitives is sketched in figure 4.9, which illustrates three canonical exemplars from each of our three main types of system. Minimally, we need the primitives in table 4.2, the use of which we will illustrate in passing. Combinations of these primitives yield a large family of systems which may be classified in the following tripartite scheme: (1) *intrinsic frame of reference*; (2) *relative frame of reference*; and (3) *absolute frame of reference*.

INTRINSIC**"He's in front of the house."****RELATIVE****"He's to the left of the house."****ABSOLUTE****"He's north of the house."****Figure 4.9**

Canonical examples of the three linguistic frames of reference.

Table 4.2
Inventory of Primitives

1. System of labeled angles

Labeled arcs are specified by coordinates around origin (language-specific); such labels may or may not form a fixed armature or template of oppositions.

2. Coordinates

a. Coordinates may be polar, by rotation from a fixed x -axis, or rectangular, by specification of two or more axes;

b. One primary coordinate system C can be mapped from origin X to secondary origin X_2 , by the following transformations:

- translation,
- rotation
- reflection
- (and possibly a combination)

to yield a secondary coordinate system C_2 .

3. Points

F = figure or referent with center point at volumetric center F_C .

G = ground or relatum, with volumetric center G_C , and with a surrounding region R

V = viewpoint

X = origin of the coordinate system, X_2 = secondary origin

A = anchor point, to fix labeled coordinates

L = designated landmark

4. Anchoring system

A = Anchor point, for example, with G or V ; in landmark systems $A = L$.

"Slope" = fixed-bearing system, yielding parallel lines across environment in each direction

Intrinsic Frame of Reference Informally, this frame of reference involves an object-centered coordinate system, where the coordinates are determined by the "inherent features," sidedness or facets of the object to be used as the ground or relatum. The phrase "inherent features," though widely used in the literature, is misleading: such "facets," as we shall call them, have to be conceptually assigned according to some algorithm, or learned on a case-by-case basis, or more often a combination of these. The procedure varies fundamentally across languages. In English, it is (apart from *top* and *bottom*, and special arrangements for humans and animals) largely functional (see, for example, the sketch in Miller and Johnson-Laird 1976, 403), so that the *front* of a TV is the side we attend to, while the *front* of a car is the facet that canonically lies in the direction of motion, and so forth. But in some languages, it is much more closely based on shape. For example, in Tzeltal the assignment of sides utilizes a volumetric analysis very similar to the object-centered analysis proposed by Marr

(1982) in the theory of vision, and function and canonical orientation is largely irrelevant (see Levinson 1994).³⁷ In many languages the morphology makes it clear that human or animal body (and occasionally plant) parts provide a prototype for the opposed sides: hence we talk about the "front," "backs," "sides," "lefts," and "rights" and in many languages "heads," "feet," "horns," "roots," etc.) of other objects.³⁸ But whatever the procedure in a particular language, it relies primarily on the conceptual properties of the object: its shape, canonical orientation, characteristic motion and use, and so on.

The attribution of such facets provides the basis for a coordinate system in one of two ways. Having found, for example, the *front*, this may be used to anchor a ready-made system of oppositions *front/back*, *sides*, and so forth.³⁹ Alternatively, in other languages, there may be no such fixed armature, as it were, each object having parts determined, for example, by specific shapes; in that case, finding *front* does not predict the locus of *back*, but nevertheless determines a direction from the volumetric center of the object through the *front*, which can be used for spatial description.⁴⁰ In either case, we can use the designated facet to extract an angle, or line, radiating out from the ground object, within or on which the figure object can be found (as in "the statue in front of the town hall").

The geometrical properties of such intrinsic coordinate systems vary crosslinguistically. Systems with fixed armatures of contrastive expressions generally require the angles projected to be mutually exclusive (nonoverlapping), so that in the intrinsic frame of reference (unlike the relative one) it makes no sense to say, "The cat is to the front and to the left of the truck." Systems utilizing single parts make no such constraints (cf. "The cat is in front of, and at the foot of, the chair"). In addition, the metric extent of the search domain designated (e.g., how far the cat is from the truck) can vary greatly. Some languages require figure and ground to be in contact, or visually continuous, others allow the projection of enormous search domains ("in front of the church lie the mountains, running far off to the horizon"). More often perhaps, the notion of a region, an object's penumbra, as it were, is relevant, related to its scale.⁴¹

More exactly An intrinsic spatial relation R is a binary spatial relation, with arguments F and G , where R typically names a part of G . The origin X of the coordinate system C is always on (the volumetric center of) G . An intrinsic relation $R(F, G)$ asserts that F lies in a search domain extending from G on the basis of an angle or line projected from the center of G , through an anchor point A (usually the named facet R), outwards for a determined distance. F and G may be any objects whatsoever (including ego), and F may be a part of G . The relation R does not support transitive inferences, nor converse inferences (see below).

Coordinates may or may not come in fixed armatures. When they do, they tend to be polar; for example, given that facet *A* is the *front* of a building, clockwise rotation in 90° steps will yield *side*, *back*, *side*. Here there is a set of four labeled oppositions, with one privileged facet, *A*. Given *A*, we know which facet *back* is. Because *A* fixes the coordinates, we call it the “anchor point.” But coordinates need not be polar, or indeed part of a fixed set of oppositions; for example, given that facet *B* is the *entrance* of a church and *G_c* its volumetric center, we may derive a line *BG_c* (or an arc with angle determined by the width of *B*)—thus “at the entrance to the church” designates a search area on that line (or in that arc), with no necessary implications about the locations of other intrinsic parts, *front*, *back*, and so on. Because *A* determines the line, we call *A* once again the “anchor point.”

Relative Frame of Reference This is roughly equivalent to the various notions of viewer-centered frame of reference mentioned above (e.g., Marr’s “2½-D sketch,” or the psycholinguist’s “deictic”), but it is not quite the same. The relative frame of reference presupposes a “viewpoint” *V* (given by the location of a perceiver in any sensory modality), and a figure and ground distinct from *V*; it thus offers a triangulation of three points and utilizes coordinates fixed on *V* to assign directions to figure and ground. English “The ball is to the left of the tree” is of this kind of course. Because the perceptual basis is not necessarily visual, calling this frame of reference “viewer-centered” is potentially misleading, but perhaps innocent enough. Calling it “deictic,” however, is potentially pernicious because the “viewer” need not be ego and need not be a participant in the speech event—take, for example, “Bill kicked the ball to the left of the goal.” Nevertheless, there can be little doubt that the deictic uses of this system are basic (prototypical), conceptually prior, and so on.

The coordinate system, centered on viewer *V*, seems generally to be based on the planes through the human body, giving us an *up/down*, *back/front* and *left/right* set of half lines. Such a system of coordinates can be thought of as centered on the main axis of the body and anchored by one of the body parts (e.g., chest). In that case we have polar coordinates, with quadrants counted clockwise from *front* to *right*, *back*, and *left* (Herskovits 1986). Although the position of the body of viewer *V* may be one criterion for anchoring the coordinates, the direction of gaze may be another, and there is no doubt that relative systems are closely hooked into visual criteria. Languages may differ in the weight given to the two criteria, for example, the extent to which occlusion plays a role in the definition of *behind*.

But this set of coordinates on *V* is only the basis for a full relative system; in addition, a secondary set of coordinates is usually derived by mapping (all or some of) the coordinates on *V* onto the relatum (ground object) *G*. The mapping involves a transformation which may be 180° rotation, translation (movement without rota-

tion or reflection), or arguably reflection across the frontal transverse plane. Thus "the cat is in front of the tree" in English entails that the cat F is between V and G (the tree), because the primary coordinates on V appear to have been rotated in the mapping onto G , so that G has a "front" before which the cat sits. Hausa (Hill 1982) and many other languages *translate* rather than rotate the coordinates, so that a sentence glossing "The cat is in front of the tree" will mean what we would mean in English by "The cat is behind the tree." But English is also not so simple, for rotation will get *left* and *right* wrong. In English, "The cat is to the left of the tree" has left on the same side as V 's left, not rotated. In Tamil, the rotation is complete; thus just as front and back are reversed, so are left and right, so that the Tamil sentence glossed "The cat is on the left side of the tree" would (on the relevant interpretation) mean "The cat is on V 's right of the tree." To get the English system right, we might suppose that the coordinates on V should be *reflected* over the transverse plane, as if we wrote the coordinates of V on a sheet of acetate, flipped it over in front of V , and placed it on G . This will get *front*, *back*, *left*, and *right* at least in the correct polar sequence around the secondary origin. But it may not be the correct solution because other interpretations are possible, and indeed more plausible.⁴² But the point to establish here is that a large variation of systems is definable, constituting a broad family of relative systems.

Not all languages have terms glossing *left/right*, *front/back*. Nor does the possession of such a system of oppositions guarantee the possession of a relative system. Many languages use such terms in a more or less purely intrinsic way (even when they are primarily used with deictic centers); that is, they are used as binary relations specifying the location of F within a domain projected from a part of G (as in "to my left," "in front of you," "at the animal's front," "at the house's front," etc.). The test for a relative system is (1) whether it can be used with what is culturally construed as a ground object without intrinsic parts,⁴³ and (2) whether there is a ternary relation with viewpoint V distinct from G , such that when V is rotated around the array, the description changes (see below). Now, languages that do indeed have a relative system of this kind also tend to have an intrinsic system sharing at least some of the same terms.⁴⁴ This typological implication, apart from showing the derivative and secondary nature of relative systems, also more or less guarantees the potential ambiguity of *left/right*, *front/back* systems (although they may be disambiguated syntactically, as in "to the left of the chair" vs. "at the chair's left"). Some languages that lack any such systematic relative system may nevertheless have encoded the odd isolated relative notion, as in " F is in my line of sight toward G ."

That some relative systems clearly use secondary coordinates mapped from V to G suggests that these mappings are by origin a means of extending the intrinsic frame of reference to cases where it would not otherwise apply. (And this may suggest that

the intrinsic system is rather fundamental in human linguistic spatial description.⁴⁵) Through projection of coordinates from the viewpoint V , we assign pseudointrinsic facets to G , as if trees had inherent fronts, backs, and sides.⁴⁶ For some languages, this is undoubtedly the correct analysis; the facets are thus named and regions projected with the same limitations that hold for intrinsic regions.⁴⁷ Thus many relative systems can be thought of as *derived intrinsic* ones—systems that utilize relative conceptual relations to extend and supplement intrinsic ones. One particular reason to so extend intrinsic systems is their extreme limitations as regards logical inference of spatial relations from linguistic descriptions. Intrinsic descriptions support neither transitive nor converse inferences, but relative ones do (Levelt 1984, chapter 3, this volume; and see below).⁴⁸

More exactly A relative relator R expresses a ternary spatial relation, with arguments V , F , and G , where F and G are unrestricted as to type, except that V and G must be distinct.⁵⁰ The primary coordinate system always has its origin on V ; there may be a secondary coordinate system with origin on G . Such coordinate systems are normally polar; for example, *front*, *right*, *back*, and *left* may be assigned by clockwise rotation from *front*. Coordinate systems built primarily on visual criteria may not be polar, but be defined, for example, by rectangular coordinates on the two-dimensional visual field (the retinal projection) so that *left* and *right* are defined on the horizontal or x -axis, and *front* and *back* on the vertical or y -axis (*back* has (the base of) F higher than G and/or occluded by G).

Terms that may be glossed *left* and *right* may involve no secondary coordinates, although they sometimes do (as when they have reversed application from the English usage). Terms glossed *front* and *back* normally do involve secondary coordinates (but compare the analysis in terms of vectors by O'Keefe, chapter 7, this volume). Secondary coordinates may be mapped from primary origin on V to secondary origin on G under the following transformations: rotation, translation, and (arguably) reflection.⁵¹ Typological variations of such systems include degree to

which a systematic polar system of coordinates is available, degree of use of secondary coordinates, type of mapping function (rotation, translation, reflection) for secondary coordinates, differing anchoring systems for the coordinates (e.g., body axis vs. gaze), and differing degrees to which visual criteria (like occlusion, or place in retinal field) are definitional of the terms.

Absolute Frame of Reference Among the many uses of the notion "absolute" frame of reference, one refers to the fixed direction provided by gravity (or the visual horizon under canonical orientation). Less obviously of psychological relevance, the same idea of fixed directions can be applied to the horizontal. In fact, many languages make extensive, some almost exclusive, use of such an absolute frame of reference on the horizontal. They do so by fixing arbitrary fixed bearings, "cardinal directions," corresponding one way or another to directions or arcs that can be related by the analyst to compass bearings. Speakers of such languages can then describe an array of, for example, a spoon in front of a cup, as "spoon to north/south/east/(etc.) of cup" without any reference to the viewer/speaker's location.

Such a system requires that persons maintain their orientation with respect to the fixed bearings at all times. People who speak such languages can be shown to do so—for example, they can dead reckon current location in unfamiliar territory with extraordinary accuracy, and thus point to any named location from any other (Lewis 1976; Levinson 1992b). How they do so is simply not known at the present time, but we may presume that a heightened sense of inertial navigation is regularly cross-checked with many environmental clues.⁵² Indeed, many such systems are clearly abstractions and refinements from environmental gradients (mountain slopes, prevailing wind directions, river drainages, celestial azimuths, etc.).⁵³ These "cardinal directions" may therefore occur with fixed bearings skewed at various degrees from, and in effect unrelated to, our "north," "south," "east," and "west." It perhaps needs emphasizing that this keeping track of fixed directions is, with appropriate socialization, not a feat restricted to certain ethnicities, races, environments, or culture types, as shown by its widespread occurrence (in perhaps a third of all human languages?) from Meso-America, to New Guinea, to Australia, to Nepal. No simple ecological determinism will explain the occurrence of such systems, which can be found alternating with, for example, relative systems, across neighboring ethnic groups in similar environments, and which occur in environments of contrastive kinds (e.g., wide open deserts and closed jungle terrain).

The conceptual ingredients for such systems are simple: the relevant linguistic expressions are binary relators, with figure and ground as arguments and a system of coordinates anchored to fixed bearings, which always have their origin on the ground. In fact, these systems are the only systems with conceptual simplicity and

elegance. For example, they are the only systems that fully support transitive inferences across spatial descriptions. Intrinsic descriptions do not do so, and relative ones do so only if viewpoint *V* is held constant (Levelt 1984). Intrinsic systems are dogged by the multiplicity of object types, the differing degrees to which the asymmetries of objects allow the naming of facets, and the problem of “unfeatured” objects. Relative systems are dogged by the psychological difficulties involved in learning left/right distinctions, and the complexities involved in mapping secondary coordinates; often developed from intrinsic systems they display ambiguities across frames of reference (like English “in front of”). The liabilities of absolute systems are not, on the other hand, logical but psychological; they require a cognitive overhead, namely the constant background calculation of cardinal directions, together with a system of dead reckoning that will specify for any arbitrary point *P* which direction *P* is from ego’s current locus (so that ego may refer to the location of *P*).

Absolute systems may also show ambiguities of various kinds. First, places of particular sociocultural importance may come to be designated by a cardinal direction term, like a quasi-proper name, regardless of their location with respect to *G*. Second, where the system is abstracted out of landscape features, the relevant expressions (e.g., “uphill” or “upstream”) may either refer to places indicated by relevant local features (e.g., local hill, local stream), or to the abstracted fixed bearings, where these do not coincide. Third, some such systems may even have relative interpretations (e.g., “uphill” may imply further away in my field of vision; cf. our interpretation of “north” as top of a map).

One crucial question with respect to absolute systems is how, conceptually, the coordinate system is thought of. It may be a polar system, as in our *north/south/east/west*, where *north* is the designated anchor and *east*, *south*, *west*, found by clockwise rotation from *north*.⁵⁴ Other systems may have a primary and a secondary axis, so that, for example, a *north-south* axis is primary, but it is not clear which direction, *north* or *south*, is itself the anchor.⁵⁵ Yet other systems favor no particular primary reference point, each half axis having its own clear anchor or fixed central bearing.⁵⁶ Some systems like Tzeltal are “degenerate,” in that they offer two labeled half lines (roughly, “north,” “south”), but label both ends of the orthogonal with the same terms. Even more confusing, some systems may employ true abstracted cardinal directions on one axis, but landmark designations on the other, guaranteeing that the two axes do not remain orthogonal when arrays are described in widely different places. Thus on Bali, and similarly for many Austronesian systems, one axis is determined by monsoons and is a fixed, abstracted axis, but the other is determined by the location of the central mountain and thus varies continuously when one circumnavigates the island. Even where systematic cardinal systems exist, the geometry of the designated angles is variable. Thus, if we have four half lines based on orthogonal

axes, the labels may describe quadrants (as in Guugu Yimithirr), or they may have narrower arcs of application on one axis than the other (as appears to be the case in Wik Mungan⁵⁷). Even in English, though we may think of north as a point on the horizon, we also use arcs of variable extent for informal description.

More exactly An absolute relator R expresses a binary relation between F and G , asserting that F can be found in a search domain at the fixed bearing R from G . The origin X of the coordinate system is always centered on G . G may be any object whatsoever, including ego or another deictic center; F may be a part of G . The geometry of the coordinate system is linguistically/culturally variable, so that in some systems equal quadrants of 90 degrees may be projected from G , while in others something more like 45 degrees may hold for arcs on one axis, and perhaps 135 degrees on the other. The literature also reports abstract systems based on star-setting points, which will then have uneven distribution around the horizon.

Just as relative relators can be understood to map designated facets onto ground objects (thus “on the front of the tree” assigns a named part to the tree), so absolute relators may also do so. Many Australian languages have cardinal edge roots, then affixes indicating, for example, “northern edge.” Some of these stems can only be analyzed as an interaction between the intrinsic facets of an object and absolute directions.

4.3.3.2 “Logical Structure” of the Three Frames of Reference We have argued that, as far as language is concerned, we must distinguish frame of reference qua coordinate system from, say, deictic center qua origin of the coordinate system. Still, the skeptical may doubt that this is either necessary or possible.

First, to underline the necessity, each of our three frames of reference may occur with or without a deictic center (or egocentric origin). Thus for the intrinsic frame, we can say, “The ball is in front of me” (deictic center); for the absolute frame we can say, “The ball is north of me”; and of course in the relative frame, we can say, “The ball is in front of the tree” (from ego’s point of view). Conversely, none of the three frames need have a deictic center. Thus in the intrinsic frame one can say “in front of the chair”; in the absolute frame, “north of the chair”; and in the relative frame, “in front of the tree from Bill’s point of view.” This is just what we should expect given the flexible nature of linguistic reference—it follows from Hockett’s (1960) design feature of displacement, or Bühler’s (1934) concept of transposed deictic center.

Second, we need to show that we can in fact define the three frames of reference adequately without reference to the opposition deictic versus nondeictic center or origin. We have already hinted at plenty of distinguishing characteristics for each of the three frames. But to collect them together, let us first consider the logical

properties. The absolute and intrinsic relators share the property that they are binary relations whereas relative relators are ternary. But absolute and intrinsic are distinguished in that absolute relators define asymmetric transitive relations (if F_1 is north of G , and F_2 is north of F_1 , then F_2 is north of G), where converses can be inferred (if F is north of G , G is south of F). The same does not hold for intrinsic relators, which hardly support any spatial inferences at all without further assumptions (see Levelt 1984 and chapter 3, this volume). In this case, absolute and relative relators share logical features because relative relators support transitive and converse inferences provided that viewpoint V is held constant.

Although this is already sufficient to distinguish the three frames, we may add further distinguishing factors. Certain important properties follow from the nature of the anchoring system in each case. In the intrinsic case we can think of the named facet of the object as providing the anchor; in the relative case we can think of the viewpoint V on an observer, with the anchor being constituted by, say, the direction of the observer's front or gaze, while in the absolute case one or more of the labeled fixed bearings establishes a conceptual "slope" across the environment, thus fixing the coordinate system. From this, certain distinct properties under rotation emerge as illustrated in figure 4.10.⁵⁸ These properties have a special importance for the study of nonlinguistic conceptual coding of spatial arrays because they allow systematic experimentation (as illustrated in section 4.1; see also Levinson 1992b; Brown and Levinson 1993b; Pederson 1993, 1994; Danziger 1993).

Altogether then, we may summarize the distinctive features of each frame of reference as in table 4.3; these features are jointly certainly sufficient to establish the nature of the three frames of reference independently of reference to the nature of the origin of the coordinate system. We may conclude this discussion of the linguistic frames of reference with the following observations:

1. Languages use, it seems, just three frames of reference: absolute, intrinsic, and relative;
2. Not all languages use all frames of reference; some use predominantly one only (absolute or intrinsic; relative seems to require intrinsic); some use two (intrinsic and relative, or intrinsic and absolute), while some use all three;
3. Linguistic expressions may be specialized to a frame of reference, so we cannot assume that choice of frame of reference lies entirely outside language, for example, in spatial thinking, as some have suggested. But spatial relators may be ambiguous (or semantically general) across frames, and often are.

4.3.3.3 Realigning Frames of Reference across Disciplines and Modalities We are now at last in a position to see how our three linguistic frames of reference align with

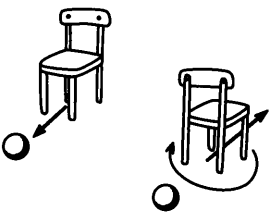
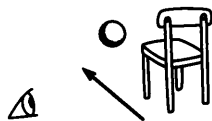
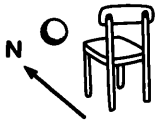
Rotation of:			
	viewer	ground object	whole array
Intrinsic "ball in front of chair" 	same description? yes	same description? no	same description? yes
Relative "ball to left of chair" 	no	yes	no
Absolute "ball to north of chair" 	yes	yes	no

Figure 4.10
Properties of the frames of reference under rotation.

Table 4.3
Summary of Properties of Different Frames of Reference

	Intrinsic	Absolute	Relative
Relation is	binary	binary	ternary
Origin on	ground	ground	viewpoint <i>V</i>
Anchored by	<i>A</i> within <i>G</i>	"slope"	<i>A</i> within <i>V</i>
Transitive?	No	Yes	Yes if <i>V</i> constant
Constant under rotation of			
whole array?	Yes	No	No
viewer?	Yes	Yes	No
ground?	No	Yes	Yes

the other distinctions in the literature arising from the consideration of other modalities (as listed in table 4.1). The motive, let us remember, is to try to make sense of the very idea of "same frame of reference" across modalities, and in particular from various kinds of nonlinguistic thinking to linguistic conceptualization.

An immediate difficulty is that, by establishing that frames of reference in language should be considered independently of the origin of the coordinate systems, we have opened up a gulf between language and the various perceptual modalities, where the origin of the coordinate system is so often fixed on some ego-center. But this mismatch is in fact just as it should be. Language is a flexible instrument of communication, designed (as it were) so that one may express other persons' points of view, take other perspectives, and so on. At the level of perception, origin and coordinate system presumably come prepackaged as a whole, but at the level of language, and perhaps more generally at the level of conception, they can vary freely and combine.

So to realign the linguistic distinctions with distinctions made across other modalities, we need to fix the origin of the coordinate system so that it coincides, or fails to coincide, with ego in each frame of reference. We may do so as follows. First, we may concede that the relative frame of reference, though not necessarily egocentric, is prototypically so. Second, we may note that the intrinsic system is typically, but not definitionally, non-egocentric. Third, and perhaps most arbitrarily, we may assign a non-egocentric origin to the absolute system. These assignments should be understood as special subcases of the uses of the linguistic frames of reference.

If we make these restrictions, then we can align the linguistic frames of reference with the other distinctions from the literature as in table 4.4.⁵⁹ Notice then that, under the restriction concerning the nature of the origin:

Table 4.4
Aligning Classifications of Frames of Reference

Intrinsic	Absolute	Relative
Origin $\neq$ ego	Origin $\neq$ ego	Origin = ego
Object-centered	Environment-centered	Viewer-centered
Intrinsic perspective		Deictic perspective
3-D model		2½-D sketch
Allocentric	Allocentric	Egocentric
Orientation-free	Orientation-bound	Orientation-bound

1. Intrinsic and absolute are grouped as allocentric frames of reference, as opposed to the egocentric relative system;
2. Absolute and relative are grouped as orientation-bound, as opposed to intrinsic, which is orientation-free.

This correctly captures our theoretical intuitions. In certain respects, absolute and intrinsic viewpoints are fundamentally similar—they are binary relations that are viewpoint-independent, where the origin may happen to be ego but need not be; they are allocentric systems that yield an ego-invariant picture of the “world out there.” On the other hand, absolute and relative frameworks are fundamentally similar on another dimension because they both impose a larger spatial framework on an assemblage, specifying its orientation with respect to external coordinates; thus in an intrinsic framework it is impossible to distinguish enantiomorphic pairs, while in either of the orientation-bound systems it is inevitable.⁶⁰ Absolute and relative frameworks presuppose a Newtonian or Kantian spatial envelope, while the intrinsic framework is Leibnizian.

The object-centered nature of the intrinsic system hooks it up to Marr's (1982) 3-D model in the theory of vision, and the nature of the linguistic expressions involved suggests that the intrinsic framework is a generalization from the analysis of objects into their parts. A whole configuration can be seen as a single complex object, so that we can talk of the leading car in a convoy as “the head of the line.” On the other hand, the viewer-centered nature of the relative framework connects it directly to the sequence of 2-D representations in the theory of vision. Thus the spatial frameworks in the perceptual systems can indeed be correlated with the linguistic frames of reference.

To summarize, I have sought to establish that there is nothing incoherent in the notion “same frame of reference” across modalities or inner representation systems. Indeed, even the existing distinctions that have been proposed can be seen in many

detailed ways to correlate with the revised linguistic ones, once the special flexibility of the linguistic systems with respect to origin is taken into account. Thus it should be possible, and intellectually profitable, to formulate the distinct frames of reference in such a way that they have cross-modal application. Notice that this view conflicts with the views of some that frames of reference in language are imposed just in the mapping from perception to language via the encoding process. On the contrary, I shall presume that any and every spatial representation, whether perceptual or conceptual, must involve a frame of reference; for example, retinotopic images just are, willy nilly, in a viewer-centered frame of reference.

But at least one major problem remains. It turns out that the three distinct frames of reference are “untranslatable” from one to the other, throwing further doubt on the idea of correlations and correspondences across sensory and conceptual representational levels. Which brings us to Molyneux’s question.

4.4 Molyneux’s Question

In 1690 William Molyneux wrote John Locke a letter posing the following celebrated question: If a blind man, who knew by touch the difference between a cube and a sphere, had his sight restored, would he recognize the selfsame objects under his new perceptual modality or not?⁶¹

The question whether our spatial perception and conception is modality-specific is as alive now as then. Is there one central spatial model, to which all our input senses report, and from which instructions can be generated appropriate to the various output systems (touch, movement, language, gaze, and so on)?

There have of course been attempts to answer Molyneux directly, but the results are conflicting. On the one hand, sight-restored individuals take a while to adjust (Gregory 1987, 94–96; Valvo 1971), monkeys reared with their own limbs masked from sight have trouble relating touch to vision when the mask is finally removed (Howard 1987, 730–731), and touch and vision are attuned to different properties (e.g., the tactile sense is more attuned to weight and texture than shape; Klatsky and Lederman 1993); on the other hand, human neonates immediately extrapolate from touch to vision (Meltzoff 1993), and the neurophysiology suggests direct cross-wirings (Berthoz 1991, 81; but see also Stein 1992), so that some feel that the answer to the question is a “resounding ‘yes’” (Eilan 1993, 237). More soberly, it seems that there is some innate supramodal system observable in monkeys and infants, but it may be very restricted, and sophisticated cross-modal thinking may even be dependent on language.⁶²

Here I want to suggest another way to think about this old question. Put simply, we may ask whether the same frames of reference can in principle operate across all

the modalities, and if not, whether at least they can be translated into one another. What we should mean by “modality” here is an important question. In what follows I shall assume that corresponding to (some of) the different senses, and more generally to input/output systems, there are specialized “central” representational systems, for example, an imagistic system related to vision, a propositional system related to language, a kinaesthetic system related to gesture, and so on (see, for example, Levelt 1989; Jackendoff 1991). Our version of Molyneux's question then becomes two related questions:

1. Do the different representational systems natively and necessarily employ certain frames of reference?
2. If so, can representations in one frame of reference be translated (converted) into another frame of reference?

Let us discount here the self-evident fact that certain kinds of information may perhaps, in principle, be modality-specific; for example, spatial representations in an imagistic mode must, it seems, be determinate with respect to shape, while those in a propositional mode need not, and perhaps, cannot be so.⁶³ Similarly, the haptic-kinesthetic modality will have available direct information about weight, texture, tactile warmth, and three-dimensional shape we can only guess at from visual information (Klatsky and Lederman 1993), while the directional and inertial information from the vestibular system is of a different kind again. All this would seem to rule out a *single* supramodal spatial representation system. What hybrid monster would a representation system have to be to record such disparate information? All that concerns us here is the compatibility of *frames of reference* across modalities.

First, let us consider question 2, translatability across frames of reference. This is the easier question, and the answer to it offers an indirect answer to question 1. There is a striking, but on a moment's reflection, self-evident fact: you cannot freely convert information from one framework to another. Consider, for example, an array, with a bottle on the ground at the (intrinsic) front side of a chair. Suppose, too, that you view the array from a viewpoint such that the bottle is to the right of the chair; as it happens, the bottle is also north of the chair (see figure 4.11). Now I ask you to remember it, and suppose you “code” the scene in an intrinsic frame of reference: “bottle in front of chair,” discarding other information. It is immediately obvious that, from this intrinsic description, you cannot later generate a relative description—if you were viewing the array so that you faced one side of the chair, then the bottle would be to the left of or to the right of the chair—depending on your viewpoint. So without a “coding” or specification of the locus of the viewpoint *V*, you cannot generate a relative description from an intrinsic description. The same holds for an absolute description. Knowing that the bottle is at the front of the chair will

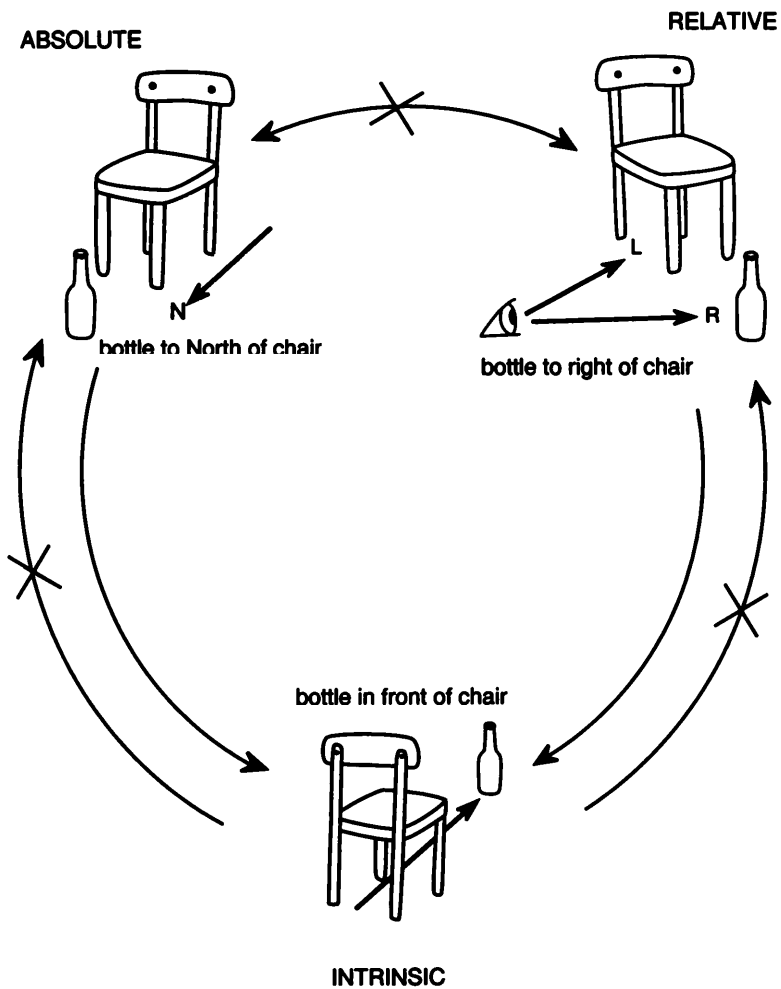


Figure 4.11
Untranslatability across frames of reference.

not tell you whether it is north or south or east or west of the chair—for that, you will need ancillary information. In short, you cannot get from an intrinsic description—an orientation-free representation—to either of the orientation-bound representations.

What about conversions between the two orientation-bound frameworks? Again, it is clear that no conversion is possible. From the relative description or coding “The bottle is to the left of the chair,” you do not know what cardinal direction the bottle lies in, nor from “the bottle is north of the chair” can you derive a viewpoint-relative description like “to the left of the chair.”

Indeed, the only directions in which you can convert frames of reference are, in principle, from the two orientation-bound frames (relative and absolute) to the orientation-free one (intrinsic).⁶⁴ For if the orientation of the ground object is fully specified, then you can derive an intrinsic description. For example, from the relative description “The chair is facing to my right and the bottle is to the right of the chair in the same plane,” and likewise from the absolute description “The chair is facing north and the bottle to the north of the chair,” you can, in principle, arrive at the intrinsic specification “The bottle is at the chair’s front.” Normally, though, because the orientation of the ground object is irrelevant to the orientation-bound descriptions, this remains a translation only in principle. By the same reasoning, translations in all other directions are in principle “out,” that is, impossible.

This simple fact about translatability across frames of reference may have far-reaching consequences. Consider, for example, the following syllogism:

1. Frames of reference are incommensurable (i.e., a representation in one framework is not freely convertible into a representation in another);
2. Each sense utilizes its own frame(s) of reference (e.g., while vision primarily uses a viewer-centered frame, touch arguably uses primarily an object-centered frame, based on the appreciation of form through three-dimensional grasping);
3. Representations from one modality (e.g., haptic) cannot be freely translated into representations in another (e.g., visual).

The syllogism suggest, then, that the answer to Molyneux’s question is no—the blind man upon seeing for the first time will not recognize by sight what he knew before by touch. More generally, we will not be able to exchange information across any internal representation systems that are not based on one and the same frame of reference.

I take this to be a counterintuitive result, a clearly false conclusion, in fact a *reductio ad absurdum*. We can indeed form mental images of contour shapes explored by touch alone, we can gesture about what we have seen, we can talk about,

or draw, what we have felt with our fingers, and so on. Because premise 1 seems self-evidently true, we must then reject premise 2, the assumption that each sensory modality or representational system operates exclusively in its own primary, proprietary frame of reference. In short, either the frame of reference must be the same across all sensory modalities to allow the cross-modal sharing of information or each modality must allow more than one frame of reference.

Intuitively, this seems the correct conclusion. On the one hand, peripheral sensory systems may operate in proprietary frames of reference; for example, low-level vision may know only of 2-D retinotopic arrays, while otoliths are restricted to a gravitational frame of reference. But, on the other hand, at a higher level, visual processing seems to deliver 3-D analyses of objects as well as 2-D ones. Thus when we (presumably) use the visual system to imagine rotations of objects, we project from 3-D models (intrinsic) to $2\frac{1}{2}$ -D (relative) ones, showing that both are available. Thus more central, more conceptual, levels of representation seem capable of adopting more than one frame of reference.

Here, then, is the first part of the answer to our puzzle. Representational systems of different kinds, specialized to different sensory modalities (like visual memory) or output systems (like gesture and language), may be capable of adopting different frames of reference. This would explain how it is that Tenejapans, or indeed Dutch subjects, can adopt the same frame of reference when utilizing different representational systems—those involved in generating gesture, those involved in tasks requiring visual memory, those involved in making spatial inferences, as well as those involved in speaking.

But to account for the facts described in section 4.2, it will not be sufficient to establish that the same frame of reference can, in principle, be used across different kinds of internal representation systems, those involved in nonverbal memory, gesture and language, and so on. To account for those facts, it will be necessary to assume that individual subjects do indeed actually utilize the same frame of reference across modalities. But now we have an explanation for this apparent fact: the untranslatability across frames of reference requires individuals to stabilize their representational systems within a limited set of frames of reference. For example, if a Tenejapan man sees an array and remembers it only in terms of a viewer-centered framework, he will not later be able to describe it—his language simply fails to provide a systematic viewer-centered frame of description. Thus the facts that (a) frameworks are not freely convertible, (b) languages may offer restricted frameworks as output, and (c) it may be desirable to describe any spatial experience whatsoever at some later point, these conspire to require that a speaker code spatial perceptions at the time of experience in whatever output frameworks the speaker's dominant language offers.

4.5 Conclusions

This chapter began with some quite unexpected findings: languages can differ in the set of frames of reference they employ for spatial description. Moreover, the options in a particular language seem to dictate the use of frames of reference in nonlinguistic tasks—there seems thus to be a cross-modal tendency to fix on a dominant frame of reference. This raises a number of fundamental puzzles: What sense does it make to talk of “same frame of reference” across modalities, or psychological faculties of quite different kinds? If it does make sense, why should it be so? What light does the phenomenon throw on how spatial information is shared across the senses, across the various “input” and “output” devices?

I have tried to sketch answers to these puzzles. The answers converge in two kinds of responses to Molyneux's question “do the senses talk to one another?” The first kind of response is an empirical argument:

1. The frame of reference dominant in a given language “infiltrates” other modalities, presumably to ensure that speakers can talk about what they see, feel, and so on;
2. Therefore, other modalities have the capacity to adopt, or adapt to, other frames of reference, which suggests a yes answer to Mr. Molyneux.

The second kind of response is an a priori argument:

1. Frames of reference cannot freely “translate” into one another;
2. Therefore, if the modality most adaptive to external influences, namely, language, adopts one frame of reference, the others must follow suit;
3. To do this, all modalities must have different frames of reference available, or be able to “annotate” experiences with the necessary ancillary information, which suggests a yes answer to Mr. Molyneux.

Actually, an affirmative answer to Molyneux's question is evidently required—otherwise we could not talk about what we see. What is deeply mysterious is *how* this cross-modal transfer is achieved. The untranslatability across frames of reference greatly increases the puzzle. It is in this light that the findings with which we began—the standardization of frames of reference across modalities in line with the local language—now seem not only less surprising, but actually inevitable.

Acknowledgments

This chapter is based on results of joint research, in particular with Penelope Brown on Tzeltal, but also with many colleagues in the Cognitive Anthropology Research Group, who have collaboratively developed the research program outlined here (see also Senft 1994; Wilkins

1993; Pederson 1994; Danziger 1994; Hill 1994). I am also indebted to colleagues in the wider Psycholinguistics Institute, who have through different research programs challenged premature conclusions and emboldened others (see, for example, in this volume Bierwisch, Levelt, and Bowerman, chapters 2, 3, and 10, respectively; the debt to Levelt's pioneering work on the typology and logic of spatial relations will be particularly evident). In addition, John Lucy, Suzanne Gaskins, and Dan Slobin have been important intellectual influences; and Bernadette Schmitt and László Nagy have contributed to experimental design and analysis. The contributions, ideas, and criticisms of other participants at the conference at which this paper was given have been woven into the text; my thanks to them and the organizers of the conference. Finally, I received very helpful comments on the manuscript from Sotaro Kita, Lynn Nadel, Mary Peterson, and David Wilkins, not all of which I have been able to adequately respond to.

Notes

1. I shall use the term *modality* in a slightly special, but I think motivated, way. When psychologists talk of "cross-modal" effects, they have in mind transfer of information across sensory modalities (vision, touch, etc.). Assuming that these sensory input systems are "modules" in the Fodorean sense, we are then interested in how the output of one module, in some particular inner representation system, is related to the output of some other module, most likely in another inner representation system appropriate to another sensory faculty. Thus cross-modal effects can be assumed to occur through communication between central, but still sense-specific, representation systems, not through peripheral representation systems specialized to modular processes. But see section 4.4.
2. Although there are phrases designating left-hand and right-hand, these are body-part terms with no spatial uses, while body-part terms for *face* and *back* are used for spatial description nearly exclusively for objects in contiguity and then on the basis of an *intrinsic* assignment, not a relative one based on the speaker's viewpoint (see Levinson 1994).
3. The design of this experiment was much improved by Bernadette Schmitt.
4. The design of this experiment is by Eric Pederson and Bernadette Schmitt, building on an earlier design described in Levinson 1992b.
5. The phenomenon of fixed bearings in gesture was first noticed for an Australian Aboriginal group by Haviland (1993), who subsequently demonstrated the existence of the same phenomenon in Zinacantan, a neighboring community to Tenejapa.
6. Rock (1992) is here commenting on Asch and Witkin 1948, which built directly on the Gestalt notions. See also Rock (1990).
7. One kind of disagreement is voiced by Paillard (1991, 471): "Spatial frameworks are incorporated in our perceptual and motor experiences. They are not however to be confused with the *system of coordinates* which abstractly represent them" (emphasis). But this is terminological; for our purposes we wish precisely to abstract out the properties of frames of reference, so that we can consider how they apply across different perceptual or conceptual systems.
8. "When places are individuated by their spatial relation to certain objects, a crucial part of what we need to know is what those objects are. As the term 'frame of reference' is commonly used, these objects would be said to provide the frame of reference" (Brewer and Pears 1993, 25).

9. I shall use the opposition *figure* versus *ground* for the object to be located versus the object with respect to which it is to be located, respectively, after Talmy 1983. This opposition is identical to *theme* versus *relatum*, *referent* versus *relatum*, *trajector* versus *landmark*, and various other terminologies.

10. Brewer and Pears (1993, 26) consider the role of coordinate systems, but what they have to say only increases our puzzlement: "Two events are represented as being in the same spatial position if and only if they are assigned the same co-ordinates. Specifying a frame of reference would have to do with specifying how co-ordinates are to be assigned to events in the world on the basis of their spatial relations to certain objects. These objects provide the frame of reference." This fails to recognize that two distinct systems of coordinates over the same objects can describe the same place.

11. There are many good sketches of parts of this intellectual terrain (see, for example, Miller and Johnson-Laird 1976; Jammer 1954; O'Keefe and Nadel 1978), but none of it all.

12. Some notion of absolute space was already presupposed by Descartes's introduction of coordinate systems, as Einstein (1954, xiv) pointed out.

13. This association was in part due to the British empiricists like Berkeley whose solipsism made egocentric relative space the basis for all our spatial ideas. See O'Keefe and Nadel 1978, 14–16.

14. Much behavioral experimentation on rats in mazes has led to classifications of behavior parallel to the notions of frame of reference. O'Keefe and Nadel's 1978 classification, for example, is in terms of body position responses (cf. egocentric frames of reference), cue responses (a kind of allocentric response to an environmental gradient), and place responses (involving allocentric mental maps). Work on infant behavior similarly relates behavioral response types to frames of reference, usually egocentric versus allocentric (or geographic—see Pick 1988, 147–156).

15. See also Brewer and Pears (1993, 29), who argue that allocentric behavior can always be mimicked through egocentric computations: "Perhaps language . . . provides the only conclusive macroscopic evidence for genuine allocentricity."

16. These distinctions are seldom properly made in the literature on mental maps in humans. Students of animal behavior, though, have noted that maps consisting of relative angles and distances between landmarks have quite different computational properties to maps with fixed bearings: in the former, but not the latter, each time landmarks are added to the map, the database increases exponentially (see, for example, McNaughton, Chen, and Markus 1990). Despite that, most rat studies fail to distinguish between these two kinds of allocentricity, relative and absolute.

17. Paillard (1991, 471–472) has a broader notion of "frames of reference" than most brain scientists (and closer to psychological ideas); he proposes that there are four such frames subserving visually guided action, all organized around the geocentric vertical: (1) a body frame, presuming upright posture for action; (2) an object frame, presumably similar to Marr's (1982) object-centered system; (3) a world frame, a Euclidean space inclusive of both body and object; and (4) a retinal frame, feeding the object and world frames. He even provides a rough neural "wiring diagram" (p. 473).

18. The age at which this switch to the non-egocentric takes place seems highly task-dependent. See Acredolo (1988), who gives sixteen months as an end point; see also Pick (1993), for a route-finding task, where the process has hardly begun by sixteen months.

19. This leap from a perspective image, or worse, a silhouette, is possible (Marr argued) only by assuming that objects can be analyzed into geometrical volumes of a specific kind (generalized cones); hence 3-D models must be of this kind, where principal axes are identified.

20. Others have suggested that what we store is a $2\frac{1}{2}$ -D image coupled with the ability to mentally rotate it (Tarr and Pinker 1989), thus giving our apparent ability to rotate mental images (Shepard and Metzler 1971) some evolutionary *raison d'être*. Yet others suggest that object recognition is achieved via a set of $2\frac{1}{2}$ -D images from different orientations (Bülthoff 1991), while some (Rock, Wheeler, and Tudor 1989) suggest we have none of these powers.

21. See Danziger 1994 for possible connections to linguistic distinctions; I am grateful to Eve Danziger for putting me in touch with this work.

22. As Kant 1768 made clear, objects differing in handedness (enantiomorphs or “incongruent counterparts” in Kant’s terminology), cannot be distinguished in an object-centered (or intrinsic) frame of reference, but only in an external coordinate system. See Van Cleve and Frederick 1991, and, for the relevance to Tzeltal, Levinson and Brown 1994.

23. For example, the cube comparisons test can be solved by (1) rotation using viewer-centered coordinates; (2) rotation around an object-centered axis imaged with viewer-centered coordinates; (3) rotation of the perspective point around the object; or (4) purely object-centered comparisons.

24. Thus Cohen and Kubovy (1993, 379) display deep confusion about frames of reference: they suggest that one can have orientation-free representations of handedness information in an orientation-free frame of reference by utilizing the notion “clockwise.” But as Kant (1768) showed, and generations of philosophers since have agreed (see Van Cleve and Frederick 1991), the notion “clockwise” presupposes an external orientation.

25. Carlson-Radvansky and Irwin’s view would seem to be subtly different from Levelt’s (1989); see below in text.

26. The equation is Tversky’s; actually, her survey perspective in some cases (e.g., outside the context of maps) may also relate to a more abstract “absolute” spatial framework where both viewer and landmarks are embedded in a larger frame of reference.

27. The conceptual system is abstract over different perceptual clues, as shown by the fact that astronauts can happily talk about, say, “above and to the left” where one perceptual clue for the vertical (namely gravity) is missing (Friederici and Levelt 1990). Levelt (1989, 154–155) concludes that the spatial representation itself does not determine the linguistic description: “There is ... substantial freedom in putting the perceived structure, which is spatially represented, into one or another propositional format.”

28. For example, there is no convincing explanation of the English deictic use of “front,” “back,” “left,” “right”: we say, “The cat in front of the tree,” as if the tree was an interlocutor facing us, but when we say, “The cat is to the left of the tree,” we do not (as, for example, in

Tamil) mean the cat is to the tree's left, therefore to our right. The reason for this explanatory gap is that the facts have always been underdescribed, the requisite coordinate systems not being properly spelled out even in the most recent works.

29. The so-called topological prepositions or relators have a complex relation to frames of reference. First, note that frames of reference are here defined in terms of coordinate systems, and many "topological" relators express no angular or coordinate information, for example, *at* or *near*. However, others do involve the vertical absolute dimension and often intrinsic features, or axial properties, of landmark objects. Thus proper analysis of the "topological" notions involves partitioning their features between noncoordinate spatial information and features of information distributed between the frames of reference mentioned below in the text. Thus English *in* as in "the money in the piggy bank" is an intrinsic notion based on properties of the ground object; *under* as in "the dust under the rug" compounds intrinsic (under surface, bottom) and absolute (vertical) information, and so forth.

30. Except in some places, like the Torres Straits, where the trade winds roar through westward and spatial descriptions can be in terms of "leeward" and "windward." Or where the earth drops away in one direction, as on the edges of mountain ranges, gravity can be naturally imported into the horizontal plane.

31. The reader may feel that the notion of "front" is different for chairs and persons (and so of course it is), and in particular that "in front of me" is somehow more abstract than "in front of the chair." But notice that we could have said "at my feet" or "at the foot of the chair"—here "feet" or "foot" clearly means something different in each case, but shares the notion of an intrinsic part of the relatum object.

32. The importance of the distinction between binary and ternary spatial relators was pointed out by Herrmann 1990.

33. For example, the Australian language Guugu Yimithirr has (derived) lexemes meaning "north side of," "south side of," and so on, which combine both intrinsic and absolute frames of reference in a single word. Less exotically, English *on* as in "the cup on the table" would seem to combine absolute (vertical) information with topological information (contact) and intrinsic information (supporting planar surface).

34. This point is important. Some psychologists have been tempted to presume, because of the ambiguity of English spatial expressions such as "in front," that frames of reference are imposed on language by a spatial interpretation, rather than being distinguished semantically (see, for example, Carlson-Radvansky and Irwin 1993).

35. We know one way in which this tripartite typology may be incomplete: some languages use conventionalized landmark systems that in practice grade into absolute systems, although there are reasons for thinking that landmark systems and fixed-bearing systems are distinct conceptual types.

36. I am indebted to many discussions with colleagues (especially Balthasar Bickel, Eric Pederson, and David Wilkins) over the details of this scheme, although they would not necessarily agree with this particular version.

37. Thus the "face" of a stone may be the bottom surface hidden in the soil, as long as it meets the necessary axial and shape conditions.

38. We tend to think of human prototypes as inevitably the source of such prototype parts, but such anthropomorphism may be ethnocentric; for example, in Mayan languages plant parts figure in human body-part descriptions (see Laughlin 1975; Levinson 1994).

39. Thus Miller and Johnson-Laird (1976, 401), thinking of English speakers: "People tend to treat objects as six-sided. If an object has both an intrinsic top and bottom, and an intrinsic front and back, the remaining two sides are intrinsically left and right." Incidentally, the possession of "intrinsic left/right" is perhaps an indication that such systems are not exclusively object-centered (because left and right cannot ultimately be distinguished without an external frame of reference).

40. For a nice contrast between two apparently similar Meso-American systems, one of which is armature-based and the other based on the location of individual facets, see MacLaury (1989) on Zapotec, and Levinson (1994) on Tzeltal.

41. Miller and Johnson-Laird (1976) suggest that the notion of intrinsic region may be linked to perceptual contiguity within 10 degrees of visual arc (p. 91), but that the conceptual counterpart to this perceptual notion of region combines perceptual information with functional information about the region drawn from social or physical interaction (pp. 387–388).

42. It may be that *left* and *right* are centered on *V*, while *front* and *back* are indeed rotated and have their origin on *G*. Evidence for that analysis comes from various quarters. First, some languages like Japanese allow both the English- and Hausa-style interpretations of *front*, while maintaining *left* and *right* always the same, suggesting that there are two distinct subsystems involved. Second, English "left" and "right" are not clearly centered on *G* because something can be to the left of *G* but not in the same plane at all (e.g., "the mountain to the left of the tree"), while English "front" and "back" can be centered on *G*, so that it is odd to say of a cat near me that it is "in front of a distant tree." Above all, there is no contradiction in "the cat is to the front and to the left of the tree." An alternative analysis of English would have the coordinates fixed firmly on *V*, and give "*F* is in front of the tree" an interpretation along the lines "*F* is between *V* and *G*" ("behind" glossing "*G* is between *V* and *F*"). My own guess is that English is semantically general over these alternative interpretations.

43. Note that, for example, we think of a tree as unfeatured on the horizontal dimension, so that it lacks an intrinsic front, while some Nilotic cultures make the assumption that a tree has a front, away from the way it leans.

44. But some languages encode relative concepts based directly on visual occlusion or the absence of it; these do not have intrinsic counterparts (as S. Kita has pointed out to me).

45. As shown by the intrinsic system's priority in acquisition (Johnston and Slobin 1978). On the other hand, some languages hardly utilize an intrinsic frame of reference at all (see, for example, Levinson 1992b on an Australian language).

46. I owe the germ of this idea to Eric Pederson.

47. This does not seem, once again, the right analysis for English *left/right*, because *F* and *G* need not be in the same plane at all (as in "the tree to the left of the rising moon"), and intuitively, "to the left of the ball" does not ascribe a left facet to the ball.

48. Although transitivity and converseness in relative descriptions hold only on the presumption that *V* is constant.
49. Conversely, other languages like Tamil use it in more far-reaching ways.
50. *F* may be a part of *G*, as in "the bark on the left (side) of the tree."
51. Rotation will have *front* toward *V*, and clockwise (looking down on *G*) from *front*: *right*, *back*, *left* (as in Tamil). Translation will have *back* toward *V*, and clockwise from *back*: *left*, *front*, *right* (as in Hausa). Reflection will have *front* toward *V*, but clockwise from *front*: *left*, *back*, *right* (as in English, on one analysis). The rotation and translation cases clearly involve secondary polar coordinates on *G*. The reflection cases can be reanalyzed as defined by horizontal and vertical coordinates on the retinal projection, or can be thought of (as seems correct for English) as the superimposition of two systems, the *left/right* terms involving only primary coordinates on *V*, and the *front/back* terms involving rotated secondary coordinates on *G*.
52. Environmental clues will not explain how some people can exercise such heightened dead reckoning abilities outside familiar territory. I presume that such people have been socialized to constantly compute direction as a background task, by inertial navigation with constant checks with visual information and other sensory information (e.g., sensing wind direction). But see Baker (1989), who believes in faint human magnetoreception.
53. Note that none of these environmental gradients can provide the cognitive basis of abstracted systems. Once the community has fixed a direction, it remains in that direction regardless of fluctuations in local landfall, drainage, wind source, equinox, and so on, or even removal of the subject from the local environment. Thus the environmental sources of such systems may explain their origins but do not generally explain how they are used, or how the cardinal directions are psychologically "fixed."
54. Our current polar system is due no doubt to the introduction of the compass in medieval times. Before, maps typically had east at the top, hence the expression "orient oneself," showing that our use of polar coordinates is older than the compass.
55. Warlpiri may be a case in point. Although such a system may be based on a solar compass, solstitial variation makes it necessary to abstract an equinoctial bisection of the seasonal movement of the sun along the horizon; it is therefore less confusing to fix the system by reference to a mentally constituted orthogonal.
56. Guugu Yimithirr would be a case in points because there are no elicitable associations of sequence or priority between cardinal directions.
57. See Peter Sutton's (1992) description of the Wik Mungan system (another Aboriginal language of Cape York).
58. I am grateful to David Wilkins, and other colleagues, for helping me to systematize these observations.
59. Table 4.4 owes much to the work of Eve Danziger (see especially Danziger 1994).
60. See Van Cleve and Frederick 1991 for discussion of this Kantian point. For the cross-cultural implications and a working out of the place of absolute systems in all this, see Danziger 1994.

61. First discussed in Locke, *Essay on Human Understanding* (book 2, ix, 8), Molyneux's question was brought back into philosophical discussion by Gareth Evans (1985: Ch. 13), and many of the papers in Eilan, McCarthy, and Brewer 1993 explicitly address it.

62. See, for example, Ettlinger 1987, 174: "language serves as a cross-modal bridge"; Dennett 1991, 194–199.

63. The issue may be less clear than it at first seems; see Tye 1991, 5–9.

64. The possibility of getting from a relative representation to an intrinsic one may help to explain the apparent inconsistency between our findings here and Levelt's (chapter 3, this volume). In Levelt's task, subjects who made ellipses always presupposed an underlying uniform spatial frame of reference, even when their spatial descriptions varied between relative and intrinsic, thus suggesting that frames of reference might reside in the mapping from spatial representation to language rather than in the spatial representation itself. But, as Levelt acknowledges, the data are compatible with an analysis whereby the spatial representation is itself in a relative frame of reference and the mapping is optionally to an intrinsic or relative description. The mapping from relative to intrinsic is one of the two mappings, in principle possible between frames of reference, as here described, whereas a mapping from intrinsic spatial representation to linguistic relative representation would be in principle impossible. This would seem to explain all the data that we currently have in hand.

References

- Acredolo, L. (1988). Infant mobility and spatial development. In J. Stiles-Davis, M. Kritchewsky, and U. Bellugi (Eds.), *Spatial cognition: Brain bases and development*, 157–166. Hillsdale, NJ: Erlbaum.
- Asch, S. E., and Witkin, H. A. (1948). Studies in space orientation 2. Perception of the upright with displaced visual fields and with body tilted. *Journal of Experimental Psychology*, 38, 455–477. Reprinted in *Journal of Experimental Psychology, General*, 121 (1992, 4), 407–418.
- Baayen, H., and Danziger, E. (Eds.). (1994). *Annual Report of the Max Planck Institute for Psycholinguistics, 1993*. Nijmegen.
- Baker, M. (1989). *Human navigation and magnetoreception*. Manchester: University of Manchester Press.
- Berthoz, A. (1991). Reference frames for the perception and control of movement. In J. Paillard (Ed.), *Brain and space*, 81–111. Oxford: Oxford Science.
- Bickel, B. (1994). Spatial operations in deixis, cognition, and culture: Where to orient oneself in Belhare. Working paper no. 28, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Bierwisch, M. (1967). Some semantic universals of German adjectivals. *Foundations of Language*, 3, 1–36.
- Bowerman, M., and Pederson, E. (1992). Cross-linguistic perspectives on topological spatial relations. Talk given at the American Anthropological Association, San Francisco, December.

- Brewer, B., and Pears, J. (1993). Frames of reference. In N. Eilan, R. McCarthy, and B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology*, 25–30. Oxford: Blackwell.
- Brown, P. (1991). Spatial conceptualization in Tzeltal. Working paper no. 6, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Brown, P., and Levinson, S. C. (1993a). “Uphill” and “downhill” in Tzeltal. *Journal of Linguistic Anthropology*, 3(1), 46–74.
- Brown, P., and Levinson, S. C. (1993b). Explorations in Mayan cognition. Working paper no. 24, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Bühler, K. (1934). The deictic field of language and deictic words. Reprinted in R. Jarvella and W. Klein (Eds.), *Speech, place and action*, 9–30. New York: Wiley, 1982.
- Bülthoff, H. H. (1991). Shape from X: Psychophysics and computation. In M. S. Landy and J. A. Movshon (Eds.), *Computational models of visual processing*, 305–330. Cambridge, MA: MIT Press.
- Campbell, J. (1993). The role of physical objects in spatial thinking. In N. Eilan, R. McCarthy, and B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology*, 65–95. Oxford: Blackwell.
- Carlson-Radvansky, L. A., and Irwin, D. A. (1993). Frames of reference in vision and language: Where is above? *Cognition*, 46, 223–244.
- Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*, 28–64. New York: Academic Press.
- Cohen, D., and Kubovy, M. (1993). Mental rotation, mental representation, and flat slopes. *Cognitive Psychology*, 25, 351–382.
- Danziger, E. (Ed.). (1993). Cognition and space kit version 1.0. Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Danziger, E. (1994). As fresh meat loves salt: The logic of possessive relationships in Mopan Maya. Working paper no. 30, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown.
- Eilan, N. (1993). Molyneux's question and the idea of an external world. In N. Eilan, R. McCarthy, and B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology*, 236–255. Oxford: Blackwell.
- Eilan, N., McCarthy, R., and Brewer, B. (1993). *Spatial representation: Problems in philosophy and psychology*. Oxford: Blackwell.
- Einstein, A. (1954). Introduction to M. Jammer, *Concepts of space: The history of theories of space in physics*. Cambridge, MA: Harvard University Press.
- Ettlinger, G. (1987). Cross-model sensory integration. In R. Gregory (Ed.), *The Oxford companion to the mind*, 173–174. Oxford: Oxford University Press.

- Evans, G. (1985). *Collected papers*. Oxford: Clarendon Press.
- Fillmore, C. (1971). Toward a theory of deixis. Paper presented at Pacific Conference on Contrastive Linguistics and Language Universals, University of Hawaii, Honolulu, January.
- Friederici, A., and Levelt, W. J. M. (1990). Spatial reference in weightlessness: Perceptual factors and mental representations. *Perception and Psychophysics*, 47(3), 253–266.
- Gregory, R. L. (1987). *Oxford companion to the mind*. Oxford: Oxford University Press.
- Haviland, J. B. (1993). Anchoring and iconicity in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology*, 3(1), 3–45.
- Herrmann, T. (1990). Vor, hinter, rechts, und links: Das 6H-Modell. *Zeitschrift für Literaturwissenschaft und Linguistik*, 78, 117–140.
- Herskovits, A. (1986). Language and spatial cognition: An interdisciplinary study of the prepositions in English. In *Studies in natural language processing*, 208 p. Cambridge: Cambridge University Press.
- Hill, C. (1982). Up/down, front/back, left/right: A contrastive study of Hausa and English. In J. Weissenborn and W. Klein (Eds.), *Here and there: Crosslinguistic studies on deixis and demonstration*, 11–42. Amsterdam: Benjamins.
- Hill, D. (1994). Spatial configurations and evidential propositions. Working paper no. 25, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203, 89–96.
- Howard, I. P. (1987). Spatial coordination of the senses. In R. L. Gregory (Ed.), *The Oxford companion to the mind*, 727–732. Oxford: Oxford University Press.
- Jackendoff, R. (1991). Parts and boundaries. *Cognition*, 41, 9–45.
- Jammer, M. (1954). *Concepts of space: The history of theories of space in physics*. Cambridge, MA: Harvard University Press.
- Johnston, J. R., and Slobin, D. (1978). The development of locative expressions in English, Italian, Serbo-Croatian, and Turkish. *Journal of Child Language*, 6, 529–545.
- Just, M., and Carpenter, P. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, 92(2), 137–172.
- Kant, E. (1768). Von dem ersten Grunde des Unterschiedes der Gegenden im Raume. Translated as On the first ground of the distinction of regions in space in J. Van Cleve and R. E. Frederick (Eds.) *The philosophy of right and left: Incongruent counterparts and the nature of space*, 27–34. Dordrecht: Kluwer, 1991.
- Klatsky, R. L., and Lederman, S. J. (1993). Spatial and nonspatial avenues to object recognition by the human haptic system. In N. Eilan, R. McCarthy and B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology*, 191–205. Oxford: Blackwell.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

- Landau, B., and Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–265.
- Laughlin, R. (1975). *The great Tzotzil dictionary of San Lorenzo Zinacantan*. Washington, DC: Smithsonian.
- Leech, G. (1969). *Towards a semantic description of English*. London: Longmans.
- Levelt, W. J. M. (1984). Some perceptual limitations on talking about space. In A. J. van Doorn, W. A. van der Grind, and J. J. Koenderink (Eds.), *Limits in perception*, 323–358. Utrecht: VNU Science Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, S. C. (1992a). Primer for the field investigation of spatial description and conception. *Pragmatics*, 2(1), 5–47.
- Levinson, S. C. (1992b). Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. Working paper no. 13, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Levinson, S. C. (1994). Vision, shape, and linguistic description: Tzeltal body-part terminology and object description. Special volume of *Linguistics*, 32(4), 791–856.
- Levinson, S. C., and Brown, P. (1994). Immanuel Kant among the Tenejapans: Anthropology as applied philosophy. *Ethos*, 22(1), 3–41.
- Lewis, D. (1976). Route finding by desert aborigines in Australia. *Journal of Navigation*, 29, 21–38.
- Lyons, J. (1977). *Semantics*. Vols. 1 and 2. Cambridge: Cambridge University Press.
- MacLaury, R. (1989). Zapotec body-part locatives: Prototypes and metaphoric extensions. *International Journal of American Linguistics*, 55(2), 119–154.
- Marr, D. (1982). *Vision*. New York: Freeman.
- McCullough, K. E. (1993). Spatial information and cohesion in the gesticulation of English and Chinese speakers. Paper presented at the Annual Convention of the American Psychological Society.
- McNaughton, B., Chen, L., and Markus, E. 1990. "Dead reckoning," landmark learning, and the sense of direction: A neurophysiological and computational hypothesis. *Journal of Cognitive Neuroscience*, 3(2), 191–202.
- Meltzoff, A. N. (1993). Molyneux's babies: Cross-modal perception, imitation, and the mind of the preverbal infant. In N. Eilan, R. McCarthy, and B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology*, 219–235. Oxford: Blackwell.
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- O'Keefe, J. (1993). Kant and the sea-horse: An essay in the neurophilosophy of space. In N. Eilan, R. McCarthy, and B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology*, 43–64. Oxford: Blackwell.

- O'Keefe, J., and Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Paillard, J. (Ed.). (1991). *Brain and space*. Oxford: Oxford Science.
- Pederson, E. (1993). Geographic and manipulable space in two Tamil linguistic systems. In A. U. Frank and I. Campari (Eds.), *Spatial information theory*, 294–311. Berlin: Springer.
- Pederson, E. (1995). Language as context, language as means: Spatial cognition and habitual language use. *Cognitive Linguistics*, 6(1), 33–62.
- Piaget, J., and Inhelder, B. (1956). *The child's conception of space*. London: Routledge and Kegan Paul.
- Pick, H. L., Jr. (1988). Perceptual aspects of spatial cognitive development. In J. Stiles-Davis, M. Kritchewsky, and U. Bellugi (Eds.), *Spatial cognition: Brain bases and development*, 145–156. Hillsdale, NJ: Erlbaum.
- Pick, H. L., Jr. (1993). Organization of spatial knowledge in children. In N. Eilan, R. McCarthy, and B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology*, 31–42. Oxford: Blackwell.
- Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.
- Rock, I. (1990). The frame of reference. In I. Rock (Ed.), *The legacy of Solomon Asch*, 243–268. Hillsdale, NJ: Erlbaum.
- Rock, I. (1992). Comment on Asch and Witkin's "Studies in space orientation. 2." *Journal of Experimental Psychology: General*, 121(4), 404–406.
- Rock, I., Wheeler, D., and Tudor, L. (1989). Can we imagine how objects look from other viewpoints? *Cognitive Psychology*, 21, 185–210.
- Senft, G. (1994). Spatial reference in Kilivila: The Tinkertoy matching games—A case study. *Language and linguistics in Melanesia*, 25, 98–99.
- Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703.
- Sorabji, R. (1988). *Matter, space, and motion: Theories in antiquity and their sequel*. London: Duckworth.
- Stein, J. F. (1992). The representation of egocentric space in the posterior parietal cortex. *Behavioral and Brain Sciences*, 15(4), 691–700.
- Sutton, P. (1992). Cardinal directions in Wik Mungan. Talk given at the 1st Australian Linguistic Institute, Sydney, July.
- Svorou, S. (1994). *The grammar of space*. Amsterdam: Benjamins.
- Takano, Y. (1989). Perception of rotated forms: A theory of information types. *Cognitive Psychology*, 21, 1–59.
- Talmy, L. (1983). How language structures space. In H. Pick and L. Acredolo (Eds.), *Spatial orientation: Theory, research, and application*, 225–282. New York: Plenum Press.

- Tarr, M., and Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology*, 21, 233–282.
- Taylor, H. A., and Tversky, B. (in press). Perspective in spatial descriptions. *Journal of Memory & Language*, 35.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189–208.
- Tversky, B. (1991). Spatial mental models. *Psychology of Learning and Motivation*, 27, 109–145.
- Tye, M. (1991). *The imagery debate: Representation and mind*. Cambridge, MA: MIT Press.
- Valvo, A. (1971). Sight restoration after long-term blindness: The problems and behavior patterns of visual rehabilitation. New York.
- Van Cleve, J., and Frederick, R. E. (Eds.). (1991). *The philosophy of right and left: Incongruent counterparts and the nature of space*. Dordrecht: Kluwer.
- Vandeloise, C. (1991). *Spatial prepositions: A case study from French*. Chicago University of Chicago Press.
- Wilkins, D. (1993). From part to person: Natural tendencies of semantic change and the search for cognates. Working paper no. 23, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.

Chapter 5

The Confluence of Space and Language in Signed Languages

Karen Emmorey

Expressed by hands and face rather than by voice, and perceived by eye rather than by ear, signed languages have evolved in a completely different biological medium from spoken languages. Used primarily by deaf people throughout the world, they have arisen as autonomous languages not derived from spoken language and are passed down from one generation of deaf people to the next (Klima and Bellugi 1979; Wilbur 1987). Deaf children with deaf parents acquire sign language in much the same way that hearing children acquire spoken language (Newport and Meier 1985; Meier 1991). Sign languages are rich and complex linguistic systems that manifest the universal properties found in all human languages (Lillo-Martin 1991).

In this chapter, I will explore a unique aspect of sign languages: the linguistic use of physical space. Because they directly use space to linguistically express spatial locations, object orientation, and point of view, sign languages can provide important insight into the relation between linguistic and spatial representations. Four major topics will be examined: how space functions as part of a linguistic system (American Sign Language) at various grammatical levels; the relative efficiency of signed and spoken languages for overt spatial description tasks; the impact of a visually based linguistic system on performance with nonlinguistic tasks; and finally, aspects of the neurolinguistics of sign language.

5.1 Multifunctionality of Space in Signed Languages

In this section, I describe several linguistic functions of space in American Sign Language (ASL). The list is not exhaustive (for example, I do not discuss the use of space to create discourse frames; see Winston 1995), but the discussion should illustrate how spatial contrasts permeate the linguistic structure of sign language. Although the discussion is limited to ASL, other signed languages are likely to share most of the spatial properties discussed here.



Figure 5.1

Example of a phonological contrast in ASL. These signs differ only in the location of their articulation.

5.1.1 Phonological Contrasts

Spatial distinctions function at the sublexical level in signed languages to indicate phonological contrasts. Sign phonology does not involve sound patternings or vocally based features, but linguists have recently broadened the term *phonology* to mean the “patterning of the formational units of the expression system of a natural language” (Coulter and Anderson 1993, 5). Location is one of the formational units of sign language phonology, claimed to be somewhat analogous to consonants in spoken language (see Sandler 1989). For example, the ASL signs SUMMER, UGLY, and DRY¹ differ only in where they are articulated on the body, as shown in figure 5.1.

At the purely phonological level, the location of a sign is articulatory and does not carry any specific meaning. Where a sign is articulated is stored in the lexicon as part of its phonological representation.² Sign languages differ with respect to the phonotactic constraints they place on possible sign locations or combinations of locations. For example, in ASL no one-handed signs are articulated by contacting the contralateral side of the face (Battison 1978). For all signed languages, whether a sign is made with the right or left hand is not distinctive (left-handers and right-handers produce the same signs—what is distinctive is a contrast between a dominant and nondominant hand). Furthermore, I have found no phonological contrasts in ASL that involve left-right in signing space. That is, there are no phonological minimal pairs that are distinguished solely on the basis of whether the signs are articulated on the right or left side of signing space. Such left-right distinctions appear to be reserved for the referential and topographic functions of space within the discourse structure, syntax, and morphology of ASL (see below). For a recent and comprehensive review of the nature of phonological structure in sign language, see Corina and Sandler (1993).

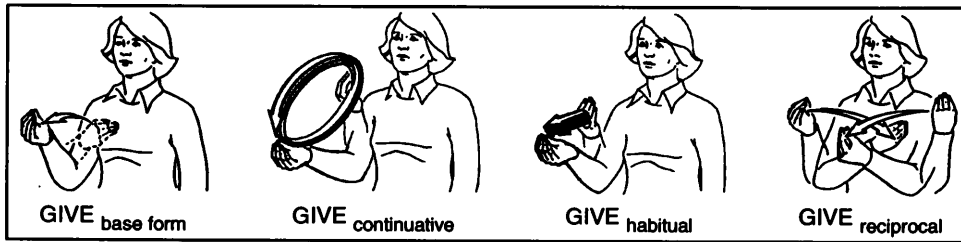


Figure 5.2
Examples of nonconcatenative morphology in ASL.

5.1.2 Morphological Inflection

In many spoken languages, morphologically complex words are formed by adding prefixes or suffixes to a word stem. In ASL and other signed languages, complex forms are most often created by nesting a sign stem within dynamic movement contours and planes in space. Figure 5.2 illustrates the base form GIVE along with several inflected forms. ASL has many verbal inflections that convey temporal information about the action denoted by the verb, for example, whether the action was habitual, iterative, or continual. Generally, these distinctions are marked by different movement patterns overlaid onto a sign stem. This type of morphological encoding contrasts with the primarily linear affixation found in spoken languages. For spoken languages, simultaneous affixation processes such as templatic morphology (e.g., in the Semitic languages), infixation, or reduplication are relatively rare. Signed languages, by contrast, prefer nonconcatenative processes such as reduplication; and prefixation and suffixation are rare. Sign languages' preference for simultaneously producing affixes and stems may have its origin in the visual-manual modality.

For example, the articulators for speech (the tongue, lips, jaw) can move quite rapidly, producing easily perceived distinctions on the order of every 50–200 milliseconds. In contrast, the major articulators for sign (the hands) move relatively slowly such that the duration of an isolated sign is about 1,000 milliseconds; the duration of an average spoken word is more like 500 milliseconds. If language processing in real time has equal timing constraints for spoken and signed languages, then there is strong pressure for signed languages to express more distinctions simultaneously. The articulatory pressures seem to work in concert with the differing capacities of the visual and auditory systems for expressing simultaneous versus sequential information. That is, the visual system is well suited for simultaneously perceiving a large amount of information, whereas the auditory system seems particularly adept at perceiving fast temporal distinctions. Thus both sign and speech have exploited the advantages of their respective modalities.



Figure 5.3

Example of the sentential use of space in ASL. Nominals (*cat*, *dog*) are first associated with spatial loci through indexation. The direction of the movement of the verb (BITE) indicates the grammatical role of subject and object.

5.1.3 Coreference and Anaphora

Another hypothesized universal use of space within sign languages is for referential functions. In ASL and other sign languages, nominals can be associated with locations in signing space. This association can be established by “indexing” or pointing to a location in space after producing a lexical sign, as shown in figure 5.3. Another device for establishing the nominal-locus association is to articulate the nominal sign(s) at a particular location or by eye gaze toward that location. In figure 5.3, the nominal DOG is associated with a spatial locus on the signer’s left and CAT is associated with a locus on the signer’s right. The verb BITE moves between these locations identifying the subject and object of the sentence “[The dog] bites [the cat].” BITE belongs to a subset of ASL verbs termed *agreeing verbs*³ whose movement and/or orientation signal grammatical role. ASL pronouns also make use of established associations between nominals and spatial loci. A pronominal sign directed toward a specific locus refers back to the nominal associated with that locus. Further description of coreference and anaphora in ASL can be found in Lillo-Martin (1991) and Padden (1988).

Recently, there has been some controversy within sign linguistics concerning whether space itself performs a syntactic function in ASL. Liddell (1993, 1994, 1995) has argued that spatial loci are not morphemic. He proposes that space in sentences like those illustrated in figure 5.3 is being used deictically rather than anaphorically. That is, the signer deictically points to a locus in the same way he would point to a physically present person. In contrast, other researchers have argued that these spatial loci are agreement morphemes or clitics that are attached to pronouns and verbs (e.g., Janis 1995; Padden 1990). As evidence for his position, Liddell (1993, 1995) argues that just as there is an unlimited number of spatial positions in which a

physically present referent could be located, there also appears to be an unlimited number of potential locations within signing space (both vertically and horizontally) toward which a verb or pronominal form can be directed (see also Lillo-Martin and Klima 1990). If this is the case, then location specifications are not listable or categorizable and therefore cannot be agreement morphemes or clitics. The syntactic role of subject or object is assigned, not by the spatial loci, but either by word order or by the orientation or the temporal end points of the verb itself.⁴ According to this view, the particular location at which a verb begins or ends serves to identify the *referent* of the subject and object roles. The space itself, Liddell has argued, is not part of a syntactic representation; rather, space is used nonmorphemically and deictically (much as deictic gesture is used when accompanying speech). This hypothesis is quite radical, and many of the details have not been worked out. For example, even if space itself does not perform a syntactic function, it does perform both a referential and a locative function within the language (see Emmorey, Corina, and Bellugi 1995). The association of a nominal with a particular location in space needs to be part of the linguistic representation at some level in order to express coreference relations between a proform and its antecedent. If this association is not part of the linguistic representation, then there must be an extremely intimate mixing of linguistic structure and nonlinguistic representations of space.

5.1.4 Locative Expressions

The spatial positions associated with referents can also convey locative information about the referent. For example, the phrase DOG INDEX_a shown in figure 5.3 could be interpreted as “the dog is there on my left,” but such an interpretation is not required by the grammar. Under the nonlocative reading, INDEX simply establishes a reference relation between DOG and a spatial locus that happens to be on the signer’s left. To ensure a locative reading, signers may add a specific facial expression (e.g., spread tight lips with eye gaze to the locus), produced simultaneously with the INDEX sign. Furthermore, ASL has a set of classifier forms for conveying specific locative information, which can be embedded in locative and motion predicates; for these predicates, signing space is most often interpreted as corresponding to a physical location in real (or imagined) space. The use of space to directly represent spatial relations stands in marked contrast to spoken languages, in which spatial information must be recovered from an acoustic signal that does not map onto the information content in a one-to-one correspondence. In locative expressions in ASL, the identity of each object is provided by a lexical sign (e.g., TABLE, T-V, CHAIR); the location of the objects, their orientation, and their spatial relation vis-à-vis one another are indicated by where the appropriate accompanying classifier sign is articulated in the space in front of the signer. The flat B handshake is

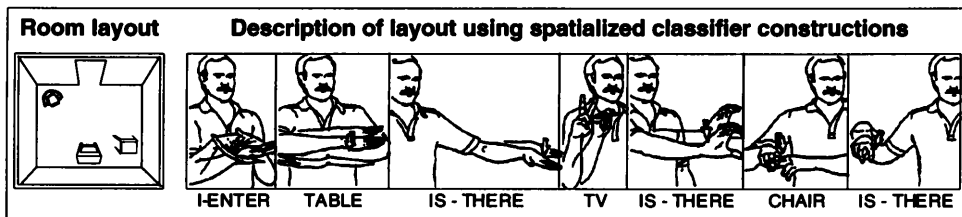


Figure 5.4

Example of an ASL spatial description using classifier constructions.

the classifier handshape for rectangular, flat-topped, surface-prominent objects like tables or sheets of paper. The C handshape is the classifier handshape for bulky boxlike objects like televisions or microwaves. The bent V is the classifier handshape for squat, “legged” objects like chairs, small animals, and seated people.

Flat B handshape: 

C handshape: 

Bent V handshape: 

These handshapes occur in verbs that express the spatial relation of one object to another and the manner and direction of motion (for moving objects/people). Figure 5.4 illustrates an ASL description of the room that is sketched at the far left. An English translation of the ASL description would be “I enter the room; there is a table to my left, a TV on the far side, and a chair to my right.” Where English uses separate words to express such spatial relations, ASL uses the actual visual layout displayed by the array of classifier signs to express the spatial relations of the objects.

Landau and Jackendoff (1993) have recently argued that languages universally encode very little information about object shape in their locative closed-class vocabulary (e.g., prepositions) compared to the amount of spatial detail they encode in object names (see also Landau, chapter 8, this volume). As one can surmise from our discussion and from figure 5.4, ASL appears to have a rich representation of shape in its locative expressions. Like the locational predicates in Tzeltal (Brown 1991; Levinson 1992a), ASL verbs of location incorporate detailed information about the shape of objects. It is unclear whether these languages are counterexamples to Landau and Jackendoff’s claims for two reasons. First, both Tzeltal and ASL express locative information through verbal predicates that form an open-class category, unlike prepositions (although the morphemes that make up these verbal predicates belong to a closed class). The distinction may hinge on whether these forms are con-



Figure 5.5
Final classifier configuration of either (2a) or (2b).

sidered grammaticized closed-class elements or not (see also Talmy 1988). Second, in ASL the *degree* of shape detail is less in classifier forms than in object names. For example, the flat B handshape classifier is used for both TABLE and for PAPER—the count nouns encode more detailed shape information about these objects than the classifier form. Thus, although the contrast is much less striking in ASL than in English, it still appears to hold.

Talmy (1983) has proposed several universal features that are associated with the figure object (i.e., the located object) and with the reference object or ground. For example, the figure tends to be smaller and more movable than the ground object. This asymmetry can be seen in the following sentences (from Talmy 1983):⁵

- (1) a. The bike is near the house.
- b. ?The house is near the bike.

In English, the figure occurs first, and the ground is specified by the object of the preposition. When a large unmovable entity such as a house is expressed as the figure, the sentence is semantically odd. This same asymmetry between figure and ground objects occurs in ASL, except that the syntactic order of the figure and ground is reversed compared to English, as shown in (2a) and (2b) (the subscripts indicate locations in space). In these examples, the classifier in the first phrase is held in space (indicated by the extended line) during the articulation of the second phrase (produced with one hand). In this way, the classifier handshape representing the figure can be located with respect to the classifier handshape representing the ground object, as illustrated in figure 5.5 (the signer's left hand shows the classifier form for



Final classifier construction for (3a).



Final classifier construction for (3b).

Figure 5.6

Thus Talmy's generalization about figure-ground complexity appears to hold even for languages that can use spatial geometry itself to encode spatial relations.

5.1.5 Frames of Reference

ASL can express spatial relations using an intrinsic, relative, or absolute frame of reference (see Levinson, chapter 4, this volume, for discussion of the linguistic and spatial properties of these reference frames).⁸ Within a relative frame of reference, scenes are most often described from the perspective of the person who is signing. In this case, the origin of the coordinate system is the viewpoint of the signer. For example, eight ASL signers were asked to describe the picture shown in figure 5.7. All but one indicated that the bowl was on their left with the banana on their right (one signer provided a description of the scene without using signing space in a topographic way, producing the neutral phrase ON S-I-D-E instead). To indicate that the banana was on their right, signers produced the classifier form for bowl on the left side of signing space, and then a classifier form for banana was simultaneously articulated on the right.

Descriptions from the addressee's viewpoint⁹ turn out to be more likely in the front-back dimension than in the left-right dimension (the signer's perspective is still the most likely for both dimensions). In describing the picture shown in figure 5.8, five of eight signers preferred their own viewpoint and produced the classifier for banana near the chest with the classifier for bowl articulated away from the chest

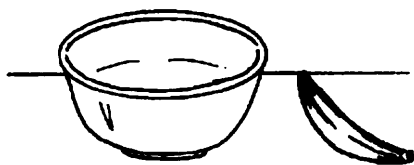


Figure 5.7

Illustration of one of the pictures that signers were asked to describe.



a. Signer's viewpoint (5/8 signers).

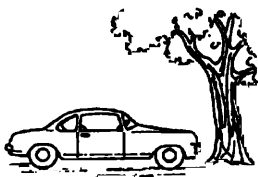


b. Addressee's viewpoint (3/8 signers).

Figure 5.8

behind the classifier for banana, as shown in figure 5.8a. This spatial configuration of classifier signs maps directly onto the view presented in figure 5.8 (remember that you as the reader are facing both the signer and the picture). In contrast, three signers described the picture from the addressee's viewpoint, producing the classifier for bowl near the chest and the classifier for banana in line with the bowl but further out in signing space, as shown in figure 5.8b. This configuration would be the spatial arrangement seen by an addressee standing opposite the signer (as you the reader are doing when viewing these figures). There were no overt linguistic cues that indicated which point of view the signer was adopting. However, signers were very consistent in what point of view they adopted. For example, when the signers were shown the reverse of figure 5.8, in which the banana is behind the bowl, all signers reversed their descriptions according to the viewpoint they had selected previously. Note that the lack of an overt marker of point of view, the potential ambiguity, and the consistency within an adopted point of view also occur in English and other spoken languages (see Levelt 1984).

Bananas and bowls do not have intrinsic front/back features, and thus signers could not use an intrinsic frame of reference to describe these pictures. In contrast, cars do have these intrinsic properties, and the classifier form for vehicles encodes intrinsic features: the front of the car is represented roughly by the tips of the index and middle fingers, which are extended. Figures 5.9 and 5.10 illustrate ASL constructions using the vehicle classifier, along with the corresponding pictures of a car in different locations with respect to a tree. Again the majority of signers expressed their own view of the picture. In figures 5.9 and 5.10, the pictured female signer adopts her own perspective (describing the picture as she sees it), while the male signer adopts the addressee's viewpoint. As noted above, lexical signs identifying the referents of the classifier signs are given first. Also as noted, the ground object (the tree) is expressed first and generally held in space while the lexical sign for car is articulated and the vehicle classifier is placed with respect to the classifier for tree. The illustrations in figures 5.9 and 5.10 represent the final classifier construction in the description. As you can see, signers orient the vehicle classifier to indicate the direction the car is facing. Note that the orientation of the car is consistent with the point of view adopted—the vehicle classifier is always oriented toward the tree.¹⁰ The majority of signers described figure 5.9 by placing the vehicle classifier to their left in signing space. Only one signer placed the car on his right and the tree on his left. Again all signers were very consistent in which point of view they adopted, although one signer switched from her own viewpoint in describing figure 5.9 to the addressee's viewpoint for figure 5.10. There were no switches in viewpoint within either the left-right or front-back dimension. Signers were also consistent within the intrinsic frame of



a. Signer's viewpoint (6/7 signers).



b. Addressee's viewpoint (1/7 signers).



a. Signer's viewpoint (5/7 signers).



b. Addressee's viewpoint (2/7 signers).

Figures 5.9 and 5.10

reference, almost always changing the orientation of the vehicle classifier appropriately (e.g., toward the left/right or away from/facing the signer).¹¹

One question of interest is whether signers can escape the relative point of view that is imposed “automatically” by the fact that signers (and addressees) view their own articulators in space and these articulators express locative relations using this space. The answer appears to be that a relative framework is not necessarily entailed in locative expressions in ASL. That is, the expressions shown in figure 5.9a and 5.9b could be interpreted as the rough equivalent of “the tree is in front of the car” without reference to the signer’s (or addressee’s) viewpoint. The car could actually be in any left-right or front-back relation with respect to the signer—what is critical to the intrinsic expression is that the vehicle classifier is oriented toward (facing) the tree. Thus the intrinsic frame of reference is not dependent upon the relative frame; in ASL these two frames of reference can be expressed simultaneously. That is, linguistic expression within an intrinsic frame occurs via the intrinsic properties of certain classifier forms, and a relative frame can be imposed simultaneously on signing space if a viewpoint is adopted by the signer. Figures 5.9 and 5.10 illustrate such simultaneous expression of reference frames. The linguistic and nonlinguistic factors that influence choice of viewpoint within a relative reference frame have not been determined, although it is likely that several different linguistic and nonlinguistic factors are involved. And just as in English (Levelt 1982a, 1984), frame of reference ambiguities can abound in ASL; further research will determine how addressee and signer viewpoints are established, altered, and disambiguated during discourse. Preliminary evidence suggests that, like English speakers (Schober 1993), “solo” ASL signers (such as those in this study) are less explicit about spatial perspective than signers with conversation partners.

Finally, ASL signers can use an absolute reference frame by referring to the cardinal points east, west, north, and south. The signs for these directions are articulated as follows: WEST: W handshape, palm in, hand moves toward left¹²; EAST: E handshape, palm out, hand moves toward right; NORTH: N handshape, hand moves up; SOUTH: S handshape, hand moves down.

N handshape: 

E handshape: 

S handshape: 

W handshape: 

These signs are articulated in this manner, regardless of where the person is standing, that is, regardless of true west or north. This situation contrasts sharply with how speakers gesture in cultures which employ absolute systems of reference such as

certain Aboriginal cultures in Australia (see Levinson 1992b and chapter 4, this volume). In these cultures, directional gestures are articulated toward cardinal points and vary depending upon where the speaker is oriented.

Although the direction of the citation forms of ASL cardinal signs is fixed, the movement of these signs can be changed to label directions within a “map” created in signing space. For example, the following directions were elicited from two signers describing the layout of a town shown on a map (from Taylor and Tversky 1992):

(4) YOU DRIVE	STRAIGHT right hand traces a path outward from the signer	EAST “e” handshape traces the same path, palm to left

“You drive straight eastward.”

(5) UNDERSTAND MOUNTAIN R-D	PATH right hand traces path toward left, near signer	NORTH “n” hand shape traces same path, palm in

“Understand that Mountain Road goes north in this direction.”

The signer who uttered (5) then shifted the map, such that north was centered outward from the signer, and the sign NORTH¹³ then traced a path similar to the one in (4), that is, centered and outward from the signer. It appears that ASL direction signs are either fixed with respect to the body in their citation form or they are used relative to the space mapped out in front of the signer. As in English, it is the direction words themselves that pick out an absolute framework within which the discourse must be interpreted.

5.1.6 Narrative Perspective

In a narrative, a spatial frame of reference can be associated with a particular character (see discussions of viewpoint in Franklin, Tversky, and Coon 1992; and Tversky, chapter 12, this volume). The frame of reference is relative, and the origin of the coordinate system is the viewpoint of that character in the story. The linguistic mechanisms used to express point of view in signed languages appear to be more explicit than in spoken languages. Both signers and speakers use linguistic devices to indicate whether utterances should be understood as expressing the point of view of the signer/speaker or of another person. Within narrative, “point of view” can mean either a visual perspective or the nonspatial perspective of a character, namely, that character’s thoughts, words, or feelings. Spoken languages have several different

devices for expressing either type of perspective: pronominal deixis (e.g., use of *I* vs. *you*), demonstratives (*here*, *there*), syntactic structure (active vs. passive), and literary styles (e.g., “free indirect” discourse). Signed languages use these mechanisms as well, but in addition, point of view (in either sense) can be marked overtly (and often continuously) by a “referential shift.” Referential shift is expressed by a slight shift in body position and/or changes in eye gaze, head position, or facial expression (for discussions of this complex phenomenon, see Loew 1983; Engberg-Pedersen 1993; Padden 1986; Lillo-Martin 1995; Poulin and Miller 1995).

The following is an example of a referential shift that would require overt marking of a spatial viewpoint. Suppose a signer were telling a story in which a boy and a girl were facing each other, and to the left of the boy was a tall tree. If the signer wanted to indicate that the boy looked up at the tree, he or she could signal a referential shift, indicating that the following sentence(s) should be understood from the perspective of the boy. To do this, the signer would produce the sign LOOK-AT upward and to the left. If the signer then wanted to shift to the perspective of the girl, he or she would produce the sign LOOK-AT and direct it upward and to the *right*. Signers often express not only a character’s attitudinal perspective, but also that character’s spatial viewpoint through signs marked for location and/or deixis. Slobin and Hoiting (1994, p. 14) have noted that “directional deixis plays a key role in signed languages, in that a path verb moves not only with respect to source and goal, but also with respect to sender and receiver, as well as with respect to points that may be established in signing space to indicate the locations and viewpoints of protagonists set up in the discourse.” That spoken languages express deixis and path through separate elements (either through two verbs or through a satellite expression and a verb) reflects, they suggest, an inherent limitation of spoken languages. That is, spoken language must linearize deictic and path information, rather than express this information simultaneously, as is easily done in signed languages. Deixis is easily expressed in signed languages because words are articulated in the space surrounding the signer, such that “toward” and “away from” can be encoded simply by the direction of motion with respect to the signer or a referential locus in space. I would further hypothesize that this simultaneous expression of deictic and other locative information within the verbs of signed languages may lead to habitual expression of spatial viewpoint within discourse.

In sum, signed languages use space in several different linguistic domains, including phonological contrast, coreference, and locatives. The visual-gestural modality of signed languages appears to influence the nature of grammatical encoding by compelling signed languages to prefer nonconcatenative morphological processes (see also Emmorey 1995; Supalla 1991; Gee and Goodhart 1988). Signed languages offer important insight into how different frames of reference are specified linguistically. A

unique aspect of the visual-gestural modality may be that intrinsic and relative reference frames can be simultaneously adopted. In addition, shifts in reference are often accompanied by shifts in visual perspective that must be overtly marked on deictic and locative verbs. Although spoken languages also have mechanisms to express deictic and locative relations, what is unique about signed languages is that such relations are directly encoded in space.

5.2 Some Ramifications of the Direct Representation of Space

In the studies reported below, I explore some possible ramifications of the spatial encoding of locative and spatial contrasts for producing spatial descriptions and solving spatial problems. Specifically, I investigate (1) how ASL signers use space to express spatial commands and directions, (2) to what extent signers use lexicalized locatives in spatial directions, (3) whether the use of sign language provides an advantage for certain spatial tasks, and (4) how differences in linguistic encoding between English and ASL affect the nature of spatial commands and directions.

5.2.1 Solving Spatial Puzzles with Spatialized Language

To investigate these questions, ten hearing English speakers and ten deaf ASL native signers were compared using a task in which they had to solve three spatial puzzles by instructing an experimenter¹⁴ where to place blocks of different colors, shapes, and sizes onto a puzzle grid (see figure 5.11). To solve the problem, all blocks must fit within the puzzle outline. The data from English speakers were collected by Mark St. John (1992), and a similar but not identical protocol was used with ASL signers.

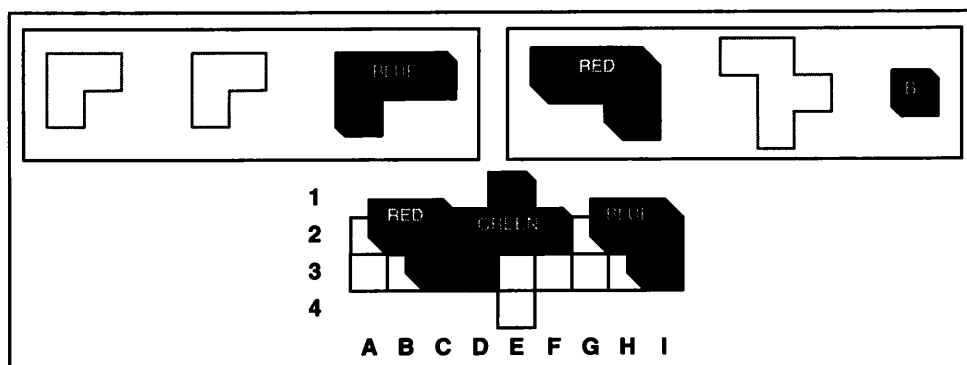


Figure 5.11

Solving a spatial puzzle: Subjects describe how to place blocks on a puzzle grid.

English speakers were instructed to sit on their hands and were not permitted to point to the puzzle or to the pieces. Of course, ASL signers could use their hands, but they were also not permitted to point to the pieces or puzzle. For both signers and speakers, the subject and experimenter sat side by side, such that each had the same visual perspective on the puzzle board.

To explore how speakers and signers use spatial language—encoded in either space or sound—we examined different types of English and ASL instructions. We hypothesized that ASL signers may be able to use signing space as a rough Cartesian coordinate system, and therefore would rely less on the coordinates labeled on the puzzle board. This prediction was confirmed: 67% of the English speakers' commands referred to the puzzle grid, whereas only 28% of the commands given by ASL signers referred to the puzzle coordinates. This difference in grid reference was statistically reliable ($F(1,18) = 9.65$; $p < .01$). The following are sample commands containing references to the puzzle grid given by English speakers:

- (6) Take the blue L piece and put it on H1 H2 G2.
- (7) Place the red block in 3G H 2G.
- (8) Green piece on E1, E2, D2, C2, and D3.

Instead of referring to grid coordinates, ASL signers used space in various ways to indicate the positions on the puzzle board—for example, by tracing a distinctive part of the board in space or by holding the nondominant hand in space, representing a part of the puzzle board (often an edge).

We also compared how signers and speakers identified the puzzle pieces to be placed for a given command (see figure 5.12a). There were no significant differences in how either ASL or English was used to label a particular block. We had hypothesized that signers might make more references to shape because shape is often encoded in classifier handshapes (see discussion above). However, the numerical difference seen in figure 5.12a was not statistically significant. Language did not appear to influence how subjects labeled the puzzle pieces within this task.

There were significant differences, however, in the types of commands used by ASL signers and English speakers (see figure 5.12b). Puzzle commands could be exhaustively divided into three categories: (1) commands referring to a *position* on the puzzle board, (2) commands expressing a *relation* between two pieces, and (3) the *orientation* of a single piece. These categories were able to account for all of the commands given by the twenty subjects. The only difference was that in ASL, two command types could be expressed simultaneously. For example, signers could simultaneously describe the orientation of a piece (through the orientation of a classifier handshape) and that piece's relation to another block through two-handed

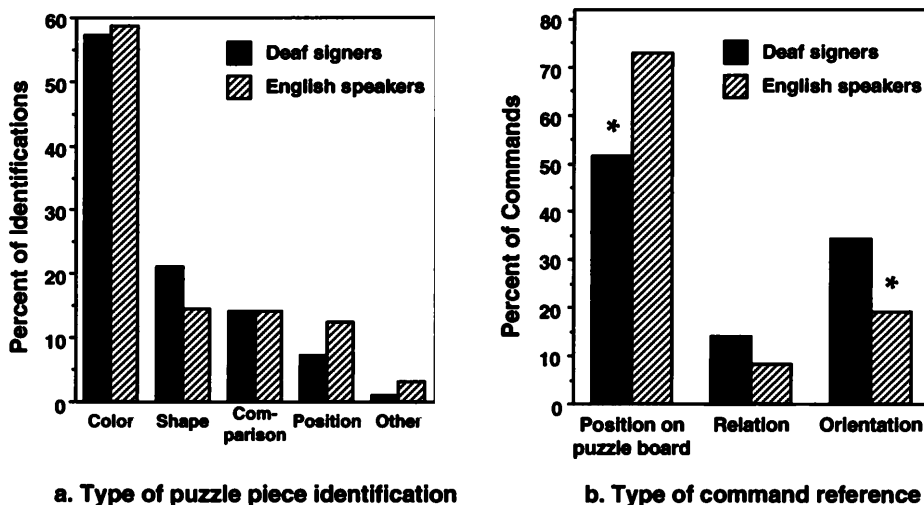


Figure 5.12

classifier constructions (see figure 5.15, as well as the constructions illustrated in figures 5.5, 5.9, and 5.10).

English speakers produced significantly more commands referring to a position on the puzzle board compared to ASL signers ($F(1,18) = 4.47$; $p < .05$). English speakers' reliance on commands involving coordinate specifications (see examples 6–8) appears to account for this difference in command type. It is interesting to note that even when ASL signers referred to grid coordinates, they often specified these coordinates within a vertical spatial plane, signing the letter coordinates moving crosswise and the number coordinates moving downward. Thus the true horizontal plane of the board laying on the tabletop was "reoriented" into a vertical plane within signing space, as if the puzzle board were set upright. The linguistic and pragmatic constraints on using a vertical versus horizontal plane to represent spatial layouts are yet to be determined, but clearly use of a vertical plane does not necessarily indicate a true vertical relation between objects.

Subjects did not differ significantly in the percentage of commands that referred to the relation of one piece to another. Examples of English relation commands are given in (9)–(11):

(9) Put the other blue L next to the green one.

(10) Put it to the left of the green piece.

(11) Switch the red and the blue blocks.

ASL signers also produced these types of commands, but generally space, rather than prepositional phrases, conveyed the relation between pieces. For example, the non-dominant hand can represent one block, and the dominant hand either points to a spatial locus to the left or right (somewhat like the construction illustrated in figure 5.6a) or the dominant hand represents another block and is positioned with respect to the nondominant hand (see figure 5.15).

Finally, ASL signers produced significantly more commands that referred to the orientation of a puzzle piece ($F(1,18) = 5.24$; $p < .05$). Examples from English of commands referring to orientation are given in (12)–(14):

(12) Turn the red one counterclockwise.

(13) Rotate it 90 degrees.

(14) Flip it back the other way.

For English speakers, a change in orientation was often inferred from where the piece had to fit on the board, given other non-orientation-specific commands. In contrast, ASL signers often overtly specified orientation. For example, figure 5.13 illustrates an ASL command that indicates a change in orientation by tracing a block's ultimate orientation in signing space (the vertical plane was often used to trace shape and orientation). Figure 5.14 illustrates a command in which orientation change is specified by a change in the orientation of the classifier handshape itself. Figure 5.15 illustrates the simultaneous production of a command indicating the

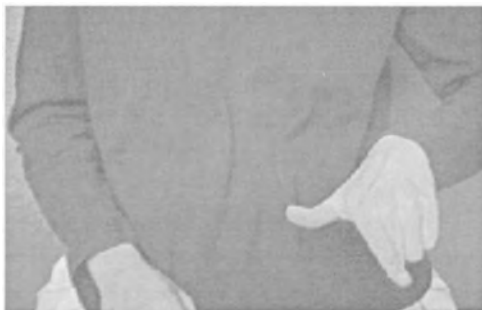


[pictured]

Figure 5.13 GREEN CL:G _____
CL:G-orientation

"Orient the green block in this way." See green block in figure 5.11; note signer's perspective.

Figure 5.13



[pictured]

Figure 5.14 BLUE L CL:L-orientation

"Move the blue L so it is oriented with the long end outward."



[pictured]

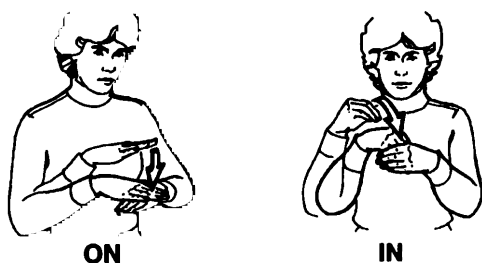
Figure 5.15 RED L CL:B _____
CL:L-orientation

"Move the red L so it is oriented lengthwise at the top of another block [the green block]."

Figures 5.14 and 5.15

orientation of an L-shaped piece and its relation to another piece. Signers also used the sign ROTATE quite often and indicated the direction of rotation by movement of the wrist (clockwise vs. counterclockwise).

ASL also has a set of lexicalized locative signs that are used much less frequently than classifier constructions in spatial descriptions. The lexicalized locatives that were produced by signers in this study included IN, ON, AGAINST, NEAR, and BETWEEN. Only about 20% of ASL commands involved lexical locatives, and these were almost always produced in conjunction with commands involving classifier constructions. The grammatical structure of these forms is not well understood—are they adpositions (see McIntire 1980) or verbs (see Shepard-Kegl 1985)?—and their

**Figure 5.16**

ASL lexicalized locative signs. Illustration by Frank Allen Paul in Newell (1983).

semantics has not been well studied either (see McIntire 1980 for some discussion of IN, UNDER, and OUT). The linguistic data from our study provided some interesting insight into the semantics of IN and ON (these signs are shown in figure 5.16).

English speakers used the prepositions *in* and *on* interchangeably to specify grid coordinates, for example, “in G2 H2” or “on G2 H2” (see sample commands 6 and 7 above). ASL signers used the lexical locative ON in this context, but never IN:

(15) PUT RED L ON G2 H2 I2 I3

(16) PUT BLUE [CL:G—*shape*]¹⁵ ON 3E 4F 3F 3G
 shape traced in
 vertical plane

(17) *PUT RED L IN G2 H2

The use of the preposition *in* for describing grid positions on the puzzle board falls under Herskovitz’s (1986) category “spatial entity in area,” namely, “the reference object must be one of several areas arising from a dividing surface” (p. 153). This particular semantic structure does not appear to be available for the ASL sign IN. Signers did use IN when aspects of the puzzle could be construed as container-like (falling under Herskovitz’s “spatial entity in a container”). For example, signers would direct pieces to be placed IN CORNER;¹⁶ in this case, two lines meet to form a type of container (see Herskovitz 1986, 149). IN was also used when a block (most often the small blue square) was placed in a “hole” created by other blocks on the board or when a part of a block was inserted into the part of the puzzle grid that stuck out (see figure 5.11). In both cases, the reference object forms a type of container into which a block could be placed. The use of the ASL lexical locative IN appears to be more restricted than English *in*, applying only when there is a clear containment relation.

horizontal planes of space itself as a rough Cartesian coordinate system. Changes in object orientation were expressed directly through changes in the spatial position of classifiers and by tracing shape and orientation in signing space. In contrast, English speakers were less likely to overtly express changes in orientation and relied heavily on direct reference to labels for coordinate positions. The heart of this different use of spatial language appears to lie in the properties of the aural-vocal and visual-manual linguistic modalities. For example, in ASL, the hands can directly express orientation by their own orientation in space—such direct representation within the linguistic signal is not available to English speakers. Finally, ASL and English differ in the semantics they assign to lexicalized locatives for the topological concepts *in* and *on*, and the semantic structure of the ASL locatives cannot be extracted from the iconic properties of the forms. In the following study, we further explore the effect modality may exert on the nature of spatial language for both spoken and signed language.

5.2.2 Room Description Study

Eight ASL signers and eight English speakers were asked to describe the layout of objects in a room to another person (“the manipulator”) who had to place the objects (pieces of furniture) in a dollhouse.¹⁷ In order to elicit very specific instructions and to eliminate (or vastly reduce) interchanges, feedback, and interruptions, “the describer” (the person giving the instructions) could not see the manipulator, but the manipulator could see the describer through a one-way mirror (see figure 5.17).

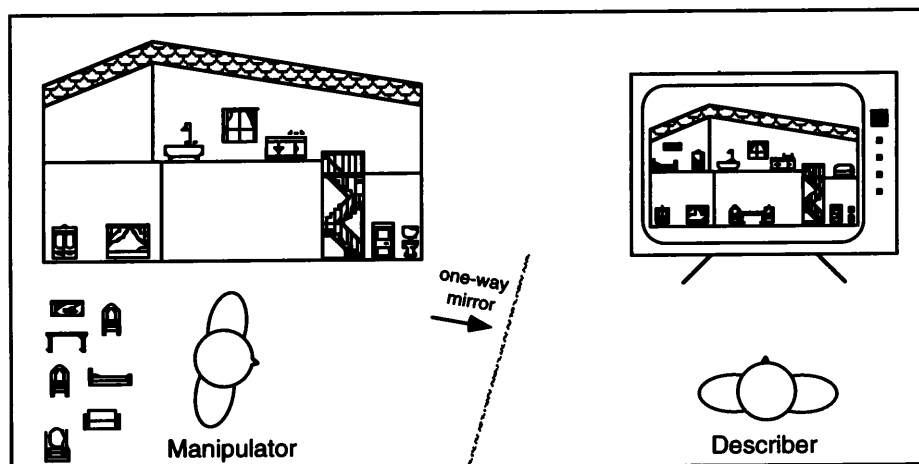


Figure 5.17
Experimental set-up for room descriptions.

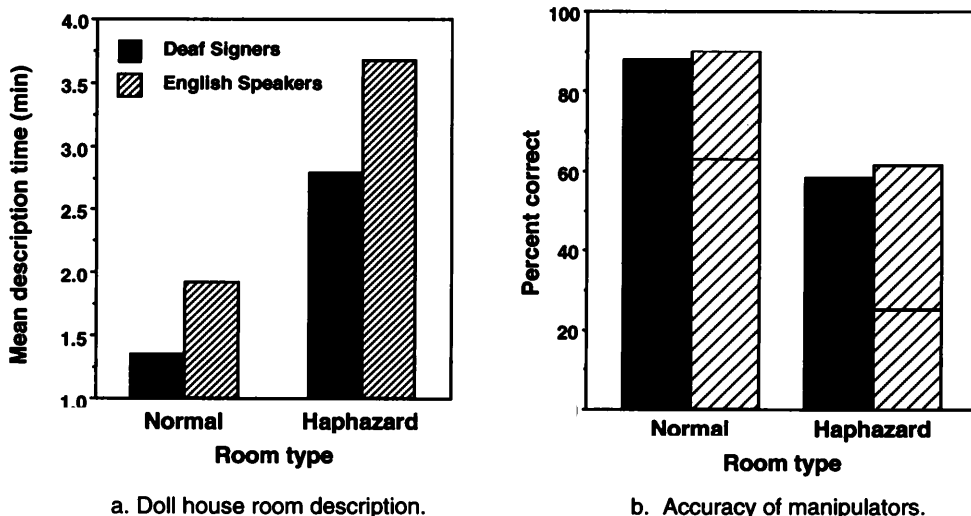


Figure 5.18

The manipulator could not ask questions but could request that the describer pause or produce a summary. Subjects described six rooms with canonical placements of furniture ("normal rooms") and six rooms in which the furniture had been strewn about haphazardly without regard to function ("haphazard rooms"). The linguistic data and analysis arising from this study are discussed elsewhere (Emmorey, Clothier, and McCullough). However, certain results emerged from the study that illuminate some ramifications of the direct representation of space for signed languages.

Signers were significantly faster than speakers in describing the rooms ($F(1,14) = 5.00$; $p < .05$; see figure 5.18a). Mean description time for ASL signers was 2 min, 4 sec; English speakers required an average of 2 min, 48 sec to describe the same rooms. In one way, the speed of the signers' descriptions is quite striking because, on average, ASL signs take twice as long as English words to articulate (Klima and Bellugi 1979; Emmorey and Corina 1990). However, as we have seen thus far in our discussion of spatial language in ASL, there are several modality-specific factors that would lead to efficient spatial descriptions and lessen the need for discourse linearization (Levelt 1982a,b), at least to some degree. For example, the two hands can represent two objects simultaneously through classifier handshapes, and the orientation of the hands can also simultaneously represent the objects' orientation. The position of the hands in space represents the position of the objects with respect to each other. The simultaneous expression of two objects, their position, and their

orientation stands in contrast to the linear strings of prepositions and adjunct phrases that must be combined to express the same information in English.

The difference in description time was not due to a speed-accuracy trade-off. Signers and speakers produced equally accurate descriptions, as measured by the percent of furniture placed correctly by the manipulators in each group (see figure 5.18b). There was no significant difference in percent correct, regardless of whether a lenient scoring measure was used (object misplaced by more than 3 cm or misoriented by 45 degrees; represented by height of the bars in figure 5.18b) or a strict scoring measure was used (object misplaced by 1 cm or misoriented by 15 degrees; shown by the line in each bar in figure 5.18b).

To summarize, this second study suggests that the spatialization of American Sign Language allows for relatively rapid and efficient expression of spatial relations and locations. In the previous study, we saw that ASL signers and English speakers focused on different aspects of objects within a spatial arrangement, as reflected by differing instructions for the placement of blocks within a coordinate plane. These differences arise, at least in part, from the spatial medium of signed languages, compared to the auditory transmission of spoken languages.

5.3 Interplay between Spatialized Language and Spatial Cognition

We now turn to the relation between general nonlinguistic spatial cognition and processing a visual-spatial linguistic signal. Does knowing a signed language have any impact on nonlinguistic spatial processing? In a recent investigation, Emmorey, Kosslyn, and Bellugi (1993) examined the relation between processing ASL and the use of visual mental imagery. Specifically, we examined the ability of deaf and hearing subjects to mentally rotate images, to generate mental images, and to maintain images in memory (this last skill will not be discussed here). We hypothesized that these imagery abilities are integral to the production and comprehension of ASL and that their constant use may lead to an enhancement of imagery skills within a non-linguistic domain. In order to distinguish the effects of using ASL from the effects of being deaf from birth, we also tested a group of hearing subjects who were born to deaf parents. These subjects learned ASL as their first language and have continued to use ASL in their daily lives. If these hearing native signers have visual-spatial skills similar to those found for deaf signers, this would suggest that differences in spatial cognition arise from the use of a visual-spatial language. On the other hand, if these signers have visual-spatial skills similar to those found in hearing subjects, this would suggest that differences in spatial cognition may be due to auditory deprivation from birth.

We hypothesized that mental rotation may play a crucial role in sign language processing because of the changes in spatial perspective that can occur during referential shifts in narrative (see above) and the shifts in visual perspective that occur between signer and addressee. As discussed earlier, during sign comprehension the perceiver (i.e., the addressee) often must mentally reverse the spatial arrays created by the signer such that, for example, a spatial locus established on the right of the person signing (and thus on the left of the addressee) is understood as on the right in the scene being described by the signer (see figures 5.9a and 5.10a). Because scenes are most often described from the signer's perspective and not the addressee's, this transformation process may occur frequently. The problem is not unlike that facing understanders of spoken languages who have to keep in mind the directions "left" and "right" with regard to the speaker. The crucial difference for ASL is that these directions are encoded spatially by the signer. The spatial loci used by the signer to depict a scene (e.g., describing the position of objects and people) must therefore be understood as the reverse of what the addressee actually *observes* during discourse (assuming a face to face interaction). Furthermore, in order to understand and process sign, the addressee must perceive the reverse of what they themselves would produce. Anecdotally, hearing subjects have great difficulty with this aspect of learning ASL; they do not easily transform a signer's articulations into the reversal that must be used to produce the signs. Given these linguistic processing requirements, we hypothesized that signers would be better than hearing subjects at mentally rotating imaged objects and making mirror image judgments. To test this hypothesis, we used a task similar to the one devised by Shepard and Metzler (1971) in which subjects were shown two forms created by juxtaposing cubes to form angular shapes. Subjects were asked to decide whether the two shapes were the same or mirror images, regardless of orientation (see figure 5.19).

Our results support the hypothesis that use of ASL can enhance mental rotation skills (see the top illustration in figure 5.19); both deaf and hearing signers had faster reaction times compared to nonsigners at all degrees of rotation. Note that the slopes for the angle of rotation did not differ between signing and nonsigning groups, and this indicates that signers do not actually rotate images faster than nonsigning subjects. Emmorey Kosslyn, and Bellugi (1993) originally suggested that ASL signers may be faster in detecting mirror reversals, particularly because they were faster even when no rotation was required (i.e., at zero degrees). However, recent research by Ilan and Miller (1994)¹⁸ indicates that different processes may be involved when mirror-same judgments are made at zero degrees within a mental rotation experiment, compared to when mental rotation is not required on any of the trials. In addition, preliminary results from Emmorey and Bettger indicate that when native ASL signers and hearing nonsigners are asked to make mirror-same judgments in a

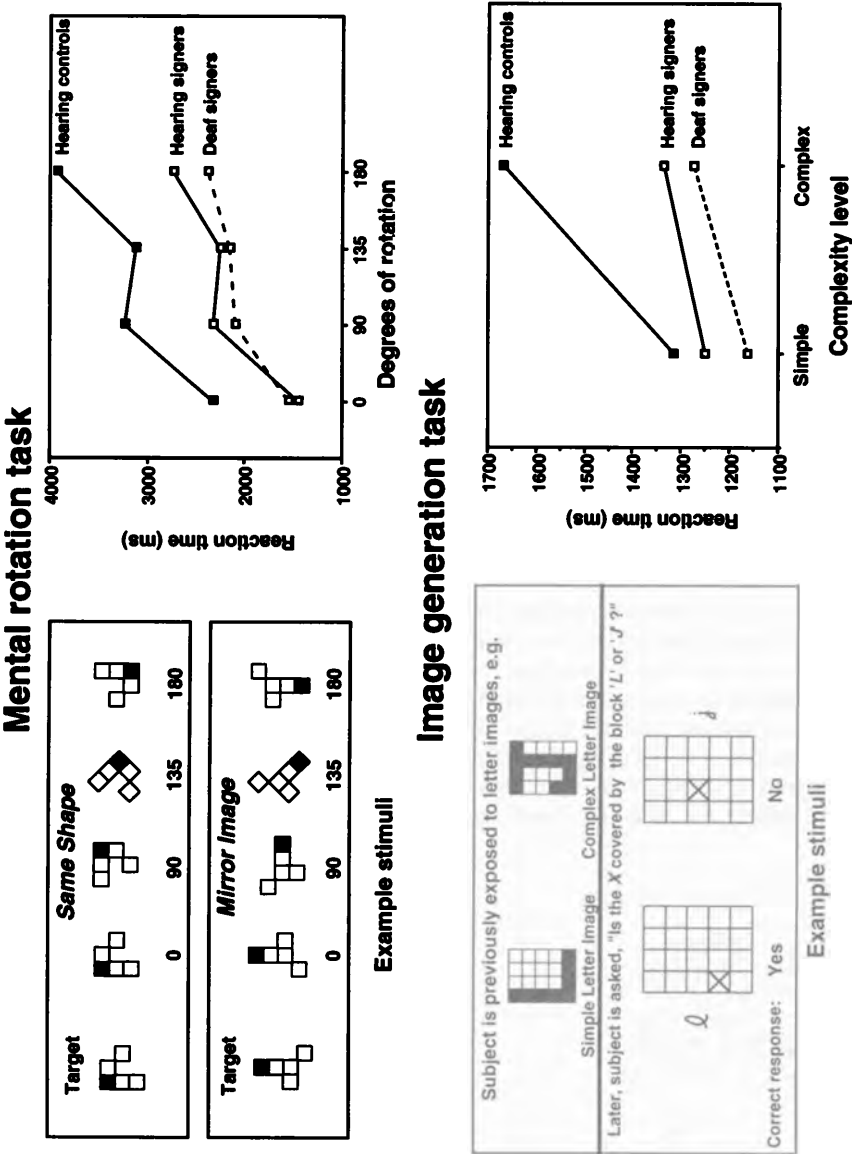


Figure 5.19
Illustration of the mental rotation and image generation tasks used by Emmorey et al. (1993).

comparison task that does not involve mental rotation, these groups do not differ in accuracy or reaction time. The faster response times exhibited by signers on the mental rotation task may reflect faster times to initiate mental rotation or faster times to generate a mental image (as suggested by the next experiment). Finally, the finding that hearing native signers performed like deaf signers indicates that enhancement on this mental rotation task is not a consequence of auditory deprivation. Rather, it appears to be due to experience with a visual language whose production and interpretation may involve mental rotation (see also Talbot and Haude 1993).

Another visual imagery skill we investigated was the ability to generate mental images, that is, the ability to create an image (i.e., a short-term visual memory representation) on the basis of information stored in long-term memory (see Kosslyn et al. 1985). In ASL, image generation may be an important process underlying aspects of referential shift. Liddell (1990) argues that under referential shift, signers may imagine referents as physically present, and these visualized referents are relevant to the expression of verb agreement morphology. Liddell gives the following example involving the verb ASK which is lexically specified to be directed at chin height (see figure 5.20):

To direct the verb ASK toward an imagined referent, the signer must conceive of the location of the imaginary referent's head. For example, if the signer and addressee were to imagine that Wilt Chamberlain was standing beside them ready to give them advice on playing basketball, the sign ASK would be directed upward toward the imaged height of Wilt Chamberlain's head (figure [5.20a]). It would be incorrect to sign the verb at the height of the signer's chin (figure [5.20b]). This is exactly the way agreement works when a referent is present. Naturally, if the referent is imagined as laying down, standing on a chair, etc., the height and direction of the agreement verb reflects this. Since the signer must conceptualize the location of body parts of

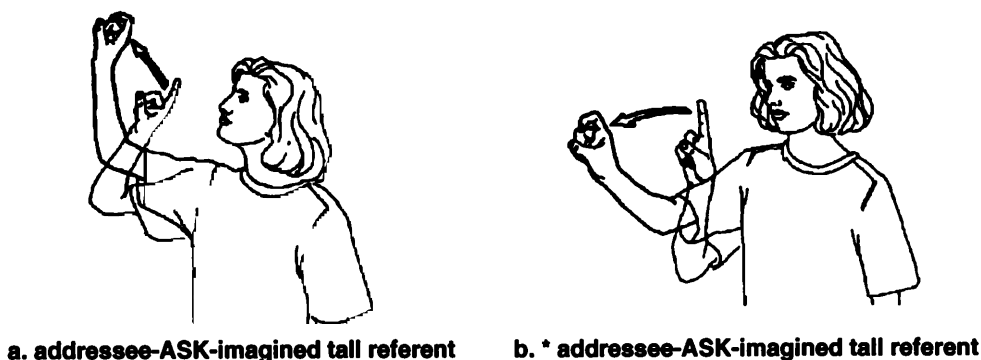


Figure 5.20

Agreement verbs and referents imagined as present. Illustration from Liddell (1990).

the referent imagined to be present, there is a sense in which an invisible body is present. The signer must conceptualize such a body in order to properly direct agreement verbs. (Liddell 1990, 184)

If deaf subjects are in fact generating visual images prior to or during sign production, then the speed of forming these images would be important, and we might expect signers to develop enhanced abilities to generate images. The image generation task we used is illustrated at the bottom of figure 5.19. Subjects first memorized uppercase block letters and then were shown a series of grids (or sets of brackets) that contained an *X* mark. A lowercase letter preceded each grid, and subjects were asked to decide as quickly as possible whether the corresponding uppercase block letter would cover the *X* if it were in the grid. The crucial aspect of the experiment was that the probe mark appeared in the grid only 500 ms after the lowercase cue letter was presented. This was not enough time for the subjects to complete forming the letter image; thus response times reflect in part the time to generate the image. Kosslyn and colleagues have used this task to show that visual mental images are constructed serially from parts (e.g., Kosslyn et al. 1988; Roth and Kosslyn 1988). Subjects tend to generate letter images segment by segment in the same order that the letter is drawn. Therefore, when the probe *X* is covered by a segment that is generated early (e.g., on the first stroke of the letter *F*), subjects have faster reaction times, compared to when the probe is located under a late-imaged segment. Crucially, this difference in response time based on probe location is not found when image generation is not involved, that is, when both the probe *X* and letter (shaded gray) are physically present.

Our results indicated that both deaf and hearing signers formed images of complex letters significantly faster than nonsigners (see figure 5.19). This finding suggests that experience with ASL can affect the ability to mentally generate visual images. Results from a perceptual baseline task indicated that this enhancement was due to a difference in image generation ability, rather than to differences in scanning or inspection—signers and nonsigners did not differ in their ability to evaluate probe marks when the shape was physically present. The signing and nonsigning subjects were equally accurate, which suggests that although signers create complex images faster than nonsigners, both groups generate equally good images. Furthermore, deaf and hearing subjects appeared to image letters in the same way: both groups of subjects required more time and made more errors for probes located on late-imaged segments, and these effects were of comparable magnitude in the two groups. This result indicates that neither group of subjects generated images of letters as complete wholes, and both groups imaged segments in the same order. Again, the finding that hearing signers performed similarly to deaf signers suggests that their enhanced image generation ability is due to experience with ASL, rather than to auditory deprivation.

This research establishes a relation between visual-spatial imagery within linguistic and nonlinguistic domains. Image generation and mental rotation appear to be deeply embedded in using ASL, and these are not processes that must obviously be involved in both visual imagery and ASL perception. Note that these experiments have focused on ASL processing; whether there is a more direct relation in sign language between linguistic *representations* (e.g., conceptual structure, see Jackendoff, chapter 1, this volume) and spatial representations is a topic for future research.

5.4 Neural Correlates for Signed and Spoken Languages

Finally, sign language exhibits properties for which each of the cerebral hemispheres of hearing people shows different predominant functioning. In general, the left hemisphere has been shown to subserve linguistic functions, whereas the right hemisphere is dominant for visual-spatial functions. Given that ASL expresses linguistic functions by manipulating spatial contrasts, what is the brain organization for sign language? Is sign language controlled by the right hemisphere along with many other visual-spatial functions or does the left hemisphere subserve sign language as it does spoken language? Or is sign language represented equally in both hemispheres of the brain? Howard Poizner, Ursula Bellugi, and Edward Klima have shown that the brain honors the distinction between language and nonlanguage visual-spatial functions (Poizner, Klima, and Bellugi 1987; Bellugi, Poizner, and Klima 1989). Despite the visual-spatial modality of signed languages, linguistic processing occurs primarily within the left hemisphere of deaf signers, whereas the right hemisphere is specialized for nonlinguistic visual-spatial processing in these signers. Poizner, Bellugi, and Klima have shown that damage to the left hemisphere of the brain leads to sign aphasia similar to classic aphasia observed in speaking patients. For example, adult signers with left-hemisphere damage may produce “agrammatic” signing, characterized by a lack of morphological and syntactic markings and often accompanied by halting, effortful signing. An agrammatic signer will produce single-sign utterances that lack the grammatically required inflectional movements and use of space (see discussion above). In contrast, right-hemisphere damage produces impairments of many visual-spatial abilities, but does *not* produce sign language aphasia. When given tests of sign language comprehension and production (e.g., from the Salk Sign Aphasia Exam; Poizner, Klima, and Bellugi 1987), signers with right-hemisphere damage perform normally, but these same signers show marked impairment on nonlinguistic tests of visual-spatial functions. For example, when given a set of colored blocks and asked to assemble them to match a model (the WAIS blocks test), right-hemisphere-damaged signers have great difficulty and are unable to capture the

overall configuration of the block design. Similar impairments on this task are found with hearing, speaking subjects with right-hemisphere damage.

Poizner, Klima, and Bellugi (1987) also reported that some signing patients with right-hemisphere damage show a selective impairment in their ability to use space to express spatial relations in ASL, for example when describing the layout of furniture in their room or apartment. Their descriptions are not ungrammatical, but they are incorrect when compared to the actual layout of objects. One hypothesis for this dysfunction following right-hemisphere damage is that, unlike spoken language, ASL requires that the cognitive representation of spatial relations be recovered from and instantiated within a spatialized linguistic encoding (i.e., cognitive spatial relations map to space, not to sound). Evidence supporting this hypothesis comes from a bilingual hearing patient with right-hemisphere damage studied by David Corina and colleagues (Corina et al. 1990; Emmorey, Corina, and Bellugi 1995; Emmorey, Hickok, and Corina 1993). The data from this case suggest that there may be more right-hemisphere involvement when processing spatial information encoded within a linguistic description for signed compared to spoken languages.

The case involves female patient D.N.,¹⁹ a young hearing signer (age 39), bilingual in ASL and English, who was exposed to ASL early in childhood. She underwent surgical evacuation of a right parietal-occipital hematoma and an arteriovenous malformation. Examination of a magnetic resonance imaging (MRI) scan done six months after the surgery revealed a predominantly mesial superior occipital-parietal lesion. The superior parietal lobule was involved, while the inferior parietal lobule was spared, although some of the deep white matter coming from this structure may also be involved. The comparison test between English and ASL spatial commands (see below and figure 5.21) was conducted by Corina approximately one year after D.N.'s surgery.

D.N. was not aphasic for either English or ASL. Her performance on the Salk Sign Diagnostic Aphasia Exam was excellent, and she showed no linguistic deficits for English. Nevertheless, she exhibited a striking dissociation between her ability to comprehend and produce spatial descriptions in English compared to ASL. Although her English description had no evident spatial distortions, she was impaired in her ability to describe the spatial layout of her room using ASL. Her ASL description showed a marked disorganization of the elements in the room. Her attempts to place one set of objects in relation to others were particularly impaired, and she incorrectly specified the orientation and location of items of furniture (see also Emmorey, Corina, and Bellugi 1995).

Corina (1989) developed a specific set of tasks to investigate D.N.'s comprehension of locative relations in English and ASL. One of these tasks required D.N. to set up

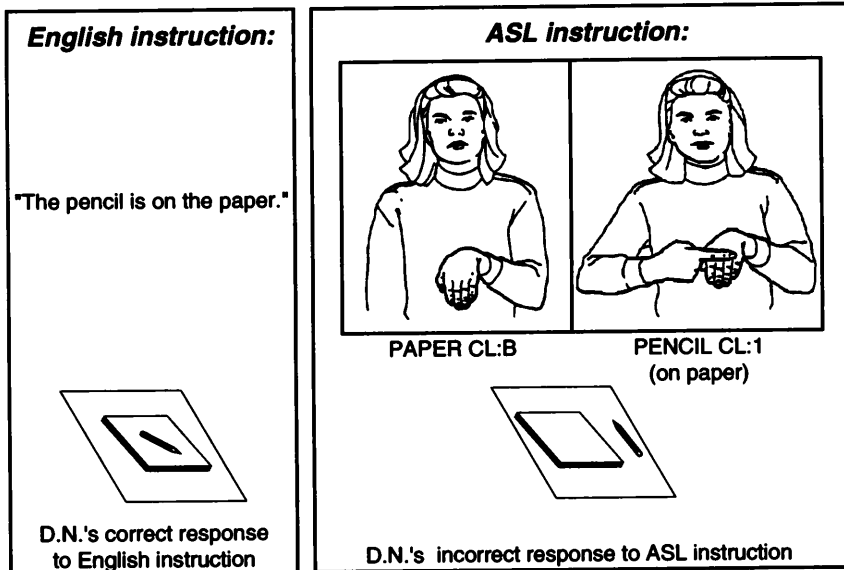


Figure 5.21

Illustration of a RHD patient's differential performance in comprehending English versus ASL spatial commands (the lexical signs PAPER and PENCIL are not shown).

real objects in accordance with spatial descriptions given in either English or in ASL. An example of a simple English instruction would be "The pen is on the paper." The English and ASL instructions along with D.N.'s responses are illustrated in figure 5.21. D.N. correctly interprets the English command, but fails with the ASL instructions. This particular example was elicited through informal testing by Corina in which the same instructions were given in both English and ASL. D.N. was later given 36 different spatial commands (18 in English and 18 in ASL) which involved from two to four objects (e.g., cup, pen, book). The instructions were matched for number of spatial relations that were encoded in each language. When D.N. was given instructions in English to locate objects with respect to one another, she performed relatively well—83% correct. Her score was worse than her normal age-matched bilingual control (100% correct), but better than other right-hemisphere-damaged subjects who were given the English test (69% correct). However, when presented with similar information in ASL—in which spatial relations are presented topographically in sign space—D.N. made many more spatial errors, scoring only 39% correct. This result is particularly striking, given the iconicity of the ASL descriptions (see figure 5.21).

We hypothesize that the dissociation between D.N.'s comprehension of English and ASL spatial commands arises because of the highly specific spatial realization of ASL classifier constructions. That is, spatial relations must be recovered from a visual-spatial signal in which much more information is encoded about the relative position and orientation of objects, compared to English. Furthermore, the requirement of reading off spatial relations directly from the orientation and position of classifier signs in space may make additional demands on spatial cognitive processes within the right hemisphere. D.N.'s comprehension impairment is not linguistic *per se*, but stems from the fact that linguistic information about spatial relations must be recovered from a representation that itself is spatialized; D.N. does not have difficulty understanding ASL spatial contrasts that do not encode information about location or orientation. Thus the case of D.N. also bears on our earlier discussion concerning referential versus topographic functions of space in ASL. D.N. exhibits a dissociation between the use of signing space as a linguistic device for marking sentence-level referential distinctions and the use of signing space as a topographic mapping device (see Emmorey et al. 1995 for a complete discussion of this dissociation and for additional evidence from language-processing experiments with normal ASL signers).

In conclusion, signed languages offer a unique window into the relation between language and space. All current evidence indicates that signed languages are constrained by the same principles that shape spoken languages. Thus far, there is no evidence that signed languages grammaticize different aspects of the spatial world compared to spoken languages (see Supalla 1982). What is different and unusual about signed languages is their visual-spatial form—the fact that space and movement can be used to linguistically represent space and movement in the world. This chapter has explored the ramifications of this spatialized encoding for the nature of linguistic structure, for language processing, for spatial cognition in general, and for the neural substrate of sign language. Future research might include investigations of the following: (1) the semantic and grammatical structure of locative constructions in different sign languages (how do sign languages vary in the way they utilize physical space to represent topological and other spatial concepts?); (2) when and how signing children acquire locative vocabulary (what is the developmental relation between spatial cognition and sign language acquisition? See Mandler, chapter 9, this volume, and Bowerman, chapter 10, this volume, for discussion of spatial cognition and spoken language acquisition); (3) spatial attention in sign language perception and nonlinguistic visual-spatial perception (do signers show differences in spatial attention that could be attributed to experience with sign language?); (4) how signers build spatial mental models (does signing space operate like a diagram? See Johnson-Laird, chapter 11, this volume); and (5) the neural substrate and psychological mechanisms

that underlie the mapping between a linguistic signal (both signed and spoken) and an amodal spatial representation. These are only some of the areas in which the study of sign language could enhance our understanding of the relation between language and space.

Acknowledgments

This work was supported by National Institutes of Health grants RO1 DC 00201, RO1 DC 00146, and R37 HD 13249. I thank David Corina, Greg Hickok, and Ed Klima for many insightful discussions about the issues presented here. Merrill Garrett and Mary Peterson provided valuable comments on an earlier draft of this chapter. I also thank Bonita Ewan and Steve McCullough, who were my primary language consultants and who were the sign language models for the figures. Mark Williams helped create many of the figures in this chapter. Finally, I am particularly grateful to the Gallaudet University students who participated in these studies.

Notes

1. Words in capital letters represent English glosses for ASL signs. The gloss represents the meaning of the unmarked, unmodulated root form of a sign. A subscripted word following a sign gloss indicates that the sign is made with some regular change in form associated with a systematic change in meaning, and thus indicates grammatical morphology in ASL (e.g., GIVE_{habitual}). Multiword glosses connected by hyphens are used when more than one English word is required to translate a single sign (e.g., LOOK-AT). Subscripts are used to indicate spatial loci; nouns, pronouns, and agreeing verbs are marked with a subscript to indicate the loci at which they are signed (e.g. INDEX_s, BITE_b). Classifier forms are abbreviated CL, followed by the handshape of the classifier and a description of the meaning in italics (CL:G—*shape*). Descriptions of how a classifier sign is articulated may be given underneath the gloss. English translations are provided in quotes.
2. Some signs such as personal pronouns may not be specified in the lexicon for location (see Lillo-Martin and Klima 1990; Liddell 1994).
3. Other terms that have been used for these verbs are *indicating* (Liddell 1995) and *inflecting* (Padden 1988).
4. Whether subject is associated with the beginning or end of the verb's movement depends upon the class of verb (cf. "backwards" verbs, Padden 1988; Brentari 1988).
5. Following traditional linguistic typography, a question mark (?) indicates that a sentence is considered marginal; a star (*) indicates that the sentence is unacceptable.
6. In this study, native signers were deaf individuals who were exposed to ASL from birth.
7. The example of drawing was suggested to me by Dan Slobin, who has made similar arguments about scene setting and the effect of modality on signed languages (Slobin and Hoiting 1994).

8. Sign linguists often use “frame of reference” in a nonspatial sense, referring to anaphoric reference in a discourse (see especially Engberg-Pedersen 1993).
9. The addressee is assumed to be facing the signer. Signers described these pictures to a video camera rather than to an actual addressee. In understanding this discussion of point of view in ASL, it might be useful for you the reader to imagine that you and the signer viewed the display from the same vantage point, and now the signer is facing you (the addressee) to describe it.
10. It should be noted that occasionally a signer may ignore the orientation features of the vehicle classifier, say, pointing the vehicle classifier toward the tree classifier, when in actual fact the car is facing away from the tree. This may occur when it is difficult to produce the correct orientation, say, pointing the vehicle classifier to the right with the right hand, palm out (try it).
11. There were only six examples (out of thirty-five) in which a signer ignored the orientation of the car because it was awkward to articulate. Also, signers did not always alternate which hand produced the classifier for TREE, as might be implied by figures 5.9 and 5.10.
12. Except for the sign LEFT, WEST is perhaps the only sign that is specified as moving toward the signer’s left rather than toward the “nondominant side.” For both left- and right-handers, the sign WEST moves toward the left, and the sign EAST moves toward the right. The direction of movement is fixed with respect to the signer’s left and right, unlike other signs. For example, right- and left-handers would articulate the signs illustrated in figure 5.1, which also move across the body, with opposite directions of motion (left to right vs. right to left, respectively). However, there is some change in articulation for left-handers, perhaps due to phonological constraints. For EAST and WEST, the orientation of the palm is reversed: outward for WEST and inward for EAST. This change in palm orientation also occurs when a right-handed signer articulates EAST or WEST with the left hand (switches in hand dominance are phonologically and discourse governed).
13. When the signs NORTH and SOUTH are used to label paths within a spatial map, they often retain some of their upward and downward movement.
14. This study was conducted in collaboration with Shannon Casey; the experimenter was either a native speaker of English (for the English subjects) or a deaf ASL signer (for the deaf subjects).
15. This is not an orientation command but a shape description, namely, a classifier construction in which the shape of the blue puzzle piece is traced in the vertical plane (see figure 5.13 for an example).
16. CORNER is a frozen classifier construction produced with nominal movement (Supalla and Newport 1978). The sign can be articulated at various positions in space to indicate where the corner is located (e.g., top left or bottom right).
17. This study was conducted with Marci Clothier and Stephen McCullough.
18. I thank Mary Peterson for bringing this work to my attention.
19. Poizner and Kegl (1992) also discuss this patient, but use the pseudonym initials A.S.

References

- Battison, R. (1978). *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linstok Press.
- Bellugi, U., Poizner, H., and Klima, E. S. (1989). Language, modality, and the brain. *Trends in Neurosciences*, 10, 380–388.
- Brentari, D. (1988). Backwards verbs in ASL: Agreement re-opened. In *Papers from the Parasession on Agreement in Grammatical Theory*, vol. 24, no. 2, 16–27. Chicago: Chicago Linguistic Society.
- Brown, P. (1991). Spatial conceptualization in Tzeltal. Working paper no. 6, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Corina, D. (1989). Topographic relations test battery for ASL. Unpublished manuscript, Salk Institute for Biological Studies, La Jolla, CA.
- Corina, D., Bellugi, U., Kritchewsky, M., O'Grady-Batch, L., and Norman, F. (1990). Spatial relations in signed versus spoken language: Clues to right parietal functions. Paper presented at the Academy of Aphasia, Baltimore.
- Corina, D., and Sandler, W. (1993). On the nature of phonological structure in sign language. *Phonology*, 10, 165–207.
- Coulter, G. R., and Anderson, S. R. (1993). Introduction to G. R. Coulter (Ed.) *Phonetics and phonology: Current issues in ASL phonology*. San Diego, CA: Academic Press.
- Emmorey, K., and Corina, D. (1990). Lexical recognition in sign language: Effects of phonetic structure and morphology. *Perceptual and Motor Skills*, 71, 1227–1252.
- Emmorey, K., Corina, D., and Bellugi, U. (1995). Differential processing of topographic and referential functions of space. In K. Emmorey and J. Reilly (Eds), *Language, gesture, and space*, 43–62. Hillsdale, NJ: Erlbaum.
- Emmorey, K., Hickok, G., and Corina, D. (1993). Dissociation between topographic and syntactic functions of space in ASL. Paper presented at the Academy of Aphasia Meeting, Tucson, AZ, October.
- Emmorey, K., Kosslyn, S. M., and Bellugi, U. (1993). Visual imagery and visual-spatial language: Enhanced imagery abilities in deaf and hearing ASL signers. *Cognition*, 46, 139–181.
- Engberg-Pedersen, E. (1993). *Space in Danish Sign Language: The semantics and morphosyntax of the use of space in a visual language*. International Studies on Sign Language Research and Communication of the Deaf, vol. 19. Hamburg: Signum.
- Franklin, N., Tversky, B., and Coon, V. (1992). Switching points of view in spatial mental models. *Memory and Cognition*, 20(5), 507–518.
- Gee, J., and Goodhart, W. (1988). American Sign Language and the human biological capacity for language. In M. Strong (Ed.), *Language learning and deafness*, 49–74. New York: Cambridge University Press.

- Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions in English*. Cambridge: Cambridge University Press.
- Ilan, A. B., and Miller, J. (1994). A violation of pure insertion: Mental rotation and choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 20(3), 520–536.
- Janis, W. (1995). A crosslinguistic perspective on ASL verb agreement. In K. Emmorey and J. Reilly (Eds.), *Language, gesture, and space*, 195–224. Hillsdale, NJ: Erlbaum.
- Klima, E. S., and Bellugi, U. (1979). *The signs of language*. Cambridge, MA: Harvard University Press.
- Kosslyn, S. M., Brunn, J. L., Cave, K. R., and Wallach, R. W. (1985). Individual differences in mental imagery ability: A computational analysis. *Cognition*, 18, 195–243.
- Kosslyn, S., Cave, K., Provost, D., and Von Gierke, S. (1988). Sequential processes in image generation. *Cognitive Psychology*, 20, 319–343.
- Landau, B., and Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–238.
- Levelt, W. (1982a). Cognitive styles in the use of spatial direction terms. In R. Jarvella and W. Klein (Eds.), *Speech, place, and action*, 251–268. New York: Wiley.
- Levelt, W. (1982b). Linearization in describing spatial networks. In S. Peters and E. Saarinen (Eds.), *Processes, beliefs, and questions*, 199–220. Dordrecht: Reidel.
- Levelt, W. (1984). Some perceptual limitations on talking about space. In A. J. van Doorn, W. A. van de Grind, and J. J. Koenderink (Eds.), *Limits in perception*, 323–358. Utrecht: VNU Science Press.
- Levinson, S. (1992a). Vision, shape, and linguistic description: Tzeltal body-part terminology and object descriptions. Working paper no. 12, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Levinson, S. (1992b). Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. Working paper no. 13, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Liddell, S. (1990). Four functions of a locus: Reexamining the structure of space in ASL. In C. Lucas (Ed.), *Sign language research: Theoretical issues*, 176–198. Washington, DC: Gallaudet College Press.
- Liddell, S. (1993). Conceptual and linguistic issues in spatial mapping: Comparing spoken and signed languages. Paper presented at the Phonology and Morphology of Sign Language Workshop, Amsterdam, August.
- Liddell, S. (1994). Tokens and surrogates. In I. Ahlgren, B. Bergman, and M. Brennan (Eds.), *Perspectives on sign language structure*. Durham, UK: ISLA.
- Liddell, S. (1995). Real, surrogate, and token space: Grammatical consequences in ASL. In K. Emmorey and J. Reilly (Eds.), *Language, gesture, and space*, 19–42. Hillsdale, NJ: Erlbaum.

- Lillo-Martin, D. (1991). *Universal grammar and American sign language: Setting the null argument parameters*. Dordrecht: Kluwer.
- Lillo-Martin, D. (1995). The point of view predicate in American Sign Language. In K. Emmorey and J. Reilly (Eds.), *Language, gesture, and space*, 155–170. Hillsdale, NJ: Erlbaum.
- Lillo-Martin, D., and Klima, E. (1990). Pointing out differences: ASL pronouns in syntactic theory. In S. D. Fischer and P. Siple (Eds.), *Theoretical issues in sign language research*, vol. 1, 191–210. Chicago: University of Chicago Press.
- Loew, R. (1983). Roles and reference in American Sign Language: A developmental perspective. Ph.D. diss., University of Minnesota.
- McIntire, M. (1980). Locatives in American Sign Language. Ph.D. diss., University of California, Los Angeles.
- Meier, R. (1991). Language acquisition by deaf children. *American Scientist*, 79, 60–70.
- Newell, W. (Ed.) (1983). *Basic sign communication*. Silver Spring, MD: National Association of the Deaf.
- Newport, E., and Meier, R. (1985). The acquisition of American Sign Language. In D. I. Slobin (Ed.), *The Crosslinguistic study of language acquisition*. Vol. 1, *The data*, 881–938. Hillsdale, NJ: Erlbaum.
- Padden, C. (1986). Verbs and role-shifting in ASL. In C. Padden (Eds.), *Proceedings of the Fourth National Symposium on Sign Language Research and Teaching*, 44–57. Silver Spring, MD: National Association of the Deaf.
- Padden, C. (1988). *Interaction of morphology and syntax in ASL*. Garland Outstanding Dissertations in Linguistics, ser. 4. New York: Garland. 1983 Ph.D. diss., University of California, San Diego.
- Padden, C. (1990). The relation between space and grammar in ASL verb morphology. In C. Lucas (Ed.), *Sign language research: Theoretical issues*, 118–132. Washington, DC: Gallaudet University Press.
- Poizner, H., and Kegl, J. (1992). Neural basis of language and motor behavior: Perspectives from American Sign Language. *Aphasiology*, 6(3), 219–256.
- Poizner, H., Klima, E. S., and Bellugi, U. (1987). *What the hands reveal about the brain*. Cambridge, MA: MIT Press.
- Poulin, C., and Miller, C. (1994). On narrative discourse and point of view in Quebec Sign Language. In K. Emmorey and J. Reilly (Eds.), *Language, gesture, and space*, 117–132. Hillsdale, NJ: Erlbaum.
- Roth, J., and Kosslyn, S. M. (1988). Construction of the third dimension in mental imagery. *Cognitive Psychology*, 20, 344–361.
- Sandler, W. (1989). *Phonological representation of the sign: Linearity and nonlinearity in American Sign Language*. Dordrecht: Foris.
- Schober, M. (1993). Spatial perspective taking in conversation. *Cognition*, 47, 1–24.

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

Shepard-Kegl, J. (1985). Locative relations in American Sign Language word formation, syntax, and discourse. Ph.D. diss., Massachusetts Institute of Technology.

Slobin, D., and Hoiting, N. (1994). Reference to movement in spoken and signed languages: Typological considerations. *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistic Society*, 1–19. Berkeley, CA: Berkeley Linguistics Society.

St. John, M. F. (1992). Learning language in the service of a task. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Supalla, S. (1991). Manually coded English: The modality question in signed language development. In P. Siple and S. D. Fischer (Eds.), *Theoretical issues in sign language research*, vol. 2, 85–109. Chicago: University of Chicago Press.

Supalla, T. (1982). Structure and acquisition of verbs of motion and location in American Sign Language. Ph.D. diss., University of California, San Diego.

Supalla, T., and Newport, E. (1978). How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In P. Siple (Ed.), *Understanding language through sign language research*, 91–132. New York: Academic Press.

Talbot, K. F., and Haude, R. H. (1993). The relationship between sign language skill and spatial visualization ability: Mental rotation of three-dimensional objects. *Perceptual and Motor Skills*, 77(3), 1387–1391.

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

Talmy, L. (1988). The relation of grammar to cognition. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics*, 165–207. Amsterdam: Benjamins.

Taylor, H., and Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, 31, 261–292.

Wilbur, R. (1987). *American Sign Language: Linguistic and applied dimensions*. Boston: Little, Brown.

Winston, E. (1995). Spatial mapping in comparative discourse frames. In K. Emmorey and J. Reilly (Eds.), *Language, gesture, and space*, 87–114. Hillsdale, NJ: Erlbaum.

Chapter 6

Fictive Motion in Language and “Ception”

Leonard Talmy

6.1 Introduction

This chapter proposes a unified account of the extensive cognitive representation of nonveridical phenomena—especially forms of motion—both as they are expressed linguistically and as they are perceived visually. Thus, to give an immediate sense of the matter, the framework posited here will cover linguistic instances that depict motion with no physical occurrence, for example: *This fence goes from the plateau to the valley*; *The cliff wall faces toward/away from the island*; *I looked out past the steeple*; *The vacuum cleaner is down around behind the clothes hamper*; and *The scenery rushed past us as we drove along*.

In a similar way, our framework will also cover visual instances in which one perceives motion with no physical occurrence, for example: the perceived “apparent motion” in successive flashes along a row of lightbulbs, as on a marquee; the perceived “induced motion” of a rod when only a surrounding frame is moved; the perception of a curved line as a straight line that has undergone processes like indentation and protrusion; the possible perception of an obliquely oriented rectangle (e.g., a picture frame) as having been tilted from a vertical-horizontal orientation; and the possible perception of a “plus” figure as involving the sequence of a vertical stroke followed by a horizontal stroke.

6.1.1 Overall Framework

Our unified account of the cognitive representation of nonveridical phenomena, just exemplified, is a particular manifestation of the “overlapping systems” model of cognitive organization. This model sees partial similarities and differences across distinct cognitive systems in the way they structure perceptual, conceptual, or other cognitive representations. We will mainly consider similarities between two such cognitive systems: language and visual perception.

The particular manifestation of overlap we address involves a major cognitive pattern: a discrepancy within the cognition of a single individual. Specifically, this discrepancy is between two different cognitive representations of the same entity, where one of the representations is assessed as being more veridical than the other. We presume that the two representations are the products of two different cognitive subsystems, and that the veridicality assessment itself is produced by a third cognitive subsystem whose general function it is to generate such assessments.

In the notion of discrepancy we intend here, the two cognitive representations consist of different contents that could not both concordantly hold for their represented object at the same time—that is, they would be inconsistent or contradictory, as judged by the individual's cognitive systems for general knowledge or reasoning. On the other hand, the individual need not have any active experience of conflict or clash between the two maintained representations, but might rather experience them as alternative perspectives. Further, in saying that the two discrepant representations differ in their assessed degree of veridicality, we use the less common term *veridical*—rather than, say, a term like *true*—to signal that the ascription is an assessment produced by a cognitive system, with no appeal to some notion of absolute or external reality.

Of the two discrepant representations of the same object, we will characterize the representation assessed to be more veridical as “factive” and the representation assessed to be less veridical as “fictive.” Adapted from its use in linguistics, the term *factive* is here again intended to indicate a cognitive assessment of greater veridicality, but not to suggest (as perhaps the word *factual* would) that a representation is in some sense objectively real. And the term *fictive* has been adopted for its reference to the imaginal capacity of cognition, not to suggest (as perhaps the word *fictitious* would) that a representation is somehow objectively unreal. As a whole, this cognitive pattern of veridically unequal discrepant representations of the same object will here be called the pattern of “general fictivity.”

In the general fictivity pattern, the two discrepant representations frequently—though not exclusively—disagree with respect to some single dimension, representing opposite poles of the dimension. Several different dimensions of this sort can be observed. One example of such a dimension is *state of occurrence*. Here, factive presence (the presence of some entity in the more veridical representation) is coupled with fictive absence (the absence of that entity from the less veridical representation) or vice versa. Another example of a dimension is *state of change*. Here, the more veridical representation of an object could include factive stasis, while the less veridical representation includes fictive change—or vice versa. One form of this last dimension when applied to a physical complex in space-time is the more specific dimension *state of motion*. Here, the more veridical representation could include

stationariness, while the less veridical representation has motion—or vice versa. Thus, frequently in conjunction with their factive opposites, we can expect to find cases of fictive presence, fictive absence, fictive stasis, fictive change, fictive stationariness, and fictive motion. In fact, to a large extent, general fictivity can accommodate any “fictive *X*.”

Of these types, the present chapter focuses on fictive motion, usually in combination with factive stationariness. It will be seen that such fictive motion occurs preponderantly more than does fictive stationariness coupled with factive motion. As will be discussed, this fact reflects a cognitive bias toward dynamism.

The general fictivity pattern can be found in a perhaps parallel fashion in both language and vision. In language, the pattern is extensively exhibited in the case where one of the discrepant representations is the belief held by the speaker or hearer about the real nature of the referent of a sentence, and the other representation is the literal reference of the linguistic forms that make up the sentence. Here the literal representation is assessed as less veridical than the representation based on belief. Accordingly, the literal representation is fictive, while the representation based on belief is factive. Given our focus on the pattern in which fictive motion is coupled generally with factive stationariness, we here mainly treat the linguistic pattern in which the literal meaning of a sentence ascribes motion to a referent one would otherwise believe to be stationary.

In vision, one main form of the general fictivity pattern is the case where one of the discrepant representations is the concrete or fully palpable percept an individual has of a scene on viewing it, and the other is a particular, less palpable percept the individual has of the same scene concurrently. Here the less palpable percept is assessed as the less veridical of the two representations. Parallel to the linguistic case, the term *factive* may be applied to the more palpable visual representation, and the term *fictive* to the less palpable representation. We will say that an individual “sees” the factive representation, but only “senses” the fictive representation (when it occurs at a particular lower level of palpability, to be discussed later). Here, too, we focus on fictive motion, where the less palpable visual representation is of motion, while the fully palpable representation is generally of stationariness.

To accommodate this account of visual representations that differ with respect to their palpability, we posit the presence in cognition of a gradient parameter of palpability. Moreover, one may identify a number of additional cognitive parameters that largely tend to correlate with the palpability parameter. All of these “palpability-related parameters” are characterized below in section 6.9.1. Further, these parameters appear to extend continuously through a cognitive domain larger than that generally associated with perception alone, one that in fact covers the combination of what is usually associated differentially with separate domains of perception and

conception. Accordingly, to accommodate the full range of each such parameter, we advance the idea of a single continuous cognitive domain, which we call “ception.”

In the present chapter we largely restrict our study of general fictivity in language to the case where both of the two discrepant representations are of a physical complex in space-time. In this way, there is generally the potential for any linguistic example to have an analogue in a visual format. Accordingly, in a cross-domain correspondence of this sort, we could expect to find two component parallels. One parallel would hold between the two factive representations; the other between the two fictive representations. In particular, one parallel would hold between the linguistic representation of a sentence believed to be veridical and the concrete, fully palpable appearance of the corresponding visual display. The other parallel would then hold between the less veridical literal reference of the sentence and a less palpable associated image perceived on viewing the display.

If we view this correspondence starting from the language end, a linguistic example of general fictivity whose representations pertain to physical entities in space-time can, in effect, be mapped onto a visual example of general fictivity. In such a mapping, the linguistic referential difference between credence and literality is then translated in the visual domain into a difference in palpability. Experimental methods are needed to determine whether the parallel between the two fictive representations holds. In fact, one aim for the present chapter is to serve as a guide and as a call for such experimental research.

The restriction of the present study to the representation of physical forms in space-time excludes treatment of nonspatial metaphor. For example, a metaphor like *Her mood went from good to bad* would be excluded; although its source domain is motion in space-time, its target domain is the nonphysical one of mood states. However, as discussed later, linguistic metaphor as a whole fits as a category within the framework of general fictivity. General fictivity can serve as the superordinate framework because, among other reasons, its concepts and terms can apply as readily to visual representations as to linguistic ones, whereas metaphor theory is cast in concepts and terms more suitable for language alone. Using the perspective and methods of cognitive linguistics, the present study of fictive motion is based in language, but extends out from there to considerations of visual perception.

6.1.2 Fictive Motion in Language

Fictive motion in language encompasses a number of relatively distinct categories (first set forth in Talmy 1990). These categories include emanation, pattern paths, frame-relative motion, advent paths (including site manifestation and site arrival), access paths, and coverage paths. This last category, perhaps the type of fictive

motion most familiar in the previous linguistic literature, was called “virtual motion” in Talmy (1983), “extension” in Jackendoff (1983), “abstract motion” in Langacker (1987), and “subjective motion” in Matsumoto. Our current term *coverage paths* is used as part of the more comprehensive taxonomy of fictive motion presented here.

Illustrating coverage paths can serve as an orientation to fictive motion in general. This category is most often demonstrated by forms like *This road goes from Modesto to Fresno* or *The cord runs from the TV to the wall*. But a purer demonstration of this type of fictive motion would exclude reference to an entity that supports the actual motion of other objects (as a road guides vehicles) or that itself may be associated with a history of actual motion (like a TV cord). The “mountain range” example in (1) avoids this problem.

- (1) a. That mountain range lies between Canada and Mexico.
- b. That mountain range goes from Canada to Mexico.
- c. That mountain range goes from Mexico to Canada.

Here (1a) directly expresses the more veridical static spatial relationships in a stative form of expression, without evoking fictive motion. But (1b) and (1c) represent the static linear entity, the mountain range, in a way that evokes a sense or a conceptualization of something in motion—respectively, from north to south and from south to north. These latter two sentences manifest the general fictivity pattern. They each involve two discrepant representations of the same object, the mountain range. Of these two representations, the fictive representation—that is, the one that is assessed and experienced as less veridical—consists of the literal reference of the words, which directly depict the mountain range as moving. The factive representation, the one assessed and experienced as more veridical, consists of our belief that the mountain range is stationary. This factive representation is the only representation present in sentence (1a), which accordingly does not manifest the general fictivity pattern.

Most observers can agree that languages systematically and extensively refer to stationary circumstances with forms and constructions whose basic reference is to motion. We can term this *constructional fictive motion*. Speakers exhibit differences, however, over the degree to which such expressions evoke an actual sense or conceptualization of motion—what can be termed *experienced fictive motion*. Thus, for the same instance of constructional fictive motion, some speakers will report a strong semantic evocation of motion, while other speakers will report that there is none at all. What does appear common, though, is that every speaker experiences a sense of motion for *some* fictive motion constructions.

Where an experience of motion does occur, there appears an additional range of differences in what is conceptualized as moving. This conceptualization can vary

across individuals and types of fictive motion; even the same individual may deal with the same example of fictive motion differently on different occasions. Included in the conceptualizations of this range, the fictive motion may be manifested by the named entity, for example, by the mountain range in (1); by some unnamed object that moves with respect to the named entity, for example, a car or hiker relative to the mountain range; in the mental imagery of the speaker or hearer, by the imagistic or conceptual equivalent of their focus of attention moving relative to the named entity; by some abstracted conceptual essence of motion moving relative to the named entity; or by a sense of abstract directedness suggesting motion relative to the named entity. The strength and character of experienced fictive motion, as well as its clarity and homogeneity, are a phenomenological concomitant of the present study that will need more investigation.

The several distinct categories of fictive motion indicated above differ from each other with respect to a certain set of conceptual features. Each category of fictive motion exhibits a different combination of values for these features, of which the main ones are shown in (2).

(2) *Principal features distinguishing categories of fictive motion in language*

1. Factive motion of some elements need not/must be present for the fictive effect;
2. The fictively moving entity is itself factive/fictive;
3. The fictive effect is observer-neutral/observer-based—and, if observer-based, the observer is factive/fictive and moves/scans;
4. What is conceived as fictively moving is an entity/the observation of an entity.

Out of the range of fictive motion categories, this chapter selects for closest examination the category of emanation, which appears to have been largely unrecognized. The other indicated categories of fictive motion will be more briefly discussed in section 6.8.¹

6.1.3 Properties of the Emanation Type as a Whole

Amid the range of fictive motion categories, *emanation* is basically the fictive motion of something intangible emerging from a source. In most subtypes, the intangible entity continues along its emanation path and terminates by impinging on some distal object. The particular values of the general fictive features of (2) that are exhibited by the emanation category are listed in (3). Specifically, the intangible entity is what moves fictively and is itself fictive, and its fictive motion does not depend on any factive motion by some tangible entity nor on any localized observer.

(3) *The feature values for emanation paths in language*

1. Factive motion of some elements need not be present for the fictive effect;
2. The fictively moving entity is itself fictive;
3. The fictive effect is observer-neutral;
4. What is conceived as fictively moving is an entity.

The category of emanation comprises a number of relatively distinct types. We present four of these emanation types in sections 6.2–6.5: orientation paths, radiation paths, shadow paths, and sensory paths. The illustrations throughout will be from English only in the present version of this chapter, but examples from other languages can be readily cited. The demonstrations of at least constructional fictive motion will rely on linguistic forms with basically real-motion referents such as verbs like *throw* and prepositions like *into* and *toward*. In the exposition, wherever some form of linguistic conceptualization is posited, we will raise the possibility of a corresponding perceptual configuration. Then, in section 6.7, we will specifically suggest perceptual analogues to the emanation types that have been discussed.

6.2 Orientation Paths

The first type of emanation we consider is that of *orientation paths*. The linguistic conceptualization—and possibly a corresponding visual perception—of an orientation path is of a continuous linear intangible entity emerging from the front of some object and moving steadily away from it. This entity may be conceived or perceived as a moving intangible line or shaft—the only characterization used below. Alternatively, though, the entity might be conceived or perceived as some intangible abstraction moving along a stationary line or shaft—itself equally intangible—that is already in place and joined at one end to the front of the object. In addition to fictive motion along the axis of such a line, in some cases the line can also be conceptualized or perceived as moving laterally.

In this characterization, the “front” of an object is itself a linguistic conceptualization or perceptual ascription based on either a particular kind of asymmetry in the object’s physical configuration; or on the object’s motion along a path, where the leading side would generally constitute the front.² In the main cases relevant here, such a front can be either a planar or “face”-type front, consisting of an approximately planar surface on a volumetric object, or a point-type front, consisting of an endpoint of a linearly shaped object.

Presented next are five subtypes of orientation paths that variously differ with respect to several factors, including whether the front is a face-type or a point-type, and whether the fictive motion of the intangible line is axial or lateral. First, though,

we note the occurrence of constructions that are sensitive to the fictive *presence* of an intangible line aligned with the front of an object, before we proceed to its fictive motion. Consider the sentences in (4):

- (4) a. She crossed in front of me/the TV.
 b. She crossed ?behind/*beside me/the TV.

The sentences here show that the verb *cross* can felicitously be used when walking transversely in front of an object with a front, but only poorly when walking behind, and not at all when walking to one side.³ This usage pattern seems to suggest there is something linear present to walk across directly in front of an object, but not elsewhere with respect to that object. We would argue that what is thus being crossed is the posited intangible line conceived to emerge from the front of an object, that will next be seen to exhibit fictive motion in a further set of construction types.

6.2.1 Prospect Paths

The first type of orientation path that we examine can be termed a *prospect path*. The orientation that an object with a face-type front has relative to its surroundings can be conceptualized linguistically—and perhaps perceived—in terms of fictive motion. With its front face, the object has a particular “prospect,” “exposure,” or “vista” relative to some other object in the surroundings. This prospect is characterized as if some intangible line or shaft emerges from the front and moves continuously away from the main object relative to the other object. The linguistic constructions, in effect, treat this line as Figure moving relative to the other object as Ground or Reference Object (in Talmy’s [1987b, 1983] terms) along a path indicated by directional adpositions. In English, such constructions generally employ verbs like *face* or *look out*.

In the example in (5), the vertical side of a cliff acts as its face-type front. The cliff’s prospect upon its surroundings is characterized in terms of a fictive course of motion emerging from its face and moving along the path specified by the preposition relative to a valley as Reference Object. Again, this example manifests the general fictivity pattern. The literal sense of its words depicts a fictive, less veridical representation in which something moves from the cliff wall along a path that is oriented with respect to the valley. But this representation is discrepant with the factive, more veridical representation consisting of our belief that all the referent entities in the scene are static and involve no motion.

- (5) The cliff wall faces toward/away from/into/past the valley.

6.2.2 Alignment Paths

The *alignment path* type of orientation involves a stationary straight linear object with a point-type front. The orientation of such a linear object is here conceptualized

linguistically—and perhaps perceived—in terms of something intangible moving along the axis of the object, emerging from its front end, and continuing straight along a prepositionally determined path relative to some distal object. As it happens, the English constructions that evoke this arrangement are not free to represent just any orientation, but are limited to the two cases where the linear object is aligned with the distal object—the front being the end either closer to or further from the distal object, the sentences in (6) illustrate this type.⁴

(6) The snake is lying toward/away from the light.

Here the snake is the linear object with its head as the point-type front, and the light is the distal object. Of note, this construction combines a verb of stationariness, *lie*, with a path preposition, *toward* or *away from*, that coerces the verb’s semantic properties. A sentence with *lie* alone would permit an interpretation of the snake as coiled and, say, pointing only its head at or away from a light. But in the normal understanding of (6), the snake’s body forms an approximately straight line that is aligned with the light. That is, the addition of a path preposition in this construction has the effect of forcing a fictive alignment path interpretation that requires a straight-line contouring of the snake’s body. The hypothesis that fictive orientation paths emerge from an object’s front and move away from the object correctly accounts for the fact that the sentence with “toward” refers to the head end of the snake as the end closer to the light, while the sentence with “away from” indicates that the head end is the further end.

6.2.3 Demonstrative Paths

The *demonstrative* type of orientation path also involves a linear object with a point-type front from which an intangible line emerges. But here the fictively moving line functions to direct or guide someone’s attention along its path. The particular orientation of the linear object can either be an independent factor that simply occasions an instance of directing someone’s attention, or can be intentionally set to serve the purpose of attentional guidance. This function of directing a person’s attention can be the intended end result of a situation. Or it can be a precursor event that is instantiated or followed by another event, such as the person’s directing his or her gaze, or moving bodily along the fictive path.

Thus, in the examples in (7), a linear object with a front end, such as an arrow or an extended index finger, seems to emit an intangible line from its front end. This line moves in the direction of the object’s orientation so as to direct someone’s attention, gaze, or physical motion along the path specified by the preposition.

- (7) a. I/The arrow on the signpost pointed toward/away from/into/past the town.
 b. I pointed/directed him toward/past/away from the lobby.

6.2.4 Targeting Paths

In a *targeting path*, an Agent intentionally sets the orientation of a front-bearing object so that the fictive line that is conceptualized or perceived as emerging from this front follows a desired path relative to the object's surroundings. This fictive motion establishes a path along which the Agent further intends that a particular subsequent motion will travel. This subsequent motion either is real or is itself fictive. Although comparatively complex, something like this sequence of intentions and actions, with a single or double fictive path, seems to underlie our concepts of aiming, sighting, or targeting. Consider the sentences in (8) in this regard.

(8) I pointed/aimed (my gun/camera) into/past/away from the living room.

Here the case of a bullet shot from the aimed gun exemplifies real motion following the preset fictive path. In contrast, the camera provides an instance of fictive motion following the fictive path, with a so-conceived photographic probe emerging from the camera's front.

One might ask why the camera example is included here under the targeting type of orientation path, rather than below under sensory paths along with "looking." The reason is that the act of looking is normally treated differently in English from the act of photographic shooting. We normally do not speak of "aiming" or "pointing" our gaze, and we do not conceive of the act of looking as involving first the establishment of a targeting path and then a viewing along that path.

6.2.5 Line of Sight

Line of sight is a concept that underlies a number of linguistic patterns, and perhaps also a component of perceptual structure. It is an intangible line emerging from the visual apparatus canonically located on the front of an animate or mechanical entity. The present discussion deals only with *lateral* motion of the line of sight, that is, with shifts in its orientation. Axial fictive motion along the line of sight will be treated in section 6.5 on sensory paths. Additional evidence for treating the shifting line of sight as an orientation path is that the sentences exhibiting this phenomenon can use not just sensory verbs like *look* but also nonsensory verbs like *turn*.

In the examples in (9), the object with the vision-equipped front—whether my head with its eyes or the camera with its lens—swivels, thus causing the lateral motion of the line of sight that emerges from that front. The path preposition specifies the particular path that the line of sight follows. Consider how fictive motion is at work in the case of a sentence like *I slowly turned/looked toward the door*. A path preposition like *toward* normally refers to a Figure object's executing a path in the direction of the Reference Object, where the distance between the two objects progressively decreases. But what within the situation depicted by the example sentence could be

exhibiting these characteristics? The only object that is physically moving is my turning head, yet that object stays in the same location relative to the door, not moving closer to it. Apparently what the preposition *toward* in this sentence refers to is the motion of the line of sight that emerges from my eyes. As I turn my head in the appropriate clockwise or counterclockwise direction, this line of sight does indeed follow a path in the direction of the door and shorten its distance from it.

- (9) I slowly turned/looked—//I slowly turned my camera—
toward the door./around the room./away from the window./
from the painting, past the pillar, to the tapestry.

We can note that English allows each linguistic form in a succession of path indications to specify a different type of fictive motion. Thus, in (10), the first path-specifying form, the satellite *down*, indicates a lateral motion of a line of sight, of the type discussed in this section. Under its specification, the likely interpretation is that my line of sight is initially horizontal (I am looking "straight ahead"), and then swivels downward so as to align with the axis of a well. The second spatial form, the preposition *into*, indicates that once my line of sight is oriented at a downward angle, then the fictive motion of my vision proceeds away from me axially along the line of sight, thus entering the well.

- (10) I quickly looked down into the well.

6.3 Radiation Paths

The second type of emanation we consider is that of *radiation paths*. The linguistic conceptualization of a radiation path is of radiation emanating continuously from an energy source and moving steadily away from it. This radiation can additionally be understood to comprise a linear shaft and to subsequently impinge on a second object. This additional particularization is the only type treated here. In this type, then, the radiating event can be characterized as involving three entities: the radiator, the radiation itself, and the irradiated object. And this radiating event then involves three processes: the (generation and) emanation of radiation from the radiator, the motion of the radiation along a path, and the impingement of the radiation upon the irradiated object. A radiation path differs from an orientation path in that the latter consists of the motion of a wholly imperceptible line. In a radiation path, though, one can often indeed detect the presence of the radiation—for example, in the case of light radiation, one can see the light. What one cannot directly detect—and, hence, what remains imperceptible—is any motion of this radiation.

The sentences in (11) reflect the preceding characterization of radiation for the particular case of light in the way they are linguistically constructed. This linguistic

construction mainly involves the choices of subject, of path-specifying preposition, and of prepositional object. In both sentences, then, the general understanding is that the visible light is a radiation; that the sun is the source of the light (perhaps its generator, but at least its locus of origination); that the light emanates from the sun and moves steadily as a beam along a straight path through space; and that the light moves into the cave or impinges on its back wall to illuminate that spot.

- (11) a. The sun is shining into the cave/onto the back wall of the cave.
- b. The light is shining (from the sun) into the cave/onto the back wall of the cave.

Now, as compelling as this characterization of light radiation may be felt to be, it is, in the end, purely a conceptualization. Although physicists may tell us that photons in fact move from the sun to the irradiated object, we certainly cannot actually see any such occurrence. Therefore, any correspondence between the scientific characterization and the conceptualization of the phenomenon must be merely coincidental. In other words, the so-conceived motion of radiation from the radiator to the irradiated must be fictive motion. Because direct sight does not bring a report of light's motion, it must be other factors that lead to a conceptualization in terms of motion away from the sun, and we will speculate on those factors in section 6.6. At this point, however, the task is to suggest a number of viable alternatives to the normal conceptualization. These alternatives show that the unique appearance of this conceptualization cannot be explained by virtue of its being the only conceptualization possible.

One alternative conceptualization is that there is a radiation path, but that it moves in the reverse direction from that in the prevailing conceptualization. Imagine the following state of affairs. All matter contains or generates energy. The sun (or a comparable entity) attracts this energy. The sun draws this energy toward itself when there is a straight clear path between itself and the matter. Matter glows when its energy leaves it. The sun glows when energy arrives at it. An account of this sort is in principle as viable as the usual account. In fact, it is necessarily so, because any phenomenon that could be explained in terms of imperceptible motion from A to B must also be amenable to an explanation in terms of a complementary imperceptible motion from B to A. However, for all its equality of applicability, the fact is that this reverse-direction scenario is absent from—even resisted by—our normal conceptual apparatus. And it is certainly absent from extant linguistic constructions. Thus English lacks any sentence like that in (12), and we suspect that any counterpart formulation is universally absent from the languages of the world.

- (12) *The light is shining from my hand onto the sun.

The conceptualization that an object like the sun, a fire, or a flashlight produces light that radiates from it to another object is so intuitively compelling that it can be of value to demonstrate the viability of the reverse-direction conceptualization in different circumstances. Consider, for example, a vertical pole and its shadow on the ground. The sun-as-Source conceptualization here has the pole as blocking the light that would otherwise proceed from the sun onto the ground directly behind the pole. But the reverse-direction conceptualization works here as well. The sun attracts energy from the side of the pole facing it, but cannot do so from the portion of the ground directly behind the pole because there is no straight clear path between that portion of the ground and the sun—the pole blocks the transit of energy in the reverse direction. Because no energy is drawn out of the portion of the ground behind the pole, it fails to glow, whereas the portions of ground adjacent to it, from which energy is being directly drawn, do glow.

Or consider a fire. Here one can see that the surfaces of oneself facing the fire are brighter than the other surfaces and, in addition, one can feel that they are warmer as well. Further, this effect is stronger the closer one is to the fire. Once again, the fire-as-Source of both light and heat is not the only possible conceptualization. The same reverse-direction conceptualization used for the sun holds as well for the fire. The additions in this example are that when the fire attracts energy from the parts of one's body facing it, the departure of that energy causes not only a glow but also the sensation of warmth. (Such warmth is of course also the case for the sun, but more saliently associated with fire, hence saved for the present example). And the one further factor here is that the attraction that the fire exerts on an object such as one's body is stronger the closer it is.

The reverse-direction conceptualization is not the only feasible alternative to the prevailing conceptualization of a radiation path, itself a constellation of factors, any one of which can be challenged. The reverse-direction alternative attempted to invert the directionality of the fictive motion in the prevailing conceptualization. But we can also test out the factor which holds that a radiation path originates at one of the salient physical objects and terminates at the other. Thus we can check the viability of a conceptualization in which light originates at a point between the two salient objects and fictively moves out in opposite directions to impinge on each of those two objects. (13) tries to capture this conceptualization. However, this sentence does not work linguistically and the conceptualization it expresses seems wholly counterintuitive.

(13) *The light shone out onto the sun and my hand from a point between us.

Another assumption in the normal conceptualization we can try to challenge is that the radiation moves at all. Perhaps the radiation does not exhibit fictive motion at all

but rather rests in space as a stationary beam. But sentences like (14) show that this conceptualization, too, has neither linguistic nor intuitive viability.

(14) *The light hung between the sun and my hand.

6.4 Shadow Paths

The third type of emanation we consider is that of *shadow paths*. The linguistic conceptualization—and perhaps also a perception—of a shadow path is that the shadow of some object visible on some surface has fictively moved from that object to that surface. Sentences like those in (15) show that English suggests a conceptualization of this sort through its linguistic construction. Thus these sentences set up the nominal that refers to the shadow as the Figure; the object whose shadow it is as the Source; the surface on which the shadow is located as the Ground object, here functioning as Goal; the predicate as a motion verb like *throw*, *cast*, or *project*; and a path preposition such as *into*, *onto*, *across*, or *against*.

- (15) a. The tree threw its shadow down into/across the valley.
 b. The pillar cast/projected a shadow onto/against the wall.

We can note that with radiation paths, the argument could conceivably be made that the direction of the fictive motion proceeds from the sun to my hand, because that is the direction that photons actually travel. But however tenable a weak argument like this may be, even this argument could not be used in the case of shadow paths. For there is no theory of particle physics that posits the existence of “shadowons” that move from an object to the silhouette of its shadow.

6.5 Sensory Paths

One category of emanation paths well represented in language is that of *sensory paths*, including *visual paths*. This type of fictive motion involves the conceptualization of two entities, the *Experiencer* and the *Experienced*, and of something intangible moving in a straight path between the two entities in one direction or the other. By one branch of this conceptualization, the Experiencer emits a *Probe* that moves from the Experiencer to the Experienced and detects it upon encounter with it. This is the Experiencer-as-Source type of sensory path. By the other branch of the conceptualization, the experienced emits a *Stimulus* that moves from the Experienced to the Experiencer and sensorily stimulates that entity on encounter with it. This is the Experienced-as-Source type of sensory path. Sight, in particular, is thus treated either as a probing system that emanates from or is projected forth by a viewer so as to

detect some object at a distance, or else as a visual quality that emanates from some distal object and arrives at an individual, thereby stimulating a visual experience.

We can first illustrate this phenomenon using a nonagentive verb lexicalized so as to take the Experiencer as subject, namely, *see*. In (16) the two oppositely directed paths of fictive motion are represented by two different path phrases:

- (16) a. The enemy can see us from where they're positioned.
- b. ?The enemy can see us from where we're standing.

Some speakers have difficulty with with an experiencer-as-source sentence like (16b), but this difficulty generally disappears for the counterpart passive sentence, as shown in (17b).

- (17) a. We can be seen by the enemy from where they're positioned.
- b. We can be seen by the enemy from where we're standing.

And generally no problem arises at all for nonvisual sensory paths, such as those for audition or olfaction shown in (18).

- (18) a. I can hear/smell him all the way from where I'm standing.
- b. I can hear/smell him all the way from where he's standing.

The bidirectional conceptualizability of sensory paths can also be seen in alternatives of lexicalization. Thus, among the nonagentive vision verbs in English, *see* is lexicalized to take the Experiencer as subject and the Experienced as direct object, thereby promoting the interpretation of the Experiencer as Source. But *show* is lexicalized to take the Experienced as subject and can take the Experiencer as the object of the preposition *to*, thereby promoting the interpretation of the Experienced as Source. We illustrate in (19).

- (19) a. Even a casual passer-by can see the old wallpaper through the paint.
- b. The old wallpaper shows through the paint even to a casual passer-by.

Despite these forms of alternative directionality, fictive visual paths may generally favor the Experiencer as Source. This is the case for English, where some forms with the Experienced as Source offer difficulty to some speakers, and the use of a verb like *show* is minimal relative to that of a verb like *see*. Further, agentive verbs of vision in English are exclusively lexicalized for the Experiencer as subject and can take directional phrases only with the Experiencer as Source. As shown in (20a), this is the case with the verb *look*, which takes the Experiencer as subject and allows a range of directional prepositions. Here the conceptualization appears to be that the Agent subject volitionally projects his line of sight as a Probe from himself as Source along the path specified by the preposition relative to a Reference Object (the Experienced

is not named in this type of construction). However, there is no (20b)-type construction with *look* in which the visual path can be represented as if moving to the Experiencer as goal.

- (20) a. I looked into/toward/past/away from the valley.
 b. *I looked out of the valley (into my eyes).
 <where I am located outside the valley>

6.6 A Unifying Principle and an Explanatory Factor for Emanation Types

So far, this chapter has laid out the first-level linguistic phenomena that manifest different types of fictive emanation. It is now time to consider the principles that govern and the context that generalizes these phenomena.

In the preceding part of the chapter, the conceptualizations associated with the different types of emanation were treated as distinct. But underlying such diversity, one may discern commonalities that unite the various types and may posit still deeper phenomena that can account for their existence. We present here a unifying principle and an explanatory factor.

6.6.1 The Principle that Determines the Source of Emanation

For the emanation types in which a fictive path extends between two objects, we can seek to ascertain a cognitive principle that determines which of the two objects will be conceptualized as the source of the emanation, while the other object is understood as the goal. On examination, the following cognitive principle appears to be the main one in operation: the object taken to be the more active or determinative of the two is conceptualized as the source of the emanation. This will be called the “active-determinative principle.”

We can proceed through several realizations of this principle that have functioned in the earlier examples. Thus, as between the sun and my hand, or the sun and the cave wall, the sun is perceived as the brighter of the two objects. This greater brightness seems to lead to the interpretation that the sun is the more active object, in particular, more energetic or powerful. By the operation of the active-determinative principle, the sun will be conceptualized, and perhaps perceived, as the source of the radiation moving through space to impinge with the other object, rather than any of the alternative feasible conceptualizations presented earlier.

Another application of the active-determinative principle can be seen in shadow paths. As between, say, a pole and the shadow of the pole, the pole is the more determinative entity, while the shadow is the more contingent or dependent entity. This is understood from such evidence as that in total darkness or in fully diffuse

light, the pole is still there but no shadow is present. Further, one can move the pole and the shadow will move along with it, whereas there is no comparable operation performable on the shadow. By the operation of the active-determinative principle, the shadow-bearing object is thus conceptualized as generating the shadow, which then moves fictively from that object to an indicated surface. That is, it is by the operation of the principle that this interpretation of the direction of the fictive motion prevails, rather than any alternative interpretation such as that the shadow moves from the indicated surface to the physical object.

A further realization of the active-determinative principle can be seen in the case of agentive sensory paths, that is, ones with an Experiencer that acts as an intentional Agent as well as with an Experienced entity. Here it seems the very property of exercised agency leads to the interpretation that the Agent is more active than the Experienced entity, which is either inanimate or currently not manifesting relevant agency. By the operation of the active-determinative principle, then, the agentive Experiencer is conceptualized as the Source of the sensory path, whose fictive motion thus proceeds from the Experiencer to the Experienced. In the visual example presented earlier, *I looked into the valley*, because the referent of "I" is understood as an agentive Experiencer, while the referent of "valley" is understood as a nonagentive Experienced entity, the active-determinative principle requires that the Experiencer be conceptualized as the Source of the fictive sensory motion, and this, in fact, is the only available interpretation for the sentence.

The active-determinative principle also holds for those types of orientation paths that are agentive, for example, targeting paths and agentive demonstrative paths, where the active and determinative entity in the situation is the agent who fixes the orientation of the front-bearing object, such as a camera or the Agent's own arm with extended index finger. With our principle applying correctly again, it will be this object, positioned at the active-determinative locus, that will be conceptualized as the source of the fictive emanation.

The fact that nonagentive sensory paths can be conceptualized as moving in either of two opposite directions might at first seem to challenge the principle that the more active or determinative entity is treated as the source of fictive emanation. But this need not be the case. It may be that either object can, by different criteria, each be interpreted as the one that is more active than the other. For example, by one set of criteria, a nonagentively acting Experiencer, from whom a detectional probe is taken to emanate, is interpreted as more active than the entity probed. But under an alternative set of criteria, the Experienced entity taken to emit a stimulus is interpreted as being more active than the entity stimulated by it. Thus the active-determinative principle is saved. The task remaining, though, is to ascertain the additional cognitive criteria that ascribe greater activity to one set of phenomena or to a competing set,

and that are in effect in the absence of the principle's already known criteria (e.g., greater agency or energeticness).

Finally, there is a remainder of emanation types to which the active-determinative principle does not obviously apply in any direct way, namely, the nonagentive orientation path types: prospect paths, alignment paths, and nonagentive demonstrative paths. Here the fictive motion emanates from only one of the two relevant entities, but this entity is not apparently the more active or determinative of the two. In these cases, however, the directionality of the fictive motion may be set indirectly by the conceptual mapping of principle-determined cases onto the configuration, as described in the next section.

6.6.2 Possible Basis of Fictive Emanation and its Types

If we have correctly ascertained that the more active or determinative entity is conceptualized as the Source of fictive emanation, the next question to ask is why this should be the case. We speculate here that the active-determinative principle is a consequence of a foundational cognitive system every sentient individual has and experiences, that of agency. Specifically, the individual's exercise of agency functions as the model for the Source of emanation. We remain agnostic on whether the connection is learned or innate. If it is learned in the course of development, then each individual's experience of agency leads by steps to the conceptualization of fictive emanation. If it is innate, then something like the same steps may have been traversed by genetically determined neural configurations as these evolved. Either way, we can suggest something of the steps and their consequent interrelationships.

The exercise of agency can be understood to have two components, the generation of an intention and the realization of that intention (cf. Talmy 1976, forthcoming). An intention can be understood as one's desire for the existence of some new state of affairs where one has the capability to act in a way that will bring about that state of affairs. The realization component, then, is one's carrying out of the actions that bring about the new state of affairs. Such exercise of agency is experienced as both active and determinative. It is active because it involves the generation of intentions and of actions, and it is determinative because it remodels conditions to accord with one's desires. In this way, the characteristics of agency may provide the model for the active-determinative principle.

The particular form of agency that can best serve as such a model is that of an Agent's affecting a distal physical object—what can be called the “agent-distal object pattern.”⁵ In this pattern an Agent, say, intending to affect the distal object must either move to it with her whole body, reach to it with a body part, or cause (as by throwing) some intermediary object to move to it. The model-relevant characteristics

of this form of agency are that the determining event, the act of intention, takes place at the initial locus of the Agent, and the ensuing activity that finally affects the distal object progresses through space from that initial locus to the object. But these are also the characteristics of the active-determinative principle: namely, the more active or determinative entity is the Source from which fictive motion emanates through space until reaching the less active or determinative entity, the distal object. Hence one can posit that the pattern of agency affecting a distal object is the model on which the active-determinative principle is based.

In particular, we can see how the agent-distal object pattern can serve as the model for the two main agentive forms of emanation, namely, agentive demonstrative paths and agentive sensory paths. To consider the former case first, the specific agent-distal object pattern of extending the arm to reach for some object may directly act as the model for agentive demonstrative paths, such as an Agent extending his arm and pointing with his finger. In both cases, the extending arm typically exhibits actual motion away from the body along a line that connects with the target object, where, when fully extended, the arm's linear axis coincides with its path of motion. Possibly some role is played by the fact that the more acute tapered end of the arm, the fingers, leads during the extension and is furthest along the line to the object when the arm is fully extended. Such an agentive demonstrative path might then in turn serve as the model for the nonagentive type, for example, one associated with a figure like an arrow, whose linear axis also coincides with the line between the arrow and the distal object, and whose tapered end is the end closest to the distal object and the end conceptualized as the Source from which the demonstrative line emanates.

Similarly, we can see parallels between the agent-distal object pattern, in which an Agent executes fictive motion toward distal object, and agentive visual sensory paths, in which an Experiencer projects a fictive line of sight from himself to the distal object. Specifically, like the Agent, the Experiencer is active and determinative; like the Agent, the Experiencer has a front; like the Agent's moving along a straight line between his front and the distal object, the intangible line of sight moves in a straight line between the front of the Experiencer and the distal object; like this line's moving away from the initial locus of the Agent, the visual sensory path moves away from the Experiencer as Source; like the Agent's motion continuing along this line until it reaches the object, the visual sensory path progresses until it encounters the distal object. Thus the perception of the Agent's motion in the physical world appears to be mapped onto the conceptualization of an intangible entity moving along a line. Again, such a mapping might either be the result of learning during an individual's development, or might have been evolutionarily incorporated into the perceptual and

conceptual apparatus of the brain. Either way, an organism's production of factive motion can become the basis for the conceptualization of fictive motion.

In turn, this agentive visual type of fictive emanation may serve as the model for several nonagentive emanation types. In particular, this modeling may occur by the conceptual mapping or superimposition of a schematized image—that of an Experiencer's front emitting a line of sight that proceeds forward into contact with a distal object—onto situations amenable to a division into comparably related components. Thus, in the prospect type of orientation path, the Experiencer component may be superimposed onto, say, a cliff, with her face corresponding to the cliff wall, with her visual path mapped onto the conceptualized schematic component of a prospect line moving away from the wall, and with the distal object mapped onto the vista toward which the prospect line progresses.⁶

In a similar way, the schema for the agentive visual path may get mapped onto the radiation situation, where the Experiencer, as the active determinative Agent, is associated with the most energetic component of the radiation scene—the brightest component in the case of light, say, the sun. The visual path is mapped onto the radiation itself, for example, onto light visible in the air (especially, say, a light beam, as through an aperture in a wall), and the distal object is mapped onto the less bright object in the scene. The direction of motion conceptualized for the visual path is also mapped onto the radiation, which is thus conceptualized as moving from the brighter object to the duller object. An association of this sort can explain why much folk iconography depicts the sun or moon as having a face that looks outward.

As for shadow paths, the model may be the situation in which the agentive Experiencer herself stands and views her own shadow from where she is located. Once again, the visual path moving from this Experiencer to the ground location of the shadow is mapped onto the conceptualization of the fictive path that the shadow itself traverses from the solid body onto the ground. A reinforcement for this mapping is that the Experiencer is determinative as the Agent and the solid object is determinative over the shadow dependent on it.

The only emanation types not yet discussed in terms of mapping are the non-agentive sensory paths that can proceed in either direction. The direction from Experiencer to Experienced is clear because that is the same as for agentive viewing. We may account for the reverse case—where the Experienced emits a Stimulus—on the grounds that it, too, can serve as a receptive frame onto which to superimpose the model of an Agent emitting a visual path. What is required is simply the conclusion that the conceptualization of an object emitting a Stimulus can be taken as active enough to be treated as a kind of modest agency in its own right, and hence to justify this conceptual superimposition of an Agent onto it.

6.7 Relation of Emanation in Language to Counterparts in Other Cognitive Systems

In this section we present a number of apparent similarities in structure or content between the emanation category of fictive motion in language and counterparts of emanation in cognitive systems other than that of language. We mainly consider similarities that language has to perception and to cultural conceptual structure, as well as to folk iconography, which may be regarded as a concrete symbolic expression of perceptual structure. A brief description of our model of cognitive organization, referred to in the introduction, will first provide the context for this comparison.

6.7.1 "Overlapping Systems" Model of Cognitive Organization

Converging lines of evidence in the author's and others' research point to the following picture of human cognitive organization. Human cognition comprehends a certain number of relatively distinguishable cognitive systems of fairly extensive compass. This research has considered similarities and dissimilarities of structure—in particular, of conceptual structure—between language and each of these other cognitive systems: visual perception, kinesthetic perception, reasoning, attention, memory, planning, and cultural structure. The general finding is that each cognitive system has some structural properties that may be uniquely its own; some further structural properties that it shares with only one or a few other cognitive systems; and some fundamental structural properties that it has in common with all the cognitive systems. We assume that each such cognitive system is more integrated and interpenetrated with connections from other cognitive systems than is envisaged by the strict modularity notion (cf. Fodor 1983). We call this view the "overlapping systems" model of cognitive organization.⁷

6.7.2 Fictive Emanation and Perception

The visual arrays that might yield perceptual parallels to the emanation type of fictive motion have been relatively less investigated by psychological methods than in the case of other categories of fictive motion (see below). One perceptual phenomenon related to orientation paths has been demonstrated by Palmer (1980) and Palmer and Bucher (1981), who found that in certain arrays consisting of co-oriented equilateral triangles, subjects perceive all the triangles at once pointing by turns in the direction of one or another of their common vertices. Moving the array in the direction of one of the common vertices biases the perception of the pointing to be in the direction of that vertex, although these experiments did not test for the perception of an intangible line emerging from the vertex currently experienced as the pointing "front" of each triangle or of the array of triangles. One might need experiments, for example,

that test for any difference in a subject's perception of a further figure depending on whether or not a fictive line was perceived to emerge from the array of triangles and pass through that figure. But confirmation of a perceptual analogue to emanation paths must await such research.

We can also note that Freyd's work on "representational momentum" (e.g., Freyd 1987) does not demonstrate perception of orientation paths. This work involved the sequential presentation of a figure in successively more forward locations. The subjects did exhibit a bias toward perceiving the last-presented figure further ahead than its actual location. But this effect is presumably due to the factively forward progression of the figure. To check for the perceptual counterpart of linguistic orientation paths, experiments of this type would need to test subjects on the presentation of a single picture containing a forward-facing figure with an intrinsic front.

The robust and extensive representation of fictive emanation in language calls for psychological research to test for parallels to this category of fictive motion in perception. That is, the question remains whether the appropriate experimental arrangements will show particular perceptions for this category that accord with the general fictivity pattern, hence with the concurrent perception of two discrepant representations, one of them more palpable and veridical than the other. Consider, for example, visual arrays that include various front-bearing objects, designed to test the perception of fictive orientation paths in their several distinct types—prospect paths, alignment paths, demonstrative paths, and targeting paths. One would need to determine whether subjects, on viewing these arrays, see the factive stationariness of the depicted objects at the fully palpable level of perception, but concurrently sense the fictive motion of something intangible emanating from the objects' fronts at a faintly palpable level of perception.

Similarly, to probe for visual counterparts of linguistic radiation paths, research will need to test for anything like a fictive and less palpable perception of motion along a light beam, in a direction away from the brighter object, that is concurrent with, perhaps superimposed on, the factive and more palpable perception of the beam as static. Similarly, to test for a visual parallel to linguistic shadow paths, experimental procedures will need to probe whether subjects, on viewing a scene that contains an object and its shadow, have some fictive, less palpable sense of the shadow as having moved from that object to the surface on which it appears, concurrently with a factive and palpable perception of everything within the scene as stationary. Finally, to check for a perceptual analogue of visual sensory paths in language, one can use either a scene that depicts someone looking or a subject's own process of looking at entities to ascertain whether subjects simply perceive a static array of entities or superimpose on that array a less palpable perception of motion along the probing line of sight.

6.7.3 Fictive Emanation and Folk Iconography

Fictive representations that are normally only sensed at a lower level of palpability can sometimes be modeled by fully palpable representations. An example to be cited below is the use of stick figure drawings or of pipe cleaner sculptures to explicitly image objects' schematic structure, which is normally only sensed. In the same way, various other aspects of fictive emanation normally only sensed have been made explicit in the concrete depictions of folk iconography.

For example, fictive sensory paths of the agentive visual type are linguistically conceptualized as intangible lines that Agents project forward from their eyes through space into contact with distal objects. But this is exactly the character of Superman's “X-ray vision” as depicted in comic books. Superman sends forth from his eyes a beam of X-rays that penetrates opaque materials to make contact with an otherwise obscured object and permits it to be seen. Note that Superman's X-ray vision is not depicted as stimuli that emanate from the obscured object and proceed toward and into Superman's eyes where they might be perceptually registered. Such an Experienced-to-Experiencer path direction might have been expected from our understanding of X-ray equipment, where the radiation moves from the equipment onto a photographic plate on which the image is registered. This plate might have been analogized to Superman's eyes, but the conceptual model in which the Agent emits a sensory Probe appears to hold sway in the cartoon imagery.

Comparable examples based on the linguistic conceptualization of an Agent emitting a visual Probe are represented not only by grammatical constructions and other closed-class forms, but also by metaphoric expressions. Thus the expression “to look daggers at,” as in *Jane looked daggers at John*, represents the notion that Jane's mien, reflecting a current feeling of hate for John, can be elaborated as the projection of weapons from her eyes to John; indeed, cartoon depictions actually show a line of daggers going from the experiencer's eyes to the body of the experienced.

The linguistic conceptualization of fictive demonstrative paths emerging from the point-type front of a linear object, as from a pointing finger, seems also to parallel a type of iconographic depiction. This is the depiction of magical power beams that an Agent can project forth from his extended fingertips. For example, movies and comic books often have two battling sorcerers raise their extended hands and direct destructive beams at each other.

Finally, it is the author's observation—though a careful study would be needed to confirm this—that in the process of drawing the sun, schematically, after completing a circle for the body of the sun, both children and adults represent its radiation with lines drawn radially outward from the circle, not inward toward it. If so, this iconographic procedure reflects the linguistic conceptualization of fictive radiation paths as emanating and moving off from the brightest object. Further, iconographic

representations of the sun and moon often depict a face on the object, as if to represent the object as containing or comprising an Agent that is emitting the radiation of light. As noted in section 6.6.2, a representation of this sort can be attributed to the mapping of the schema of an agentive visual sensory path onto the radiation situation, much as it may be mapped onto other fictive motion types.

6.7.4 Relation of Fictive Emanation to Ghost Physics and Other Anthropological Phenomena

We can discern a striking similarity between fictive motion—in particular, orientation paths—and the properties exhibited by ghosts or spirits in the belief systems of many traditional cultures. The anthropologist Pascal Boyer (1994) sees these properties as a culturally pervasive and coherent conceptual system, which he calls “ghost physics.” Boyer holds that ghost and spirit phenomena obey all the usual causal expectations for physical or social entities, with only a few exceptions that function as “attention attractors.” Certain of these exceptions have widespread occurrence across many cultures, such as invisibility or the ability to pass through walls or other solid objects, but other kinds of potential exceptions, which on other grounds might have seemed just as suited for conceptualization as special properties, instead appear never to occur. An example of this is temporally backward causality; that is, cultural belief systems seem universally to lack a concept that a ghost can at one point in time bring about some state of affairs at a prior point of time.

Boyer has no explanation for the selection of particular exceptions that occur in ghost physics and may even find them arbitrary. However, we can suggest that the pattern of standard and exceptional properties is structured and cognitively principled. In fact, the findings reported in this chapter may supply the missing account. The exceptional phenomena found to occur in ghost physics may be the same as certain cognitive phenomena that already exist in other cognitive systems, and then are tapped for service in cultural spirit ascriptions. The linguistic expression of fictive demonstrative paths and its gestural counterpart may well afford the relevant properties.

To consider gesture first, if I, for example, am inside a windowless building and am asked to point toward the next town, I will not, through gesticulations, indicate a path that begins at my finger, leads through the open doorway, out the exit of the building, turns around and then moves in the direction of the town. On the contrary, I will simply extend my arm with pointed finger in the direction of the town, regardless of the structure around me. That is to say, the demonstrative path, effectively conceptualized as an intangible line emerging from the finger, itself has the following crucial properties: (1) it is invisible, and (2) it passes through walls—the very same properties ascribed to spirits and ghosts.

These same properties hold for the conceptualization that accompanies the linguistic expression of fictive demonstrative paths. For example, in the set of sentences *this arrow points to/toward/past/away from the town*, the use of any of the directional prepositions suggests the conceptualization of an intangible line emerging from the front end of the arrow, following a straight course coaxial with the arrow's shaft, and moving along the path represented by the preposition. Once again, this imaginal line is invisible and would be understood to pass through any material objects present on its path.

In addition to such demonstrative paths, we can observe further relations between cultural conceptualizations and another type of fictive emanation, that of agentive visual paths. Consider the notion of the "evil eye," found in the conceptual systems of many cultures. In a frequent conception of the evil eye, an agent who bears malevolent feelings toward another person is able to transmit the harmful properties of these feelings along the line of his gaze at the other person. This is the same schema as for a fictive visual path: the Agent as Source projecting forth something intangible along his line of sight to encounter with a distal object.

Relations between fictive motion and cultural conceptualizations extend still further. One may look to such broadly encountered cultural concepts as those of mana, power, fields of life force, or magical influence emanating from entities; these forms of imagined energy—just like the fictive emanations of linguistic construals—are conceptualized (and perceived?) as being invisible and intangible, as being (generated and) emitted by some entity, as propagating in one or more directions away from that entity, and in some forms as then contacting a second distal entity that they may affect. The structural parallel between such anthropological concepts of emanation and the emanation type of fictive motion we have here described for language is evident and speaks to a deeper cognitive connection.

It thus seems that the general fictivity pattern generates the imaginal schemas of fictive motion in the cognitive systems not only of language and of visual perception, but also of cultural cognition, specifically in its conceptualizations of spirit and power. That is, in the cognitive culture system, the structure of such conceptions as ghost phenomena, harmful influence, and magical energy appears not to be arbitrary. Nor does it exhibit its own mode of construal or constitute its own domain of conceptual constructs of the sort posited, for example, by Keil (1989) and Carey (1985) for other categories of cognitive phenomena. Rather, it is probably the same as or a parallel instance of conceptual organization already extant in other cognitive systems. In terms of the "overlapping systems" framework outlined above, general fictivity of this sort is thus one area of overlap across at least the three cognitive systems of language, visual perception, and cultural cognition.

6.8 Further Categories of Fictive Motion

As indicated earlier, language exhibits a number of categories of fictive motion beyond the emanation type treated thus far. We here briefly sketch five further categories; for each, we suggest some parallels in visual perception that have already been or might be examined.⁸ The purpose of this section is to enlarge both the linguistic scope and the scope of potential language-perception parallelism. In the illustrations that follow, the fictive motion sentences are provided, as a foil for comparison, with factive motion counterpart sentences, shown within brackets.

6.8.1 Pattern Paths

The *pattern paths* category of fictive motion in language involves the fictive conceptualization of some configuration as moving through space. In this type, the literal sense of a sentence depicts the motion of some arrangement of physical substance along a particular path, while we factively believe that this substance is either stationary or moves in some way other than along the depicted path. For the fictive effect to occur, the physical entities must factively exhibit some form of motion, qualitative change, or appearance/disappearance, but these in themselves do not constitute the fictive motion. Rather, it is the pattern in which the physical entities are arranged that exhibits the fictive motion. Consider the example in (21).

(21) Pattern paths

As I painted the ceiling, (a line of) paint spots slowly progressed across the floor.

[cf. As I painted the ceiling, (a line of) ants slowly progressed across the floor.]

Here each drop of paint does factively move, but that motion is vertically downward in falling to the floor. The fictive motion, rather, is horizontally along the floor and involves the linear pattern of paint spots already located on the floor at any given time. For this fictive effect, one must in effect conceptualize an envelope located around the set of paint spots or a line located through them. The spots thus enclosed within the envelope or positioned along the line can then be cognized as constituting a unitary Gestalt linear pattern. The appearance of a new paint spot on the floor in front of one end of the linear pattern can then be conceptualized as if that end of the envelope or line extended forward so as now to include the new spot. Such is the forward fictive motion of the configuration. By contrast, if the sentence were to be interpreted literally—that is, if the literal reference of the sentence were to be treated as factive—one would have to believe that the spots of paint physically slid forward along the floor.

In one respect, the pattern paths type of fictive motion is quite similar to the emanation type. In both these categories of fictive motion, an entity that is itself fictive—an imaginal construct—moves fictively through space. One difference, though, is that the emanation type does not involve the factive motion of any elements within the referent scene. Accordingly, it must depend on a principle—the active-determinative principle—to fix the source and direction of the fictive motion. But the pattern paths type does require the factive motion or the change of some components of the referent situation for the fictive effect to occur; indeed, this determines the direction of the fictive motion, so that no additional principle need come into play.

The perceptual phenomena generally termed *apparent motion* in psychology would seem to include the visual counterpart of the pattern paths type of fictive motion in language. But to establish the parallel correctly, one may need to subdivide apparent motion into different types. Such types are perhaps largely based on the speed of the process viewed and, one may speculate, involve different perceptual mechanisms. Most research on apparent motion has employed a format like that of dots in two locations appearing and disappearing in quick alternation. Here, within certain parameters, subjects perceive a single dot moving back and forth between the two locations. In this fast form of apparent motion, the perceptual representation most palpable to subjects is in fact that of motion, and thus would not correspond to the linguistic case.

On the other hand, there may exist a slower type of apparent motion that can be perceived and that would parallel the linguistic case. One example might consist of a subject viewing a row of light bulbs in which one after another bulb is briefly turned on at consciously perceivable intervals. Here, it may be surmised, a subject would have an experience that fits the general fictivity pattern. The subject will perceive at a higher level of palpability, that is, as factive, the stationary state of the bulbs, as well as the periodic flashing of a bulb at different locations. But the subject would concurrently perceive at a lower level of palpability—and assess it as being at a lower level of veridicality—the fictive motion of a seemingly single light progressing along the row of bulbs.

6.8.2 Frame-Relative Motion

With respect to a global frame of reference, a language can factively refer to an observer as moving relative to the observer's stationary surroundings. This condition is illustrated for English in (20a) and is diagrammed in figure 6.1a. But a language can alternatively refer to this situation by adopting a local frame around the observer as center. Within this frame, the observer can be represented as stationary and her surroundings as moving relative to her from her perspective. This condition is

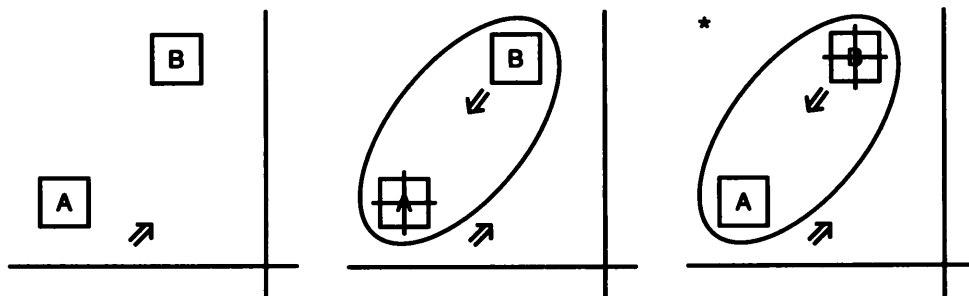


Figure 6.1
Frame-relative motion: global and local.

illustrated in (20b) and diagrammed in figure 6.1b. This condition is thus a form of fictive motion, one in which the factively stationary surroundings are fictively depicted as moving. In a complementary fashion, this condition also contains a form of fictive stationariness, for the factively moving observer is now fictively depicted as stationary. Stressing the depiction of motion, we term the fictive effect here *observer-based frame-relative motion*.

Further, a language can permit shifts between a global and a local framing of a situation within a single sentence. For instance, (22C) shifts from the global frame to the local frame and, accordingly, shifts from a directly factive representation of the spatial conditions to a fictive representation. But one condition no language seems able to represent is the adoption of a conceptualization that is part global and part local, and accordingly, part factive and part fictive. Thus English is constrained against sentences like (22D), which suggests the adoption of a perspective point midway between the observer and her surroundings.

(22) Frame-relative motion: With factively moving observer

A. *Global frame: Fictive motion absent*

I rode along in the car and looked at the scenery we were passing through.

B. *Local frame: Fictive motion present*

I sat in the car and watched the scenery rush past me.

[cf. I sat in the movie set car and watched the backdrop scenery rush past me.]

C. *Shift in midreference from global to local frame, and from factive to fictive motion*

I was walking through the woods and this branch that was sticking out hit me.

[cf. I was walking through the woods and this falling pinecone hit me.]

D. *Lacking: Part-global, part-local frame with part-factive, part-fictive motion*

***We and the scenery rushed past each other.**

[cf. We and the logging truck rushed past each other.]

In the preceding examples, the observer was factively in motion while the observed (e.g., the scenery) was factively stationary—properties expressed explicitly in the global framing. In a complementary fashion, a sentence can also express a global framing in which, factively, the observer is stationary while the observed moves. This situation is illustrated in (23 Aa, Ab). However, this complementary situation differs from the earlier situation in that it cannot undergo a local reframing around the stationary observer as center. If such a local frame were possible, one could find acceptable sentences that fictively depict the observer as moving and the observed as stationary. But sentences attempting this depiction—for example, (23 Ba) with a uniform local framing and (23 Bb) with a shift from global to local framing—are unacceptable. The unacceptable fictive local framing that they attempt is diagrammed in figure 6.1c.

(23) Frame-relative motion: With factively stationary observer

A. *Global frame: Fictive motion absent*

a. The stream flows past my house.

b. As I sat in the stream, its water rushed past me.

B. *Local frame: Blocked attempt at fictive motion*

a. *My house advances alongside the stream.

b. *As I sat in the stream, I rushed through its water.

We can suggest an account for the difference between moving and stationary observers in their acceptance of fictive local framing. The main idea is that stationariness is basic for an observer. Accordingly, if an observer is factively moving, a sentence is free to represent the situation as such, but a sentence may also “ratchet down” its representation of the situation to the basic condition in which the observer is stationary. However, if the observer is already stationary, that is, already in the basic state, then a sentence may only represent the situation as such, and is not free to “ratchet up” its representation of the situation into a nonbasic state.

If this explanation holds, the next question is why it should be that stationariness is basic for an observer. We can suggest a developmental account. An infant experiences optic flow from forward motion while being held by a parent long before the stage at which it locomotes, a stage at which it will agentively bring about optic flow itself. That is, before the infant has had a chance to integrate its experience of moving into its perception of optic flow, it has months of experience of optic flow without an experience of motion. This earlier experience may be processed in terms of the

surrounding world as moving relative to the self fixed at center. This experience may be the more foundational one and persist to show up in subtle effects of linguistic representations like those just seen.

One possible corroboration of this account can be cited. Infants at the outset do have one form of agentive control over their position relative to their surroundings, namely, turning the eyes or head through an arc. Rather than the forward type of optic flow just discussed, this action brings about a transverse type, although not extended rotation. Because the infant can thus integrate the experience of motor control in with experience of transverse optic flow at a foundational level, we should not expect to find a linguistic effect that treats observer stationariness as basic relative to an observer's arc-sized turning motion. Indeed, English, for one language, typically permits only factive representations of such turning by an observer, for example, *As I quickly turned my head, I looked over all the room's decorations*. It does not typically ratchet down to a fictive stationary state for the observer, as in **As I quickly turned my head, the room's decorations sped by in front of me*. A sentence of the latter sort would be used only for special effect, not in the everyday colloquial way the forward motion case is treated. On the other hand, as still further corroboration, because extended spinning is not part of the infant's early experience, it should behave like forward translational motion and permit a linguistic reframing. Indeed, this is readily found, as in English sentences like *As our space shuttle turned, we watched the heavens spin around us*, or *I rode on the carousel and watched the world go round*.

Psychological experiments have afforded several probable perceptual parallels to frame-relative motion in language. One parallel is the "induced motion" of the "rod and frame" genre of experiments. Here, prototypically, while a rectangular shape that surrounds a linear shape is factively moved, some subjects fictively perceive this frame as stationary while the rod moves in a complementary manner. However, this genre of experiments is not observer-based in our sense because the observer is not one of the objects potentially involved in motion. Closer to our linguistic case is the "motion aftereffect," present where a subject has been spun around and then stopped. The subject factively knows that he is stationary, but concurrently experiences a perception—assessed as less veridical, hence fictive—of the surroundings as turning about him in the complementary direction. Perhaps the experimental situation closest to our linguistic type would in fact be a subject's moving forward through surroundings, much as when riding in a car. The question is whether such a subject will concurrently perceive a factive representation of herself as moving through stationary surroundings, and a fictive representation of herself as stationary with the surroundings as moving toward and past her.

6.8.3 Advent Paths

An *advent path* is a depiction of a stationary object’s location in terms of its arrival or manifestation at the site it occupies. The stationary state of the object is factive, whereas its depicted motion or materialization is fictive and, in fact, often wholly implausible. The two main subtypes of advent paths are *site arrival*, involving the fictive motion of the object to its site, and *site manifestation*, which is not fictive motion but fictive change, namely the fictive manifestation of the object at its site. This category is illustrated in (22).

(24) Advent paths

A. Site arrival

1. *With active verb form*

- a. The palm trees clustered together around the oasis.

[cf: The children quickly clustered together around the ice cream truck.]

- b. The beam leans/tilts away from the wall.

[cf: The loose beam gradually leaned/tilted away from the wall.]

2. *With passive verb form*

- c. Termite mounds are scattered/strewn/spread/distributed all over the plain.

[cf. Gopher traps were scattered/strewn/spread/distributed all over the plain by a trapper.]

B. Site manifestation

- d. This rock formation occurs/recurs/appears/reappears/shows up near volcanoes.

[cf. Ball lightning occurs/recurs/appears/reappears/shows up near volcanoes.]

For a closer look at one site arrival example, (24a) uses the basically motion-specifying verb *to cluster* for a literal but fictive representation of the palm trees as having moved from some more dispersed locations to their extant neighboring locations around the oasis. But the concurrent factive representation of this scene is contained in our belief that the trees have always been stationary—located in the sites they occupy. Comparably, the site manifestation example in (24d) literally represents the location of the rock formation at the sites it occupies as the result of an event of materialization or manifestation. This fictive representation is concurrent with our believed factive representation of the rock formation as having stably occupied its sites for a very long time.

We can cite two psychologists who have made separate proposals for an analysis of visual forms that parallels the linguistic site arrival type of fictive motion. Pentland

(1986) describes the perception of an articulated object in terms of a process in which a basic portion of the object, for example, its central mass, has the remaining portions moved into attachment with it. An example is the perception of a clay human figure as a torso to which the limbs and head have been affixed. Comparably, Leyton (1992) describes our perception of an arbitrary curved surface as a deformed version of a simple surface; for example, a smooth closed surface is described as the deformation of a sphere, one that has undergone protrusion, indentation, squashing, and resistance. He shows that this set of processes corresponds to the psychologically salient causal descriptions that people give of shapes, say, of a bent pipe or a dented door. In a similar way, as described in the tradition of Gestalt psychology, certain forms are regularly perceived not as original patterns in their own right, but rather as the result of some process of deformation applied to an unseen basic form. An example is the perception of a Pac-Man-shaped figure as a circle with a wedge-shaped piece removed from it.

To consider this last example in terms of our general fictivity pattern, a subject looking at such a Pac-Man shape may concurrently experience two discrepant perceptual representations. The factive representation, held to be the more veridical and perceived as more palpable, will be that of the static Pac-Man configuration *per se*. The fictive representation, felt as being less veridical and perceived as less palpable, will consist of an imagined sequence that starts with a circle, proceeds to the demarcation of a wedge shape within the circle, and ends with that wedge exiting or being removed from the circle.

6.8.4 Access Paths

An *access path* is a depiction of a stationary object's location in terms of a path that some other entity might follow to the point of encounter with the object. What is factive here is the representation of the object as stationary, without any entity traversing the depicted path; what is fictive is the representation of some entity traversing the depicted path, whether this is plausible or implausible. Though it is not specified, the fictively moving entity can often be imagined as being a person, some body part of a person, or the focus of a person's attention, depending on the particular sentence, as can be seen in the examples of (25).

(25) Access paths

- a. The bakery is across the street from the bank.
[cf. The ball rolled across the street from the bank.]
- b. The vacuum cleaner is down around behind the clothes hamper.
[cf. I extended my arm down around behind the clothes hamper.]
- c. The cloud is 1,000 feet up from the ground.
[cf. The balloon rose 1,000 feet up from the ground.]

In greater detail, (25a) characterizes the location of the bakery in terms of a fictive path that begins at the bank, proceeds across the street, and terminates at the bakery. This path could be followed physically by a person walking, or perceptually by someone shifting the focus of his gaze, or solely conceptually by someone shifting her attention over her mental map of the vicinity. The depicted path can be reasonable for physical execution, as when I use (25a) to direct you to the bakery when we are inside the bank. But the same depicted path may also be an improbable one, as when I use (25a) to direct you to the bakery when we are on its side of the street—it is unlikely that you will first cross the street, advance to the bank, and then recross to find the bakery. Further, a depicted access path can also be physically implausible or impossible. Such is the case for referents like that in *That quasar is 10 million light-years past the North Star*. Apart from the use of fictive access paths such as these, an object’s location can generally also be directly characterized in a factive representation, as in *The bakery and the bank are opposite each other on the street*.

Does the fictivity pattern involving access paths occur perceptually? We can suggest a kind of experimental design that might test for the phenomenon. Subjects can be shown a pattern containing some point to be focused on, where the whole can be perceived factively as a static geometric Gestalt and/or fictively as involving paths leading to the focal point. Perhaps an example would be a “plus” figure with the letter *A* at the top point and, at the left-hand point, a *B* to be focused on. A subject might factively and at a high level of palpability perceive a static representation of this figure much as just described, with the *B* simply located on the left. But concurrently, the subject might fictively and at a lower level of palpability perceive the *B* as located at the endpoint of a path that starts at the *A* and, say, either slants directly toward the *B*, or moves first down and then left along the lines making up the “plus.”

6.8.5 Coverage Paths

A *coverage path* is a depiction of the form, orientation, or location of a spatially extended object in terms of a path over the object’s extent. What is factive here is the representation of the object as stationary and the absence of any entity traversing the depicted path. What is fictive is the representation of some entity moving along or over the configuration of the object. Though it is not specified, the fictively moving entity can often be imagined as being an observer, the focus of attention, or the object itself, depending on the particular sentence, as can be seen in the examples of (26). Note that in (26a) the fictive path is linear, in (26b) it is radially outward over a two-dimensional plane, and in (26c) it is the lateral motion of a line (a north-south line advancing eastward), that is further correlated with a second fictive change (increasing redness).

(26) Coverage paths

- a. The fence goes/zigzags/descends from the plateau to the valley.
[cf. I went/zigzagged/descended from the plateau to the valley.]
- b. The field spreads out in all directions from the granary.
[cf. The oil spread out in all directions from where it spilled.]
- c. The soil reddens toward the east.
[cf. (1) The soil gradually reddened at this spot due to oxidation.
(2) The weather front advanced toward the east.]

Consider the fictivity pattern for (26a). On the one hand, we have a factive representation of the fence as a stationary object with linear extent and with a particular contour, orientation, and location in geographic space. Concurrently, though, we have the fictive representation evoked by the literal sense of the sentence, in which an observer, or our focus of attention, or perhaps some image of the fence itself advancing along its own axis, moves from one end of the fence atop the plateau, along its length, to the other end of the fence in the valley.

We can ask as before whether the general fictivity pattern involving coverage paths has a perceptual analogue. The phenomenon might be found in a visual configuration perceived factively at a higher level of palpability as a static geometric form and, concurrently, perceived fictively at a lower level of palpability in terms of pathways along its delineations. For example, perhaps a subject viewing a “plus” configuration will see it explicitly as just such a “plus” shape, while implicitly sensing something intangible sweeping first downward along the vertical bar of the plus and then rightward along the horizontal bar (cf. Babcock and Freyd 1988).

6.9 “Ception”: Generalizing over Perception and Conception

In this section, we suggest a general framework that can accommodate the visual representations involved in general fictivity, together with representations that appear in language.

Much psychological discussion has implicitly or explicitly treated what it has termed *perception* as a single category of cognitive phenomena. If further distinctions have been adduced, they have been the separate designation of part of perception as *sensation*, or the contrasting of the whole category of perception with that of *conception/cognition*. One motivation for challenging the traditional categorization is that psychologists do not agree on where to draw a boundary through observable psychological phenomena such that the phenomena on one side of the boundary will be considered “perceptual,” while those on the other side will be excluded from that designation. For example, as I view a particular figure before me, is my identification

of it as a knife to be understood as part of my perceptual processing of the visual stimuli, or instead part of some other, perhaps later, cognitive processing? And if such identification is considered part of perception, what about my thought of potential danger that occurs on viewing the object? Moreover, psychologists not only disagree on where to locate a distinction boundary, but also on whether there is a principled basis on which one can even adduce such a boundary.

Accordingly, it seems advisable to establish a theoretical framework that does not imply discrete categories and clearly located boundaries, and that recognizes a cognitive domain encompassing traditional notions of both perception and conception. Such a framework would then further allow for the positing of certain cognitive parameters that extend continuously through the larger domain (as described below). To this end, we here adopt the notion of "ception" to cover all the cognitive phenomena, conscious and unconscious, understood by the conjunction of perception and conception. While perhaps best limited to the phenomena of current processing, *ception* would include the processing of sensory stimulation, mental imagery, and ongoingly experienced thought and affect. An individual currently manifesting such processing with respect to some entity could be said to "ceive" that entity.⁹

The main advantage of the ception framework in conjoining the domains of perception and conception is not that it eliminates the difficulty of categorizing certain problematic cognitive phenomena. Though helpful, that characteristic, taken by itself, could also be seen as throwing the baby out with the bathwater, in that it by fiat discards a potentially useful distinction simply because it is troublesome. The strength of the ception framework, rather, is precisely that it allows for the positing or recognition of distinction parameters that extend through the whole of the new domain, parameters whose unity might not be readily spotted across a gerrymandered category boundary. Further, such parameters are largely gradient in character and so can reintroduce the basis of the discrete perception-conception distinction in a graduated form.

We here propose thirteen parameters of cognitive functioning that appear to extend through the whole domain of ception and to pertain to general fictivity. Most of these parameters seem to have an at least approximately gradient character—perhaps ranging from a fully smooth to a merely rough gradience—with their highest value at the most clearly perceptual end of the ception domain and with their lowest value at the most clearly conceptual end of the domain. It seems that these parameters tend to covary or correlate with each other from their high to their low ends, that is, any particular cognitive representation will tend to merit placement at a comparable distance along the gradients of the respective parameters. Some of the parameters seem more to have discrete regions or categorial distinctions along their lengths than to involve continuous gradience, but these, too, seem amenable to alignment with the

other parameters. One of the thirteen parameters, the one that we term *palpability*, appears to be the most centrally involved with vision-related general fictivity. Given that the other twelve parameters largely correlate with this one, we term the whole set that of the *palpability-related parameters*.

This entire proposal of palpability-related parameters is heuristic and programmatic. It will require adjustments and experimental confirmation with regard to several issues. One issue is whether the set of proposed parameters is exhaustive with respect to palpability and general fictivity (presumably not), and, conversely, whether the proposed parameters are all wholly appropriate to those phenomena. Another issue is the partitioning of general visual fictivity that results in the particular cognitive parameters named. Thus perhaps some of the parameters presented below should be merged or split. More generally, we would first need to show that our proposed parameters are in synchrony—aligned from high end to low end—sufficiently to justify their being classed together as components of a common phenomenon. Conversely, though, we would need to show that the listed parameters are sufficiently independent from each other to justify their being identified separately, instead of treated as aspects of a single complex parameter.

6.9.1 Palpability and Related Parameters

The parameter of palpability is a gradient parameter that pertains to the degree of palpability with which some entity is experienced in consciousness, from the fully concrete to the fully abstract. To serve as reference points, four levels can be designated along this gradient: the (fully) concrete, the semiconcrete, the semiaabstract, and the (fully) abstract. These levels of palpability are discussed the next four sections and illustrated with examples that cluster near them. In this section, we present the thirteen proposed palpability-related parameters. As they are discussed here, these thirteen parameters are treated strictly with respect to their phenomenological characteristics. There is no assumption that levels along these parameters correspond to other cognitive phenomena such as earlier or later stages of processing.

1. The parameter of *palpability* is a gradient at the high end of which an entity is experienced as being concrete, manifest, explicit, tangible, and palpable. At the low end, an entity is experienced as being abstract, unmanifest, implicit, intangible, and impalpable.
2. The parameter of *clarity* is a gradient at the high end of which an entity is experienced as being clear, distinct, and definite. At the low end, an entity is experienced as being vague, indistinct, indefinite, or murky.
3. The parameter of *strength* is a gradient in the upper region of which an entity is experienced as being intense or vivid.¹⁰ At the low end, an entity is experienced as being faint or dull.

4. The *ostension* of an entity is our term for the overt substantive attributes that an entity has relative to any particular sensory modality. In the visual modality, the ostension of an entity includes its “appearance” and motion—thus, more specifically, including its form, coloration, texturing, and pattern of movements. In the auditory modality, ostension amounts to an entity’s overt sound qualities, and in the taste modality, its flavors. As a gradient, the parameter of ostension comprises the degree to which an entity is experienced as having such overt substantive attributes.
5. The parameter of *objectivity* is a gradient at the high end of which an entity is experienced as being real, as having autonomous physical existence, and as having its own intrinsic characteristics. Such an entity is further experienced as being “out there,” that is, as external to oneself—specifically, to one’s mind, if not also one’s body. At the low end of the gradient, the entity is experienced as being subjective, a cognitive construct, a product of one’s own mental activity.¹¹
6. The parameter of *localizability* is the degree to which one experiences an entity as having a specific location relative to oneself and to comparable surrounding entities within some spatial reference frame. At the high end of the gradient, one’s experience is that the entity does have a location, and that this location occupies only a delimited portion of the whole spatial field, can be determined, and is in fact known. At mid-range levels of the gradient, one may experience the entity as having a location but as being unable to determine it. At the low end of the gradient, one can have the experience that the concept of location does not even apply to the ceived entity.
7. The parameter of *identifiability* is the degree to which one has the experience of recognizing the categorial or individual identity of an entity. At the high end of the gradient, one’s experience is that one recognizes the ceived entity, that one can assign it to a familiar category or equate it with a familiar unique individual, and that it thus has a known identity. Progressing down the gradient, the components of this experience diminish until they are all absent at the low end.
8. The *content/structure* parameter pertains to whether an entity is assessed for its content as against its structure. At the content end of this gradient—which correlates with the high end of other parameters—the assessments pertain to the substantive makeup of an entity. At the structure end of the parameter—which correlates with the low end of other parameters—the assessments pertain to the schematic delineations of an entity. While the content end deals with the “bulk” form of an entity, the structural end reduces or “boils down” and regularizes this form to its abstracted or idealized lineaments. A form can be a simplex entity composed of parts or a complex entity containing smaller entities. Either way, when such a form is considered overall in its entirety, the content end can provide the comprehensive summary or Gestalt of the form’s character. On the other hand, the structure end can

reveal the global framework, pattern, or network of connections that binds the components of the form together and integrates them into a unity.

9. The *type of geometry* parameter involves the geometric characterization imputed to an entity, together with the degree of its precision and absoluteness of one's characterization. At the high end of this parameter, the assessments pertain to the content of an entity and are (amenable to being) geometrically Euclidean, metrically quantitative, precise as to magnitude, form, movements, and so on, and absolute. At the low end of the parameter, the assessments pertain to the structure of an entity, and are (limited to being) geometrically topological or topology-like, qualitative or approximative, schematic, and relational or relativistic.

10. Along the gradient parameter of *accessibility to consciousness*, an entity is accessible to consciousness everywhere but at the lowest end. At the high end of the parameter, the entity is in the center of consciousness or in the foreground of attention. At a lower level, the entity is in the periphery of consciousness or in the background of attention. Still lower, the entity is currently not in consciousness or attention, but could readily become so. At the lowest end, the entity is regularly inaccessible to consciousness.

11. The parameter of *certainty* is a gradient at the high end of which one has the experience of certainty about the occurrence and attributes of an entity. At the low end, one experiences uncertainty about the entity—or, more actively, one experiences doubt about it.

12. The parameter of *actionability* is a gradient at the high end of which one feels able to direct oneself agentively with respect to an entity—for example, to inspect or manipulate the entity. At the low end, one feels capable only of receptive experience of the entity.

13. The parameter of *stimulus dependence* is the degree to which a particular kind of experience of an entity requires current on-line sensory stimulation in order to occur. At the high end, stimuli must be present for the experience to occur. In the midrange of the gradient, the experience can be evoked in conjunction with the impingement of stimuli, but it can also occur in their absence. At the low end, the experience does not require, or has no relation to, sensory stimulation for its occurrence.

The terms for all the above parameters were intentionally selected so as to be neutral to sense modality. But the manner in which the various modalities behave with respect to the parameters—in possibly different ways—remains an issue. We briefly address this issue later. But for the sake of simplicity, the first three levels of palpability presented next are discussed only for the visual modality. Our characterization of each level of palpability below will generally indicate its standing with respect to each of the thirteen parameters.

6.9.2 Concrete Level of Palpability

At the concrete level of palpability, an entity that one looks at is experienced as fully manifest and palpable, as clear and vivid, with the ostensive characteristics of precise form, texture, coloration, and movement, and with a precise location relative to oneself and to its surroundings, where this precision largely involves a Euclidean-type geometry and is amenable to metric quantification. The entity is usually recognizable for its particular identity, and is regarded as an instance of substantive content. The entity is experienced as having real, physical, autonomous existence—hence not as dependent on one’s own cognizing of it. It is accordingly experienced as being “out there,” that is, not as a construct in one’s mind. The viewer can experience the entity with full consciousness and attention, has a sense of certainty about the existence and the attributes of the entity, and feels volitionally able to direct his or her gaze over the entity, change position relative to it, or perhaps manipulate it to expose further attributes to inspection. Outside of abnormal psychological states (such as the experiencing of vivid hallucinations), this concrete experience of an entity requires currently on-line sensory stimulation—for example, in the visual case, one must be actually looking at the entity. In short, one experiences the entity at the high end of all thirteen palpability-related parameters.

Examples of entities experienced at the concrete level of palpability include most of the manifest contents of our everyday visual world, such as an apple, or a street scene. With respect to general fictivity, a representation ceived at the concrete level of palpability is generally experienced as factive and veridical. It can function as the background foil against which a discrepant representation at a lower level of palpability is compared.

6.9.3 Semiconcrete Level of Palpability

We can perhaps best begin this section by illustrating entities ceived at the semiconcrete level of palpability, before outlining their general characteristics. A first example of a semiconcrete entity is the grayish region one “sees” at each intersection (except the one in direct focus) of a Hermann grid. This grid consists of evenly spaced vertical and horizontal white strips against a black background and is itself seen at the fully concrete level of palpability. As one shifts one’s focus from one intersection to another, a spot appears at the old locus and disappears from the new one. Another example of a semiconcrete entity is an afterimage. For example, after staring at a colored figure, one ceives a pale image of the figure in the complementary color when looking at a white field. Comparably, after a bright light has been flashed on one spot of the retina, one ceives a medium grayish spot—an “artificial scotoma”—at the corresponding point of whatever scene one now looks at. An apparently further

semiconcrete entity is the phosphene effect—a shifting pattern of light that spans the visual field—which results from, say, pressure on the eyeball.

In general, an entity ceived at the semiconcrete level of palpability, by comparison with the fully concrete level, is experienced as less tangible and explicit, as less clear, and as less intense or vivid. It has the quality of seeming somewhat indefinite in its ostensive characteristics, perhaps hazy, translucent, or ghostlike. Although one has the experience of directly “seeing” the entity, its less concrete properties may largely lead one to experience the entity as having no real physical existence or, at least, to experience doubt about any such corporeality. Of the semiconcrete examples cited above, the grayish spots of the Hermann grid may be largely experienced as “out there,” though perhaps not to the fullest degree because of their appearance and disappearance as one shifts one’s focus. The “out there” status is still lower or more dubious for afterimages, artificial scotomas, and phosphenes because these entities move along with one’s eye movements. The Hermann grid spots are fully localizable with respect to the concretely ceived grid and, in fact, are themselves ceived only in relation to that grid. But an afterimage, artificial scotoma, or phosphene image ranks lower on the localizability parameter because, although each is fixed with respect to one’s visual field, it moves about freely relative to the concretely ceived external environment in pace with one’s eye movements. The identifiability of a semiconcrete entity is partially preserved in some afterimage cases, but the entity is otherwise largely not amenable to categorization as to identity.

Generally, one may be fully conscious of and direct one’s central attention to such semiconcrete entities as Hermann grid spots, afterimages, scotomas, and phosphenes, but one experiences less than the fullest certainty about one’s ception of them, and one can only exercise a still lower degree of actionability over them, being able to manipulate them only by moving one’s eyes about. The ception of Hermann grid spots requires concurrent on-line sensory stimulation in the form of viewing the grid. But, once initiated, the other cited semiconcrete entities can be ceived for a while without further stimulation, even with one’s eyes closed.

With respect to general fictivity, a representation ceived at the semiconcrete level of palpability on viewing a scene is generally experienced as relatively more fictive and less veridical than the concrete-level representation that is usually being ceived at the same time. The type of discrepancy present between two such concurrent representations of a single scene is generally not that of fictive motion against factive stationariness, as mainly treated so far. Rather, it is one of fictive presence, as against factive absence; that is, the fictive representation, for example, of Hermann grid spots, of an afterimage, of an artificial scotoma, or of phosphenes, is assessed as being present only in a relatively fictive manner, while the factive representation of the scene being viewed is taken more veridically as lacking any such entities.

6.9.4 Semiabstract Level of Palpability

An entity at the semiabstract level of palpability is experienced as present in association with other entities that are seen at the fully concrete level, but it itself is intangible and nonmanifest, as well as vague or indefinite and relatively faint. It has little or no ostension, and with no quality of direct visibility. In viewing a scene, one's experience is that one does not "see" such an entity explicitly, but rather "senses" its implicit presence. In fact, we will adopt *sensing* as a technical term to refer to the ception of an entity at the semiabstract level of palpability, while engaging in on-line viewing of something concrete.¹² One experiences an entity of this sort as "out there," perhaps localizable as a genuinely present characteristic of the concrete entities viewed, but not as having autonomous physical existence. Insofar as such a sensed entity is accorded an identity, it would be with respect to some approximate or vague category.

A sensed entity is of relatively low salience in consciousness or attention, seems less certain, and is difficult to act on. Often a sensed entity of the present sort is understood as a structural or relational characteristic of the concrete entities viewed. Its type of geometry is regularly topology-like and approximative. Such sensed structures or relationships can often be captured for experiencing at the fully concrete level by schematic representations, such as line drawings or wire sculptures, but they lack this degree of explicitness in their original condition of ception.

Because the semiabstract level of palpability is perhaps the least familiar level, we present a number of types and illustrations of it, characterizing the pattern of general fictivity that holds for three of these types. General fictivity works in approximately the same way for all three types: object structure, reference frames, and force dynamics. In order to characterize the general fictivity pattern for these three types together, we refer to them here collectively as "structurality." The representation of structurality one senses in an object or an array is generally experienced as more fictive and less veridical than the factive representation of the concrete entities whose structurality it is. The representation of structurality is a case of fictive presence rather than of fictive motion. This fictive presence contrasts with the factive absence of such structurality from the concrete representation. Unlike most forms of general fictivity, the representation of concrete content and that of sensed structurality may seem so minimally discrepant with each other that they are rather experienced as complementary or additive. (The type in section 6.9.4.4 involving structural history and future has its own fictivity pattern, which will be described separately.) Much of visually sensed structure is similar to the structure represented by linguistic closed-class forms, and this parallelism will be discussed later in section 6.9.11.

6.9.4.1 Sensing of Object Structure One main type of sensed entity is the structure we sense to be present in a single object or over an array of objects due to its arrangement in space. To illustrate first for the single-object case, when one views a certain kind of object such as a vase or a dumpster, one sees concretely certain particulars of ostension such as outline, delineation, color, texture, and shading. But in addition, one may sense in the object a structural pattern comprising an outer portion and a hollow interior. More precisely, an object of this sort is sensed—in terms of an idealized schematization—as consisting of a plane curved in a way that defines a volume of space by forming a boundary around it. A structural schema of this sort is generally sensed in the object in a form that is abstracted away from each of a number of other spatial factors. This “envelope/interior” structuring can thus be sensed equally across objects that differ in magnitude, like a thimble and a volcano; in shape, like a well and a trench; in completeness of closure, like a beachball and a punchbowl; and in degree of continuity/discontinuity, like a bell jar and a birdcage. This pattern of ception shows—as appropriate to the semiabstract level of palpability—that the type of geometry (parameter 9) here sensed in the structure of an object is topological or topology-like. In particular, object structure sensed as being of the envelope-interior type is magnitude-neutral and shape-neutral, as well as being closure-neutral and discontinuity-neutral.

For a more complex example, on viewing a person, one sees at the fully concrete level of palpability that person’s outline and form, coloration and shading, textures, the delineations of the garments, and so on. However, one does not see but rather senses the person’s bodily structure in its current configuration, for example, when in a squatting or leaning posture. A sensed structural schema of this sort can be made concretely visible, as when a stick figure drawing or a pipe cleaner sculpture is shaped to correspond to such a posture. But one does not concretely see such a schema when looking at the person—one only senses its presence. The Marrian abstractions (Marr 1982) that represent a human figure in terms of an arrangement of axes of elongation provide one theoretization of this sensed level of ception.

A comparable sensing of structure can occur for an array of objects. For example, a person may ceive one object as located at a point or points of the interior space of another object that she senses as having the envelope/interior structure described above. The person may sense in this object complex a structural schema—what she may categorize as the “inside” schema—wherein the first object is inside the second. As in the single-object case, this object array also exhibits a number of topology-like characteristics. Thus not only can the first object and the second object themselves each vary in magnitude and shape, but in addition the first object can exhibit any orientation relative to the second object and can be located throughout any portion

or amount of the second object's interior space, while still being sensed as manifesting the "inside" schema.

For a more intricate example, when one views the interior of a restaurant, one senses a hierarchically embedded structure in space that includes the schematic delineations of the dining hall as the largest containing frame and the spatial pattern of tables and people situated within this frame. Perhaps one can see some of the hall's framing delineations concretely, for example, some ceiling-wall edges; but for the most part, the patterned arrangement in space seems to be sensed. Thus, if one were to represent this sensed structure of the scene in a schematic drawing, one might include some lines to represent the rectilinear frame of the hall, together with some spots or circles for the tables and some short bent lines for the people that mark their relative positions within the frame and to each other. However, though it can be so represented, this is an abstraction for the most part not concretely seen as such, but rather only sensed as present.

Further cases perhaps also belong in this object structure type of sensing. Thus parts of objects not concretely seen but known or assumed to be present in particular locations may be sensed as present at those locations. This may apply to the part of an object being occluded by another object in front of it, or to the back or underside of an object not visible from a viewer's current perspective.

6.9.4.2 Sensing of Path Structure When one views an object moving with respect to other objects, one concretely sees the path it executes as having Euclidean specifics such as exact shape and size. But in addition, one may sense an abstract structure in this path. The path itself would not be a case of fictive motion, for the path is factive. But the path is sensed as instantiating a particular idealized path schema, and it is this schema that is fictive. Thus one may sense as equal instantiations of an "across" schema both the path of an ant crawling from one side of one's palm to the opposite side and the path of a deer running from one side of a field to the opposite side. This visually sensed "across" schema would then exhibit the topological property of being magnitude-neutral. Comparably, one may equally sense an "across" schema in the path of a deer running in a straight perpendicular line from one boundary of a field to the opposite boundary, and in the path of a deer running from one side of the field to the other along a zigzag slanting course. The visually sensed "across" schema would then also exhibit the topological property of being shape-neutral.

6.9.4.3 Sensing of Reference Frames Perhaps related to the sensing of object/array structure is the sensing of a reference frame as present amid an array of objects. For example, in seeing the scenery about oneself at the concrete level, one can sense a grid of compass directions amid this scenery. One may even have a choice of

alternative reference frames to sense as present (as described in Talmy 1983). For example, consider a person who is looking at a church facing toward the right with a bicycle at its rear. That person can sense within this manifest scene an earth-based frame, in which the bike is west of the church. Or she can sense the presence of an object-based frame, in which the bike is behind the church. Or she sense the presence of a viewer-based frame radiating out from herself, in which the bike is to the left of the church. Levinson (1996) and Pederson (1993) have performed experiments on exactly this issue, with findings of strong linguistic-cultural biasing for the particular type of reference frame sensed as present.

One may also sense the presence of one or another alternative reference frame for the case of a moving object executing a path. Thus, on viewing a boat leaving an island and sailing an increasing distance from it, one can sense its path as a radius extending out from the island as center within the concentric circles of a radial reference frame. Alternatively, one can sense the island as the origin point of a rectilinear reference frame and the boat's path as an abscissal line moving away from an ordinate.

6.9.4.4 Sensing of Structural History and Future Another possible type of sensed phenomenon also pertains to the structure of an object or of an array of objects. Here, however, this structure is sensed not as statically present but rather as having shifted into its particular configuration from some other configuration. In effect, one senses a probable, default, or pseudohistory of activity that led to the present structure. A sensed history of this sort is the visual counterpart of the fictive site arrival paths described for language in section 6.8.3. The examples of visual counterparts already given in that section were of a figurine perceived as a torso with head and limbs affixed to it; of an irregular contour perceived as the result of processes like indentation and protuberance; and of a Pac-Man figure perceived as a circle with a wedge removed.

In addition to such relatively schematic entities, it can be proposed that one regularly senses certain complex forms within everyday scenes not as static configurations self-subsistent in their own right but rather as the result of deviation from some prior, generally more basic, state. For example, on viewing an equal-sided picture frame that is hanging on the wall at an oblique angle, one may not perceive the frame as a static diamond shape, but may rather sense it as a square manifesting the result of having been tilted away from a more basic vertical-horizontal orientation. Another example is the sensing of a dent in a fender not as a *sui generis* curvature but as the result of a deformation. One senses a set of clay shards not as an arrangement of separate distinctively shaped three-dimensional objects but as the remains of a flowerpot that had been broken. One may even sense toys that are lying over the floor not simply as

comprising some specific spatial static pattern but rather as manifesting the result of having been scattered into that configuration from a home location within a box.

Viewing an entity may lead one to sense not only a history of its current configuration, but also to sense a potential or probable future succession of changes away from its current configuration. Such a sensed future might involve the return of the entity to a basic state that it had left. For example, on viewing the previous picture frame hanging at an angle, one may sense its potential return to the true (probably as part of imagining one's manipulations to right it).

In terms of general fictivity, the sensing of an entity's structural history or future is a less veridical representation of fictive motion in a sensory modality. It is superimposed on the factively and veridically seen static representation of the entity. Thus, with respect to the picture frame example, the difference between the factive and the fictive modes of ceiving the frame is the difference between seeing a static diamond and sensing a square with a past and a future.

6.9.4.5 Sensing of Projected Paths Another type of sensed ception can be termed *projected paths*. One form of path projection is based on motion already being exhibited by a Figure entity, for example, a thrown ball sailing in a curve through the air. A viewer observing the concretely occurrent path of the object can generally sense—but not palpably see—the path that it will subsequently follow. Here we do not refer simply to unconscious cognitive computations that, say, enable the viewer to move to the spot at which she could catch the ball; rather, we refer to the conscious experience a viewer often has of a compelling sense of the specific route that the object will traverse. One may also project backward to sense the path that the ball is likely to have traversed before it was in view. Path projection of this sort is thus wholly akin to the sensing of structural history and future discussed in the preceding section. The main difference is that there the viewed entity was itself stationary, whereas here it is in motion. Accordingly, there the sensed changes before and after the static configuration were largely associations based on one's experience of frequent occurrence, whereas here the sensed path segments are projections mostly based on one's naive physics applied to the viewed motion.

Another form of projected path pertains to the route that an agentive viewer will volitionally proceed to execute through some region of space. It applies to a viewer, say, standing at one corner of a restaurant crowded with tables who wants to get to the opposite corner. Before starting out, such a viewer will often sense at the semi-abstract level of palpability an approximate route curving through the midst of the tables that he could follow to reach his destination. The viewer might sense the shape of this path virtually as if it were taken by an aerial photograph. It may be that the initially projected route is inadequate to the task, and that the route-sensing process

is regularly updated and reprojected as the viewer moves along his path. But throughout such a process, only the physical surroundings are seen concretely, whereas the path to follow is sensed. This form of projected path is akin to the fictive access paths described in section 6.8.4.

6.9.4.6 Sensing of Force Dynamics Also at the semiabstract level of palpability is the sensing of force interrelationships among otherwise concretely seen objects. Included in such sensed *force dynamics* are the interactions of opposing forces such as an object's intrinsic tendency toward motion or rest; another object's opposition to this tendency; resistance to such opposition; the overcoming of resistance; and the presence, appearance, disappearance, or absence of blockage. (See Talmy 1988b for an analysis of the semantic component of language that pertains to force dynamics.)

To illustrate, Rubin (1986) and Engel and Rubin (1986) report that subjects perceive (in our terms, *sense*) forces at the cusps when viewing a dot that moves along a path like a bouncing ball. When the bounce is progressively heightened, then the perception is that a force has been added at the cusps. Complementarily, when the ball's bounce is reduced, the force is perceived as being dissipated. Jepson and Richards (1993) also note that when a block is drawn with one face coplanar to and in the middle of the vertical face of a larger block, then the percept is as if the smaller block is "attached" or glued to the larger block, analogously to what is sensed in the viewing of an object stuck to a wall. But there is no such perception of an "attaching force" when the same small block is similarly positioned on the top face of the larger block (i.e., when the original configuration is rotated 90 degrees). In this latter case, only contact, not attachment, is perceived, just as would be expected in viewing an object resting on a horizontal surface.

For a less schematic example, consider a scene in which a large concrete slab is leaning at a 45° angle against the outer wall of a rickety wooden shed. A person viewing this scene would probably not only see at the concrete level the slab and the shed in their particular geometric relationship, but also would sense a force dynamic structure implicit throughout these overt elements. This sensed force structure might include a force (manifested by the shed) that is now successfully but tenuously resisting an unrelenting outside force impinging on it (manifested by the slab), and that is capable of incrementally eroding and giving way at any moment.

6.9.4.7 Sensing of Visual Analogues to Fictive Motion in Language Finally, the fictive motion types presented before this section on ception can now be recalled for their relevance to the present discussion. Most of the visual patterns suggested as counterparts of the linguistic fictive motion types seem to fit at the semiabstract level of palpability—that is, they are sensed. Further, in terms of general fictivity, these

visual analogues have involved the sensing of fictive *motion*; they do not involve the sensing of fictive *presence* (as was the case for the representations of “structurality” just seen). As a summary, we can list here the fictive types from sections 6.2–6.5 and 6.8, all of which participate in this phenomenon. Thus, we may sense at the semi-abstract level of palpability the fictive motion of the visual counterparts of orientation paths (including prospect paths, alignment paths, demonstrative paths, and targeting paths), radiation paths, shadow paths, sensory paths, pattern paths, frame-relative motion, advent paths, access paths, and coverage paths. With the addition of the cases of structural history/future and projected paths characterized just above, this is a complete list of the fictive types proposed, in this chapter, to have a visual representation sensed as fictive motion.

6.9.5 Abstract Level of Palpability

The cases cited thus far for the first three levels of palpability have all depended on concurrent on-line sensory stimulation (with the exception that afterimages, artificial scotomas, and phosphenes require stimulation shortly beforehand). But we can adduce a level still further down the palpability gradient, the (fully) abstract level. At this level, one experiences conceptual or affective entities that do not require on-line sensory stimulation for their occurrence and may have little direct relation to any such stimulation. Largely clustering near the lower ends of the remaining palpability-related parameters, such entities are thus largely impalpable, abstract, vague, and perhaps faint, lacking in ostensive characteristics, and not amenable to localization in space or identification as to category. They are often experienced as subjective, hence in oneself rather than “out there.” They do seem to exhibit a range across the remaining palpability-related parameters. Thus, they can range from full salience to elusiveness or virtual inaccessibility to consciousness; one can range from certainty to puzzlement over them, and from a capacity to manipulate them in one’s mind to an experience of being only a passive receptor to them. Finally, they can exhibit either content or structure, and, insofar as they manifest a type of geometry, this, too, can exhibit a range, though perhaps tending toward the approximative and qualitative type.

Such abstract entities may be ceived as components in the course of general on-going thought and feeling. They might include not only the imagined counterparts of entities normally ceived as a result of on-line stimulation—for example, the experience only in imagination of the structure one would otherwise sense on-line while viewing an object or array in space—but also phenomena that cannot normally or ever be directly ascribed as intrinsic attributes to entities ceived as the result of on-line sensory stimulation. Such phenomena might include the following: the awareness of relationships among concepts within one’s knowledge representation; the experience

of implications between sets of concepts, and the formation of inferences; assessments of veridicality; assessments of change occurring over the long term; experiences of social influence (such as permissions and requirements, expectations and pressures); a wide range of affective states; and “propositional attitudes” (such as wish and intention).

Many cognitive entities at the abstract level of palpability are the semantic referents of linguistic forms and thus can also be evoked in awareness by hearing or thinking of those forms. These forms themselves are fully concrete when heard, and of course less concrete when imagined in thought, but the degree of concreteness they do have tends to lend a measure of explicitness to the conceptual and affective phenomena associated with them. And with such greater explicitness may come greater cognitive manipulability (actionability) and access to consciousness. However, these are phenomena that, when experienced directly without association with such linguistic forms, may be at the fully abstract level of palpability. Despite such upscaling lent by linguistic representation, it is easiest to give further examples of conceptually abstract phenomena by citing the meanings of certain linguistic forms. Because open-class forms tend to represent more contentful concepts, while closed-class forms tend to represent more structural—and hence, more abstract—concepts, we next cite a number of closed-class meanings so as to further convey the character of the fully abstract end of the palpability gradient, at least insofar as it is linguistically associated.¹³

First, a schematic structure one might otherwise sense at the semiabstract level of palpability through on-line sensory stimulation—as by looking at an object or scene—can also beceived at the fully abstract, purely ideational level in the absence of current sensory stimulation by hearing or thinking of a closed-class linguistic form that refers to the same schematic structure. For example, on viewing a scene in which a log is straddling a road, one might sense the presence of a structural “across” schema in that scene. But one can alsoceive the same “across” schema at the abstract level of palpability by hearing or thinking of the word *across* either alone or in a sentence like *The log lay across the road*.

We can next identify a number of conceptual categories expressed by linguistic closed-class forms that are seemingly never directly produced by on-line sensory stimulation. Thus the conceptual category of *tense*, with such specific member concepts as past, present, and future, pertains to the time of occurrence of a referent event relative to the present time of speaking. This category is well represented in the languages of the world but has seemingly scant homology in the forms of ception higher on the palpability scale that are evoked by current sensory stimulation. A second linguistically represented category can be termed *reality status*—a type largely included under the traditional linguistic term *mood*. For any event being referred to,

this category would include such indications as that the event is actual, conditional, potential, or counterfactual, and would also include the simple negative (e.g., English *not*). Again, aspects of situations that are currently seen, heard, smelled, and so on at the concrete level or sensed at the semiabstract level are seemingly notceived as having any reality status other than the actual. Similarly, the linguistically represented category of *modality*, with such member notions as those expressed by English *can*, *must*, and *should*, has little concrete or sensed counterpart.

To continue the exemplification, a further set of categories at the abstract level of palpability that can be evoked by closed-class forms pertain to the cognitive state of some sentient entity; these categories, too, seem unrepresented at the higher levels of palpability. Thus a conceptual category that can be termed *speaker's knowledge status*, represented by linguistic forms called "evidentials," particularizes the status of the speaker's knowledge of the event that she is referring to. In a number of languages (e.g., in Wintu, where it is expressed by inflections on the verb), this category has such member notions as: "known from personal experience as factual," "accepted as factual through generally shared knowledge," "inferred from accompanying evidence," "inferred from temporal regularity," "entertained as possible because of having been reported," and "judged as probable." Another linguistic category of the cognitive state type can be termed the *addressee's knowledge status*. This is the speaker's inference as to the addressee's ability to identify some referent the speaker is currently specifying. One common linguistic form representing this category is that of determiners that mark definiteness—for example, the English definite and indefinite articles *the* and *a* or *an*. Further grammatically represented cognitive states are intention and volition, purpose, desire, wish, and regret.

For some final examples, a linguistic category that can be termed *particularity* pertains to whether an entity in reference is to be understood as unique (*That bird just flew in*), or as a particular one out of a set of comparable entities (*A bird just flew in*), or generically as an exemplar standing in for all comparable entities (*A bird has feathers*). But the on-line ception of an entity at the concrete or semiabstract level may not accommodate this range of options. In particular, it apparently tends to exclude the generic case—for example, looking at a particular bird does not tend to evoke the ception of all birds generically. Thus the ception of genericness in human cognition may occur only at the abstract level of palpability. Finally, many linguistic closed-class forms specify a variety of abstract relationships, such as kinship and possession. The English ending *-s* expresses both of these relationships, as in *John's mother* and *John's book*. Again, on-line ception, such as viewing John in his house and Mrs. Smith in hers, or viewing John in the doorway and a book on the table, may not directly evoke the relational concepts of kinship and possession the linguistic forms do.¹⁴

6.10 Further Types and Properties of Ception

The full structure of the entire system of ception certainly remains to be characterized, but some brief notes here will sketch in a few lineaments of that structure. We cite some further types of ception, some forms of dissociation across the palpability-related parameters outlined above, and some differences among the various sensory modalities as to their parametric behavior.

6.10.1 Imagistic Forms of Ception

What can be termed *imagistic* forms of ception include mental imagery, whether related to vision or to other sensory modalities. Along the gradient parameter of stimulus dependence, imagistic ception seems to fall in the midrange. That is, it can be evoked in association with an entity ceived at the concrete level during on-line stimulation by that entity. For example, on seeing a dog, one can imagine the sight and sound of it starting to bark, as well as the sight and kinesthesia of one's walking over and petting it. But imagistic ception can also occur without on-line stimulation, as during one's private imaginings. It needs to be determined whether imagistic ception can also occur at the low end of the stimulus dependence parameter, that is, whether aspects of it are unrelated to sensory attributes, as in the case of many conceptual categories of language.

6.10.2 Associative Forms of Ception

What can be termed *associative forms* of ception pertain to ceptual phenomena evoked in association with an entity during on-line sensory stimulation by it, but not ascribed to that entity as intrinsic attributes of it. Such associated phenomena could include the following type: (1) mental imagery, as just discussed; (2) actions one might undertake in relation to the entity; (3) affective states experienced with respect to the entity; (4) particular concepts or aspects of one's knowledge one associates with the entity; and (5) inferences regarding the entity.

Having already discussed mental imagery, we can here illustrate the remaining four of these types of associative ception. As examples of associated action (2), on viewing a tilted picture frame, one might experience a motoric impulse to manipulate the frame so as to right it. Or, on viewing a bowling ball inexorably heading for the side gutter, one might experience or execute the gyrations of "body English" as if to effect a correction in the ball's path. In fact, with respect to such kinesthetic effects, there may be a gradient of palpability—parallel to what we have posited for ception—that applies to motor control. Proceeding from the least to the most palpable, at the low end would be one's experience of intending to move; in the midrange would be one's

experience of all-but-overt motion, including checked movement and covert body English; and at the high end would be one's experience of one's overt movements.

Associated affect (3) has such straightforward examples as experiencing pleasure, disgust, or fear at the sight of something, e.g., of a child playing, of roadkill, or of a mugger. Associated knowledge or concepts (4) could include examples like thinking of danger on seeing a knife, or thinking of one's childhood home on smelling fresh bread. And examples of associated inference (5) might be gathering that Mrs. Smith is John's mother from the visual apparency of their ages and of their resemblance, or inferring that a book on a table belongs to John from the surroundings and John's manner of behaving toward it.

6.10.3 Parameter of Intrinsicity

Associative forms of ception like those just adduced may be largely judged to cluster near the semiabstract level of palpability. In fact, the phenomena described in section 6.9.4 as “sensed” at the semiabstract level and the associative phenomena reported here may belong together as a single group ceived at the semiabstract level of palpability. But the sensed type and the associative type within this group would still differ from each other with respect to another gradient parameter, what might be termed *intrinsicity*. At the high end of this gradient, the sensed phenomena would be experienced as intrinsic to the entity being ceived at the concrete level, that is, they would be ceived as actually present and perhaps inherent attributes—such as structure and patterns of force impingement—that the ceiver is “detecting” in the concretely seen entity. But at the lower end of the intrinsicity gradient, the associative phenomena presented here would be experienced as merely associated with the concretely ceived entity, that is, they would be experienced as incidental phenomena the ceiver brings to the entity.

This intrinsicity parameter, however, is actually just the objectivity gradient (parameter 5) when applied to phenomena connected with an entity rather than to the entity itself. To be sure, where a particular phenomenon is placed along the intrinsicity gradient varies according to the type of phenomenon, the individual, the culture, and the occasion. For a classical example, if one ceives beauty in conjunction with seeing a particular person, one may experience this beauty as an intrinsic attribute of the person seen, much like the person's height, or, alternatively, as a personal interpretive response by the beholder.

6.10.4 Dissociations among the Palpability-Related Parameters

While the thirteen palpability-related parameters listed in section 6.9.1 generally tend to correlate with one another for the types of ception that had been considered, some

dissociations can be observed. For example, with respect to the imagistic forms of ception, visual mental imagery can have a fairly high degree of ostension (parameter 4), for instance, having relatively definite form and movement. At the same time, however, it may rank somewhere between the semiconcrete level and the semiabstract level along the palpability gradient (parameter 1) and at a comparably midrange level along the clarity gradient (parameter 2). For another case of dissociation, already noted, the cognitive phenomena expressed by closed-class linguistic forms are generally at the most abstract level of the palpability gradient (parameter 1). But the conscious manipulability of the linguistic forms expressing these conceptual phenomena ranks them near the high end of the actionability gradient (parameter 12). Or again, some affective states may rank quite low on most of the parameters—for example, intangible on the palpability gradient (parameter 1), murky on the clarity gradient (parameter 2), and nonostensive on the ostension gradient (parameter 4)—while ranking quite high on the strength gradient (parameter 3) because they are experienced as intense and vivid. The observation of further dissociations of this sort can argue for the independence of the parameters adduced and ultimately justify their identification as distinct phenomena.

6.10.5 Modality Differences along the Palpability Gradient

In the discussion on ception, we have mostly dealt with phenomena related to the visual modality, which can exhibit all levels along the palpability gradient except perhaps the most abstract. But we can briefly note that each sensory modality may have its own pattern of manifestation along the various palpability-related parameters adduced. For example, the kinesthetic modality, including one's sense of one's current body posture and movements, may by its nature seldom or never rank very high along the palpability, clarity, and ostension gradient (parameters 1, 2, and 4), perhaps hovering somewhere between the semiconcrete and the semiabstract level. The modality of smell, at least for humans, seems to rank low with respect to the localizability gradient (parameter 6). And the modalities of taste and smell, as engaged in the ingestion of food, may range more over the content region than over the structure region of the content/structure gradient (parameter 8). Comparison of the sensory modalities with respect to ception requires much further investigation.

6.11 Content/Structure Parallelisms between Vision and Language

The analysis to this point permits the observation of two further parallelisms between vision and language.

6.11.1 Complementary Functions of the Content and Structure Subsystems in Vision and Language

First, both cognitive systems, vision and language, have a content subsystem and a structure subsystem. Within on-line vision, for example, in the viewing of an object or array of objects, the content subsystem is foremost at the concrete level of palpability, while the structure subsystem is foremost at the semiabstract level of palpability. In language, the referents of open-class forms largely manifest the content subsystem, while the referents of closed-class forms are generally limited to manifesting the structure subsystem. The two subsystems serve largely distinct and complementary functions, as will be demonstrated next, first for vision and then for language. A number of properties from both the content/structure gradient (parameter 8) and the type-of-geometry gradient (parameter 9) align differentially with the distinctive functioning of these two subsystems. Included are properties pertaining to bulk as against lineaments, Euclidean geometry as against topology, absoluteness as against relativity, precision as against approximation, and, holistically, a substantive summary as against a unifying framework.¹⁵

We can first illustrate the properties and operations of the two subsystems in vision. For a case involving motor planning and control, as in executing a particular path through space, the content subsystem is relevant for fine-grained local calibrations, while the structure subsystem can project an overall rough-and-ready first approximation. Thus, to revisit an earlier example, a person wanting to cross the dining area of a restaurant will likely plot an approximate, qualitative course curving through the tables, using the sensed semiabstract level of structure in a spatial array. But in the process of crossing, the person will attend to the Euclidean particulars of the tables, using the concrete level of specific bulk content, so as not to bump into the tables' corners. If such were possible, a person operating without the overall topology-like subsystem would be reduced to inching along, using the guidelines of the precision subsystem to follow the sides of the tables and the curves of the chairs, without an overarching schematic map for guidance. On the other hand, a person lacking the precision subsystem might set forth on an approximate journey but encounter repeated bumps and blockages for not being able to gauge accurately and negotiate the local particulars. The two subsystems thus perform complementary functions and are both necessary for optimal navigation, as well as other forms of motor activity.

We can next illustrate the two subsystems at work in language. To do this, we can observe the distinct functions served by the open-class forms and by the closed-class forms in any single sentence. Thus, consider the sentence *A rustler lassoed the steers*. This sentence contains just three open-class forms, each of which specifies a rich complex of conceptual content. These are the verb *rustle*, which specifies notions of

illegality, theft, property ownership, and livestock; the verb *lasso*, which specifies a rope looped and knotted in a particular configuration that is swung around, cast, and circled over an animal's head in a certain way; and the noun *steer*, which specifies notions of a particular animal type, the institution of breeding for human consumption, and castration.

On the other hand, the sentence contains a number of closed-class forms that specify relatively spare concepts serving a structuring function. These include the suffix *-ed* specifying occurrence before the time of the current speech event; the suffix *-s*, specifying multiple instantiation, and the "zero" suffix (on *rustler*), specifying unitary instantiation; the article *the*, specifying the speaker's assumption of ready identifiability for the addressee, and the article *a*, specifying the opposite of this; the suffix *-er*, specifying the performer of an action; the grammatical category of noun (for *rustler* and *steers*), indicating an object and that of verb (for *lassoed*) indicating a process; and the grammatical relation of subject, indicating an Agent, and that of direct object, indicating a Patient.

The distinct functions served by these two types of forms can be put into relief by alternately changing one type of form in the above sentence, while keeping the other constant. Thus we can change only the closed-class forms, as in a sentence like *Will the lassosers rustle a steer?* Here, all the structural delineations of the depicted scene and of the speech event have been altered, but because the content-specifying open-class forms are the same, we are still in a Western cowboy landscape. But we can now change only the open-class forms, as in *A machine stamped the envelopes*. Here, the structural relationships of the scene and of the speech event are the same as in the original sentence, but with the content-specifying forms altered, we are now transposed to an office building. In sum, then, in the referential and discourse context of a sentence, the open-class forms of the sentence contribute the majority of the content, whereas the closed-class forms determine the majority of the structure.

Thus, both in ceiving and motorically negotiating a visual scene and in cognizing the reference of a sentence, the two cognitive subsystems of content and of structure are in operation, performing equally necessary and complementary functions as they interact with each other.

6.11.2 Comparable Character of the Structure Subsystem in Vision and in Language

The structural subsystems in vision and in language exhibit great similarity. First, recall that section 6.9.4 on ception at the semiaabstract level of palpability proposed that we can sense the spatial and force-related structure of an object or an array of objects when viewing it. It was suggested that any structure of this sort is sensed as consisting of an idealized abstracted schema with a topology-like or other qualitative type of geometry. With respect to language, the preceding section has shown that the

system of closed-class forms is dedicated to specifying the structure of the whole or some part of a conceptual complex in reference. We can now point out that when such linguistically specified structure pertains to space or force, it, too, consists of idealized abstracted schemas with topology-like properties. In fact, the character of the structuring yielded by visual sensing and that yielded by the linguistic closed-class system appear to be highly similar. If we can heuristically hypothesize that some particular neural system is responsible for processing schematic structure in general, then we can suppose that both visual sensing and linguistic closed-class representation are connected with, or "tap into," that single neural system for this common characteristic of their mode of functioning.

The structure subsystems of vision and language exhibit a further parallel. Recall the observation in section 6.9.4 that the structural schemas one semiabstractly senses to be present in an object or array are assessed as being fictive, relative to the factive status of the way one concretely sees the object or array. Now, the structural schemas expressed by linguistic closed-class forms—here, specifically, those pertaining to space and force—are also fictive representations, relative to the factive character of the objects and arrays that a language user understands them to pertain to. That is, all these cases of abstracted or conceptually imposed schemas—whether sensed visually or specified by linguistic closed-class forms—can be understood as a form of fictivity. They constitute not fictive motion but fictive presence—here, the fictive presence of structure. Accordingly, the extensive body of linguistic work on spatial schemas (e.g., Talmy 1975, 1983 and Herskovits 1986, 1994, among much else) constitutes a major contribution to fictivity theory. In particular, Herskovits has made it a cornerstone of her work to treat the spatial schemas she describes as "virtual structures" (previously called "geometric conceptualizations"), which are to be distinguished from the "canonic representations" of objects "as they are." If we can now extend the hypothesis of a neural system responsible for processing schematic structure, we can add that the products of its processing have ascribed to them the character of being fictive, relative to the products of other neural systems for processing the concrete ostensions of ceived entities.

Proceeding now to demonstrations of similarity, we consider several parallel vision-language cases. With respect to the structure of an array of objects, it was proposed in section 6.9.4.1 that one can visually sense the presence of an "inside" type of structural schema on viewing a two-object complex in which one object is sensed as located at a point or points of the interior space defined by the other object. This schema can be topologically or qualitatively abstracted away from particulars of the objects' size, shape, state of closure, discontinuity, relative orientation, and relative location. Now, the spatial schema specified by the English preposition *in* exhibits all these same properties. This closed-class form can thus be used with equal

appropriateness to refer to some object as located *in a thimble, in a volcano, in a well, in a trench, in a beachball, in a punchbowl, in a bell jar, or in a birdcage*. Further, it can be said that in abstracting or imposing their schema, the structure subsystems of both vision and language produce a fictive representation, relative to the concreta of the object array.

Comparably, section 6.9.4.2 addressed the topology-like properties of the structure sensed in the path of a viewed moving object. But this type of visually sensed structure also has linguistic closed-class parallels. Thus the English preposition *across*—which specifies a schema prototypically involving motion from one parallel line to another along a perpendicular line between them—exhibits the topological property of being magnitude-neutral. This is evident from the fact that it can be applied both to paths of a few centimeters, as in *The ant crawled across my palm*, as well as to paths of thousands of miles, as in *The bus drove across the country*. In a related way, the preposition *through* specifies (in one sector of its usage) a schema involving motion along a line located within a medium. But, topology-like, this schema is shape-neutral; thus *through* can be applied equally as well to a looped path, as in *I circled through the woods*, as to a jagged path, as in *I zig-zagged through the woods*. And, again, the topological schemas thus visually sensed in or linguistically imputed to a path are fictive representations relative to the Euclidean particulars seen or believed to be present.

For a final case, section 6.9.4.3 suggested that, on viewing certain scenes, one may sense the presence of either a rectilinear or a radial reference frame as the background against which an object executes a path. But these two alternate schemas can also be represented by closed-class forms. Thus English *away from* indicates motion from a point on an ordinate-type boundary progressing along an abscissa-type axis within a rectilinear grid. But *out from* indicates motion from a central point along a radius within a radial grid of concentric circles. These alternative conceptual schematizations can be seen in sentences like: *The boat drifted further and further away/out from the isle*, or *The sloth crawled 10 feet away/out from the tree trunk along a branch*. Here, both reference frames are clearly fictive cognitive impositions upon the scene, whether this scene is viewed visually or referred to linguistically.

6.11.3 Structural Explicitness in Vision and Language

The cognitive system pertaining to vision in humans has another feature that may have a partial counterpart in language. It has a component for representing in an explicit form the kinds of schematic structures generally sensed only implicitly at the semiabstract level of palpability. We here call this the component for “schematic pictorial representation.”

In iconographic representation, a full-blown pictorial depiction manifests the content subsystem. But the structure subsystem can be made explicit through the component of schematic pictorial representation by schematic depictions involving the use of points, lines, and planes, as in both static and filmic cartoons and caricatures, line drawings, wire sculptures, and the like. The very first pictorial depictions children produce—their "stick figure" drawings—are of this schematic kind. For example, a child might draw a human figure at an early phase as a circle with four lines radiating from it, and later as a circle atop a vertical line from which two lines extend laterally right and left at a midpoint and two more lines slope downward from the bottom point. Thus, in depicting an object or scene viewed, a child represents not so much its concrete-level characteristics as the structure that he or she can sense in it at the semiabstract level of palpability.

It must be emphasized that such schematizations are not what impinges on one's retinas. What impinges on one's retinas are the particularities of ostension: the bulk, edges, textures, shadings, colorings, and so on of an entity looked at. Yet what emerges from the child's hand movements are not such particulars of ostension, but rather one-dimensional lines forming a structural schematic delineation. Accordingly, much cognitive processing has to occur between the responses of the retinas and these hand motions. This processing in a principled fashion reduces, or "boils down," bulk into delineations. As proposed in this chapter, such structural abstractions are in any case necessary for the ception of visual form, both of single objects and of object arrays (cf. Marr 1982); they constitute a major part of what is sensed at the semiabstract level of palpability. It then appears that the component of the visual system involved in producing external depictions taps specifically into this same abstractional structuring system, a mechanism already in place for other functions—where this mechanism may be the same as the earlier heuristically hypothesized neural system for schematic structure in general. In fact, in the developmentally earliest phase of operation, a child's iconographic capacity would appear to be linked mainly to this structuring mechanism, more so than to the cognitive systems for concretelyceiving the full ostension of objects.

The component of language that may partially correspond to this representational explicitness is the closed-class system itself, as characterized in the preceding section. The linguistic linkage of overt morphemes to the structural schemas they represent lends some concreteness to those cognitive entities, otherwise located at the fully abstract level of palpability. These morphemes constitute tangible counterparts to the abstract forms, permit increased actionability upon them, and perhaps afford greater conscious access to them. The form of such morphemes, however, does not reflect the form of the schemas they represent, and in this way, this language component differs

crucially from the pictorial schematic representations, which do correspond in structure to what they represent.

Although this section has pointed to content-structure parallelisms between vision and language, it remains to chart their differences. It may be expected that the structure subsystems in vision and language differ in various respects as to what they treat as structural, their degree and type of geometric abstraction, the degree and types of variation such structural features can exhibit across different cultural groups, and the times and sequences in which these structural features appear in the developing child.

6.11.4 Some Comparisons with Other Approaches

The present analysis raises a challenge to the conclusions of Cooper and Schacter (1992). They posit “explicit” and “implicit” forms of visual perception of objects—apparently the concepts in the literature closest to this chapter’s concepts of the concrete and semiabstract levels of palpability. But they claim that their implicit form of perception is inaccessible to consciousness. We would claim instead, first, that entities such as structural representations sensed at the semiabstract level of palpability (like those treated in section 6.9.4) can in fact be experienced in awareness at least at a vague or faint degree of clarity, rather than being wholly inaccessible to consciousness. And, second, the fact that vision and language—both largely amenable to conscious control—can generally render the structural representations of the structure subsystem explicit suggests that these representations were not inaccessibly implicit in the first place.

Separate cognitive systems for representing objects and spaces have been posited by Nadel and O’Keefe (1978), by Ungerleider and Mishkin (1982), and by Landau and Jackendoff (1993), who characterized them as the “what” and the “where” systems. To be sure, these systems fit well, respectively, into the content and structure subsystems posited in Talmy (1978a, 1988a) and here. However, the “where” system would seem to comprise only a part of the structure subsystem because the former pertains to the structural representation of an extended object array—the field with respect to which the location of a figure object is characterized—whereas the latter also includes the structural representation of any single object.

6.12 Relation of Metaphor to Fictivity

Metaphor theory, in particular as expounded by Lakoff and Johnson (1980), accords readily with general fictivity. The source domain and the target domain of a metaphor supply the two discrepant representations. The representation of an entity within the target domain is understood as factive and more veridical. The representation from the source domain that is mapped onto the entity in the target domain, on the other hand, is understood as fictive and less veridical.

For example, linguistic expressions often involve space as a source domain mapped onto time as a target domain. This can be seen in sentences like *The ordeal still lies ahead of us*, and *Christmas is coming*, where the static spatial relation of "frontality" is mapped onto the temporal relation of "subsequence," while the dynamic spatial relation of "approach" is mapped onto temporal "succession." In terms of general fictivity, factive temporality is here expressed literally in terms of fictive spatiality.

One observation arising from the fictivity perspective, perhaps not noted before, is that any of the Lakoff and Johnson's (1980) three-term formulas—for example, "Love is a journey," "Argument is war," "Seeing is touching"—is actually a cover term for a pair of complementary formulas, one of them factive and the other fictive, as represented in (27).

(27) Fictive: *X is Y* Factive: *X is not Y*

Thus, factively, love is *not* a journey, while in some fictive expressions, love *is* a journey. The very characteristic that renders an expression metaphoric—what metaphoricality depends on—is that speakers or hearers have somewhere within their cognition a belief about the target domain contrary to their cognitive representation of what is being stated about it, and have somewhere in their cognition an understanding of the discrepancy between these two representations.

One reason for choosing to adopt fictivity theory over metaphor theory as an umbrella aegis is that it is constructed to encompass cognitive systems in general rather than just to apply to language. Consider, for example, a subject viewing a round and narrow-gapped *C*-like figure. In terms of general fictivity, the subject will likely see a *C* at the concrete level of palpability—its factive representation. Concurrently for the same figure, she will sense a complete circle at the semiabstract level of palpability—its fictive representation. She will experience the former representation as more veridical and the latter one as less so, and may experience a degree of discrepancy between the two representations. This, then, is the way that the framework of general fictivity would characterize the Gestalt phenomenon of closure.

As for the framework of linguistic metaphor, if its terms were to be extended to cover vision, they might characterize the perception of the *C* figure as involving the mapping of a source domain of continuity onto a target domain of discontinuity, so that the subject experiences a visual metaphor of continuity. An extension of this sort should indeed be assayed. But at present, both psychologists and linguists might balk at the notion of closure as a metaphor. Meanwhile, the outline of a general framework for addressing such phenomena across cognitive systems is here in place.

6.13 Cognitive Bias toward Dynamism

As we have noted above, phenomena other than motion—notably, stationariness—can have fictive status in both language and vision; fictive stationariness has already been seen in frame-relative motion. In the examples given, when the scenery is fictively treated as moving toward the observer, the observer is fictively treated as stationary. In addition, certain linguistic formulations treat motion as if it were static. For example, instead of saying *I went around the tree*, which explicitly refers to my progressive forward motion, I can say *My path was a circle with the tree at its center*, which confines the fact of motion to the noun *path* and presents the remainder of the event as a static configuration.

Visual counterparts of fictive stationariness can be found in viewing such phenomena as a waterfall or the static pattern of ripples at a particular location along a flowing stream. Here one ceives a relatively constant configuration while all the physical material that constitutes the configuration constantly changes, that is, the physical material is factively moving, while the fictive pattern that it forms is stationary. This situation is the reverse of the pattern paths of section 6.8.1. There the physical substance was for the most part factively stationary, while the fictive pattern that it formed moved.

We can now compare the relative occurrence of fictive motion and fictive stationariness in language and, perhaps also, in vision. In terms of metaphor theory, fictive motion in language can be interpreted as the mapping of motion as a source domain onto stationariness as a target domain. A mapping of this sort can be seen as a form of cognitive “dynamism.” Fictive stationariness, then, is the reverse: the mapping of stationariness as a source domain onto motion as a target domain. This sort of mapping, in turn, can be understood as a form of cognitive “staticism.” Given this framework, it can be observed that, in language, fictive motion occurs preponderantly more than fictive stationariness. That is, linguistic expressions manifesting fictive motion far outnumber ones manifesting fictive stationariness. In other words, linguistic expression exhibits a strong bias toward conceptual dynamism as against staticism.

The cognitive bias toward dynamism in language shows up not only in the fact that stationary phenomena are fictively represented as motion more than the reverse. In addition, stationary phenomena considered by themselves can in some cases be represented fictively as motion even more than factively as stationariness. The factive representation of a stationary referent directly as stationary is what Talmy (1988a) calls the “synoptic perspectival mode”; in a related way, it is what Linde and Labov (1975) call a “map” and what Tversky (chapter 12, this volume) calls the “survey” form of representation. This is illustrated in (28a). Correspondingly, its fictive representation in terms of motion exemplifies Talmy’s “sequential perspectival mode,”

and, comparably, what both Linde and Labov and Tversky call the "tour" form of representation, as illustrated in (28b).

- (28) a. There are some houses in the valley.
 b. There is a house every now and then through the valley.

While this example allows both modes of representation, other examples virtually preclude a static representation, permitting only a representation in terms of fictive motion for colloquial usage, as seen in (29).

- (29) a. ?The wells' depths form a gradient that correlates with their locations on the road.
 b. The wells get deeper the further down the road they are.

In a similar way, factively static phenomena in cognitive systems other than language may also be more readily cognized in fictively dynamic terms than in static terms. For example, in vision, on viewing a picture hanging on a wall at an angle, a person may more readily ceive the picture as a square that has been tilted out of true and calls for righting, whereas he may require a special effort to ceive the picture statically as a diamond. Comparably, in the cognitive system of reasoning, one usually progresses through a proof step by step rather than seeing the full complement of logical relationships all at once.

In fact, cognitive dynamism is so much more the normal mode that the cognizing of staticism is often regarded as a special and valued achievement. Thus an individual who suddenly ceives all the components of a conceptual domain as concurrently copresent in a static pattern of interrelationships is said to have an "aha" experience, while an individual who ceives a succession of one consequent event after another through time as a simultaneous static pattern of relationships is sometimes thought to have had a visionary experience.

Acknowledgments

I am grateful to Lynn Cooper, Annette Herskovits, Kean Kaufmann, Stephen Palmer, and Mary Peterson for much valuable discussion. And my thanks to Karen Emmorey for corroborative data on fictive motion in American Sign Language, which unfortunately could not be included in the present version of this chapter for lack of space.

Notes

1. This chapter is planned as the first installment on a more extensive treatment of all the fictive categories.
2. Bucher and Palmer (1985) have shown that, when in conflict, configuration can prevail over motion as a basis for ascription of "front" status. Thus, if an equilateral triangle moves along

one of its axes of symmetry, then that line is seen as defining the front-back. Whether the triangle's vertex leads along the line of motion or trails, the line is still seen as the front. Where the vertex trails, the triangle is simply seen as moving backward.

3. Note that the notion of crossing behind a front-bearing object may be partially acceptable, possibly due to a conceptualization like this: the posited intangible line, though more salient in front, actually extends fully along the front-back axis of the object.

4. Due to the constraint noted above, this construction cannot refer to nonaligned fictive paths, for example, **The snake is lying past the light* cannot refer to a snake lying straight with its head pointing past the light. Still needing explanation, however, is why this construction cannot also be used for aligned arrangements with path geometries other than "toward" or "away from," as in **The snake is lying into/out of the mouth of the cave* to refer to a snake lying straight with its head pointing into or out of a cave mouth.

5. Probably poorer as models are such other forms of agency as an Agent's affecting some cognitive state that she herself has or some physical object that she is already in contact with.

6. This mapping may be reinforced by the fact that the prospect path ascribed to an inanimate configuration, such as a cliff wall or a window, is often associated with an actual viewer located at that configuration and directing her or his visual path along the same path as the prospect line. Thus, in (i), one readily imagines a viewer standing at the cliff edge or in the bedroom looking out along the same path as is associated with the cliff wall or the window.

(i) a. The cliff wall faces/looks out toward the butte.

b. The bedroom window faces/looks out/opens out toward the butte/onto the patio.

7. Comparisons of language structure to the structure in visual perception appear in Talmy (1978, 1983, 1988a, and this chapter) and in Jackendoff (1987). Comparisons of language structure to the structure of the reasoning system appear in Talmy (1988a); to the structure of kinesthetic perception, in Talmy (1988b); to the structure of the cognitive culture system, in Talmy (1995 and this chapter); and to the attentional system, in Talmy (1995a). And the most extensive identification and analysis to date of the foundational structural properties common to all the cognitive systems appears in the "Parameters" section of Talmy. In this work, the analysis is presented primarily with reference to a putative cognitive subsystem underlying the structure of narrative, but the analysis is intended to be quite general across the range of cognitive systems.

8. To note the correspondences, Jackendoff (1983) has abstracted a concept of pure "directedness" with four particularizations: "actual motion," "extension" (e.g., *The road goes from New York to L.A.*), corresponding to our coverage paths, "orientation" (e.g., *The arrow points to/toward the town*), corresponding to our demonstrative paths, and "end location" (e.g., *The house is over the hill*), corresponding to our access paths.

9. The term and perhaps the basic concept of *ception* derive from a short unpublished paper by Stephen Palmer and Eleanor Rosch titled "Ception: Per- and Con-". But the structuring of the ception concept found here, as well as the parameters next posited to extend through it, belong to the present approach.

Already in common usage are other terms that are neutral to any perception-conception distinction, though perhaps without much recognition of conferring that advantage. Such

terms include *representation*, *experience*, *cognize*, and sometimes *cognition*. All these terms have their particular applications and will be used in this chapter, but the new term *ception* is specifically intended to emphasize the continuity across the larger domain and the existence of largely gradient parameters that span it.

10. Perhaps alone out of the thirteen, the parameter of strength has an open-ended upper region, allowing increasingly greater degrees of intensity. Thus the point along this parameter that would tend to correlate with the high ends of the other parameters should be located within its upper region.

11. The parameter of objectivity, like the others, is intended as a phenomenological parameter. An entity is assigned to the high end of this gradient because it is *experienced* as being "out there," not because it fits a category of a theoretical ontology according to whose tenets the entity "*is*" out there.

Insofar as it is concluded in our scientific ontology that an entity is in fact located external to one's body, note further the following. Once stimuli from the entity impinge on the body's sensory receptors, the neural processing of the stimuli, including the portion that leads to conscious experiencing of the entity, never again leaves the body. Despite this fact, we experience the entity as external. We lack any direct conscious experience that our processing of the entity is itself internal. In physiological terms, we apparently lack brain-internal sense organs or other neural mechanisms that register the interior location of the processing and that transmit that information to the neural consciousness system. On the contrary, the processing is specifically organized to generate the experience of the entity's situatedness at a particular external location.

12. The adoption of the verb *to sense* as a term for this purpose is derived from its everyday colloquial usage, not from any other uses this word may have been put to in the psychological literature.

13. As treated extensively in Talmy (1988a), open-class forms are categories of forms that are large and easily augmented, consisting primarily of the roots of nouns, verbs, and adjectives. Closed-class forms are categories of forms that are relatively small and difficult to augment. Included among them are bound forms like inflectional and derivational affixes; free forms like prepositions, conjunctions, and determiners; abstract forms like grammatical categories (e.g., "nounhood" and "verbhood" *per se*), grammatical relations (e.g., subject and direct object), and word order patterns; and complexes like grammatical constructions and syntactic structures.

14. Linguistic categories like the preceding have been presented only to help illustrate the abstract end of the palpability parameter, not because that parameter is relevant to general fictivity in language. It should be recalled that the palpability gradient has here been introduced mainly to help characterize general fictivity in vision. Though linguistic reference can be located along it, this parameter is not suitable for characterizing general fictivity for language. As discussed, general fictivity in language involves the discrepancy between the representation of one's belief about a referent situation and the representation of a sentence's literal reference. The mapping of two such language-related representations into the visual modality does tend to involve a palpability contrast, but the original two representations do not.

15. Talmy (1978a, 1988a) first observed the homology between vision and language as to a content/structure distinction. These papers also present an expanded form of the linguistic demonstration synopsized in the text below.

References

- Babcock, M., and Freyd, J. (1988). Perception of dynamic information in static handwritten forms. *American Journal of Psychology*, 101, 111–130.
- Boyer, P. (1994). Cognitive constraints on cultural representations: Natural ontologies and religious ideas. In L. A. Hirschfeld and S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.
- Bucher, N. M., and Palmer, S. E. (1985). Effects of motion on the perceived pointing of ambiguous triangles. *Perception and Psychophysics*, 38, 227–236.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Cooper, L. A., and Schacter, D. L. (1992). Dissociations between structural and episodic representations of visual objects. *Current Directions in Psychological Science*, 1(5), 141–146.
- Engel, S. A., and Rubin, J. M. (1986). Detecting visual motion boundaries. In *Proceedings of the Workshop on Motion: Representation and Analysis*, IEEE Computer Society, Charleston, SC, 7–9 May.
- Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- Freyd, J. (1987). Explorations of representational momentum. *Cognitive Psychology*, 19(3), 369–401.
- Herskovits, A. (1986). Language and spatial cognition: An interdisciplinary study of the prepositions in English. Cambridge: Cambridge University Press.
- Herskovits, A. (1994). “Across” and “along”: Lexical organization and the interface between language and spatial cognition. Unpublished manuscript.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1987). On beyond zebra: The relation of linguistic and visual information. *Cognition*, 26, 89–114.
- Jepson, A., and Richards, W. (1993). *What is a Percept?* Technical report RBCV-TR-93-43. Toronto: University of Toronto Department of Computer Science.
- Keil, F. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Lakoff, G., and Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Landau, B., and Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2), 217–238.
- Langacker, R. (1987). *Foundations of cognitive grammar*. Stanford: Stanford University Press.

- Levinson, S. (1996). Relativity in spatial conception and description. In J. J. Gumperz and S. C. Levinson (Eds.), *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.
- Leyton, M. (1992). *Symmetry, causality, mind*. Cambridge, MA: MIT Press.
- Linde, C., and Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Language*, 51, 924–939.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Matsumoto, Y. (in prep.). Subjective motion and English and Japanese verbs. *Cognitive Linguistics*.
- Nadel, L., and O'Keefe, J. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Palmer, S. E. (1980). What makes triangles point: Local and global effects in configurations of ambiguous triangles. *Cognitive Psychology*, 12, 285–905.
- Palmer, S. E., and Bucher, N. M. (1981). Configural effects in perceived pointing of ambiguous triangles. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 88–114.
- Pederson, E. (1993). Geographic and manipulable space in two Tamil linguistic systems. In A. U. Frank and I. Campari (Eds.), *Spatial information theory*, Berlin: Springer.
- Pentland, A. (1986). Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28, 293–331.
- Rubin, J. M. (1986). Categories of visual motion. Ph.D. diss., Massachusetts Institute of Technology.
- Talmy, L. (1975). Semantics and syntax of motion. In J. P. Kimball (Ed.), *Syntax and semantics*, vol. 4, 181–238. New York: Academic Press.
- Talmy, L. (1976). Semantic causative types. In *Syntax and semantics*. Vol. 6, M. Shibatani (Ed.), *The grammar of causative constructions*, 43–116. New York: Academic Press.
- Talmy, L. (1978a). The relation of grammar to cognition: A synopsis. In D. Waltz (Ed.), *Proceedings of TINLAP-2* (Theoretical Issues in Natural Language Processing). Urbana: University of Illinois.
- Talmy, L. (1978b). Figure and ground in complex sentences. In *Universals of human language*. Vol. 4, J. H. Greenberg (Ed.), *Syntax*, 625–649. Stanford, CA: Stanford University Press.
- Talmy, L. (1983). How language structures space. In H. L. Pick, Jr., and L. P. Acredolo (Eds.), *Spatial orientation: Theory, research, and application*, 225–282. New York: Plenum Press.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In *Language typology and syntactic description*. Vol. 3, T. Shopen (Ed.), *Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.
- Talmy, L. (1988a). The relation of grammar to cognition. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics*, 165–205. Amsterdam: Benjamins.
- Talmy, L. (1988b). Force dynamics in language and cognition. *Cognitive Science*, 12, 49–100.

Talmy, L. (1990). Fictive motion and change in language and cognition. Plenary address at Conference of the International Pragmatics Association, Barcelona. July, 1990.

Talmy, L. (1995). The cognitive culture system. *Monist*, 78(1), 81–116.

Talmy, L. (1995a). The windowing of attention in language. In M. Shibatani and S. Thompson (Eds.), *Grammatical constructions: Their form and meaning*, Oxford: Oxford University Press.

Talmy, L. (1995b). Narrative structure in a cognitive framework. In G. Bruder, J. Duchan, and L. Hewitt (Eds.), *Deixis in narrative: A cognitive science perspective*, 421–460. Hillsdale, NJ: Erlbaum.

Ungerleider, L. G., and Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. H. W. Mansfield (Eds.), *Analysis of visual behavior*, Cambridge, MA: MIT Press.

Chapter 7

The Spatial Prepositions in English, Vector Grammar, and the Cognitive Map Theory

John O'Keefe

7.1 Introduction

In this chapter I wish to return to a subject that Lynn Nadel and I first addressed in our book *The Hippocampus as a Cognitive Map* (1978) nearly two decades ago. The gist of the argument presented there was as follows. Evidence from animal experiments proves strong evidence that the hippocampus, a cortical area in the mammalian forebrain, is involved in the construction of an allocentric spatial representation of the environment, what Tolman (1948) called “a cognitive map.” Constructed and modified during exploration (a cognitive behavior), this map provides the animal with a representation centered on the environment and locates it within that environment. During the initial exploration of an environment and subsequently, places of interest are labeled in the map and their label and locations stored for future use; these locations can subsequently be retrieved into the map and used as goals to direct behavior. For example, if a satiated animal notices food in a location during its initial exploration of an environment, it can on a subsequent occasion use that information to satisfy a hunger need. Upon finding itself in the same environment it can retrieve the location of the food and use it to direct its behavior toward that location.

This theory can account for a substantial part of the experimental literature on the infrahuman hippocampus. In order to extend the theory to account for the human data, however, we needed to extend it in two ways. First, we had to incorporate a temporal sense into the basic map to account for the ability of humans to process and store spatiotemporal, or *episodic*, information. Second, we had to allow for the impressive lateralization of function that has been repeatedly demonstrated in the human brain. Neuropsychological studies had suggested that while much of the right cerebral hemisphere is specialized for “visuospatial” processing, the left side has been given over to language function. Following her dramatic demonstration with Scoville of a memory function for structures in the mesial temporal lobe (Scoville and Milner 1957), Milner showed that this memory function respected the general lateralization

of function: patients with damage restricted to the right mesial temporal lobe were amnesic for visuospatial material, whereas those with left-sided damage were amnesic for linguistic material. Evidence gathered since has strengthened this conclusion (Smith and Milner 1981, 1989; Frisk and Milner 1990).

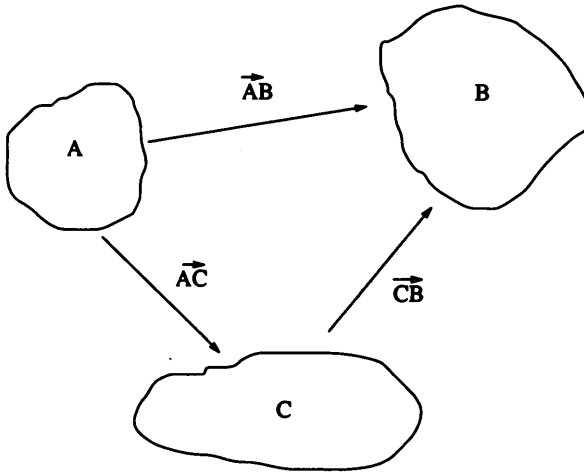
Nadel and I suggested that this lateralization of function might be due primarily to differences between the inputs to the hippocampal map on the two sides of the human brain and not necessarily to any fundamental differences in principles of operation. The right human hippocampus would receive inputs about objects and stimuli derived from the sensory analyzers of the neocortex and attributable to inputs from the external physical world. It would operate in the same way as both right and left infrahuman hippocampuses. In contrast, the left human hippocampus would receive a new set of inputs, which would come primarily from the language centers of the neocortex and would consist of the names of objects and features and not of their sensory attributes. In addition, this "semantic map" would incorporate linear temporal information and in consequence would serve as the deepest level of the linguistic system, providing the basis for narrative comprehension and narrative memory.

However, language is clearly not reducible to the set of spatial sentences; therefore we sought to create a more general framework by following the work of Gruber (1965, 1976) and Jackendoff (1976), who noted the similarity in structure between sentences such as "The message went from New York to Los Angeles," "The book went from Mary to the library," "The rock went from smooth to pitted," "The librarian went from laughing to crying." They proposed that the parallels in surface structure reflected parallels in underlying meaning, in this case the substitution of possessional sense, identificational sense, and a circumstantial sense for the positional sense of the prototype. Nadel and I interpreted this to mean it might be possible to envisage nonphysical spaces that located items, not according to their physical location, but according to their location in these other dimensions. We suggested one such dimension might be that of *influence* but did not develop this notion any further.

In this chapter I would like to develop further this idea of the semantic map. In the years that have intervened since the first publication of the idea, we have learned a great deal about the working of the infrahuman cognitive map at the physiological level, and there are now several computational models available. I intend to explore the adequacy of one of these in particular (O'Keefe 1990) as the basis for a semantic map.

Before returning to the semantic map idea, it will be helpful if I elaborate some of the details of the basic theory as developed for physical space. In the cognitive map theory, *entities* are located by their spatial relationships to each other. *Spatial relationships* are specified in terms of three variables: places, directions, and distances (figure 7.1). *Places* are patches of an environment that can vary in size and shape

ELEMENTS FOR A MAP



MAP =

PLACES	A	B	C
DIRECTIONS	$\angle AB$	$\angle AC$	$\angle CB$
DISTANCES	$ AB $	$ AC $	$ CB $

Figure 7.1

Cognitive maps consist of a set of place representations and the distances and directions between them. Distances and directions can be represented by vectors drawn from one place to another. In animals such as the rat, they are computed in real time on the basis of actual movements, whereas in higher mammals they may become autonomous from actual movements.

depending on the size of the environment and the distribution of features in that environment. They are located in terms of the spatial relations among the invariant features of the environment; they can also be located by their direction and distance from other places. The *place code* is carried by the pattern of firing of the place cells in the cortical region called the "hippocampus." *Directions* are specified as a set of parallel vectors. As with places, these can be identified in one of several ways: either as the local gradient of a universal signal such as gravity, geomagnetism, or olfactory currents, as the vector originating at a place or object and passing through another place or object (or passing through two places), or as having a specified angle to a

previously identified direction (e.g., through updating the current direction on the basis of angular head movements). For every direction there is an opposite direction, which can be marked by the negative of that vector. The *direction code* is carried by the pattern of firing of the head direction cells in the postsubiculum (see, for example, Taube, Muller, and Ranck 1990), another cortical region that neighbors on the hippocampal region and is anatomically connected to it. Distances between objects or places are given by a metric. The basic unit of this metric might be derived from one of two sources: either there is a reafference signal from the motor system which estimates the distance that a given behavior should translate the animal *or* use is made of environmental or interoceptive inputs which result from such movements. An example of an environmental input would be a change in retinal location of visual stimuli, and an example of an interoceptive input would be a vestibular signal. In either case, the geodesic distance between two objects or places needs to be computed by, for example, gating the metric signals arising from such sources by the head-direction signals so that only movements when the animal is heading in the same direction are integrated.

A *path* is an ordered sequence of places and the translation vectors between them. Paths can be identified by their end places or by a distinct name. Conversely, places along the path can be identified and associated with the path. A path may be marked by a continuous feature such as an odor trail or a road but need not be.

Within this spatial framework, translations of position in an environment are specified as *translation vectors* whose tail begins at the origin of movement and whose head ends at the destination. Vector addition and subtraction allow journeys with one or more subgoals to be represented and integrated. Furthermore, on a journey with more than one destination the optimal or minimal path can be calculated. It is still not clear whether the spatial coordinate framework is a rectilinear or a polar one and whether the metric is Euclidean or otherwise. In recent papers, I have explored Euclidean polar models (O'Keefe 1988, 1990, 1991).

If the cognitive map theory is on the right track in its contention that the left human hippocampus is basically a spatial mapping system that has been modified to store linguistic as opposed to physical information, then it might be possible to learn something about the structure of the system by analyzing the way it represents space, linguistically. A long tradition in linguistics, recently revived within case grammar theory, postulates that the deep semantic structure of language is intrinsically spatial and that other, nonspatial, propositions are in some way parasitical on these prototypical formulas, perhaps by means of metaphorical extension of their core spatial meanings. This is the contention of a group of linguists called "locationists" or "localists" (Anderson 1971; Bennett 1975). These localist theories (see Cook 1989 for a recent review) suggest that the basis for spatial sentences consists in a verb and its

associated cases. Typical cases might be agent, object, and locative, identifying the initiator of the action, the thing acted on, and the place or places of the action, respectively. In an uninflected language such as English, many of the spatial relations described in spatial sentences are conveyed by the prepositions. As Landau and Jackendoff (1993) have pointed out in their recent article, there are only a limited number of these. If this be the case then it is possible that a description of the representations set up by the spatial prepositions might provide the basis for a more general linguistics. Nadel and I speculated that the origin of language might have been the need to transmit information about the spatial layout of an area from one person to another (O'Keefe and Nadel 1978, 401n). This view suggests that at some point in their evolution hominids began to elaborate the basic cognitive map by substituting sounds for the elements in the map or for some of the sensory aspects of these elements. These maps were probably primarily transmitted as drawings in the sand or dirt with different icons standing for different environmental objects. In this way one group of a family could forage a patchy environment and report back the locations of foods to the rest of the family. Different grunts would enrich the detailed information in the map and might serve the additional purpose of acting as an encrypting device. Over time, an increase in vocabulary would eventually obviate the need for the externalized map entirely, but the neural substrate would retain the structure of the original mapping function.

In this chapter I will set out the basic framework of vector grammar and show how it accounts for many of the spatial meanings of the spatial prepositions. My thesis is that the primary role of the prepositions is to provide the spatial relationships among a set of places and objects and to specify movements and transformations in these relationships over time; these spatial relationships and their modifications can be represented by vectors.

The location of an entity within this notation is given by a vector that consists of a direction and a distance from a known location. Much of the work of the locative prepositions involves the identification of these two variables. In some cases (for example, with vertical prepositions; see below), the direction is given by an environmental signal such as the force of gravity. In most cases, however, it needs to be calculated from the spatial relationships between two or more objects or places, which specify the origin and termination (or the tail and the head) of the vector or a point along the vector. By contrast, distances are less well specified; in most cases, the metric is an interval one. One of the roles of the preposition *for* is to supply the necessary metric information. The space coded by the locative prepositions is a mixed polar-rectilinear one.

In this chapter I will assume (following the locationists; see above) that the prepositions in English have a spatial (or in one or two instances, temporal) sense as

their basic meaning and that the other meanings are derived by metaphor. I will concentrate on the locative prepositions and in particular those dealing with the vertical dimension, although others will also be covered. I will then extend the analysis to show how the temporal prepositions code for a fourth dimension, which differs only slightly from the three spatial ones, and how changes in state or location can be coded by the translational meanings of the same prepositions. If time can be coded by a fourth dimension, is it possible to incorporate other nonspatial relationships by higher dimensional axes as well? As a preliminary exploration of this question, I will conclude with a discussion of the metaphorical uses of the vertical stative prepositions to represent the nonphysical relations of status and control.

My primary concern in this chapter is to set out the premise that a vector notation can capture many of the basic meanings of the spatial prepositions in English. Consequently, I will not address in any detail the role of syntax in this kind of grammar. In general, the syntax of such a system will consist of a set of rules for relating the spatial structure of the deep semantic narrative to the temporal structure of the surface information transmission system. Thus, just as there is an associated motor programmer that translates information from the spatial map into instructions to the motor planning systems so that the animal can approach places containing desirable objects or avoid ones with undesirable objects, so also there is a production system for generating sentences from the map narrative. The syntactic rules specify, among other things, the order in which the elements of the narrative are to be read and how the different parts of the vector system are to be translated into surface elements as a function of the way that they are read. For example, the difference between the active and the passive voice in the surface sentence depends on the direction of travel along the underlying vector (head to tail or vice versa) relating an agent and its actions.

7.2 Physical Spatial Meanings of the Vertical Prepositions

In this section I shall analyze the spatial meanings of four related prepositions: *below*, *down*, *under*, and *beneath* (or *underneath*). Although these have antonyms (*above*, *up*, *over*, and *on top of*), I shall refer to these latter only when they contribute something extra to the discussion. The four prepositions have in common that they denote spatial relationships between entities¹ in one linear dimension, which I shall call the "Z-dimension." They differ from each other in interesting ways that will allow us to explore the properties of the space they depict.

7.2.1 Below

Let us begin with what I believe to be the most basic of the four prepositions, *below*. On my reading, *below* relates two entities (*A* and *B*) in terms of their relative location along the Z-direction. Consider the simple deictic sentence

(1) John is below.

Because *below* is a bipolar preposition, there must be a second suppressed term, which I shall argue is the place occupied by the speaker or the listener. John or his place is *A*, the speaker's (or listener's) place is *B*, and the relationship between them is as follows: the magnitude of the component of *A*'s place in the $-Z$ -direction is greater than the magnitude of *B*'s place. In order to make the assertion in (1), or to assess its validity, we need a notation for specifying the Z -direction, a way of locating *A* and *B* along that direction, and means for assessing whether *A* or *B* has a larger component along that direction. The most convenient notation for accomplishing these is vector algebra.

In this notation a direction is designated by a set of parallel vectors of unlimited magnitude and unspecified metric. The location of each entity is specified by a vector drawn from an observer to the entity. This vector can be specified by a magnitude *R* and an angle ϕ with the direction vector through the point observer (figure 7.2). The component of the vector *A* along the Z -axis can be computed by calculating the inner product of *A* and Z given by the formula:

$$A_z = A \cos \phi,$$

where *A* is the magnitude of *A* and ϕ is the angle that *A* makes with the Z -direction vector at observer (obs).

In the deictic example of sentence 1, *A* is *below* the observer if $A_z < \text{obs}$, and *above* the observer if $A_z > \text{obs}$. The same considerations allow the observer to decide whether *A* is below *B* when neither is located at the observer (figure 7.3 shows this situation). Again, the question of whether *A* is below or above *B* can be assessed by comparing their relative magnitudes along the Z -axis.

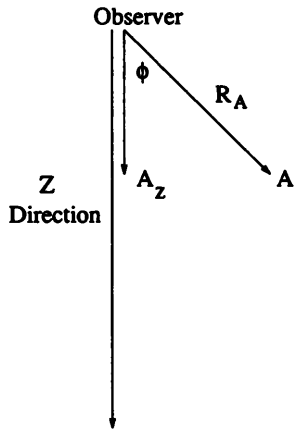
If $A_z - B_z > 0$, *A* above *B*;

If $A_z - B_z < 0$, *A* below *B*.

Note that the relationship between *A* and *B* is perfectly symmetrical and that neither *A* nor *B* can be considered a reference entity in the deep structural representation of the relationship. Choice of one or other as the referent in the surface sentence may depend more on the topic of the discussion, the previous sentences, which of the two entities has already been located, which is easier to locate perceptually, and other considerations. The *below* relationship is a transitive one. By simple transitivity of arithmetical relations on the Z -dimension,

if $A_z > B_z$ and $B_z > D_z$,

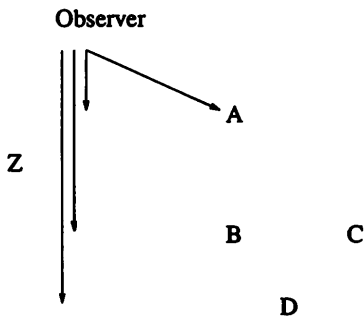
$\therefore A_z > D_z$.



A BELOW Observer

Figure 7.2

Vector location and the *below* relation. The location of an entity *A* can be represented by a vector drawn from the observer to that entity. The vector is characterized by a distance *R* and an angle ϕ measured with respect to a direction *Z*. The projection of the vector onto the *Z* direction is shown as A_z .



B,C,D Below A

Figure 7.3

Each item *A*, *B*, *C*, and *D* has a projection onto the *Z*-axis. The relative lengths of the projection onto this axis determine which items are below which. In the example, *B* and *C* have identical projections and are therefore both equally below *A*.

In figures 7.2 and 7.3, I chose to represent entities *A–D* in an allocentric framework; that is, I assumed that they existed in an environmental framework independent of the location of the observers and that their relationship within the framework could be assessed independently of the locations of the observers. Further, I assumed that the distances from each observer to the entities was known or could be computed, for example, by movement parallax. Does this imply that the spatial relationship denoted by *below* can be computed only within an allocentric framework? Can we say anything about the constraints on frameworks that can be used?

In general, the use of *below* relies on the availability of a direction vector shared between the speaker and listener; in the case of the allocentric framework, this is provided by the universal gravity signal. There are, however, other, more limited uses of *below* that employ egocentric and object-centered directional vectors. Egocentric vectors are fixed to the body surface of the observer, and object-centered vectors are fixed to the entity or entities related. Sentences 2–5 are examples.

- (2) The new planet appeared below the moon.
- (3) Below this line on the page.
- (4) Hitting below the belt.
- (5) The label below the neck of the bottle.

The egocentric use occurs under circumstances (a) where the entities are very far away from the observer and therefore do not change relative locations with observer location; or (b) where the entities are constrained to lie on the *XZ*-plane, as on a page or a video display unit. In the former case, the conversants must ensure that they are similarly aligned to each other relative to the entities or that there is a conventional orientation relative to the gravity signal that enables the *Z*-direction to be labeled conventionally. This is most obvious with the specialized case of the parts of the human body, which are probably labeled by reference to their canonical orientation relative to gravity (see Levelt, chapter 3, this volume). The case of the bottle and similar manufactured objects that refer to body parts (back of a chair, leg of a stool) would seem to follow the same rule. In general, then, normal conversation would seem to require the use of an allocentric framework for most purposes, for the reasons pointed out by other contributors to the present volume (Levelt, chapter 3; Levinson, chapter 4). Even the ability to see things from another's point of view would appear to involve computations based on an underlying knowledge of the two observers' locations in allocentric space.

A second conclusion can be drawn about the underlying framework on the basis of our discussion of *below*. Where it is used to describe an allocentric relationship, the framework cannot be a simple polar coordinate system, but must have at least one

rectilinear axis. This follows from the simple observation that in a polar coordinate system the *below* relation cannot be specified by one variable alone, but requires two variables: a distance and an angle (see figure 7.2). It follows therefore that the most parsimonious theory would specify the *Z*-direction by a single dimension in all usages. As we shall see, this does not necessarily imply that the other two (non-*Z*) dimensions are also rectilinear.

We have, then, evidence for a single dimension along which entities can be located. Can we say anything more about the metric at this stage, and if so, how are distances specified along this dimension? Scales² differ in the type of metric employed. Roughly, this describes the relationship of the observations or measurements to the system of real numbers. The usual categories of scales are the nominal, ordinal, interval, ratio, and absolute; they differ in the number of properties of the real number system they respect. This is most easily characterized by the types of transformations that can be applied to the assigned values without transforming the relationship of the scale to the thing measured. *Nominal scales* are simple classification scales in which the labels stand for the names of classes. For the purposes of the scaling, the elements within each class are considered equivalent and different from all the elements in the other classes. No other relationship among the elements is implied, and only transforms equivalent to the relabeling of the classes are allowed. Clearly, the *below* relationship satisfies a nominal scale. *Ordinal scales* consist of a series of numbers such that observations equal to each other are assigned the same number and an observation larger than another is assigned a larger number, but no significance is attached to the interval between the numbers. The relationship between numbers is transitive because $m > n$ and $n > p$ implies $m > p$, and all mathematical transformations that maintain the monotonic ordering of the numerical assignments are permissible. Because it is possible to say that *B below A* and *C below B* implies *C below A*, we are dealing with at least an ordinal scale. *Interval scales* are ordinal scales that, in addition, provide information about the differences between the scale values. In particular, they assert that some differences are equal to each other. For example, $m - n = p - q$. Transformations that preserve the differences between values as well as their ordering are permissible. Specifically, the values of one scale can be multiplied by a positive constant and added to another constant without consequence to relationships.

$$Z_2 = aZ_1 + b, a > 0$$

In this linear transform, a changes the gain of the metric, and b the origin. It would appear that the *below* directional scale comes close to fulfilling the requirements for an interval scale. One way of testing this is to ask whether it is possible to apply the comparative operator *more* to the preposition and thus to derive equivalent intervals of *belowness*. The question is whether the comparative notion is an intrinsic part of

the meaning of *below* or merely an extension of it. I would argue that because it is always legitimate to ask for the relationships set out in (8), the scale is an interval one. Indeed, it may not be possible to compute the vector calculations suggested in this chapter on material ordered on less than an interval scale.

- (6) *A* and *B* are below *C*. (nominal)
- (7) *A* is more below *C* than *B*. (ordinal)
- (8) *A* is as far below *B* as *C* is below *D*. (interval)

Compare these to

- (6a) *A* and *B* are brighter than *C*.
- (7a) **A* is more brighter compared to *B* than to *C*.
- (8a) **A* is more brighter than *B* by the same amount as *C* is more brighter than *D*.

Ratio scales are interval scales that do not have an arbitrary origin. Here the only permissible transform is the gain of the scale

$$Z_2 = aZ_1, a > 0.$$

In *absolute scales*, the final category we shall consider, no transfers are allowed and the underlying assumption is that the real number system uniquely maps onto the observations

$$Z_2 = Z_1.$$

As we have seen already, the metric of the *below* relationship is at least ordinal, and probably interval. But is it ratio? Here the fact that the *below* relationship can be assessed from any arbitrary observation point and can use any origin suggests that it does not rely on a fixed origin but is invariant under arbitrary translations. Furthermore, it is intuitively obvious that changes in scale do not affect the relationship either. These suggest that it falls short of a ratio scale. It can, however, be elevated into a ratio or even an absolute scale by the provision of explicit metric information.

- (9) a. *A* is twice as far below *B* as *C* is.
- b. *A* is three feet below the surface.

7.2.2 Down (and Up)

The locative meaning of *down* is related to that of *below* in that it specifies the direction of the entity as lying in the $-Z$ -direction. In addition, however, it requires a line or surface that is not orthogonal to the Z -direction and on which the entity is located. This line or surface is the object of the preposition *down*. As with *below*, the directional component of *down* is relative to another entity, which in this case is

governed by the preposition *from*. In general the preposition *from* identifies the source or tail of a direction vector. If this information is not supplied explicitly, it is assumed that the referent is the deictic location *here*.

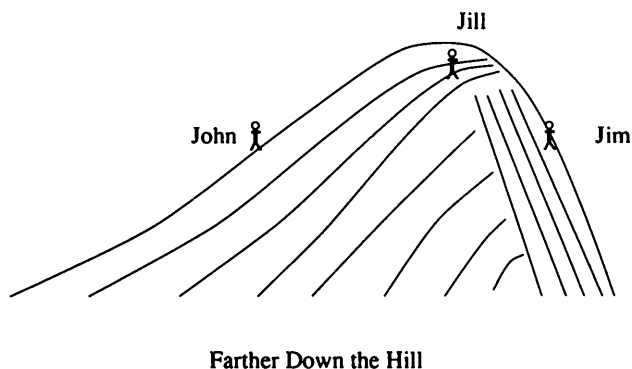
- (10) The house is down the hill (from here).
- (11) Just down the tree from Sam was a large tiger.
- (12) *The boat was down the ocean.

Thus there are two reference entities: a plane or line that I shall call the "reference plane" and a place or object that I shall call the "reference entity." As long as the extended reference entity is not horizontal (perpendicular to the *Z*-axis) as in (12), it can be a one-dimensional line or a two-dimensional surface. Intuitively, this reference entity should be a linear or at least monotonically decreasing function of *Z* over the relevant range. Someone on the other side of the hill, regardless of the person's relative $-Z$ -coordinate, is not *down* the hill from you. Similarly, a local minimum on the slope of the hill between the entity located and the reference entity disrupts the use of *down*. To put it another way, the preposition *down* can only take as direct objects entities that have or can be treated as having monotonic slopes in the nonhorizontal plane. Applying our comparative *more* to the preposition *down*, we find, as we did with *below*, that its primitive sense is to operate on the *Z*-component of the relationship.

- (13) John is more (farther) down the hill than Jill.

John and Jill are both located on the hill, the hill has a projection onto the *Z*-dimension, and John has a larger $-Z$ than Jill. There is no interaction between the steepness of the reference plane and the sense of the preposition. This can be tested by asking the question of the three people in figure 7.4 Who is more (farther) down the hill from Jill? John or Jim?

My sense of the meaning of *down* is that neither John nor Jim is more down from Jill than the other, indicating that the non-*Z*-dimensions are irrelevant. However, the ability to extract the *Z*-component from a sloping line or surface suggests either that these can be decomposed into two orthogonal components (*Z* and non-*Z*) or that their projections onto the *Z*-axis can be computed. It seems, then, on the basis of our analysis of *down*, that we are dealing with at least a two-dimensional coordinate system in which one dimension is vertical and the other one or more dimensions, orthogonal to this. As with the *below/above* direction, the difference between *down* and its antonym *up* is merely a change of sign and there are no obvious asymmetries. If *A* is *down* from *B*, then *B* is *up* from *A*. The measurement scale of the *Z*-axis would appear to be an interval one and there is clear evidence of the absence of a true 0 or

**Figure 7.4**

Down measures the relationship in the *Z*-direction. John and Jim are equally far *down* the hill from Jill, despite different lateral displacements.

origin (this is relative to the reference point identified by *from*), and therefore the scale is not a ratio one. The scale of the other two dimensions is not clear from the two prepositions *below* and *down* because the use of the comparative operator *more* in conjunction with these only operates on the *Z*-component of the meaning. Evidence about these other dimensions can, however, be garnered from an analysis of the third of our prepositions, *under*.

7.2.3 Under

Under is similar to *down* and *below* in that it also codes for the spatial relationship between two entities in the *Z*-direction. In addition, however, it places restrictions on the location of these entities in one or two directions orthogonal to the *Z*-direction. If *B* is *under* *A*, then it must have a more negative value in the *Z*-dimension. In addition, however, it must have one or more locations in common in at least one orthogonal dimension (let us call them *X* and *Y* for the moment without prejudice to the question of the best representation of relationships in this plane). The projection of the entity onto the *X*-direction is determined in the same way as that onto the *Z*-direction by calculating the inner product of the vector drawn to the entity from an observer. Figure 7.5 shows this relationship for three pointlike objects. The relation depicted is conveyed by the sentences

(14) *C* is under *A* but not under *B*; *B* is not under *A*.

When one or more of the entities is extended in one or more of the non-*Z*-directions, the *under* relationship can be assessed by the same algorithm. For example, if the entities are extended in the *XY*-plane, then an overlap in any location in the

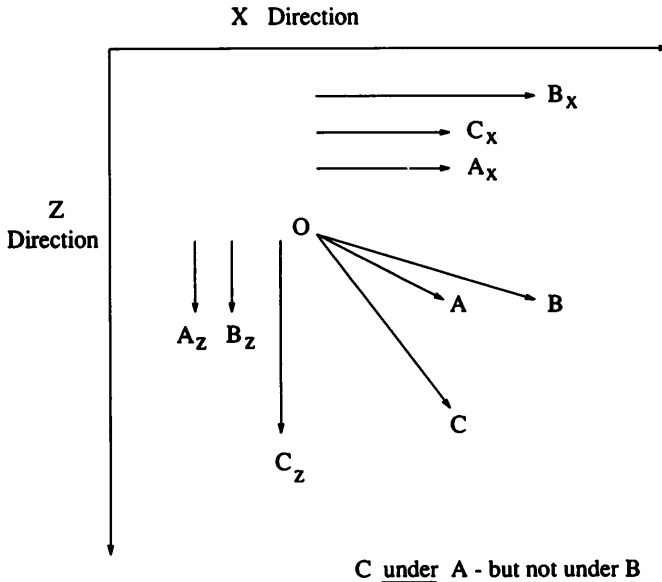


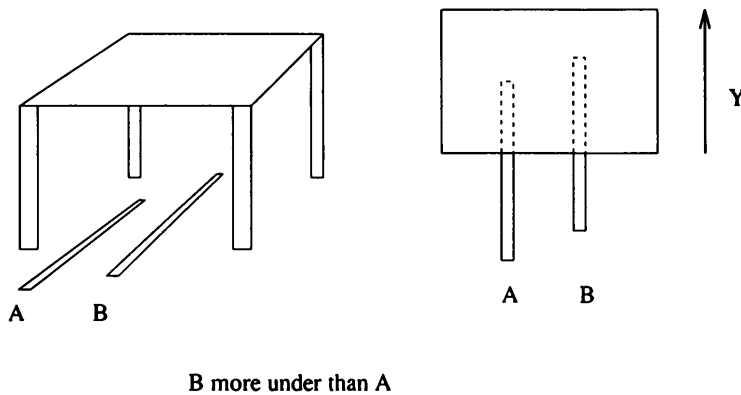
Figure 7.5

Under represents a spatial relationship in the *XY*-plane as well as the *Z*-direction. *C* is *under A* because it has the same *X*-length and a greater *−Z*-length. *C* is not *under B* because the B_x and C_x lengths differ.

XY-plane suffices. Note that unlike *below*, *under* is not transitive when applied to entities that are extended in the *XY*-plane. *B under A* and *C under B* does not mean that *C* is *under A*. Another interesting difference between *under*, on the one hand, and *down* and *below*, on the other, arises when we examine the locus of operation of the comparator *more*. Recall that when applied to *below* and *down*, *more* acted to increase the length of the *Z*-component of the vector to the entity. When applied to *under*, the effect of the comparator is not fixed but depends on the relative dimensions of the two entities. Let us leave aside for the moment the small number of usages that seem to mean that there is no intervening entity between the two relata:

- (15) Under the canopy of the heavens.
- (16) Under the widening sky.

The comparator cannot be applied to these usages, which I shall designate *under*¹. In the more frequent usage of *under*, the comparator is more often found to operate on the orthogonal *X*-dimension than on the primary *Z*-dimension. Compare the following two sentences:

**Figure 7.6**

Stick *B* is *farther (more) under* the table than stick *A* because there is a greater length of overlap with the projection onto the *XY*-plane.

(17) The wreck was farther under the water than expected.

(18) The box was farther under the table than expected.

Ignoring the metonymic uses of table and water, it is clear that the first usage, (17), implies a greater depth or *Z*-dimension, while the second, (18), implies a greater length in the *X*-dimension. In the first usage, which I shall designate *under*², *under* acts as a synonym for *below*, and the substitution can usually be made transparently. These usages may be confined to situations in which the upper entity is very long relative to the lower one and completely overlaps with it. It follows that any change in the lateral location of the lower one will not affect the amount of overlap, and there is no information contained in the preposition about the lateral variable. In contrast, where both relata have a limited extension in the *XY*-plane, *under*² is responsive to these dimensions. We can use this fact to explore the properties of the second and third dimensions of spatial language and the relations between these and the *Z*-dimension. Consider sentence (19) and related figure 7.6:

(19) Stick *A* was under the table, but stick *B* was even farther under it.

I read sentence (19) to mean that both sticks *A* and *B* and the table (top) have projections onto the *XY*-plane and these projections overlap, that is, have locations in common. Further, the magnitude of some aspect of the projection of *B* onto the table is greater than that of *A*. In general, this magnitude will be a length along some vector (e.g., *Y* in figure 7.6) measured from the edge of the table to the farthest edge

of the object projection. Furthermore, any differences in the projections of the objects in the Z -direction are irrelevant. Thus

(20) Box A was farther below the shelf than box B and farther under it.

Applying the comparative test to the preposition *under*² reveals that the metric is the same as that for the $-Z$ -direction, that is, an interval scale.

(21) Chair A was as far under the table as chair B .

Note that this sentence can be used even when the chairs are at right angles to each other, in which case each distance is measured from the edge of the table intersected by the chair. The sentence also confirms that both measurements are on an interval scale and that the same metric applies to each. This conclusion is strengthened by the fact that it makes sense to say

(22) Chair A was as far under the table as it was below it.

This last sentence also suggests that the meaning of *under*² in the XY -plane is a distance and not an area. Evidence for this can be gained by imagining the same or different objects of different projection sizes and exploring the meaning of

(23) A farther under than B ,

as these objects are positioned in different ways *under* a constant-size table (see figure 7.7). Figure 7.7 shows that the judgment of which objects are *more under* (or *more under*²) does not depend on the relative proportion of the length that intersects with the reference object (B *more under* than A); the orientation of the objects need not necessarily be the same because the relevant length is taken from the intersection of the object with the edge of the table or from the nearest edge (C is *as far under* as B).

My claim that A *more under*² refers to the absolute length of A might appear to be contradicted by sentences such as

(24) Mary got more under the umbrella than Jane and thus got less wet.

This clearly implies that Mary got more of herself (i.e., a greater proportion) under the umbrella. In this usage, however, it is clear that “more” modifies “Mary” rather than “under,” and does not constitute a refutation of the present proposal.

Finally, D *more under*² than C in figure 7.7 suggests that when an object has two dimensions either of which could be taken into consideration, the distance *under*² is taken from the longer length. It is interesting to note that, unlike the antonyms *up* (for *down*) and *above* (for *below*), *over* does not show complete symmetry with *under*². In some subtle sense, the table is less *over* the chair than the chair is *under*² the table. This slight asymmetry appears not to relate so much to size as to relative mobility. Consider (25) and (26):

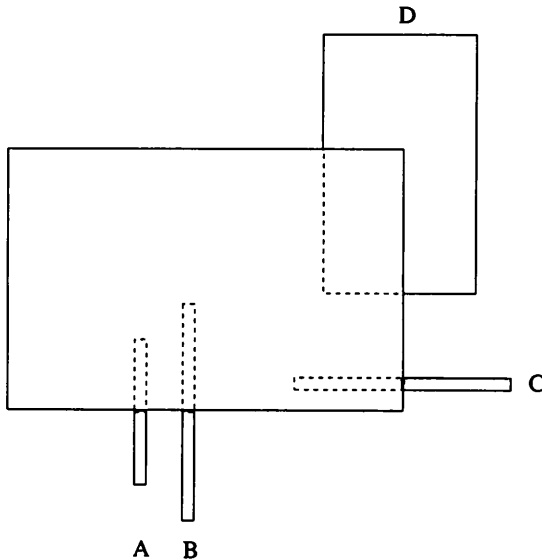


Figure 7.7

The relationship *more under* is determined by the total length of the overlap between the two objects in the *XY*-plane and not by the proportion of the total object which is *under* ($B > A$), or the orientation of the object ($C > A$). When two objects differ in more than one dimension, *farther under* is determined by the largest dimension of each and not by the total area ($D > C$).

(25) The red car was under the street lamp.

(26) The street lamp was over the red car.

Sentence (26) is not incorrect, but less likely in most contexts. The reason for this, at least in part, may be that the places in the cognitive map are specified primarily by the invariant features of an environment and only secondarily and transiently by objects which occupy them.

7.2.4 Beneath (or Underneath)

Beneath (or *underneath*) has a meaning that is close to that of *under* but differs in two ways. First, it has a more restricted sense in the *XY*-plane. Whereas *under* means an overlap between the projections of the reference entity and the target entity, *beneath* means that the target entity is wholly contained within the limits of the reference entity projection. It follows that the projection of the lower entity in the *XY*-plane must be smaller than the upper. Furthermore, and in part as a consequence of this restriction, the application of the comparator *more* (or *farther*) to *beneath* operates on the *Z*-direction and not on the *XY*-plane.

(27) The red tray was farther beneath the top of the stack than the blue one.

Beneath then means that the target element is contained within the volume of space defined by its *XY*-projection through a large (or infinite) distance in the $-Z$ -direction. *Underneath* seems to have a slightly more restricted meaning in the sense of limiting the projection in the $-Z$ -direction. *More underneath* sounds less acceptable than *more beneath* and might indicate that *underneath* is a three-dimensional volume of space restricted to the immediate proximity of the $-Z$ or *under* surface of the reference element.

7.3 Distance Prepositions

Distances are given by the preposition *for* and the adverbials *near (to)* and its antonym *far (from)* as in (28) and (29).

(28) This road goes on for three miles.

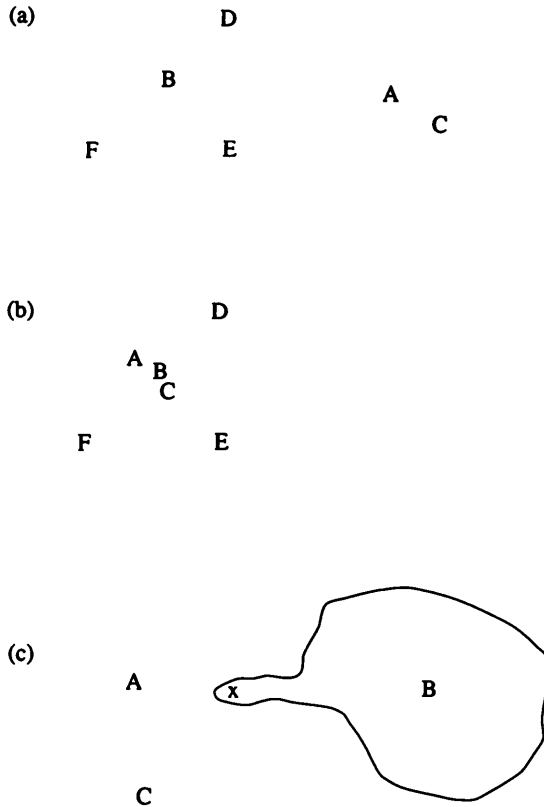
(29) The house was near (far from) the lake.

For gives the length of a path; *near* and *far from* give relative distances that are contextually dependent. In some cases, one or more of the contextual referents have been omitted. Let us begin by examining the meaning of *near* when points are being related. O'Keefe and Nadel (1978, 8) observed that the meaning of *near* was context-dependent, and I will pursue that line here. It follows that, with only two points, neither is *near* (or *far from*) the other. Three points, *A*, *B*, and *C*, provide the necessary and sufficient condition for use of the comparatives *nearer* and *farther*. Note that the directions of the points from each other are not confined to the same dimension but are free to vary across all three dimensions, and that the distance is measured along the geodesic line determined by the Euclidean metric. *Near* is not simply derived from *nearer* but contains in addition a sense of the proportional distances among the items in question.

(30) *A* is not near *B* but it is nearer to *B* than *C* is.

The distance measure incorporated in *near* seems to be calibrated relative to distances between the items with the smallest and largest Euclidean distance separation in the set. These items act as anchor points that control the meaning of the terms for all the others. Changing the relations of other items in the set can alter whether two items are near to or far from each other. Thus, in figure 7.8a, *B* and *E* are near each other, but in figure 7.8b, they are not.

Consideration of the *near/far* relationship of two- or three-dimensional entities shows it is the surface points that are important and not any other aspect of their

**Figure 7.8**

Nearness is context-dependent. In (a) *A* is not *near B* but *nearer* than *C*. *E* is *near B* in (a) but not in (b). In (c), *B* is *nearer A* than *C* is by virtue of point *x*.

shape (e.g., centroid) or mass (center of gravity). If we inspect figure 7.8c and ask which is nearer to *A*, shape *B* or shape *C*, we will see that *B* is, by virtue of point *x*. Finally, the presence of barriers seems not to influence our judgment of *near* or *far*, because (31) is permissible.

(31) The house is nearby, but it will take a long time to get there since we have to go the long way around.

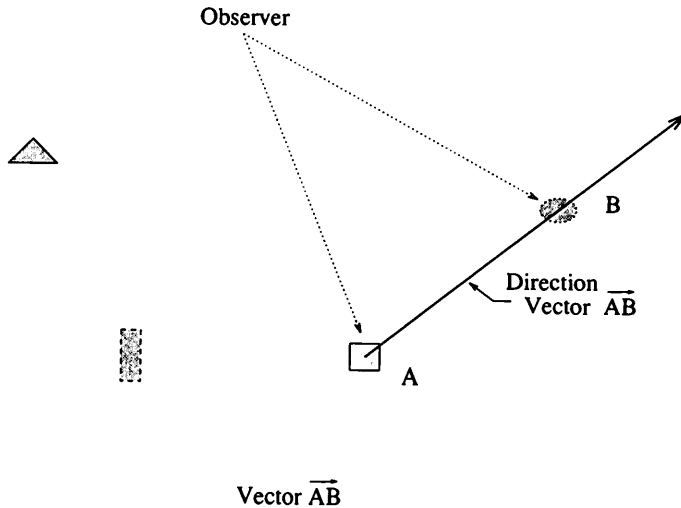
7.4 Vertical Prepositions: Reprise

These considerations of the meanings of the vertical prepositions suggest the following conclusions:

1. Prepositions identify relationships between places, directions, and distances, or combinations of these. Static locative prepositions relate two entities; static directional prepositions relate three entities because there is always an (often implied) origin of the directional vector; and static distance prepositions also relate three entities because this is the minimum required to give substance to the comparative judgment that they imply.
2. The space mapped by the prepositions is at least two-dimensional and rectilinear in the vertical direction. The nonvertical dimension (if present) may be rectilinear, but there are also circumstances in which the two nonvertical dimensions may be expressed in polar (or other) coordinates.
3. The metric of vertical and nonvertical axes is identical because it is possible to compare distances along orthogonal axes. Interestingly, the distance between objects is calculated from the nearest surface of each entity and not from some alternative derived location such as the geometric centroid or center of mass.
4. The scale is an interval scale with a relative origin determined by one of the reference entities of the directional prepositions (usually the vector source or tail).
5. In the vertical dimension, direction can be given by the universal gravity signal, which is constant regardless of location. In the horizontal plane, nothing comparable to this signal is available and the direction vectors must be computed from the relative positions of environmental cues.³

7.5 Horizontal Prepositions

The original cognitive map theory suggested that, in the horizontal plane, places could be located in several ways. Foremost among these was their relation to other places as determined by vectors that coded for distance and direction (figure 7.1). In a recent paper (O'Keefe 1990) I have suggested that the direction component of this vector is carried by the head direction cells of the postsubiculum. These cells are selective for facing in specific directions relative to the environmental frame, irrespective of the animal's location in that environment. The direction vector originating in one place or entity and running through a second can be computed by vector subtraction (see figure 7.9) of the two vectors from the observer to each of the entities, and this computation is independent of the observer's location. The resultant direction vector functions in the same way in the horizontal plane as the gravitational signal in the vertical direction. The primary difference is that, whereas the latter is a universal signal, the horizontal direction vectors are local and need to be coordinated relative to each other. This is achieved by mapping them onto the global directional system. Locative horizontal prepositions, in common with their vertical cousins, specify places in terms of directions and distances. The directions are given relative to the

**Figure 7.9**

The direction vector through two objects *A* and *B* can be computed by taking the difference between the vectors *A* and *B*.

direction vector, and distances are given relative to the length of a standard vector drawn between the two reference entities along the reference direction.

7.5.1 *Beyond*

Let us begin with an analysis of the spatial meaning of the preposition *beyond*. As shown on the left side of figure 7.10, this specifies a three-dimensional region located by the set of vectors with a specific relationship to the reference direction and a pair of reference vectors ($\vec{AB}$, $\vec{AC}$) terminating on different parts of the reference object or place. The region *beyond* the mound is specified by the set of vectors originating at *A* whose projection onto the direction vector (inner product) has a greater length than the larger of the two reference vectors coincident with the direction vector ($\vec{AC}$). According to this definition, it acts in a manner analogous to *below* in the vertical dimension. No restriction is placed on the location of the entity in the vertical direction, as can be seen from sentence (32):

(32) Jane camped beyond and above the woods.

Furthermore, the effect of the comparator *more* is to act on the length of the vector in the horizontal plane:

(33) The tower was farther beyond the mound than the castle.

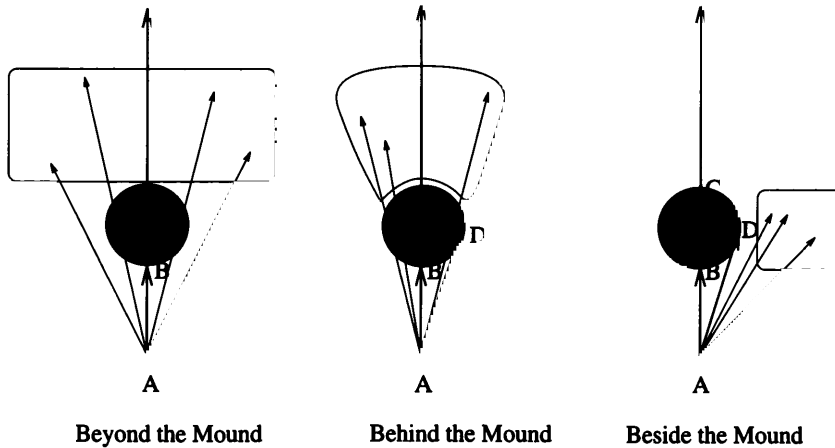


Figure 7.10

Beyond, *behind*, and *beside* can be represented as places determined by their relation to the direction vector drawn through two reference entities and a set of reference vectors ($\overrightarrow{AB}$, $\overrightarrow{AC}$, $\overrightarrow{AD}$). *Beyond* is the set of all places with a length greater than AC . *Behind* is a restricted subset of *beyond* and includes only the places with location vectors greater than AC and angle with the direction vector smaller than AD . *Beside* represents those places having a projection onto the reference direction of magnitude greater than AB and less than AC . In addition the angle with the direction vector must exceed that of AD .

The opposite of *beyond* is the seldom-used *behither*, and this simply means that the location vector has a length less than the reference vector $\overrightarrow{AB}$.

7.5.2 Behind

Behind functions in a manner analogous to *under* in that it places greater restrictions on location than does *beyond*. An object *behind* a reference entity is located by the set of vectors with a larger magnitude than the reference vector ($\overrightarrow{AC}$) but with an angle less than vector $\overrightarrow{AD}$ (figure 7.10, center). As with *under*, an entity can be partially *behind* the reference entity, and the test for this is an overlap in the projections of the two in the XZ -plane. This need for overlap accounts for the awkwardness in using *behind* with referents that are not extended in the vertical dimension.

(34) ?The tree was *behind* the trench.

(35) ?The cottage was *behind* the lake.

The application of the comparator test shows further similarities. In the same way that *farther under* can refer to the amount of overlap in the XY -plane between two

entities separated in the vertical dimension, so farther *behind* can refer to greater overlap in the *XZ*-plane of entities separated along a horizontal reference direction.

(36) The red toy was pushed farther behind the box than the blue ball.

The source of the direction vector can be specified explicitly as the object of the preposition *from*.

(37) From where Jane stood, James was hidden behind the boulder.

More usually, the source is implicit, being inferable from the previous context. In sentence (37), for example, it would be legitimate to omit the first clause if the previous narrative had established that Jane had been looking for James. More often, the source of the direction vector is the implicit deictic *here*. In a pool game it might be the cue ball:

(38) The last red was behind the eight ball.

Familiar objects have “natural” *behinds* established by a vector drawn from one differential part to another, as, for example, the front to the back of a car. However, this is easily overridden by the motion of the vehicle:

(39) The car careered backward down the hill, scattering pedestrians in front of it and leaving a trail of destruction behind it.

The opposite of *behind* is *before*, or more usually *in front of*.

7.5.3 *Beside*

Beside identifies a region at the end of the set of vectors whose projections onto the reference direction fall between the reference vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ but whose angle with the reference direction is greater than that of reference vector $\overrightarrow{AD}$ (figure 7.10, right).

7.5.4 *By*

By is the generalized horizontal preposition and includes the meanings of *before*, *behind*, *beyond*, and *beside* with a slight preference for the latter.

7.6 Omnidirectional Prepositions

At, *about*, *around*, *between*, *among* (*amid*), *along*, *across*, *opposite*, *against*, *from*, *to*, *via*, and *through* locate entities in terms of their relationships to other entities irrespective of their direction in a coordinate reference framework and therefore can be used in any of the three directions. *At* is the general one-to-one substitution operator that locates the entity in the same place as the reference entity. *About* relaxes the precision

of the localization and introduces a small uncertainty into the substitution. *About* is equivalent to *at* plus contiguous places. In the cognitive map theory the size of the place fields is a function of the overall environment, and this would appear to apply to *about* as well. Therefore the area covered by *about* is relative to the distribution of the other distances in the set under consideration in the same way that the meaning of *near* depends on the distribution of the entities within the set. *Around* has at least two distinct meanings, both related to the underlying figure of a circle (i.e., the set of vectors of a constant R originating at an entity) with the reference entity at its center. The first meaning is that the located entity is somewhere on that circle. If it is extended, it lies on several contiguous places along the circle; if more compact, it lies at one place on the circle perhaps at the end of an arc of the circle.

(40) The shop was around the corner.

Because in almost all instances the radius of the circle is left undefined, except that it be small relative to the average interentity distances of the other members of the set, there is little to choose between the use of *about* and *around* when single entities are located. When multiple entities are located, there is the weak presumption that they all lie on the same circle when *around* is used, but not when *about* is used.

(41) Those who could not fit around the table sat scattered about the room.

Between locates the entity on the geodesic connecting the two reference entities. The computation is the same as that for deriving a direction vector from the subtraction of two entity vectors (see above discussion in section 7.5), except that the order in which these are taken is ignored. An equivalent definition of *between* is that the sum of the distances from each of the reference entities to the target entity is not greater than the distance between the two reference entities. Alternatively, the angle made by the vectors joining the target to each of the references should be 180° . *Among* increases the number of reference entities to greater than the two of *between*. The interesting issue here, as with many of these prepositions that use multiple reference entities, is how the reference set is defined. *Among* roughly means that the target entity is within some imaginary boundary formed by the lines connecting the outermost items of the set. But clearly the membership of the reference set itself is not immediately obvious. Consider a cluster of trees with an individual outlier pine tree some distance from the main group.

(42) He was not among the trees, but stood between the thicket and the lone pine.

This suggests that the application of the preposition *among* depends on a prior clustering operation that is necessary to determine the numbers of the reference set. *Amid* is a stronger version of *among* that conveys the sense of a location near to the center

of the reference entities. One possibility is that the centroid or geometrical center of the cluster is computed, and *amid* denotes a location not too far from this. The centroid is a central concept in one computational version of the cognitive map theory (O'Keefe 1990).

Across, *along*, and *opposite* are like *down* in that they situate an entity in terms of its relationship to a reference entity and a one- or two-dimensional feature. Two-dimensional features are usually more extended in one direction than the other. *Across* specifies that the vector *from* the reference entity to the target intersects the reference line or plane an odd number of times. *Along* specifies an even number (including 0) of intersections. In addition, there is the weak presumption that the distance from the target entity to the last intersection is roughly the same as from the reference entity to the first intersection; that is, both are roughly the same distance from the reference line or plane. *Opposite* restricts the number of intersections to one and the intersection angle to 90°.

Against specifies that the entity is in contact with the surface of the reference entity at at least one point. It is, however, not attached to it but is supported independently in the vertical dimension. In the present scheme, *from* and *to* mark places at the beginning and end of a path that consists of a set of connected places, and *via* and *through* specify some of the places along the way.

- (43) Oxford Street goes from Tottenham Court Road to Marble Arch via Bond Street but doesn't pass through Hyde Park.

7.7 Temporal Prepositions and the Fourth Dimension

The incorporation of time into the mapping system is accomplished through various grammatical and lexical features. The primary grammatical features are tense, aspect, and the temporal prepositions. Because my emphasis in this chapter is on the prepositional system, I will mention tense and aspect only in passing (see Comrie 1975, 1976/1985 for detailed discussions).

In the present system, time is represented as a set of vectors along a fourth dimension at right angles to the three spatial ones. Each event is represented as a vector that is oriented with its tail to the left and its head to the right, this constraint being due to the fact that changes in time can take place in only one direction (from past to future). The location of these time events is also based on vectors and these can be oriented in either direction from a reference point, which can be the present moment of the utterance or some other time. Times future to the reference point have vectors of positive length, times past have vectors of negative length, and the present, a vector of 0 length. These different times are represented by the tenses of the verb.

The choice of the present time as a 0 reference point is traditionally called "absolute tense" while that of a nonpresent reference point, "relative tense" (see Comrie 1985 for further discussion). Because the vectors representing time are all unidimensional, lying parallel to the fourth axis, we will expect that the senses of the temporal prepositions are also unidirectional. For example, most of the temporal prepositions are similar to (diachronically borrowed from?) their homophonic spatial counterparts, but not all spatial prepositions can be so employed. The general rule seems to be that only spatial prepositions that can operate in the single, nonvertical dimension of the line can be borrowed in this way (but see the special cases *around* and *about*). As we shall see, this leaves the nonphysical vertical prepositions free to represent specialized relationships between entities.

The temporal prepositions, then, specify the location, order, and direction within the fourth dimension of the entities and events of the other three dimensions. In my brief summary I will classify them according to whether they use one or more reference points. Because the temporal dimension appears to be confined to a single axis orthogonal to the spatial axes, in the latter cases the two references are confined to that axis and are therefore collinear. My discussion of the meanings of the temporal prepositions will be based on the abstract events portrayed in figure 7.11. The upper event shows a state of affairs in which an entity occupies a vertical location before time *A*, then jumps to a new location and remains there for a short period *AB*, after which it returns to the previous location. The lower event shows a process of movement over a period of time. Let us use the sentences 44 and 45 as examples of the process *CD* and the state *AB*, respectively.

- (44) Mary moved from an apartment on the top floor to one on the floor beneath.
 (45) Sarah, Mary's roommate, dropped down to tidy up the new apartment for an hour during the move.

The projection of these sequences of events onto the time axis is shown at the bottom of the figure. The punctate events *A* and *B*, the beginning and end of the dropping down, are marked as points on the time axis. These points can be located in three ways. First, they can be placed in isolation independently of any other representation, as might occur at the beginning of a story. Second, they can be related to the present time of the speaker/listener or, third, to some other previously identified time. In these latter instances, the location vector is drawn with the tail at the reference point and the head at the located time, that is, from right to left (with a negative magnitude) if the event occurred prior to the reference point, and from left to right (with a positive magnitude) if it occurred later than the reference point.

The events themselves are states (dropping down) or processes (Mary's move) and are represented as vectors that must move from left to right (no time reversal). The

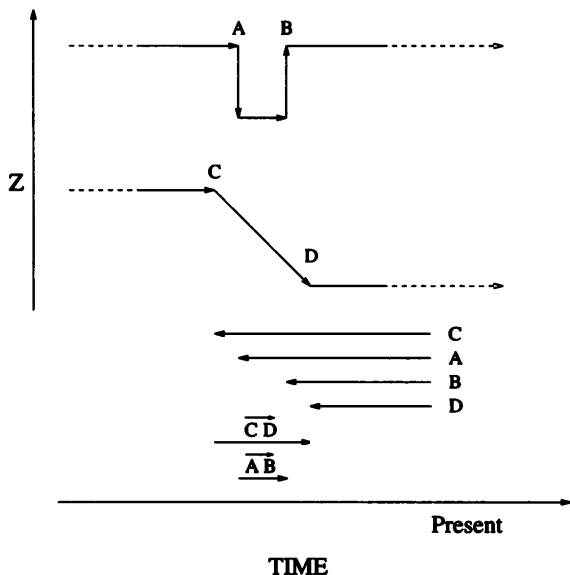


Figure 7.11

Temporal prepositions as relationships in a fourth dimension. An event such as “Sarah dropped down” is represented by a physical movement on the Z -axis that begins at time A , ends at time B , and is represented by vector $\overrightarrow{AB}$ on the time axis. A process such as “Mary moved” has a similar representation on the time axis. The representation assumes that the events occurred in the past, but other 0 reference points could have been adopted.

three events of the top sequence (the dropping down and the presuppositions of being in and returning to the upstairs apartment, are represented on the T -axis by vectors $\overrightarrow{AB}$, $-\overrightarrow{TA}$, and $+\overrightarrow{BT}$, respectively. The tail of the second and head of the third are left indeterminate. Here I am assuming that all events have some projection in the time domain, but that this can be ignored, for example, when the length of the event vector is short in comparison to the length of the location vectors.

The process of moving represented by vector $\overrightarrow{CD}$ has a similar representation on the time line, the difference between a state and a process residing in changes in the nontime dimensions.

Referring to figure 7.11, I suggest that the meaning of the temporal prepositions is as follows. The usual representation of a process such as CD is

(46) The move took place *from* noon *to* 2 P.M.

The event CD has a time vector which begins at T_C (noon) and ends at T_D (2 P.M.). $T(\overrightarrow{CD}) = T_D - T_C$, where D and C are the respective location vectors.

(47) The move lasted *for* two hours

sets the length of vector $\overrightarrow{CD}$.

(48) Sarah dropped down *after* Mary began moving, *before* Mary finished moving,
by the end of the move

sets $T_A > T_C$, $T_A < T_D$, $T_A \leq T_D$.

(49) Sarah visited the new apartment during the move

sets $T_C < T_A \leq T_B < T_D$.

Since and *until* are two temporal prepositions that do not have spatial homologues. *Until* specifies the time at which a state or process ended, whereas *since* specifies the time at which it began. *Since* has the additional restriction that the temporal reference point acting as the source of the location vectors for the event in question must be later than the event, that is, the location vectors must have negative magnitudes. This is to account for the acceptability of (50) but not (51).

(50) Mary has (had) been moving since noon.

(51) ?By 2 P.M. tomorrow Mary will have been moving since noon.

The simple temporatives *at*, *by*, *in* locate an entity by reference to a single place on the fourth axis. *At* operates in the same way as it does in the spatial domain by substituting the place of the referent for the entity. *By* fixes the location of the reference point as the maximum of a set of possible places. *In* suggests that there is an extent of time that is considered as the referent and that contains the entity. *On* is somewhat more difficult; it would seem to introduce the notion of a second temporal dimension, a vertical dimension that would place the entity at a location above or alongside of the time point. *About* and *around* also suggest a second dimension. In general, however, the temporal use of *on* seems to be restricted to the days of the week (*on Friday*) and to dates (*on the first of April*) and is not used in any general sense. It may therefore be an idiosyncratic use to distinguish these from the pointlike hours of the day (*at 5 o'clock*) on the one hand and the extended months of the year (*in May*).

Other simple temporal prepositions give the location of the event or duration of the condition by reference to a time marker that fixes the beginning or end of the time vector. Whereas *by* and *to* set the head of the temporal vector at the reference place, *before* sets it to the first place to the left of that place. In neither case is the origin or tail of the vector specified. This is given as the object of *from*. *During* specifies both the head and tail of the temporal vector. An event that occurs *after* one time and *before* another occurs *during* the interval. The length of the vector is given by the preposition *for*.

As with the spatial prepositions, some of the temporal prepositions require two reference points for their meaning. These include *between*, *beyond*, *past*, *since*, and *until*. *Between* two times locates the start of the event later than the first time and the end of the event before the second. The referent in *beyond* denotes the value that the head of an event vector exceeds. Because the time axis is basically a unidimensional one, the important distinction between *past* and *beyond* in the location of the entity in the orthogonal axis of the spatial domain does not apply, and the two prepositions appear to be interchangeable in most expressions.

7.8 Translation and Transformation Vectors

Once one has a temporal framework, it is possible to incorporate the notion of changes into the semantic map. These take two forms: changes in location and changes in state. The second of these relates to the circumstantial mode of Gruber (1976) and Jackendoff (1976). Both changes are represented by vectors. Changes in location of an object are represented by a vector whose tail originates at the object in a place at a particular time and ends at the same object in a different place at a subsequent time. Changes in state are represented by a vector drawn from an object at time t to itself in the same location at time $t + 1$. The change is encoded in the attributes of the object. In both types of change, the origin or tail of the vector is the object of the locative preposition *from*, and the head or terminus of the vector is the location identified by the locative preposition *to*.

(52) The icicle fell from the roof to the garden.

The representation of this is shown in figure 7.12. It consists of a four-dimensional structure with time as the fourth dimension. In the figure, I have shown two spatial dimensions and one temporal dimension. The left side of the representation shows the unstated presupposition that the icicle was on the roof for some unstated time prior to the event of the sentence. As Nadel and I noted (O'Keefe and Nadel 1978), the relationship between an object and its location is read as

- (53) a. The icicle was on the roof (before time t).
 b. The roof had an icicle on it.

The middle of the figure shows the translation vector that represents the event of the sentence, and the right hand the postsupposition that the icicle continues in the garden for some duration after the event.

- (53) c. The icicle was in the garden (after time t).

The representation of the second type of change, the circumstantial change, also involves a vector, this time a transformation vector, where there is no change in the

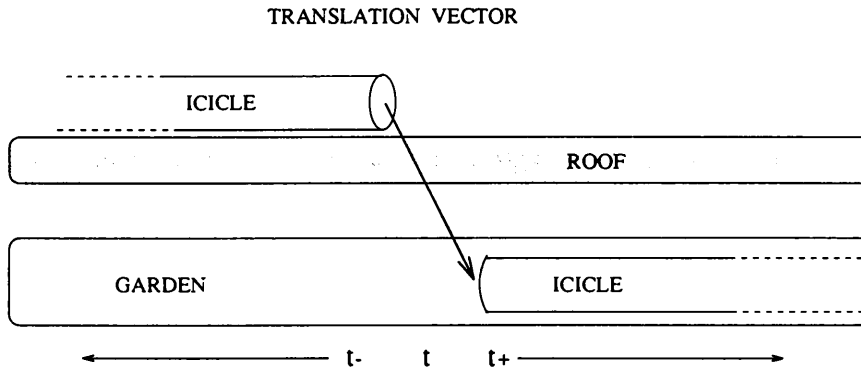


Figure 7.12

Change in location of an object in the semantic map at a particular time t is represented by a translation vector. In addition to the time axis, one spatial axis (Z) is shown. The four-dimensional object, labeled "icicle," is shown on the place labeled "roof" at all times prior to t ($t-$) and in the place labeled "garden" at all times after t ($t+$). The vertical movement between the two places at t is represented by a translation vector drawn between the two places.

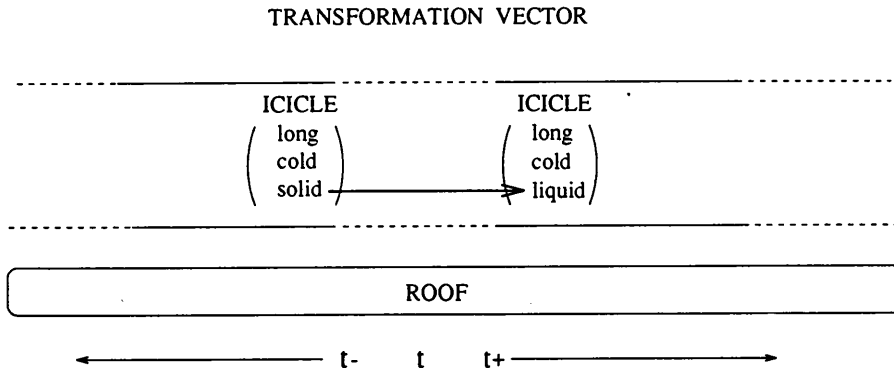
location of the object, but a change in one of the attributes assigned to the object. Objects are formed from the collection of inputs that occupy the same location in the map and that translocate as a bundle (see O'Keefe 1994 for a discussion of this Kantian notion of the relationship between objects and spatial frameworks). Thus each object has associated with it a list of attributes. In a circumstantial change, a vector represents the change in one of these attributes at a time t . Figure 7.13 shows the map representation of sentence 54.

- (54) The icicle melted (= changed from hard to soft at time t , or changed from solid to liquid).

7.9 Metaphorical Uses of Vertical Prepositions

In the following sections, I shall explore the metaphorical uses of the vertical stative prepositions. I hope to show that they apply to two restricted domains: influence (including social influence) and social status. In the course of this discussion I shall ask some of the same questions about these metaphorical uses as I did for their physical uses: what are the properties of the spaces represented, what type of scale is used, and so on?

Section 7.9.1 will explore the metaphorical meanings of *below* and *beneath* as used within the restricted domain of social status. Section 7.9.2 will deal with *under*, whose

**Figure 7.13**

Changes in state of an object in the semantic map are represented by a transformation vector whose tail originates in the old property before t and whose head ends in the new property after t .

semantics is more complex, but appears to be restricted to the domain of influence or control. In general, the representation of ideas such as causation, force and influence in the semantic map presents a problem. The basic mapping system appears to be a kinematic one which does not represent force relations. The closest one comes in the physical domain is the implicit notions that an entity which is vertical to another and in contact with it might exert a gravitational force on it or that an entity inside another might be confined by it. This might explain why the prepositions that convey these relationships, such as *under* and *in*, are used to represent influence in the metaphorical domain.

7.9.1 *Below, Beneath, and Down*

Contrast the following legitimate and illegitimate metaphorical uses of *below* and *under*:

- (55) She was acting below (beneath) her station.
- (56) She was acting under his orders.
- (57) *She was acting under her station.
- (58) *She was acting below his orders.

When looking at *below* and *beneath* within the domain of social status, the first thing to notice is that people are ranked or ordered in terms of their social status on a vertical scale. One person has a higher or lower status than another, and that status would appear to be transitive: if A has a higher status than B and B than C , it follows

that *A* has a higher status than *C*. I am ignoring here the possibility that status might be context-specific because I do not think this is reflected in the semantics of the prepositions. Now within the vertical scale of status, one can have a disparity between the value assigned to an individual act and the longer-term status. This gives rise to sentences such as

(59) John acted in a manner beneath him.

(60) That remark was below you.

A sequence of such actions, however, will result in a status change, so that

(61) Until recently that remark would have been beneath you, but now it is quite in character.

The antonym of *below/beneath* in this context is *above*, although it is not much used.

(62) Sally was getting above her station,

but not

(63) *That remark was above you.

The use of *below* and *beneath* in this sense is restricted to reflexive status, and thus one could not say

(64) John acted in a way beneath Sally (Sally's station).

Thus the best model (see figure 7.14) seems to be one in which each status token is confined to a vertical line in the status dimension, but these are free to vary in the other dimensions such that John can move so as to be beneath himself but not beneath Sally, but at the same time can be compared in the vertical dimension with Sally, "His status is below hers." Finally, note that there is no vantage point (egocentric point) from which these judgments are made or which would change them (i.e., the speaker's status is not relevant).

The stative preposition *down* seems to have almost no use in the nonphysical sense. The closest one comes are colloquial forms of verbal ranking such as

(65) Put him down.

7.9.2 Under

Under has perhaps the most interesting use of the vertical prepositions in the metaphorical domain. It seems to be confined to the domain of influence or control. In *The Hippocampus as a Cognitive Map* (1978), Nadel and I suggested that one of the metaphorical domains would be that of influence. Here I will pursue the idea that this relationship is represented by an additional "vertical" dimension (figure 7.15).

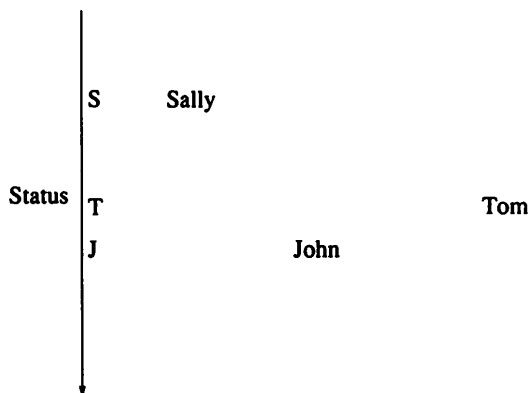


Figure 7.14

Social status is conveyed by the metaphorical use of *below* and is represented as a location on the metaphorical vertical axis.

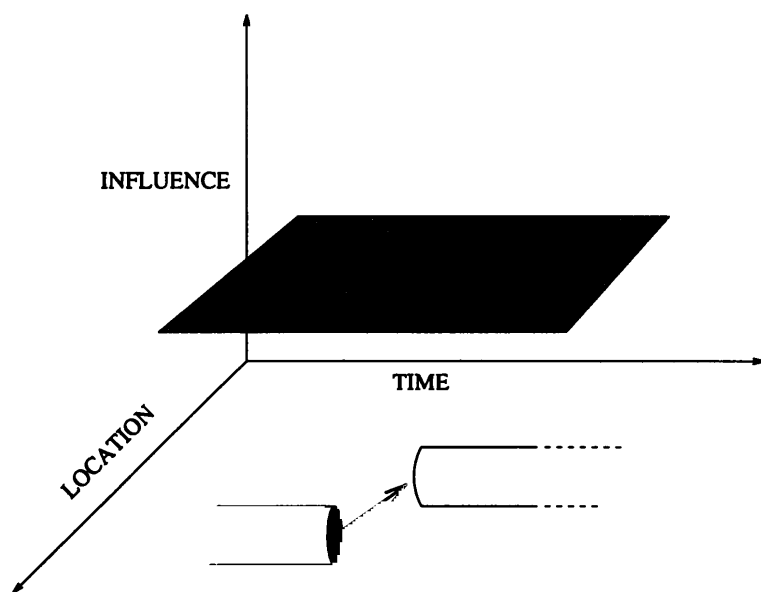


Figure 7.15

Influence of one entity, usually an agent, over another entity or an event is represented by a superior location of the first on the vertical influence axis.

There are two homophones (*under*¹ and *under*²), which follow different rules and which are derived from the two meanings in the physical domain:

(66) under a widening sky

(67) under the table

Compare

(68) Under the aegis of

with (66), and

(69) a. under John's influence

b. under Sally's control

with (67).

The first meaning of *under* cannot take a comparative form.

(70) *More under the aegis of the King

is not transitive, and has no antonym.

(71) *He was above, outside of, free from the aegis of the King.

In contrast, the second meaning follows all the rules for the second physical *under*².

(72) More under her influence every day.

But surprisingly the antonym of this *under*² is not *over* in many examples, but varies with the direct object.

(73) She was free from stress.

(74) The car was out of control.

(75) He was out from under the control of his boss.

As the last examples suggest, the referent in this meaning of *under* has an extent in the vertical dimension, and to be more *under* a cloud than *X* has the same sense of a greater overlap in the projection onto (one or more) horizontal dimension as in the physical meaning. To increase or decrease this influence requires a movement or expansion of one or the other entity in the horizontal plane, and this may require force in that direction.

(76) John was more under control than Sam.

(77) John was more under the influence of Mary than Sam.

(78) She slowly extricated Sam from Harry's influence.

There are two types of relationships that conform to this pattern, control and influence, and these vary in the amount of freedom left to the referent object.

(79) Jane increased her influence over Harry until she had complete control.

The antonym of *under*² is *over*.

(80) Jane's influence over John

(81) Jane lords it over John.

(82) Jane holds sway over John.

(83) a. *The King's aegis was over John.

b. *The King held his aegis over John.

Notice that the *under* relationship is not transitive. John can be under Jane's influence and Jane can be under Joe's, but John is not necessarily under Joe's.

Finally, I wish to remark briefly on the fact that there appear to be two nonphysical vertical dimensions that are orthogonal to each other and to the physical vertical one. On the face of it, it does not seem obvious how they could be reduced to a single dimension because one wishes to preserve the possibility of the following types of relationship.

(84) Jack felt it necessary to act below his station in order to maintain control over Jane.

Perhaps here one should consider the possibility that overlapping representations symbolize a control or influence relationship while nonoverlapping ones stand for a status one in the same 2-D space. If this were the case, what would the Z-axis be? Perhaps the higher the status, the more possibility for control?

Finally, in terms of the scaling of the metaphorical vertical prepositions, they appear to have the same interval scale as their physical counterparts. Thus one can say:

(85) Jane is as far below Mary in status as John is above

(86) John is less under Sam's control than Jim is

and it will be easier to extricate John.

Note that, unlike the three dimensions of physical space, we cannot compare the Z-axis and the non-Z-axis directly.

(87) *John is more under Sam's control than Sam acted below himself.

Now we come to the most difficult part of the theory: the relationship between control and causation. Causation, on this reading, would be the occurrence of an event underneath the control of an agent's influence.

(88) The book went to the library.

(89) John caused the book to go to the library.

7.10 Causal Relations in the Semantic Map

Our analysis of the metaphorical use of *below* and *under* has led to the suggestion that the causal influence of one item in the map over another might be represented by relationships in the fifth dimension. If the influence of an agent over another agent or object can be represented by the location of the first above the second, then it might be possible to represent the influence of an agent over an event such as that portrayed in (90) or (91) by an action or movement along the influence dimension. Consider the closely related sentences:

(90) Mary made (caused) the icicle fall from the roof to the garden.

(91) Mary let (did not prevent) the icicle fall from the roof to the garden.

According to the present analysis, these are five-dimensional sentences, which differ in the control exerted by the agent over the event. As we saw in the previous section, influence is represented by an *under*² relationship between the influencer and the influenced. The lateral overhang between the two represents the amount of control exerted, and the distance between them on the vertical dimension, the amount of influence exerted. On the simplest reading, causation is represented as a pulsatile increase in influence coincident with the physical spatial event. Figure 7.16 shows this as a momentary increase in Mary's influence to symbolize an active role in the event, while figure 7.17 shows a continuing influence but no change to symbolize a passive role in the event. The sentence

(92) Mary did not cause *X*

is ambiguous, with two possible underlying structures: one in which Mary has influence but the event did not happen; and the other in which the event did happen but the causal influence was not exerted by Mary. This type of representation can also capture some of the more subtle features of causal influence, because it can show how influence can selectively act on parts of the event as well as on the whole. For example, the sentence

(93) Mary made John throw down the icicle

means that both Mary and John had agentive roles in the event, but that Mary's was the superior one. This can be represented by placing Mary at a higher level than John in influence space and showing momentary synchronous changes in their locations at the time of the event. The complex influence relationship also allows for the following sentences:

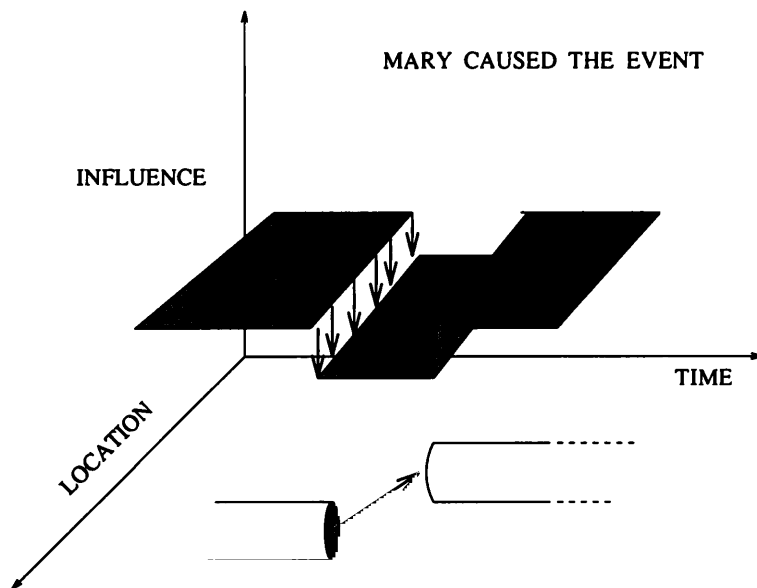


Figure 7.16

Causal influence is represented by a pulsatile change in the vertical influence dimension at the same time t as the physical event.

(94) Mary allowed John to throw down the icicle.

(95) Mary allowed John to drop the icicle.

(96) Mary made John drop the icicle.

It also permits one to represent relative degrees of influence over an event in a manner analogous to that over agents or objects, as in

(97) Mary had more influence over the course of events than John,

or the idea that an event of continuing duration can have variable amounts of control at different times,

(98) Mary took over control of the event from John on Monday.

7.11 Syntactic Structures in Vector Grammar

Thus far, I have said very little about the way that surface sentences and paragraphs could be generated from the static semantic map. Nadel and I (O'Keefe and Nadel

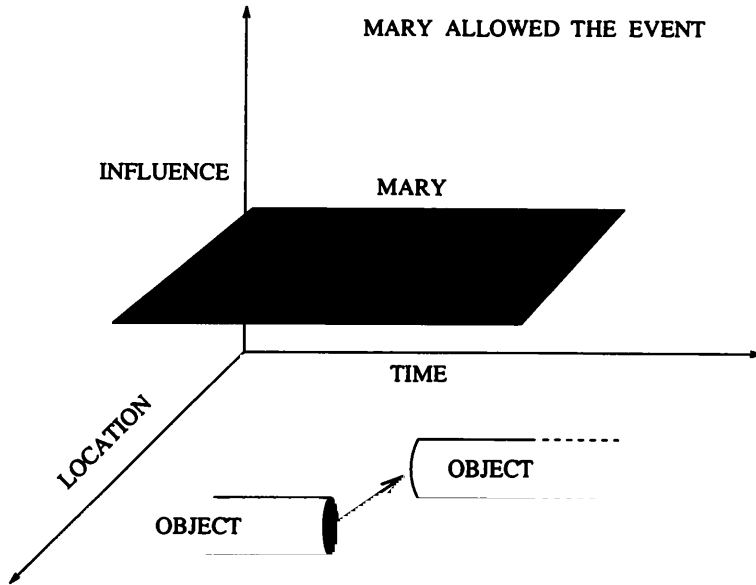


Figure 7.17

Permissive influence is represented by the absence of change in the vertical influence dimension of the influencer during the event.

1978) likened this operation to the way in which an infinite number of routes between two places could be read off a map. Recall that the cognitive map system in animals includes a mechanism for reading information from the map as well as for writing information into the map. In particular, we postulated a system that extracts the distance and direction from the current location to the desired destination. This information can be sent to the motor programming circuits of the brain to generate spatial behaviors. The corresponding system in the semantic map would comprise the syntactic rules of the grammar. The syntactic rules operate on both the categories of the deep structures and the direction and order in which they are read. For example, reading the relationship between an influencer and the object or event influenced determines whether the active or passive voice will be used. In an important sense there are no transformation rules for reordering the elements of sentences because these are read directly from the deep structure. Given a particular semantic map, a large number of narrative strings can be generated depending on the point of entry and the subsequent route through the map. Economy of expression is analogous to the optimal solution to the traveling salesman problem.

Acknowledgments

I would like to thank Miss Maureen Cartwright for her extensive help and substantive contributions to this chapter. Neil Burgess made comments on an earlier version. The experimental research that forms the basis for the cognitive map model was supported by the Medical Research Council of Britain.

Notes

1. I have deliberately chosen the term *entities* to refer to the relationships because I do not wish to limit my discussion to objects, but wish to include places, features, and so on.
2. In what follows, I have relied heavily on the classic discussion by Torgerson (1958).
3. I am assuming the geomagnetic sense is absent or so weak in humans that it is not available for spatial coding. As far as I am aware, there is no evidence for it in the prepositional system of any language.

References

- Anderson, J. M. (1971). *The grammar of case: Towards a localistic theory*. Cambridge: Cambridge University Press.
- Bennett, D. C. (1975). *Spatial and temporal uses of English prepositions: An essay in stratificational semantics*. London: Longmans.
- Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.
- Comrie, B. (1985). *Tense*. Cambridge: Cambridge University Press.
- Cook, W. A. (1989). *Case grammar theory*. Washington, DC: Georgetown University Press.
- Frisk, V., and Milner, B. (1990). The role of the left hippocampal region in the acquisition and retention of story content. *Neuropsychologia*, 28, 349–359.
- Gruber, J. (1965). Studies in lexical relations. Ph.D. diss., Massachusetts Institute of Technology.
- Gruber, J. (1976). *Lexical structures in syntax and semantics*. Amsterdam: North Holland.
- Jackendoff, R. (1976). Toward an explanatory semantic representation. *Linguistic Inquiry*, 7, 89–150.
- Landau, B., and Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–265.
- O’Keefe, J. (1988). Computations the hippocampus might perform. In L. Nadel, L. A. Cooper, P. Culicover, and R. M. Harnish (Eds.), *Neural connections, mental computation*, 225–284. Cambridge, MA: MIT Press.
- O’Keefe, J. (1990). A computational theory of the hippocampal cognitive map. In O. P. Ottersen and J. Storm-Mathisen (Eds.), *Understanding the brain through the hippocampus*, 287–300. Progress in Brain Research, vol. 83. Amsterdam: Elsevier.

- O'Keefe, J. (1991). The hippocampal cognitive map and navigational strategies. In J. Paillard (Ed.), *Brain and space*, 273–295. Oxford: Oxford University Press.
- O'Keefe, J. (1994). Cognitive maps, time and causality. *Proceedings of the British Academy*, 83, 35–45.
- O'Keefe, J., and Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11–21.
- Smith, M. L., and Milner, B. (1981). The role of the right hippocampus in the recall of spatial location. *Neuropsychologia*, 19, 781–793.
- Smith, M. L., and Milner, B. (1989). Right hippocampal impairment in the recall of spatial location: Encoding deficit or rapid forgetting? *Neuropsychologia*, 27, 71–81.
- Taube, J. S., Muller, R. U., and Ranck, J. B. (1990). Head direction cells recorded from the postsubiculum in freely moving rats. 1. Description and quantitative analysis. *Journal of Neuroscience*, 10, 420–435.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55, 189–208.
- Torgerson, W. (1958). *Theory and methods of scaling*. New York: Wiley.

Chapter 8

Multiple Geometric Representations of Objects in Languages and Language Learners

Barbara Landau

Central to our understanding of how young children learn to talk about space is the question of how they represent objects. Linguistically encoded spatial relationships most often represent relationships between two objects, the one that is being located (the “figure” object, in Talmy’s 1983 terms) and one that serves as the reference object (Talmy’s “ground” object). Crucially, learning the language of even the plainest spatial preposition—say, *in* or *on*—requires that the child come to represent objects in terms of geometrical descriptions that are quite abstract and quite distinct from each other.

Consider the still life arrangement in figure 8.1. If we were to describe this scene, we might say any of the following:

- (1) a. There is a bowl.
- b. The bowl has flowers painted *on* it.
- c. It has some fruit *in* it.
- d. There is a cup *in front of* the bowl and a vase *next to* it.

What are the geometric representations underlying these different spatial descriptions? In calling each object by its name—“bowl,” “cup,” “vase”—we distinguish among three containers that have rather different shapes (and functions), suggesting that we are recruiting relatively detailed descriptions of the objects’ shapes. Such descriptions could be captured within a volumetric framework such as that described by modern componential theories in which object parts and their spatial relationships are represented (e.g., Binford 1971; Lowe 1985; Marr 1982; Biederman 1987). This is one kind of representation. However, in describing the spatial relationships between or among objects, we seem to recruit representations of a quite different sort. When we say, “The bowl has some fruit *in* it,” we recruit a relatively global representation of the object’s shape, in which its status as a volume—a “container”—is critical, but no further details are. When we say, “The bowl has flowers painted *on* it,” we seem to recruit a different representation, one in which the surface of the object is relevant,



Figure 8.1

Each object in this scene can be represented as a number of different geometric types.

but nothing else is. When we say, “There is a cup *in front of* the bowl,” we recruit yet a different representation—one in which the principal axes of the bowl are relevant. The region “in front of” the bowl spreads out from one of its half axes (and whether these axes are object-centered or environment-centered depends on a variety of factors; see Levelt, chapter 3, this volume).

These few examples show that learning the meanings of spatial terms requires learning the mapping between spatial terms and their corresponding regions—where the relevant regions are defined with reference to geometrically idealized or “schematized” representations of objects (Talmy 1983). Therefore a crucial part of learning the mappings is properly representing objects in terms of their distinct relevant geometrical descriptions—for example, representing an object as a volume in the case of the term *in*, as a surface in the case of the term *on*, and as a set of axes in the case of *in front of* and *behind*. In fact, learners must possess these object representations before learning the correct mapping; if the objects cannot be represented properly, the terms cannot be learned.

The brief analysis just given suggests that there is a variety of object representations underlying spatial language—the language of objects and places. Objects must be represented at a fairly detailed level of shape, they must also be represented at a skeletal level—simply as a set of axes—and they must be represented at a level that is quite coarse (as volumes, surfaces, or simply “blobs”). That we can talk easily about bowls, cups, and vases, and the kinds of spatial relationships into which they enter suggests that we possess a cognitive system that allows for flexible “schematizing” of objects (cf. Talmy 1983). Central to the present discussion, the early acquisition of spatial terms among children suggests that these multiple representations of objects may exist early in life and may be used to guide the learning of spatial language.

The idea that very young children might possess such rich and flexible representations of objects is at odds with traditional theories of spatial development, which posit substantial changes in spatial knowledge over the first six years of life. According to Piaget’s theory, the first two years of life are devoted to constructing a system of knowledge that can support the general permanence of objects in the face of continually changing perceptual and motor interactions between the infant and objects in the world (Piaget and Inhelder 1948; Piaget 1954). Once such knowledge has developed, the child is said to possess true “representations” of objects—representations that go beyond perception. However, the child’s knowledge of space is still incomplete. Piaget hypothesized that from around age two, the development of spatial knowledge would proceed through a sequence of stages in which children would first represent only topological properties of space—highly general properties such as connectedness and openness versus closedness. Although even infants might be capable of discriminating between objects having different metric properties (e.g., a square vs. a triangle), Piaget proposed that the child possessing a topological representation of space would only be capable of *representing* the difference between a line and a closed loop, but not the difference between a square and a triangle. For Piaget, such impoverished representations were evidenced, for example, by the fact that two- and three-year-olds draw a variety of geometric figures as simple open versus closed figures, possessing no specific metric properties. Later, projective properties would develop, such as the straight line, or a relationship specified by location along such a line; metric properties such as angles and distances would come to be represented even later, sometime during later childhood.

Extending Piaget’s view to the realm of spatial relationships, a topological representation could support understanding of a contact or attachment relationship between two objects, but could not support the representation of a distinction between contact with a vertical *versus* a horizontal surface. Similarly, relationships such as

that encoded by the terms *in front of* or *behind* would require at least projective representations of space, emerging during late childhood.

While topological properties might seem congenial to the analysis of spatial locational terms (Talmy 1983), a variety of evidence suggests that a topological representation of objects and relationships is too weak to characterize young children's knowledge. For example, the child who was limited to representing objects topologically would be incapable of using precise object shape for naming bowls or cups, would be unable to represent objects in terms of their axes in order to learn such basic spatial terms as *in front of* or *behind*, and would be unable to learn the distinction between German *auf* and *an* (attachment to vertical vs. horizontal surface).

In this chapter I review evidence showing that such nontopological representations are indeed accessible to young children learning the language of space. Further, it appears that young children possess multiple representations of objects that can support acquisition of different parts of the spatial lexicon. I focus on three different kinds of representations: (1) "coarse," bloblike representations of objects, which eliminate all details of shape information; (2) "axial" representations, which eliminate all details of shape except the relative length and orientations of the three principal axes; and (3) "fine-grained" representations, which preserve a considerable degree of shape detail. The evidence I will describe is primarily based on studies of young children learning English, although evidence from children learning other languages is consistent. The evidence indicates that both coarse and axial representations of objects can be elicited by engaging children's knowledge of known and novel spatial terms (in English, spatial prepositions). The axial representations in particular illustrate that young children naturally represent objects in terms of skeletal descriptions in which the object's principal axes are the major components of its "shape." The studies also indicate that, although the representations underlying spatial terms appear to "strip away" details of shape (as suggested by Talmy 1983), fine-grained, shape-based representations of objects are also accessible to young children. These representations tend to emerge when children are engaged in learning object names.

In the following sections, I first outline how objects are represented when they are encoded by noun phrase arguments of spatial prepositions in English (e.g., the "cat" or "mat" in the sentence "The cat is on the mat"), and how these object descriptions differ from those relevant to similar spatial terms in other languages. Particular emphasis will be placed on comparing English to other languages whose locational terms appear to incorporate much more shape information than those in English. Next I present evidence showing that young children learning English show strong biases to ignore fine-grained shape when learning novel spatial terms or when interpreting known English spatial terms, but that they show equally strong biases

to *attend* to fine-grained shape when learning novel object names. This empirical evidence will raise a number of questions, which I will outline, including issues of possible structures and mechanisms underlying this gross difference in object representation.

8.1 Ways of Representing Objects in Places

How are objects represented when they serve as figure or reference object in a locational expression? In English, spatial locations—places—are encoded canonically by prepositional phrases headed by spatial prepositions. In a simple sentence such as “The flowers are on the vase,” the “flowers” play the role of figure, the “vase” is the reference object, and the spatial preposition “on” maps a region of space onto the reference object. Although the upper surface of an object may be the preferred reading for *on* in English, the relevant region is actually any portion of the surface of the vase: The sentence will be true regardless of where in particular the flowers are located, as long as they are somewhere contiguous with the surface of the vase.¹

Note that spatial prepositions do not exhaust the possibilities for talking about spatial location, even in English, where places are canonically encoded this way. For example, there exist verbs that describe posture, a kind of static spatial relationship: *stand* represents the vertical posture of an object; *recline* represents horizontal posture; *crouch* and *kneel* other postures; etc. However, because spatial prepositions in English encode location only, they provide a well-defined domain within which to intensively examine the kinds of spatial relationships that languages encode. With that knowledge, one can compare these meanings to those encoded by other spatial terms in English (e.g., nouns such as *top* and *bottom*; adjectives such as *long* and *wide*; verbs such as *stand* and *recline*) and to locational terms in other languages.²

8.1.1 English Spatial Prepositions

The spatial prepositions in English form a relatively small closed class numbering somewhere above eighty (not considering compounds such as *right next to*). A sample list is given in table 8.1. Most of these prepositions are two-place predicates, although there are some with a greater number of arguments, for example, *among*, *amidst*. Other languages contain as few as one general locational marker (e.g., *ta* in Tzeltal; Levinson 1992), and there is variability in the precise relationships that are encoded by spatial terms in other languages: Considering prepositions only, some languages collapse several English distinctions into broader categories (e.g., Spanish *en* covers English *in* and *on*), while others split a single English distinction into several finer categories (e.g., German *auf* and *an* cover English *on* but distinguish between vertical

Table 8.1
English Prepositions

at	in	on	off
out	by	from	to
up	down	over	under
with	within	without	near
nearby	into	onto	toward
through	throughout	underneath	along
across	among(st)	against	around
about	above	amid(st)	before
behind	below	beneath	between
beside	outside	inside	alongside
via	after	upon	
Compounds			
on top of	in between	to the right of	parallel to
in back of	to the left of	to the side	perpendicular to
in front of			
Temporal only			
during	since	until	ago
Intransitives			
here	outward	backward(s)	south
there	afterward(s)	away	east
upward	upstairs	apart	west
downward	downstairs	together	left
inward	sideways	north	right

and horizontal attachment, respectively; Korean *ahn* and *sok* cover English *in* but distinguish between “loose” and “deep” or “tight” containment, respectively).

Despite this variability, however, there appear to be universals in how figure and reference objects are geometrically schematized and in the kinds of spatial relationships that are encoded. These universals can be revealed by considering the geometric restrictions imposed by a spatial term on its arguments (see, for example, Miller and Johnson-Laird 1976; Talmy 1983; Jackendoff 1983; Herskovits 1986). As one example, the preposition *in* requires a reference object that can be construed as having an interior: If one object is *in* another, the latter must have some volume or area within which the object can be located. Phrases such as “in the bowl” or “in the house” are easily understood because bowls and houses are easily construed as volumes. However, the abstract nature of these geometric descriptions can be seen through other cases, in which the preposition will coerce one’s reading of the reference object. For

example, in a phrase such as “in the dot” or “in the mat” the dot or mat will be construed as a 2-D area or even a 3-D volume (e.g., “dirt *in* the mat”). Thus, although the term *in* seems to express straightforward “containment” (with the reference object some sort of “container”), we can use it equally well for “coffee in a cup” (where the reference object is a physical container), “birds in a tree” (a virtual volume), or “customers in a line” (a virtual line). Such semantically motivated restrictions appear comparable to restrictions imposed by verbs on their arguments. For example, the verb *to drink* requires an argument construable as a continuous quantity (centrally, a liquid), the verb *eat* requires an argument construable as an edible (hopefully, food), and so forth. Given coercion by the verb, we can interpret a sentence such as “John drank marbles,” where marbles are taken as a continuous stream (cf. **“John drank a marble”*).

This process of “schematizing” objects has been described by Talmy (1983) in his seminal work on the geometry of figure and reference object where he suggested strong universal constraints on the geometric properties relevant to the figure and reference object. Specifically, he proposed an asymmetry in the geometric descriptions of figure and reference object, with the figure often represented as a relatively shapeless blob, and the reference object represented more richly, often in terms of the object’s three principal orthogonal axes.

8.1.2 Geometry of the Figure Object

Taking examples from English, the prepositions listed in table 8.1 show very few constraints on the figure object. Terms such as *in*, *on*, *above*, *below*, and many others do not impose any special geometrical requirements on the figure object—any object of any shape, size, or type can play the role without violating the meanings of the majority of prepositions. There do exist, however, a few restrictions for certain terms. Terms such as *across* and *along* represent relationships of intersection and parallelism, respectively; and these relationships appear to require a figure and reference object that can be construed as a “linear” object.³ Thus sentences (2a, b) both are easily understood, whereas sentence (2c) is marginal because it is difficult to construe a ball as a “linear” object. Note, however, that sentence (2d) is completely natural; in this case, the ball’s path (as it bounces) becomes the figure.

- (2) a. A snake lay along the road.
 b. Trees stood along the road.
 c. ?A ball lay along the road.
 d. A ball bounced along the road.

One further distinction mentioned by Talmy is the figure object’s distribution in space: *through* is used for nondistributed objects, while *throughout* expresses

distribution of the object in the ground (compare “There were raisins throughout the pudding” to “?There were raisins through the pudding”).

Aside from these few distinctions, there do not appear to be any other requirements on the geometry of the figure object for spatial prepositions in English. Nor do I know of any in the spatial prepositions of other languages, although other languages have locational *verbs* that do impose shape restrictions on the figure object. For example, there is only one basic spatial preposition in Tzeltal (*ta*, a general relational marker), but information about an object’s axial structure (specifically, aspect ratio, or the ratio of height to width) can appear as part of different spatial predicates used for locating objects (see Brown 1993; Levinson 1992). Thus *waxal-ta* is predicated of objects whose opening is smaller than their height, *pachal-ta* of objects whose opening is larger than their height, *chepel-ta* of flexible bulging bags (Brown 1993). As another example, Atsugewi possesses a considerable number of figure object distinctions in locational verbs, including roots meaning “small, shiny, spherical object to move/be located,” “slimy, lumpish object to move/be located,” “limp, linear object suspended by one end to move/be located,” and “runny, icky material to move/be located” (Talmy 1985). English makes similar distinctions in certain verbs (e.g., *to rain*, *to spit*), although this particular pattern of conflation is not dominant in English, according to Talmy.

These examples—in which a greater amount of geometric information is incorporated into the figure object—are challenging because they raise the question of whether there are universal biases in the kinds of information typically incorporated into the figure object in locational expressions. At this point, it should be noted that the degree of shape information exhibited in, say, Tzeltal locational predicates, is greater than that shown by English prepositions. It remains to be determined, however, exactly how fine-grained these shape descriptors are, and what role they play in the overall system of spatial language.

8.1.3 Geometry of the Reference Object

Like the figure, the reference object tends to be represented fairly coarsely. For certain terms, it is represented as a shapeless point or blob (e.g., terms such as *near* or *at* do not require that any specific geometric information be preserved). For other terms, the reference object is represented as a volume (*in*, *inside*) or as a surface (*on*), and for still other terms, the number of reference objects is distinguished (*between* for two reference objects, *among* or *amid* for more than two). In other languages, the orientation of the ground is distinguished (German *auf* vs. *an*), the openness of the ground (Korean has two separate terms for English *through*), and direction toward or away from the speaker (German *her* vs. *hin*), among others.

Most critically, however, a number of spatial prepositions require that the reference object be construed in terms of its three principal axes. The vertical axis (*above/below*), and the two sets of horizontal axes (*right/left* or *beside*; *in front of/behind*). These axes are also engaged by certain spatial nouns and adjectives in English: *top/bottom*, *front/back*, and *side* express regions defined by reference to the axes, and *tall*, *long*, *thin*, and *wide* express size differences along different axes. The spatial nouns are marked not only for different axes, but also for different ends of the axes (*top/bottom*, *front/back*, *right/left*, with the viewpoint varying application of the latter being quite difficult to learn).

These spatial terms appear to be insensitive to reference system. For example, "The star is above the flagpole" can be used to describe a location with respect to an object-centered framework (the region near the top of the *flagpole*, regardless of its orientation) or an environment-centered framework (the region adjacent to the *gravitational* top). However, people do appear to have biases to interpret these terms with regard to different reference systems under different conditions (Levelt, chapter 3, this volume; Carlson-Radvansky and Irwin 1993). At least one language possesses different sets of terms to refer to the object-centered versus environment-centered application of these terms. The Tzeltal body-part system utilizes one set of terms to refer to object parts, and another to refer to (environmentally determined) regions adjacent to the object (Levinson 1992).

The axial representations as a whole appear to be the richest geometric representations required by English spatial prepositions; they also play a major role in the spatial terms of other languages. For example, the Tzeltal body-part system is massively dependent on the object axial system, which specifies an object's principal dimensions, the ends of which are often labeled with locational terms, such as "at the head of," "at the butt of," "at the nose of," and so on (Levinson 1992). English also has such expressions (e.g., "at the head of the table," the foot of the bed," "the arm of the chair"), but the Tzeltal system is richer in its range of locational terms. Each of these terms, however, depends on very much the same kind of analysis into principal object axes. Levinson suggests that the assignment of body-part terms depends on a strict object-centered algorithmic assignment that analyzes the object into its principal and secondary axes, and then decides on markedness using detailed shape information (e.g., for *top* vs. *bottom*). For example, a novel object might possess a clear principal axis for which "head of" and "foot of" would be relevant, but if one end of the axis has a distinct protrusion, then that would be marked "head," or perhaps "nose," consistent with its shape. The rough shape parameters required for such assignment provide a challenge to the generalization that ground objects are stripped of detailed shape elements, even though there is still quite a broad range of shape variation sufficient for assigning "nose of" to an object part.

The axial system thus appears to be critical to the representation of reference objects in English and in other languages. Interestingly, this system has also been posited to be developmentally complex, with children coming to represent projective geometric properties such as straight lines only during middle childhood (Piaget and Inhelder 1948; Piaget, Inhelder, and Szeminska 1960). Based on this proposal for the development of nonlinguistic (axial) representations, a number of investigators have proposed that the spatial prepositions recruiting axial representations may be relatively difficult to learn (see Johnston 1985 for review). I return to this issue in section 8.2.2.

8.1.4 Summary

The geometries of both figure and reference object are relatively coarse, incorporating distinctions such as volume, surface, number, and most critically, principal axes (of either the figure object, the reference object, or both). As Talmy (1983) suggested, there appears to be an asymmetry between the figure and reference object, with the figure incorporating relatively less geometric specification than the reference object. If we consider the degree of geometric specification to be a dimension, English appears to incorporate the least information in figure objects, disregarding almost all shape specification of the figure object. At the other end of the dimension, languages such as Tzeltal appear to include more shape information, for example, grouping together objects by the relative proportions of the object's principal dimensions (e.g., *pachal* vs. *waxal*). However, even Tzeltal incorporates relatively little shape information, when compared with the much richer information available to identify objects. As for the reference object, English again incorporates very little shape information; at the most, it engages an axial representation of the reference object in order to describe the relevant region. Other languages also recruit the axial representation, but apparently, not much more.

These geometric descriptions appear quite different from those which might be engaged during object naming. The basic vocabulary for object names in English includes proper nouns (e.g., Fred, Mother) and count nouns (a *dog*, a *tree*). To the extent that these terms are linked with schemes for object recognition, they would seem to require geometric representations that preserve much more fine-grained spatial information than the ones so far described.

How do young children appear to represent objects (both figure and reference) when learning spatial terms, and how do they represent the same objects when learning object names? Are young children flexible in their representations? Can they represent objects as coarse, as axial, as fine-grained? The following empirical evidence provides positive evidence for each of these types of representation in young learners.

8.2 Empirical Evidence for Different Kinds of Object Representation among Young Learners

In order to determine whether young learners possess the different kinds of object representation underlying figure and reference object, we have conducted a variety of studies examining children's treatment of objects when learning novel spatial prepositions and when comprehending familiar prepositions. These studies have shown that children can ignore shape information altogether and that they can treat objects in terms of their axial representations. In addition, we have conducted a separate line of investigation to determine how children treat objects when they are learning a novel name for the object, independent of its location. These studies have shown that relatively fine-grained shape information can be used to assign objects to named categories.

8.2.1 Coarse Representations: Schematizing the Figure Object

Recall that in English, the figure object is generally treated quite coarsely—either as a shapeless point or blob, or (for terms such as *along* and *across*) as a linear object, focusing on the object's principal axis. Recall also that other languages may incorporate somewhat more detailed shape information into the figure object. Spatial predicates in Tzeltal include terms that incorporate information about the figure's aspect ratio (height-to-width proportions), flexibility, and curvature, for example. Two questions arise. One is whether young children learning a novel English spatial preposition will tend to ignore shape entirely (or perhaps, attend only to the object's principal axis). If the answer to this question is positive, then one might wonder whether English-speaking children could readily learn to incorporate the somewhat more detailed (axial) information captured, for example, in Tzeltal spatial predicates.

8.2.1.1 Ignoring the Shape of the Figure Object Landau and Stecker (1990) posed the first question by modeling a novel spatial preposition for young English-speaking children and then asking to what new figure objects and locations children would generalize this term. Three-year-olds and adults were shown a novel object (the "standard") being placed on the top of a box in the front right-hand corner (the "standard" location: see figure 8.2). As the object was placed, subjects heard, "See this? This is *acorp* my box," using the novel term *acorp* in a syntactic and morphological context compatible with interpretation as a novel preposition. The entire display then was set aside, and subjects saw each of three different objects being placed in each of five different locations on and around a second box. One of the objects was identical to the standard, and the other two were different from it in shape only (see figure 8.2 for objects). Each time subjects viewed an object being placed on the second

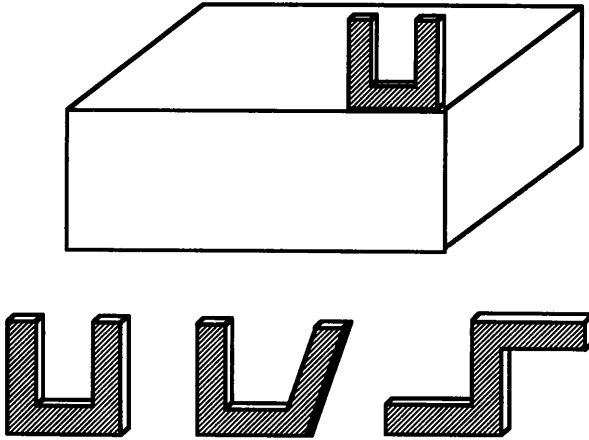


Figure 8.2

Objects and layout used by Landau and Stecker (1990). Children and adults were shown a novel object being placed on the top of a box, as shown. They heard either “See this? This is *acorp* my box” (novel preposition) or “See this? This is *a corp*” (novel count noun). Then they were shown the three different objects each being placed one at a time on and around the box in different locations. Each time, they were asked either “Is this *acorp* the box?” or “Is this a *corp*?” Subjects hearing the novel preposition ignored the object’s shape and generalized on the basis of its location. Subjects hearing the novel count noun ignored the object’s location and generalized on the basis of its shape.

box, they were asked, “Is this *acorp* your box?” The question was how children would generalize the meaning of the novel term. Would they generalize only to the standard in its standard location? Or would they generalize the term in a way consistent with the general pattern of English spatial prepositions—ignoring the particular shape of the standard object, and generalizing to a range of locations?

In this condition, both children and adults ignored the shape of the standard, accepting all three objects equally (summed over locations). However, they did attend to the object’s location. Having been told that the object was “*acorp* the box” (when placed on the top front right-hand corner of the box), children then generalized to all locations on the top of the box, rejecting all locations off the box. Adults showed a similar pattern, also rejecting all locations that were off the box, although they were somewhat more conservative than the children. Some of them confined their generalization to any object in the standard location only (top front right-hand corner).⁴

One might wonder whether the context of the experiment—in which objects are being placed in various locations—might itself predispose subjects to ignore object shape. We found evidence against this interpretation in a second experimental condi-

tion. In this condition, we followed the same procedures as above, with one critical exception. This time, as the standard was being placed on the box, we told subjects, "See this? This is a *corp*," using a the same phonological sequence as for the novel preposition (*acorp*), but placing the new word in a syntactic and morphological context appropriate to a count noun interpretation. Subjects then were shown the same test objects placed in the same test locations as in the first condition, but each time they observed a test object being placed in one of the locations, they were asked, "Is this a *corp*?" With this syntactic context serving as a mental pointer to a count noun reading, subjects now generalized only to the standard object, regardless of its location, rejecting both of the objects that were not identical to the standard. That is, while subjects hearing a novel preposition ("*acorp* the box") ignored shape and attended to location, subjects hearing a novel count noun ("a *corp*") ignored location, and attended to the object itself.

This pattern of findings shows that young children are capable of representing the figure object at a very coarse-grained level, completely ignoring shape. But they do not show that children are *incapable* of incorporating any elements of the figure object's shape when learning a new spatial term. Even in English, certain terms require attention to the figure object's principal axis—for example, *along* requires a roughly linear figure object as does *across*. And, as mentioned above, some Tzeltal terms appear to require even more shape information.

Thus one might ask, How readily will young children incorporate shape information into the figure object? We have approached this question through two sets of experiments. In both, we have modeled novel spatial prepositions using figure objects that possess a very clear principal axis. The question is whether such modeling might more strongly elicit at least an axial representation of the figure object.

8.2.1.2 Incorporating Axial Information into the Figure Object One experiment was exactly like the one just described, except that different objects and locations were used (Landau and Stecker 1990; see figure 8.3 for standard object and standard location). The standard object was now a 7-inch straight rod, and the test objects included a replica of the standard, a wavy rod of the same extent as the standard, and a $2'' \times 2'' \times 1''$ block. As subjects heard, "See this? This is *acorp* my box," the standard object was placed perpendicular to the box's main axis. Test locations included this same location as well as one slightly to the left of it, one parallel to the box's principal axis, and one diagonal to it.

The results of this experiment again showed that subjects tended to ignore shape and generalize primarily on the basis of the object's demonstrated location. In fact, many of the three-year-olds tested behaved just as they had in the first experiment, ignoring object shape and generalizing solely on the basis of location. However,

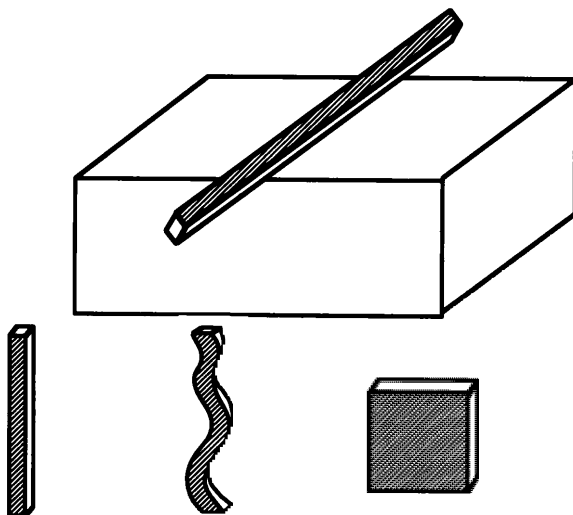


Figure 8.3

Objects and layout used by Landau and Stecker (1990) in a second study using the same method as described in figure 8.2. Subjects hearing the novel preposition ignored the object's detailed shape, and generalized on the basis of its location *and* its principal axis. Subjects hearing the novel count noun generalized on the basis of the object's exact shape.

some three-year-olds and most five-year-olds and adults accepted both the standard and the wavy object while rejecting the block. That is, they showed some attention to an abstract component of shape, accepting objects that were sufficiently long to intersect the box (when placed perpendicular to its main axis). In doing so, these subjects treated the two objects as similar with respect to their principal axis, whereas they disregarded the details of their very different shapes. These subjects also tended to generalize to the two locations in which the test object was at perpendicular intersection with the box; the horizontal and diagonal locations were considerably less favored (see note 4).

Thus, when we modeled with a standard object possessing a more salient principal axis, younger subjects (three-year-olds) still tended to completely ignore detailed shape, although some did attend to the axis. Older children (five-year-olds) and adults tended to attend to one skeletal component of shape—the principal axis. All subjects in this preposition condition also attended to location. This contrasts markedly with the pattern shown by subjects in a second condition of this experiment. These subjects were shown the same objects and locations, but heard the novel term in the count noun context, that is, “See this? This is a corp.” When asked, “Is this a corp?”

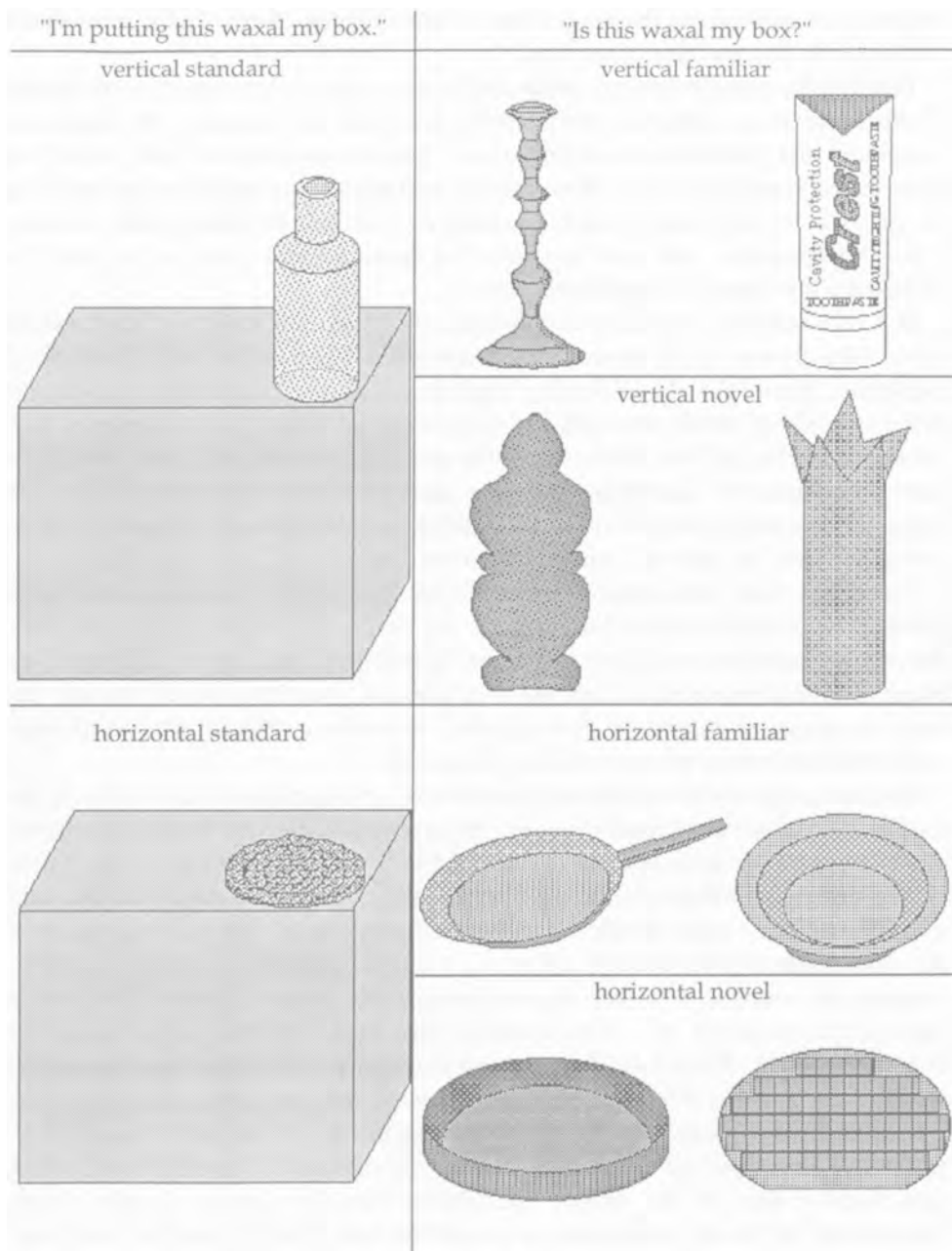
subjects now generalized the novel count noun to objects of exactly the same shape as the standard, regardless of location.

Thus the dissociation between shape and location that we had found in the first set of experiments was replicated with entirely new objects and locations. This illustrates once more that children's responses were not forced by salience (or lack thereof) of either object shape or location. Both children and adults were capable of generalizing on the basis of the object shown, ignoring its location. However, when learning a novel preposition, they tend to ignore the figure object's shape, or, at best, to schematize it in terms of its principal axis.

In a relatively new approach to this issue, we have been modeling novel spatial terms using figure objects whose shape properties are represented in Tzeltal spatial predicates. Figure 8.4 shows displays appropriate to the two terms *waxal-ta* and *lechel-ta*, each of which describes the location of an object. The locative *ta* is a relational marker and the predicates *waxal* and *lechel* each is used when locating a particular geometric figure type. *Waxal* is used for vertically oriented objects, for example, a tall oblong-shaped container or solid object canonically "standing"; *lechel* is used for wide flat objects "lying flat" (Brown 1993).

Given that these terms are found in a natural language, the conflation of specific geometry with location must be learnable. All children learning Tzeltal must learn the range of application of these two terms, as well as quite a number of others that encode different geometric distinctions. Our question, therefore, was not whether the terms are learnable, but rather, how difficult it would be for English speakers to infer such meanings from a relevant modeling situation.

In order to answer this question, we conducted an experiment quite similar to the studies of novel spatial prepositions described above (Landau and Hsieh in progress). We introduced the experiment by telling subjects that we were interested in how people speaking a different language, Tzeltal, might talk about locating objects, and that we would use some words that Tzeltal speakers might use. We then modeled two different locational situations. For one group of three-year-olds and adults, we modeled the meaning of *waxal*. As we placed a tall, oblong-shaped bottle on the top right hand corner of a box, we said, "See this? I'm putting this *waxal* my box" (see top left, figure 8.4). For a second group of three-year-olds and adults, we modeled the meaning of *lechel*. As we placed a wide, flat disk in the same location on a box, we told subjects, "See this? I'm putting this *lechel* my box" (bottom left, figure 8.4). The object on its box was then moved aside, and a second, identical box was placed in front of the subject. All subjects then saw a series of eight objects being placed in various locations on or around the box. Half of the objects were tall, oblong-shaped objects, and half of them were wide, flat objects (see right column, figure 8.4). As each test object was placed in its location, subjects were asked, "What



about now? Am I putting this waxal (lechel) the box?" After the object was placed, they were asked again, "Is this waxal (lechel) the box?"

If subjects attended to the overall shape (verticality or horizontality) of the figure object as well as its location, then we should expect them to generalize to a compound of shape and position. If they had heard "waxal," they should generalize to all vertical objects in the relevant location; if "lechel," then to all horizontal objects in that location. (And this region might be the top surface of the box, as it had been in the previous studies.) Alternatively, subjects might ignore the object's overall shape, generalizing to all objects located in the relevant region, as subjects had done in the previous studies.

The overall pattern of results was consistent with previous findings. Subjects tended to generalize the novel term to new locations and to new objects, with children showing an overall tendency to say yes to novel object/position combinations more frequently than adults. Generalization to novel positions was consistent with previous results. Locations on the top of the box were accepted more frequently than those off the box, and adults tended to be more conservative than children, saying yes to the standard position and no to the position off the box more frequently than children. Most crucial to the design of the experiment, there was an interaction between the modeling condition subjects observed and the test objects to which they generalized. Subjects who saw the vertical standard (and heard, "This is waxal the box") generalized more often to other vertical test objects, while subjects who saw the horizontal standard ("This is lechel the box") generalized more often to other horizontal test objects. However, this effect was small, and there was no reliable interaction reflecting differential effects of the standard in both object shape *and* position.

Examination of the individual response patterns shows that few subjects actually generalized on the *compound* basis of object shape and position. Of the twenty adults tested, nine generalized to all objects located in the standard position, and nine more generalized to all objects located on the top surface of the box. Only one subject

Figure 8.4

Objects and layout used by Landau and Hsieh (in progress). Subjects were shown either a vertical object or a horizontal object being placed on the top of a box, as shown in the left column. Subjects shown the vertical object (upper left) were told, "I'm putting this *waxal* my box" (using the Tzeltal spatial predicate for tall oblong objects "sitting canonically"). Subjects shown the horizontal object (lower left) were told, "I'm putting this *lechel* my box" (the Tzeltal predicate for flat objects lying on a surface). All subjects then were shown four vertical and four horizontal objects (right column) being placed on or around the box, and were asked whether each was waxal/lechel the box. Adults entirely ignored the vertical/horizontal aspect of the objects whereas three-year-olds tended to generalize on the basis of the object's principal axis, sometimes in combination with its location.

responded in terms of both shape and position, and this subject said yes to only the standard object in its standard position—that is, he did not generalize beyond the modeled context. This overall pattern is quite different from that shown by the three-year-olds. Removing from consideration the children who said yes to all queries left seventeen children. Of these, three children accepted all objects on the top of the box, and fourteen responded on the basis of the standard object's axis. Of the latter, seven children accepted either vertical or horizontal objects (but not both), four accepted the standard object (vertical or horizontal) in the standard position, and three accepted either vertical or horizontal objects that were on the top surface of the box. Thus, while only 2 of 20 adults had considered the object's axis at all relevant to the novel spatial term, 14 of 17 children (who did not show a “yes” bias) did so. Not a single adult had actually *generalized* on the basis of the compound axis-plus-position, while three children did so.

While these results are only suggestive, the general pattern is intriguing. In this study children, but not adults, tended to conflate the direction of the object's axis with position. Why should children have been more likely to conflate axial information and location in this study when they had shown strong biases in the other studies to ignore axial information? At this point, we do not know, but it is possible that the contrast between vertical and horizontal objects in this experiment led to relatively strong weighting of this object property, while the contrast between two long and one short object (all of which were horizontal) in the previous study could have diminished attention to the axis. If so, this would suggest that the parameters of the contrast set used in such studies might lead to different conjectures about which object dimensions are important. In real language learning, the *linguistic* contrast between such parameters might readily serve to partition the geometric space so as to respect the verticality or horizontality of the object's axis. For example, because Tzeltal contrasts include vertical objects (*waxal*), flat objects (*lechel*), flexible objects (*pachal*), and so forth, they might lead children to partition the geometric object descriptions in a different way from those invited by the partitioning of the object space in English. That even a small number of young English-speaking children are willing to conflate vertical/horizontal axis together with location suggests that the learning process is not over by age three. English-speaking adults appear to be firmer in their conviction that object shape simply should not be conflated with position for novel spatial terms.

8.2.2 Axial Representations: Schematizing Reference Object

That young English-speaking children resist incorporating axial information into the figure object raises the question of whether they show similar limitations for the reference object. As described in section 8.1, languages tend to incorporate a greater

degree of geometric detail in the reference than the figure object. In English, terms such as *in front of/behind*, *above/below*, and *right/left* represent regions surrounding an object, with the particular region defined in terms of the object's three principal orthogonal axes. Identifying such a region and mapping it to its respective term might seem simple. The observer can derive the three axes, extend them outward from the object, and establish regions centered on these virtual axes.

In fact, establishing the relevant regions for such terms requires considerable structure on the part of the observer—that is, representations and rules to ensure that the correct axes are found and that they are extended in a linear fashion from the object itself (see Narissiman 1993, reported in Jackendoff, chapter 1, this volume, and Levinson 1993 for some rules of application). The object axes are not given directly in the stimulus, although many theories of visual object representation suggest that recovering an object's axes is critical to reconstructing an object's shape, hence to recognizing it (Marr 1982; Leyton 1992). The axial representations that must be extended outward from the object are not directly given in the stimulus either; here it would seem critical to acknowledge the role of spatial representation in constructing these extended axes.

These represented axes might be difficult for the learner to construct. According to Piaget (Piaget and Inhelder 1948), the representation of axes does not emerge until well into middle childhood. Moreover, a number of studies have shown that terms such as *in front of* and *behind* are not completely mastered until around age four or even later; this compares to terms such as *in* or *on*, which appear much earlier and do not appear to undergo much developmental change. A prominent view of this difference in acquisition time is that object axes are difficult to represent; in addition, mastering the changing use of reference systems might be quite difficult (compared to using *in* or *on*, which do not engage such systems; see Levelt, Tversky, and Logan and Sadler, chapters 3, 12, and 13, respectively, this volume, for discussion of the complexities of reference system usage). Consistent with this view is Piaget's argument that representation of the straight line is not achieved until middle childhood, and that sensitivity to viewpoint differences is not complete until this time (Piaget and Inhelder 1948; Piaget, Inhelder, and Szeminska 1960). Both of these limitations would impose serious restrictions on the child's ability to learn terms requiring representation of the object axis, and in particular, terms that require extension of the axis outward into space (see, for example, Johnston and Slobin 1978).

The empirical results from acquisition studies have indeed suggested that these terms appear later than other terms not requiring axial representation. It is not obvious, however, that this is the result of a representational problem in the child. They could be due to more data-driven causes such as morphological complexity, form-meaning transparency (e.g., the difference between *in back of* and *behind*) or even

input frequency. In English, *in* and *on* are ranked among the 20 most frequent words, while *behind* is ranked 450th (Francis and Kucera 1982).

Perhaps more to the point, two separate studies have shown that very young children—who have not completely mastered *in front of/behind*—nevertheless appear to possess representations congenial to the mature understanding of these terms. Levine and Carey (1982) gave two-year-olds a linguistic task in which they were to place objects “in front of” another, and a nonlinguistic task in which they were to place dolls and toy animals on a table such that they could either “talk” to each other or follow each other in a parade. Even the youngest children tended to orient the toys properly, suggesting that they recognized the fronts and backs of the objects and knew how to align them with each other. In a separate set of experiments, Tanz (1980) showed that when young children make errors placing one object in front of or behind another, these errors tend to cluster around “cardinal” points, that is, the endpoints of the objects’ principal axes. This again suggests that very young children may possess axial representations of objects which can be accessed for learning spatial language.

These observations motivated us to investigate in detail the nature of young children’s representations underlying the single spatial relationship encoded in English by “in front of.” We had two principal questions. First, we asked whether young children possess an axial representation of objects that could support learning of these axis-based spatial terms, and critically, whether this axial representation permitted extension of the object’s axes to the larger region surrounding the reference object. Second, we asked whether certain structural (shape-based) properties of the reference object might more readily invite an axial interpretation. A number of studies have found that young children are especially poor at assigning fronts or backs to objects that themselves have no intrinsic orientation (e.g., trees, balls, etc.; see Kuczaj and Maratsos 1975; Tanz 1980). If children do possess axial representations that can be accessed for spatial term learning, then there still may be conditions under which this access is impeded, for example, for objects whose principal axes are not clearly accessible from a geometric analysis of their shape.

In the experiment, we showed two-, three-, and five-year-olds and adults one of three different reference objects placed flat and directly in front of them on a table (Landau, in progress; see figure 8.5). One reference object was U-shaped, and because of its proportions and symmetry, it possessed a clear principal axis; a second reference object was round and therefore possessed no principal axis; and a third reference object was identical to the second, but was marked with “eyes” and a “tail” (simple pieces of fabric glued to the surface). These latter properties might induce assignment of a principal axis, and might therefore induce better performance than the round object. Subjects were tested on one of these reference objects; comparison across

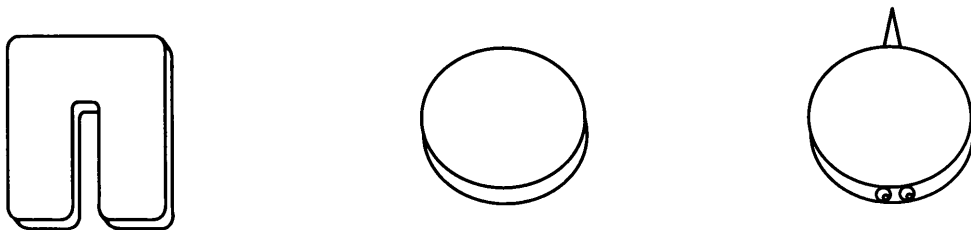


Figure 8.5

Three reference objects used in study of the structure of regions. Objects were presented in the horizontal plane in front of subjects. Subjects were asked to place objects “in front of” each reference object and to judge what locations were acceptable instances of being “in front of” each. Reference objects varied in how clear an axis they exhibited. The U-shaped object possessed a clear principal axis, the round object possessed no such axis, and the round object with “eyes” and “tail” possessed cues to indicate the probable location of the principal axis. Variation in these properties affected the nature of young children’s and adults’ judgments of the region “in front of” each.

reference objects would determine whether cues to the location of the principal axis (in the case of the U-shaped or “eyes” objects) might induce better performance among the youngest children.

Three- and five-year-olds and adults performed in a yes/no task in which they were shown a range of small novel objects (the figure) placed in a variety of locations around the reference object and were asked to judge whether the figure was “in front of” the reference object. Each of the figure objects were placed one at a time in each of the four cardinal locations plus a fifth, directly on top and in the center of the reference object (see figure 8.6). Each time the small object was placed in a location, subjects were asked, “Is this (figure) in front of this (reference object)?” (indicating each object at the relevant moment). Following ten such trials, the object was placed in (up to 16) additional locations in the region fanning out from the side of the object facing the subject.⁵ Figure 8.6 shows all 21 locations, separated into regions that correspond to (A) the broad rectangular region following from the object’s principal axis and (B) the broad triangular region surrounding this. Locations were probed in a particular order, as indicated in figure 8.6, in order to ensure obtaining responses for critical areas such as the region closest to the object (locations 6, 7, 8, 9), the region extending directly from the object’s principal axis (locations 10, 11), the regions surrounding the axis (locations 12, 13, 18, 19) and the regions farther away from the reference object (locations 14, 15, 16, 17, 20, 21).

Following the entire yes/no procedure, subjects were assigned a placement task in which they were given a series of four small objects with no distinguishing features

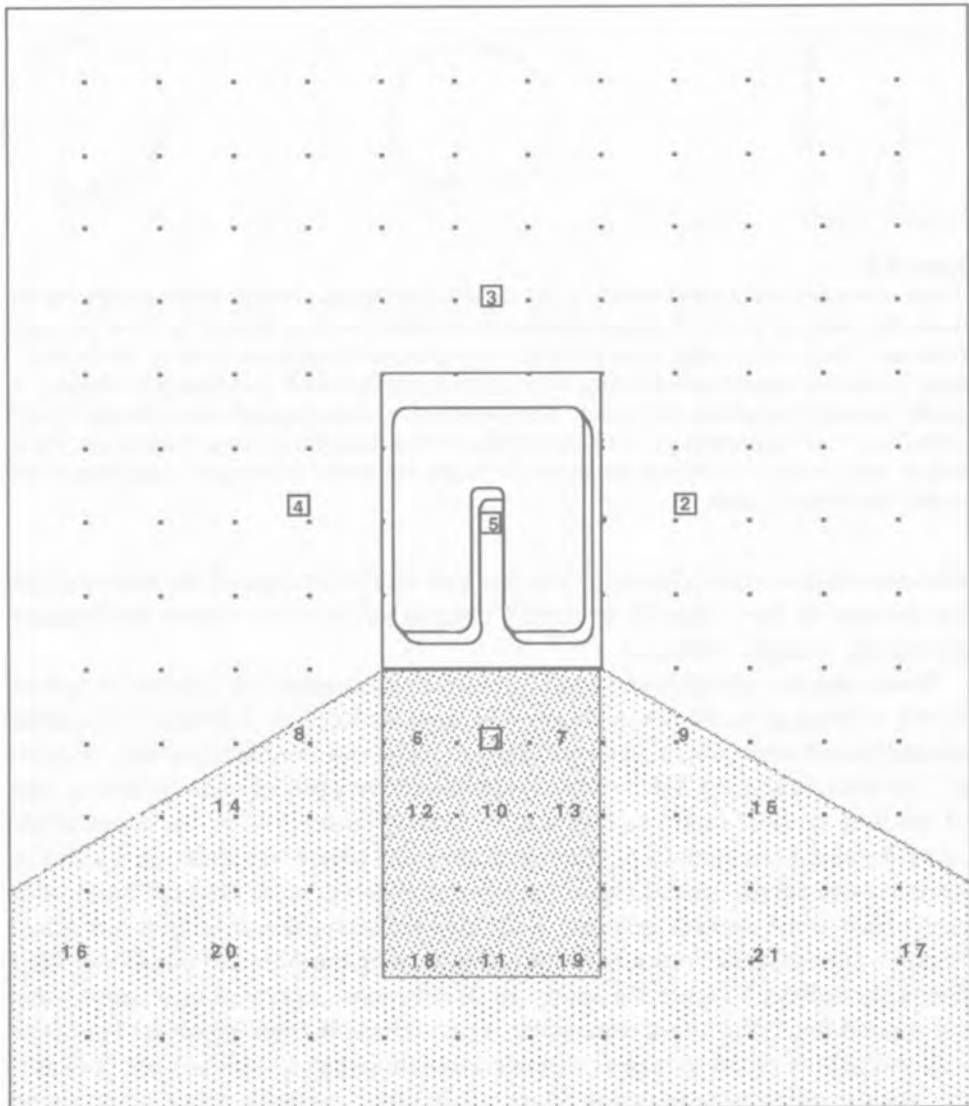


Figure 8.6

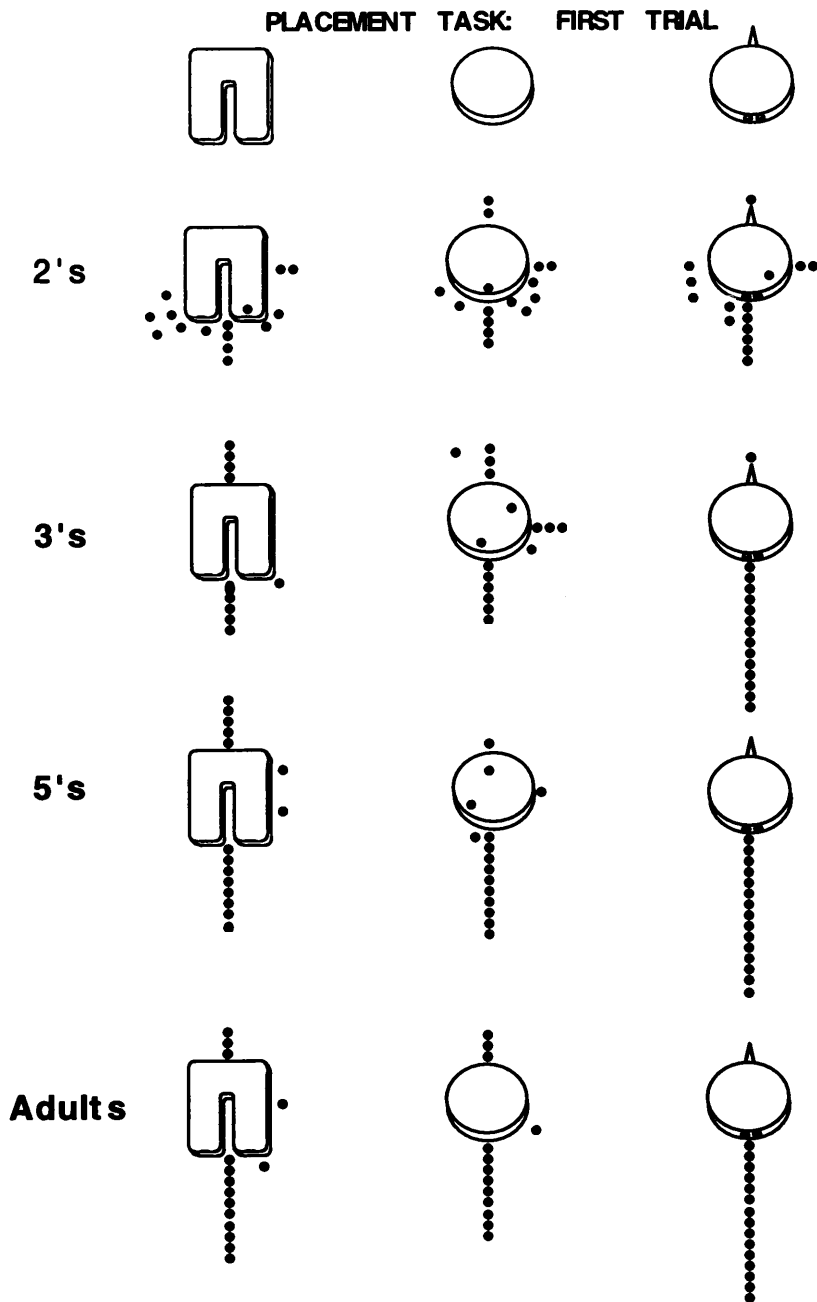
Layout of locations probed in regions task. Subjects first were queried on locations 1–5, each time twice, followed by one query each on locations 6–20, in numerical order. The locations shaded within the block represent the proposed “canonical” region for the term *in front of* and were widely accepted by three-year-olds, five-year-olds and adults. The locations shaded within the triangular area surrounding that block (the “external” region) tended to be less preferred by children and adults, except for the “eyes” reference object, which elicited a high proportion of acceptance by adults (see figure 8.8 for comparison of canonical and external regions).

and were asked to place each “in front of” the reference object. Three separate groups of two-year-olds, one group for each reference object, were also assigned this placement task.

The results for the placement task are shown in figure 8.7 for each of the reference objects. Individual dots represent different subjects' placement of the first object they were given. The most frequent response across all ages was to place the object in line with the reference object at one of its cardinal points, that is, the point at which its front-back or side-side axis would have projected into the space surrounding the object. For ease of representation, this is shown in the figure as subjects lined up adjacent to each other. The pattern becomes stronger over age; but the major developmental change appears to occur between the ages of two and three. At age two most children place the object at one of the cardinal locations, especially favoring both ends of the object's front/back axis. The second most common pattern occurs at both ages two and three, and finds children locating objects at the end of the side/side axis. Although there is some diffuseness in the responses of the two-year-olds, this disappears by the age of three.

Note that both children and adults do vary somewhat in their preferred location for this initial placement. Even some adults considered the far end of the object to be their first choice for locating the figure in front of the reference object (a pattern that is the preferred one in Hausa; see Hill 1975). This variability occurred only for the U-shaped and round reference objects, however. Adding eyes appeared to drive subjects of all ages to locate the figure object directly along the half axis extending outward from the eyes. This is consistent with previous findings suggesting that young children more often correctly place objects in front of or behind objects with clear fronts and backs (Kuczaj and Maratsos 1975).

The results of the yes/no method tell a similar story about the cardinal locations. When subjects were asked to judge whether a small object was in front of the reference object, they tended to say yes to locations 1 and 3—the locations falling at the two ends of the front/back axis. This pattern of accepting *both* 1 and 3 (in English, the canonical locations for “in front” and “behind”) occurred almost always with the U-shaped and the round reference objects. The round object with eyes elicited “yes” responses only to 1, the location directly adjacent to (i.e., in front of) the eyes. Trailing behind 1 and 3 as the subjects' second choices were locations 2 and 4—the locations falling at the ends of the side/side axis. This pattern was most prominent among three-year-olds judging locations around the round reference object, and least prominent among adults judging locations around the “eyes” reference object. The relatively high acceptance of locations 2 and 4 among three-year-olds judging the round object suggests that lack of a clear object axis invited subjects to entertain more than one axis as the relevant one for determining when one object was



in front of another. That is, the strong axis-based responses for the U-shaped and “eyes” reference objects suggest that young children are quite capable of representing an object’s axis; their relatively poor performance with the round object suggests that objects lacking a clear principal axis may be less effective in allowing young subjects to show their knowledge.

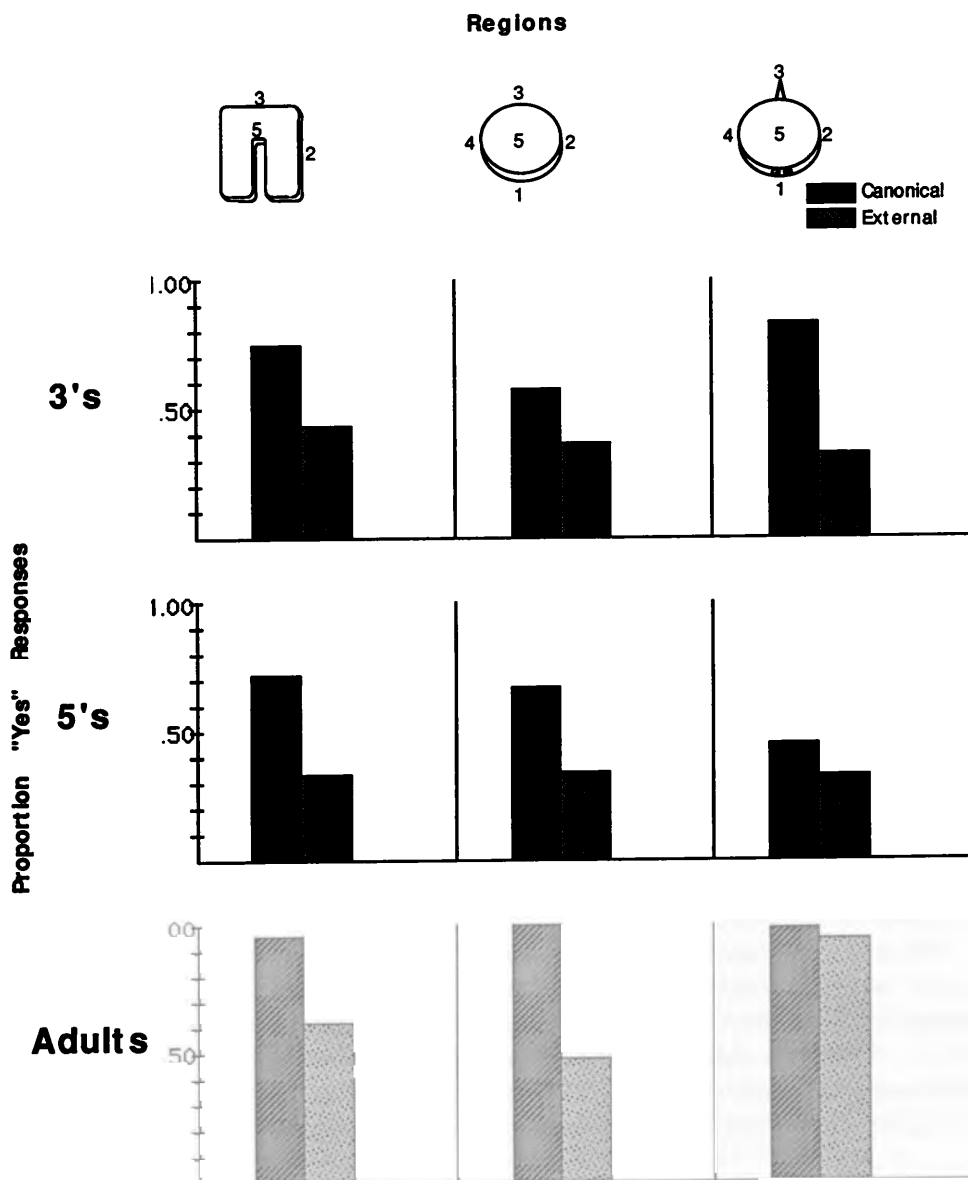
Finally, the analysis of the regions surrounding the cardinal points showed that subjects of all ages generalized their interpretation of “in front of” to regions with a well-defined geometry that was based on extension outward of the object’s principal axis. Figure 8.8 shows the proportions of “yes” responses to the two different regions. The “canonical” region represents those locations falling within the rectangular region extending directly outward from the front edge of the reference object (see figure 8.6). The “external” region represents the triangular region extending outward from the front edge of the reference object and surrounding the canonical region.

Three noteworthy trends appear in figure 8.8. First, subjects at all ages and for all reference objects accept the canonical regions more frequently than the external regions. This suggests that even the youngest subjects represent the region directly adjacent to the reference object as the preferred region for the spatial relationship encoded by “in front of.” Second, there appears to be growth in the size of the canonical region over age. Adults accept a greater proportion of the locations as legitimate cases of “in front of” than do children. The cause of this difference cannot be ascertained from the figure; but is due to expansion *around the principal axis* rather than by a random increase in the acceptance of locations or by expansion outward from the front edge of the object. That is, younger children tend to accept positions directly along the axis while adults accept the entire canonical block. Over development, the region for “in front of” does not become larger by seeping outward from the object’s edge; rather, it becomes larger by expanding outward from the object’s extended (virtual) axis.

The third significant development concerns the subjects’ treatment of the reference object with eyes, relative to the other objects. Although the regions appear to be similar for the U-shaped object and the round object at all ages, the pattern differs for “eyes.” Three-year-olds show the same pattern for eyes as they do for the other two reference objects, although their preference for the canonical over the external regions is slightly more pronounced for eyes. Five-year-olds show an overall dampening

Figure 8.7

Subjects’ first placements in response to the question “Can you put *X* in front of the (reference object)?” Subjects at all ages showed a preference for one of the four cardinal locations, each location arising from extension of axes mentally imposed on the object.

**Figure 8.8**

Proportions of "yes" responses to the question "Is this in front of (the reference object)?" Locations are within the "canonical" and "external" regions surrounding each reference object (see figure 8.6 for details of these locations). Subjects at all ages preferred the canonical locations to the external ones, except for adults judging the "eyes" reference object. There, they

of “yes” responses, both for the canonical and external region. Inspection of individual responses by the five-year-olds suggests that the critical change is in the canonical region, where a number of subjects insist that the only acceptable locations are those falling *along the extension of the axis*. This conservatism causes an overall reduction in “yes” responses. Finally, adults show an overall increase in the size of the external region for the “eyes” reference object. This appears to be due to their assumption that the relevant region is affected by the object’s status as an animate. A number of subjects were quite explicit on this issue, remarking that the object could “look over this way” (indicating a location in the external region). The idea that the geometry of a region in front of an animate object might be different from that in front of an inanimate one is reminiscent of Michotte’s (1963) notion. There may be a region of reactivity wherein perceivers represent to themselves not only an object’s region of influence when static, but also the region from which it *potentially* could influence another. Whether or not such regions are also part of young children’s representations is an intriguing question.

To summarize, the foregoing studies strongly suggest that young children can and do represent the principal axes of reference objects by the age of two. The geometric structure of the reference object itself has some effect during the early years, but by and large, young children appear to be capable of setting up object axes even in cases where the perceptual clues to the location of the axes are weak. By the age of three, these axial representations can be extended outward from the object and can serve as organizing reference frames for setting up regions relevant to basic spatial terms such as *in front of* or *behind*. These regions seem remarkably similar to those described by Tversky and by Logan and Sadler (chapters 12 and 13, this volume) among adults participating in imagery and attention tasks (see also Hayward and Tarr 1994). To the extent that they differ from those of adults, the children’s regions appear to be more narrowly defined with respect to the object’s axis. Thus, contrary to the pattern predicted by Piaget’s theory, the development of regions appears to *begin* with the axis, and broaden with development. Although the geometric and conceptual nature of the reference object may modulate the geometric details of the relevant region, these effects seem to be imposed on a basic pattern in which any reference object can be represented in terms of its axes and surrounding regions. This basic pattern

Figure 8.8 (continued)

accepted all locations fanning outwards from the reference object, as if they object was mobile and could explore the environment. Five-year-olds’ depression in acceptance of the canonical region stems from their reluctance to say “yes” to any locations **except** those along the extended axis itself (locations 1, 10, and 11 in figure 8.6).

appears to exist quite early in life and is mapped onto the corresponding spatial terms between the ages of two and three.

8.2.3 Fine-Grained Representations: Schematizing Object Kinds

Although the focus of this chapter so far has been the young child's ability to schematize objects in terms of either skeletal (axis-based) or coarse (blob-like) geometric descriptions, there is strong evidence that children are not limited to these descriptions but can also represent objects in terms of rather detailed shape. However, these representations tend to emerge when children are engaged in learning the *names* of objects, rather than their locations.

Recall that in some of the experiments described in section 8.2.1, children and adults were shown an object being placed in a location on a box. In one condition, the scene was described using a novel term in a syntactic and morphological context suitable to a preposition. In this context, neither children nor adults generalized the novel term on the basis of the object's shape. However, in another condition, the scene was described using a novel term in a syntactic context suitable to a count noun, as if the object itself (and not its location) was being named. In this context, both children and adults generalized the novel word only to objects of the same shape as the modeled one, regardless of its location. Attention to object shape during object naming has been demonstrated in a variety of other kinds of studies.

In many of these studies, young children (two-, three-, and five-year-olds) and adults are shown a novel object and they hear it labeled, for example, "See this? This is a *dax*." Then, with the novel object still in view, subjects are shown a series of additional novel objects and asked for each, "Is this a *dax*?" In another version of this study, subjects are shown the novel object, and hear it labeled, but then they are shown pairs of objects and are asked, "Which one is a *dax*?" The results from these two methods tend to converge and suggest that object shape is a privileged perceptual property that is engaged when young children are learning object names.

A variety of evidence over the past twenty years has hinted at this pronounced role of object shape. Clark (1973) reported that children's early overgeneralizations tended to be based on shape, as, for example, when a child calls the moon "ball," or a dog "kitty-cat." In another context, Rosch et al. (1976) argued that our basic level categories—sets of objects named by count nouns in response to the question "What is that?"—are organized in terms of certain key properties, including shape. A number of developmental studies have shown that children find it easy to learn names for shape-based categories (Bornstein 1985).

A systematic attempt to assess the role of object shape in the development of object naming was reported by Landau, Smith, and Jones (1988), who compared children's weighting of shape, size, and texture in generalizing the novel count noun to new

objects. In the basic experiment, children were shown a novel object and heard it labeled, then they were shown objects varying from the standard in either shape, size, or texture. Each time, they were asked whether the test object was also a member of the named category (e.g., “Is this a dax?”) Both children and adults tended to weight shape more strongly than either size or texture. For example, when told a novel object was “a dax,” subjects then generalized the word *dax* to other objects having the same shape as the original object, even if they were much larger than the original or possessed a quite different surface texture (see figure 8.9).

In this study and several follow-ups, a developmental pattern emerged: The “shape bias” appears to be weak among two-year-olds, moderate among three-year-olds, and quite strong among five-year-olds and adults (see, for example, Landau, Smith,

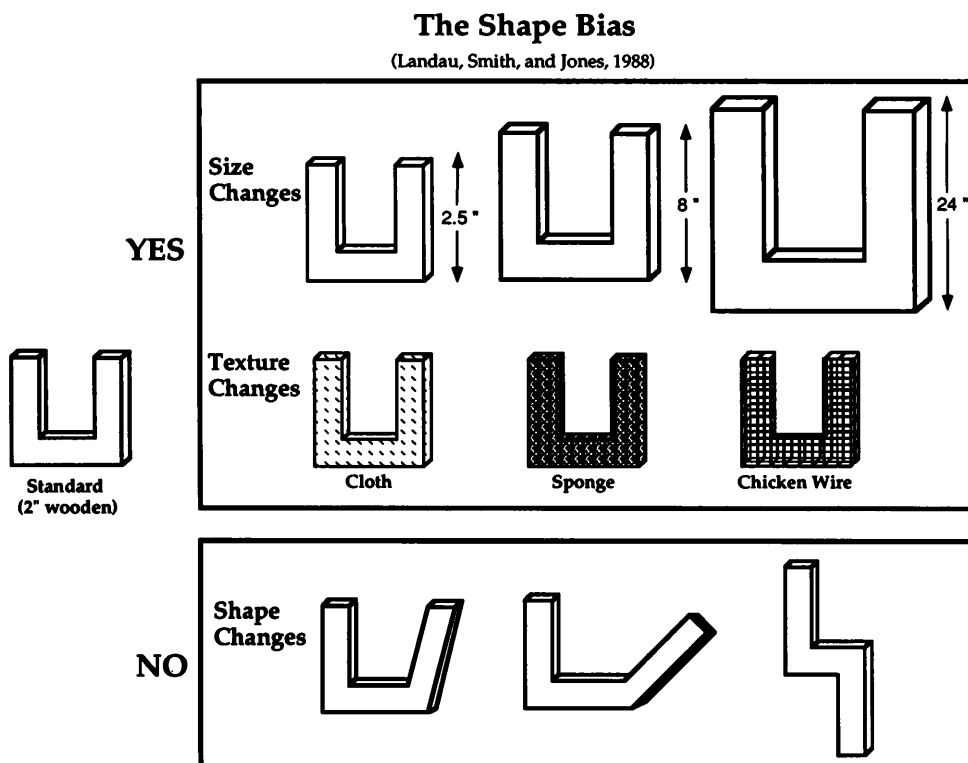


Figure 8.9

When children and adults are shown a novel object and hear it labeled, they tend to generalize the object's name to others of similar shape, regardless of size or texture. After Landau, Smith, and Jones (1988).

and Jones 1992). For example, adults reject even a small change in shape from the original but accept an object of the same shape that is ten times as large. The younger the child, the more willing he or she is to accept objects of different shape; although by the age of about two, children show a reliable tendency to generalize on the basis of same shape in the naming context (Landau, Smith, and Jones 1988). Recent studies indicate that the growth in the shape bias is correlated with children's productive vocabulary. The bias appears to begin around the time when children have fifty words in their vocabulary, suggesting that the bias may become sharper as children learn more about which properties are the best basis for generalization (Jones et al. 1992). That is, because many words do indeed refer to objects sharing the same shape (e.g., the basic level terms common in maternal input), an input bias may act in concert with the children's own representational biases; the children may sharpen their conjectures as they learn that words for objects may safely be generalized to other objects sharing the same shape. Same-shape objects are often of the same kind, hence such a generalization would in general be safe.⁶

Although the shape bias emerges quite early in development, the preference for same shape is highly context-dependent among both children and adults. The particular pattern of context dependence suggests that the bias is closely linked to the representation of objects and, in turn, object naming. By the age of two, the shape bias appears most robustly in the object-naming context, while in other contexts, young children show different preferences. For example, Soja, Carey, and Spelke (1991) found that two-year-olds showed a strong shape preference when shown a rigid object, but a very weak shape preference when shown a mass of gooey substance (see also Subrahmanyam 1993). This suggested to Soja, Carey, and Spelke that young children bring to the language-learning task certain *a priori* categories—in this case, object and substance—whose existence might constrain the range and type of inferences they can project from a single exemplar. As another example, children who have learned a bit more syntax can be guided by syntactic and morphological information to attend to properties other than object shape. Landau, Smith, and Jones (1992) showed that some three-year-olds and most five-year-olds tended to generalize on the basis of object shape when instructed with a count noun ("This is a dax"), but tended to generalize on the basis of surface texture when instructed with a novel adjective ("This is a *daxy* one"; see also Smith, Jones, and Landau 1992). Subrahmanyam (1993) showed similar effects among three- and five-year-olds using count nouns that guided attention to shape and mass nouns ("This is some dax") that guided children's attention away from shape, often to substance.

What are the geometric properties of these shape-based representations? One striking fact is their level of detail, compared to those representations recruited for locating objects. When an object is being named, its shape-based representation appears

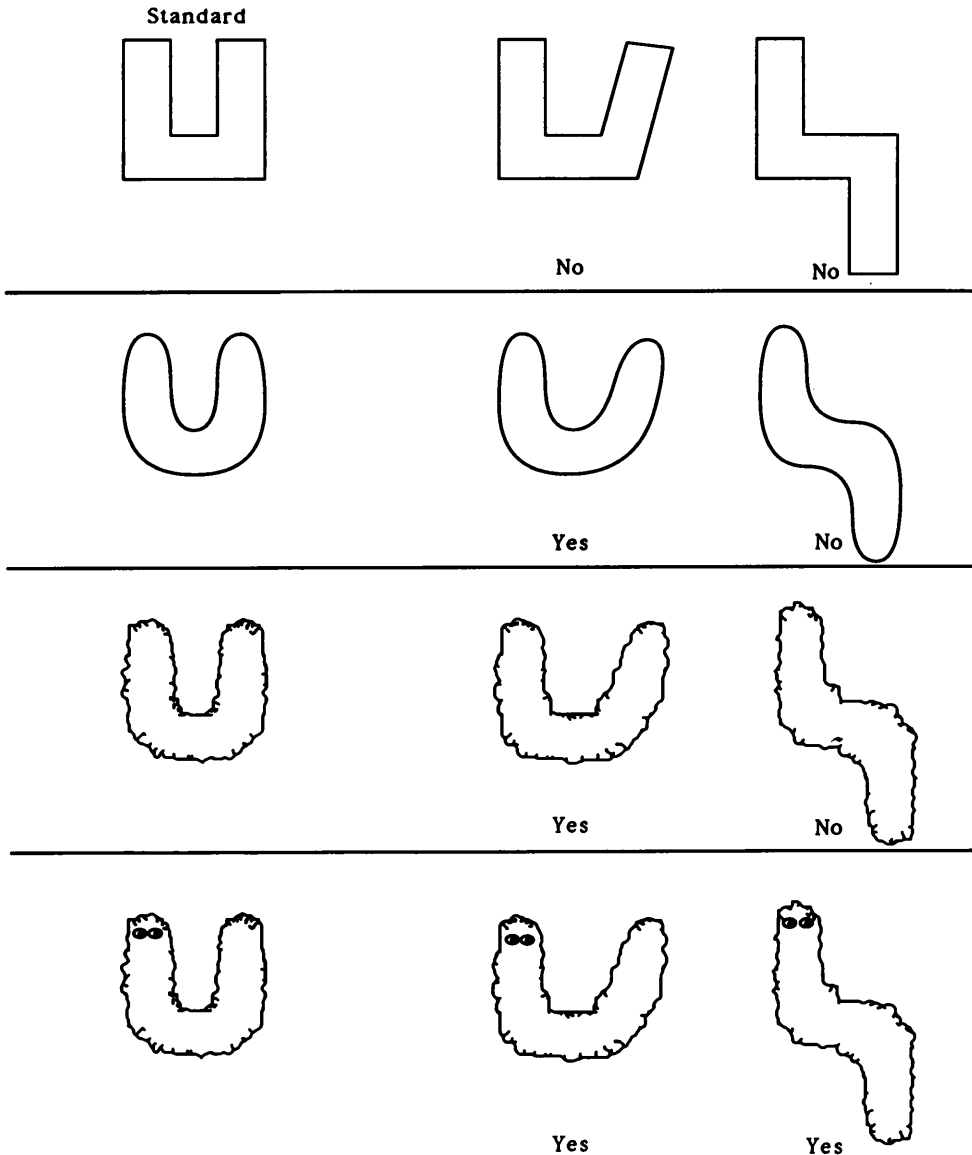
to preserve a good deal of detailed information about its part structure and arrangement (while such elements seem to have only limited relevance for locating objects). For example, in the studies described above, adults who were shown a U-shaped object tended to reject objects with even a small deformation in overall shape caused by bending one of the object parts slightly outward. In the studies described in section 8.2.1, children and adults who were shown a straight rod and heard it named tended to reject a roughly linear object of the same linear extent; this object did not qualify as a member of the same named category, apparently because its overall contour was wavy (as compared to the straight contours of the modeled object).

The range and degree of detail necessary to include objects into the same named category is as yet unknown. However, many modern theories of object recognition propose a strong role for object parts as components of object shape (Hoffman and Richards 1984; Leyton 1992; Biederman 1987); and it is the arrangement of these parts that would seem to capture what we call an object's "shape." Further, the specific arrangement of parts will be subject to some variability or range, as many objects possess parts that regularly undergo motion. Although little is known about the range of object-internal motions that must be captured by theories of shape, there do exist models for characterizing limited classes of motions (see, for example, Marr and Vaina 1982).

Because of the importance of both part articulation and part motion in the characterization of object shape and in theories of object recognition, one might expect that young children would respect both in their generalization of object names. Several recent studies suggest they do.

In one series of studies, Landau et al. (1992) sought to determine whether children would make different predictions about the range of acceptable shape transformations relevant to a named object, depending on its part structure, especially as it interacts with imputed malleability. Three-year-olds and adults were shown a novel line-drawn object, heard it named, and then were asked which of a set of shape and size transformations were also members of the named category. In a first experiment, subjects were shown a rigid-looking object with sharply delineated part boundaries (figure 8.10). They were tested on three successively more extreme shape changes, and three successively larger size changes. As in most of the studies on object shape and object naming, both children and adults extended the novel label to objects of different sizes from the standard, but did not extend the label to objects of different shape.

In subsequent experiments, subjects were shown standard objects comparable to those from the first study, but whose part structure and suggested rigidity was altered. For example, subjects in one experiment saw an object identical to the standard of the first experiment, except that it possessed curved edges, which weakened the object's

**Figure 8.10**

Children and adults' judgments of which objects belong in the same named category are affected by subtle details of object shape. When angular objects were shown and named, subjects tended to reject even small shape changes (top panel). However, as part structure was weakened, either by curving edges or adding "wrinkle," subjects tended to accept more shape changes (middle two panels). When eyes were added, subjects accepted quite substantial shape changes, as if they now assume that they object can easily be deformed (Landau et al. 1992).

part boundaries and suggested malleability (figure 8.10). In another experiment, a separate group of subjects saw an object identical to that of the second, except that it possessed massively textured edges, further suggesting nonrigidity. And in a fourth experiment, different subjects saw a curved and "wrinkled" object with "eyes" placed at one end. This last type of object was meant to test whether certain powerful cues to object kind (in this case, a cue to animacy) would affect subjects' judgments of which shape-changed objects could still be members of the named category.

The results of the four experiments showed massive effects of part structure and suggested rigidity. Although subjects had generalized solely to size changes in the first experiment, progressive weakening of the part boundaries (and correlated destruction of cues to rigidity) led them to generalize to shape changes as well. Both the curved and curved/wrinkled objects led subjects to accept a moderate number of shape changes. When eyes were added to the objects, subjects generalized to *all* shape changes, as if they now assumed that the object was a tubelike, nonrigid object capable of internal motion. Thus, as rigidity and strong part boundaries were successively destroyed, subjects became more and more willing to accept a larger range of shape changes. This suggests that a bias for "same-shape" objects must engage object representations that admit of flexibility in the face of varying rigidity and changing part structure (see also Becker and Ward 1991).

In a different series of experiments, we have been investigating children's inferences about the kinds of shape changes that might obtain under mechanical transformations. In one of these experiments, we showed children a novel object that was composed of distinct parts arranged in a particular configuration (see figure 8.11). One group of subjects was shown each of the three standard objects, heard it labeled (e.g., "This is a dax"), and then was shown a set of new objects whose configuration would obtain if the standard object's parts were capable of motion. Another group of subjects was shown objects of the same configuration, but this time, subjects also saw one part of each object undergoing a small motion. (Objects undergoing rotation now had hinges at their joints). All subjects were then shown the same set of test objects, which were possible motion-based shape changes of the standard. Children and adults who saw the standard object with no motion generalized to few shape changes. However, those who saw the standard undergoing a small amount of part motion generalized quite freely to the novel configurations, each of which was consistent with a more extensive range of object part motion.

These studies begin to suggest that the spatial system underlying object naming incorporates information about object shape. In particular, it must incorporate a relatively detailed (possibly hierarchical) representation of object shape, in which the object's parts, their spatial relationships, and their ranges of relative motion are present. These representations could provide a powerful system that would allow the

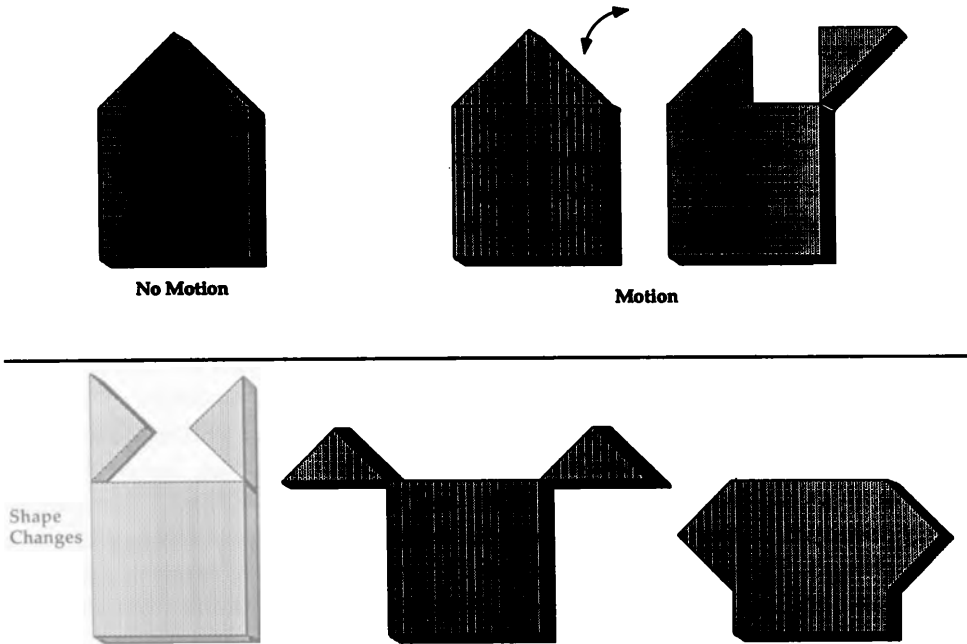


Figure 8.11

Children and adults are sensitive to the range of shape transformations likely when an object has permanently fixed versus moving parts. Subjects who viewed an object with no motion (left) and heard it named then tended to reject objects with even small changes in configuration as an instance of the named category. In contrast, subjects who viewed an object undergoing a small motion then accepted shape changes that would be the product of more extensive motions.

young child to link up an object name with its shape and to generalize to classes of transformed shapes consistent with certain principles of object constancy. These representations seem quite different from those engaged when young children are learning or using terms that describe an object's location.

8.3 Some Crosslinguistic Notes and a Possible Challenge: Tzeltal may be Exception that Proves Rule

The empirical evidence reviewed suggests that young children possess different kinds of object representations, each of which is selectively active when children are engaged in tasks involving different parts of the vocabulary. The detailed representations of object shape that seem to be engaged when young children are learning object

names do not appear to be engaged when children are learning words for object locations.

But one might object that all of the experimental evidence reviewed thus far has concerned children speaking English. Considering the variation in how locations are expressed over languages, one should be suspicious of conclusions based only on one language. Moreover, evidence on the structure and acquisition of other languages suggests that very young children—well before the age of two years—have begun to form spatial-linguistic categories consistent with those found in their native language. For example, Choi and Bowerman (Bowerman 1991 and chapter 10, this volume; Choi and Bowerman 1991) have found that children learning Korean are likely to respect distinctions between “tight fit” and “loose fit” contact and containment relations ignored by children learning English spatial prepositions (though of course English-speaking children must respect such distinctions when they learn adjectives such as *tight* and *loose* or verbs such as *to fit*).

Such crosslinguistic differences point to a strong role for early learning, but they do not invalidate the search for the universals that underlie the expression of spatial language. Continuing with the examples provided by Bowerman (chapter 10, this volume), children learning Korean, Dutch, and English all differ somewhat in the range of object types that are included in basic spatial notions expressed in English by the action of “putting in” compared to “putting on.” Korean distinguishes between “degree of fit” and among actions involving “putting on” different types of clothing. Dutch distinguishes between various types of attachment all covered by the English preposition *on*. Other languages make yet further distinctions that are not found in English. For example, as noted earlier, German distinguishes two types of “contact” (on) relationships by the orientation of the reference object (*auf* for gravitational contact, usually horizontal, and *an* for nongravitational contact, such as attachment to a wall). A number of languages collapse the distinction between locational and directional terms that is made in English by *in* versus *into*; for example, Russian has a single term, *v* (*vo* before certain consonant clusters), covering both, as does Italian. In other cases, it is English that collapses locational and directional meanings (e.g., English *over* can be locational, as in “The plane was over the house” or directional, as in “The plane flew over the house”), whereas other languages split the two (e.g., German *über* can be either directional or locational, but *oben* can be locational only).⁷

Yet none of these differences seem to provide major counterexamples to the claim that the figure object tends to be represented as a point, blob, or line, that the reference object tends to be represented as those or as a set of orthogonal axes, and that the geometries of both figure and reference object are considerably “sparser” (in terms of shape detail) than the representations of these same objects as members

of categories, named by nouns. Could this be a universal, as first suggested by Talmy (1983)?

Some recent evidence from Tzeltal might appear to provide counterexamples to such a universal. This language has often been described by investigators as particularly "visual" in that it appears to encode a large range of shape distinctions in its closed-class items, including locational terms.⁸ For example, there are predicates that apparently describe "bulging bags, sitting" and others that describe "horizontal things, lying" (Brown 1993). Tzeltal has therefore been offered as a counterexample to the notion that very little shape information is encoded in the figure and ground for purely locational terms (Levinson 1992).

The evidence comes primarily from the Tzeltal body-part system, which uses animal and human body-part terms to assign names to the spatial regions of objects, regions that are encoded in English by terms *top*, *bottom*, *front*, *back*, and *side*. For example, the term for "head" is used to describe the tops of objects and the term for "bottom" is used for the bottoms of objects. So much is actually quite similar to English. We often refer to the "head" or "foot" of the table, or the "arm" or "leg" of a chair. However, Levinson claims that much more so than in English, the Tzeltal body-part system uses specific elements of shape to assign particular terms to particular locations. For example, according to Levinson, the term for "nose" would be used to locate something at an object part with a particularly sharp protrusion, whereas the term for "mouth" would be used to locate at an object part with an edge or orifice, and the term for "tail" would be used for long thin protrusions.

Does this mean that fine-grained shape information is part of the spatial meaning of the locational term; and that this therefore erodes the shape distinction between objects-as-named and objects-as-located? I believe not. The shape distinctions do not appear to be part of the spatial meaning of the term, but rather, are distinctions used to identify particular regions relevant for the term's meaning. To put it another way, the body-part terms do not appear to refer to the distinctive shapes of, say, nose or tail poised on some object (though they would if they were used as nominals). Rather, when used as locatives, the terms refer to spatial regions whose locations are defined with respect to some salient geometric property. The meaning of the term (i.e., what region of the object it maps onto) is separate from the geometric algorithms used to assign the term to the object. To take an example from English, the "head" of the table is a region at the end of the table's principal axis; which end is usually decided by a variety of criteria (e.g., where the Queen sits). Just because the term *head* is used to name the region does not mean that each and every "head" of a table must be similar to a real head in shape.

The case of Tzeltal seems analogous. According to Levinson (1992), Tzeltal assigns most body-part terms not by a metaphorical extension, but rather, by a strictly

geometric algorithm that analyzes the object into its major components and their relative orientations. Thus the location of the region “head of” an object is defined with respect to the object’s principal axis; the axis is found by using properties such as elongation, protrusion, flatness, and symmetry—properties that are likely to be universally important in such assignments (see Jackendoff, chapter 1, this volume; Leyton, 1992). As a (perhaps necessary) bonus, such properties are in general likely to remain robust over a variety of viewing conditions (e.g., blurring, rapid exposure), thereby supporting the assignment of axes and directions to objects during a wide variety of spatial tasks that humans normally perform.⁹

What kinds of counterexamples would disconfirm the hypothesis that both figure and reference object do not contain any particular shape information necessary for describing the region? As suggested by Landau and Jackendoff (1993), one should not expect to find any spatial terms that correspond to spatial relationships holding between specific volumetric components or specific arrangements of components. Such examples might be found, of course, in some languages, and this would necessarily lead to modification of the hypothesis, possibly suggesting a restricted set of shape properties that is relevant to spatial terms. As it stands, however, the evidence from Tzeltal does not suggest that spatial terms map onto specific component shapes. Rather, it suggests that using spatial terms requires being able to locate the relevant region (usually dependent on the object’s axes). This is as true of Tzeltal as it is of English, and presumably of all natural languages. Thus Tzeltal, rather than providing a striking counterexample to the general claim that the specifics of shape are absent from the figure or reference object, may provide a particularly compelling example of how vast apparent crosslinguistic differences may ultimately rest on deep similarities in how language maps onto spatial representation.

8.4 Structure, Function, and Mechanism: Some Possibilities, More Questions

What causes the differences in geometric representation between figure and reference object on the one hand, and objects as named on the other hand? Several kinds of explanation suggest themselves.

One possibility is that this difference reflects nothing particularly interesting about either language or spatial cognition, but rather, is a direct consequence of how the world is. Objects in the world actually do come in an astounding variety of shapes, and objects in particular named categories happen to share greater overlap in shape than they do with objects in different categories (Rosch et al. 1976). Locations in the world do not possess shape themselves, but they do possess a three-dimensional structure that demands encoding in terms of three principal axes. Perhaps object shape does not matter to location because spatial locations do not demand such information.

As stated, this possibility seems wrongheaded. Although it is certainly true that objects come in many shapes, and also true that locations come specified metrically in three dimensions, it is not a foregone conclusion that all organisms will encode objects and spatial relationships in just this way. Why not encode objects in terms of relative size, rather than shape? Why not encode location in terms of general proximity to oneself—things close enough to reach, far enough not to, without regard to the three axes? Given that there are different possibilities for how we represent objects and places, the question is, What gives rise to the particular way in which we do represent these aspects of space for the purposes of language? Why do we attend to shape when naming objects, but (basically) ignore it when locating those same objects? The structure of the world surely imposes some constraints on our representational system; but these systems are not direct reflections of some objective description of “the world out there.” More plausible is the possibility that our representational systems have evolved in response to constraints on both the physical world and on the tasks we must achieve.

How, then, do we *represent* the world? The study of spatial language can tell us how we represent the world linguistically; but does this have any bearing on how we represent the world *nonlinguistically*? Are there any communalities between the representations underlying the language of objects and places, and their nonlinguistic counterparts? Is the structure of spatial language driven at all by the structure of spatial cognition?

There appear to be several intriguing parallels between spatial language and spatial cognition that suggest possible relationships. One parallel concerns the separation between object and place in language, on the one hand, and that found through neurological and cognitive studies of the “what” and “where” systems, on the other (Ungerleider and Mishkin 1982; see Landau and Jackendoff 1993 for fuller discussion of this parallel). A variety of evidence suggests the existence of two systems in monkeys and in humans, one specialized for the task of object identification (“what”) and the other for object localization (“where”). For example, experiments on monkeys have shown selective deficits in the two tasks. Damage to the inferior temporal cortex appears to disrupt object identification (but not object localization), whereas damage to the posterior parietal cortex disrupts various localization tasks (but not object recognition). These cortical areas contain neurons with quite different receptive field properties. Those in the inferior temporal lobe have a large receptive field falling within the fovea and are driven by complex sets of features; those in the posterior parietal lobe have a receptive field that does not include the fovea and are insensitive to such features (see Schneider 1960; and Ungerleider and Mishkin 1982 for review).

Converging evidence from human psychophysical studies suggests two streams of processing that may reflect a similar bifurcation. The “parvo” system is specialized

for color and shape, whereas the “magno” system is insensitive to color but is specialized for properties relevant to localization—motion, depth, and location (Livingstone and Hubel 1989; but see Van Essen, Anderson and Felleman 1992 for evidence that the systems are coordinated at relatively early stages of processing). Human clinical evidence indicates that object recognition functions can be spared without localization, and vice versa (Fara et al. 1988; Levine, Warach, and Farah 1985). Recently, evidence has appeared for a functional separation between object and color naming on the one hand, and spatial (locational) language on the other (Breedin, Saffran, and Coslett 1994).

Why is this evidence relevant to the structure of spatial language? Landau and Jackendoff (1993) suggested that the different properties of these systems might serve as one pressure in the design of spatial language. For example, the fact that object shape and color (but not location) are represented in the “what” system, whereas object location (but not shape or color) is represented in the “where” system is reminiscent of the distinctions uncovered by linguistic analysis and documented through experimentation among young children. It is possible that the relative lack of shape information in locational terms across languages is due to the lack of shape information in the cognitive and neurological systems underlying object location. Similarly, the lack of locational information in object names may be due to the lack of such information in the systems underlying object recognition. While intriguing, this parallel between spatial language and spatial cognition will undoubtedly undergo revision as we learn more about the coordination of the “what” and “where” systems. For example, a variety of evidence points to the necessity of coordinating information at levels likely to precede linguistic encoding. Objects must be assembled from parts (and this requires assignment of relative location), certain named locations must be supported by quite specific and detailed perceptual representations (e.g., “Dodger Stadium,” “Lincoln Center”), and perception of certain kinds of motion (a “where” system problem) may be constrained by the specifics of object identification (Shiffrar 1994).

A second intriguing parallel, not inconsistent with the first, is that there are different functional consequences for the tasks of object identification and object location, and that these functional differences give rise to differences in the kinds of properties most readily processed in the two tasks. A recent study by Schyns and Oliva (1994) illustrates how this might occur. Subjects were shown a target scene followed by a mask and a rapidly presented image that was a hybrid of two of different kinds of scenes (each a possible target), for example, a combination of a city scene and a highway scene. In different conditions, the hybrids were created from a low-pass filter of one scene (say, the city) and a high-pass filter of the other scene (say, the highway). The low-pass filter preserved only “coarse” information about the scene; for example

it preserved the scene's overall geometry but eliminated all "fine-grained" boundary and edge information such as would be required for identifying particular buildings or vehicles. The high-pass filter preserved fine-grained information. Thus one city-highway hybrid might contain the overall geometry of the city with the fine details of the highway vehicles; the reverse hybrid would contain the overall geometry of the highway with the fine details of city buildings. The question was whether subjects would identify the hybrids on the basis of coarse or fine-grained information, and how this would vary with exposure time.

The results showed that at the fastest presentation times (30 ms), subjects tended to identify the scene represented with low-pass filter (coarse) information; at slower presentation times (150 ms), they tended to identify the scene represented by high-pass information. Schyns and Oliva (1994) interpreted this pattern as evidence for two different processing schemes that operate in sequence. One scheme operates earlier by extracting only coarse information about scene geometry, while the other operates later by extracting the finer information. While both might be used to identify scenes, sequential operation would allow the perceiver to extract information about general geometric composition first, followed by focused attention to the details of an identified scene; this would be most beneficial when the scene was unknown and the perceiver had to categorize it quickly. Schyns (personal communication) comments that if coarse-grained information is indeed processed more rapidly than fine-grained, then the "where" system might be incapable of doing anything but selecting coarse information about objects and their general geometric relationships.

These two parallels between linguistic and nonlinguistic systems place the burden of explanation on the design of systems that presumably evolved independent of language. Does it make sense to attribute the design of spatial language to such causes? And certain facts about spatial language must surely be learned (or ignored)—children learning Tzeltal must learn to attend to an object's "bulginess" or "flatness" when describing its location, while children learning English must learn to ignore these attributes. What are we to make of this? Note that none of these possibilities is inconsistent with the others. Any learning device that begins with some broad set of distinctions is likely to converge on a solution more quickly than an unconstrained device—as long as the set of universals is correct.

Indeed, it is highly likely that universal predispositions in object representation interact with learning quite early in life. Consider object shape and object name. It is a fact that the human visual system can distinguish among an enormous variety of object shapes. It is also a fact that sameness in object shape is strongly correlated with sameness in name; this is most likely because object shape is an excellent predictor of other properties held in common by members of many object "kinds"

(though clearly not all; see Bloom 1994). Because object names often do apply to objects that are similar in shape, children learning all languages should learn terms (for object kinds) that are correlated with these same-shape objects. In this way, they could learn that shape is important to object naming. Similarly, because locational terms such as spatial prepositions tend to apply across objects that vary enormously in shape, children should learn to discount the particular shape of an object when learning those terms. A role for learning would seem to be crucial, given that some languages do incorporate somewhat more object information than English in their stock of basic spatial terms. For example, the child learning Korean will have to learn the difference between *ahn* and *sok*, corresponding roughly to loose- and tight-fit versions of the English term *in*.

It is possible, of course, that the distinctions between figure and ground geometry, and the kinds of distinctions that appear relevant across all languages are completely unrelated to the facts about structure and processing of objects compared to places. It is also possible that the facts about spatial language derive not from causes external to language, but from the requirements of a communication system that must rapidly convey complex meanings. But if this is true, we are still left with a puzzle of why figure and ground do possess comparatively little fine-grained detail, while the same objects obviously can be and are represented in detail when they are recognized or named as object kinds. From the perspective of learning, it would be reasonable to assume that the possibilities outlined above are all mutually reinforcing. That is, there may exist different systems, based on structure or function, that differentially select information relevant to naming objects and to locating them; the differential representation of shape-based information in these systems may propagate up to the highest level, appearing as differences in the coding of objects in linguistic representations of “what” and “where.”

8.5 Concluding Comments, Remaining Puzzles

More puzzles than answers remain. Although it seems clear that objects can be represented in terms of very different geometric descriptions (for different purposes), it remains unclear just what the status of these descriptions is, with respect to at least four different issues.

First, what is the status of these descriptions with respect to dividing up spatial language? If detailed shape is really a function of the “what” system, whereas coarse shape is a function of the “where” system, then we might see direct repercussions in different portions of spatial language. Objects (usually named by count nouns) preserve detailed shape, and places (more precisely, place-functions, usually named by

spatial prepositions) preserve only coarse or axial descriptions. So far, so good. But can we really connect the object/place representations to different form classes? Even within English, precise shape is encoded in certain verbs (posture verbs such as *to kneel* and *to crouch*, and perhaps manner verbs such as *to undulate* and *to spin*) and axial representation are encoded in spatial adjectives (e.g., *long*, *wide*, *thin*; see Bierwisch, chapter 2, this volume). In other languages, relatively detailed object shape can be encoded in verbs (Japanese positional verbs; see Sinha et al. 1993) and coarse or axial shape can be encoded in classifiers (see, for example, Allan 1977). Should we expect the different shape descriptions to cleave neatly along lines of form class, or along lines of some other distinction such as “what” and “where”? And if so, what do we do with the persistent appearance of the same “coarse” shape descriptors—“round,” “thin,” “long,” “flat”—that show up in classifiers, verbs, and spatial predicates?

A second puzzle concerns the status of these different object descriptions relative to visual representations. Is the three-part division (detailed, coarse, axial) to be found in any principled sense within the visual system? Or does that system give rise to a variety of different descriptions, some of which are selected as “special” by languages?

Third, what is the status of object descriptions relative to representation in the brain? Do the different object descriptions enjoy different status in the “what” and “where” systems, for example? Can we find evidence for the existence of axial and coarse descriptions in one system but not the other? A recent study by Breedin, Saffran, and Coslett (1994; Breedin and Saffran in preparation) may shed some light on this issue, at least with respect to language. One of their patients sustained damage to the inferotemporal lobe and possessed a severe object naming deficit. The deficit was specific to object naming—the patient could recognize objects. Despite the naming deficit, this patient showed no impairment on spatial prepositions, nor on object-part terms, which require labeling the ends of the object axes. Thus the axis-based terms are functionally separate from object names, supporting the functional separation between the detailed and coarse/axial descriptions outlined in this chapter.

Fourth and finally, what is the status of these descriptions as they articulate with learning and development? In this chapter, I have presented evidence suggesting that multiple representations of objects exist early in development, probably prior to language learning. The existence of these different object representations, and the flexible access to them early in life may serve as a critical cornerstone for learning. Discovering precisely how these representations become coordinated with different parts of vocabulary and how they become modified by learning remains a challenge for future research.

Acknowledgment

This work was supported by Social and Behavioral Sciences Award 12-FY93-0723 from the March of Dimes and by National Institutes for Health grant RO1 HD-28675. I thank Paul Bloom and Manish Singh for thoughtful comments on previous versions of the chapter; Jennifer Nolan and Jessie Yim for help preparing figures.

Notes

1. If the flowers are real (rather than painted), then pragmatic constraints would force the interpretation that they are on the upper surface of the bowl. See Herskovits (1986) for discussion of many other contextual constraints.
2. This chapter will focus on spatial prepositions in English. This focus does not entail that spatial information is coded only in these terms. This is clearly not the case, even for English. However, following Talmy (1983), I assume that the closed-class, grammaticized portion of the language is likely to represent the “fine” semantic structure of a language, while the open class (including spatial verbs) may represent a wider range of meanings. Should this assumption prove wrong, the analysis of English spatial prepositions can still provide a framework within which we can build richer theories of the kinds of spatial meanings encoded in languages.
3. The term *across* is described by Talmy (1983) as requiring a “ribbonal” figure and ground object. An experiment by Williams (1991) showed that people judging the acceptability of a display as an instance of *across* found circles intersecting rectangles much less acceptable than ellipses intersecting rectangles. This suggests that the figure must have a clear principal axis (making it a “linear” figure) in order to best satisfy the requirements for this term.
4. It is worth noting that neither children nor adults were simply translating known prepositions. A separate series of questions probed subjects’ generalization patterns for known terms such as *across*; the patterns were not the same as those found in the learning study (see Landau and Stecker 1990 for details).
5. This procedure was modified for the few children who said “yes” only to locations other than the one directly in front of them. Probe trials were conducted using the same span of locations, but with each surrounding the single location most frequently accepted by the child.
6. There are several possible explanations for the sharpening in the shape bias with vocabulary growth. One possibility (described in the text) is that children begin with a representational bias in which objects are represented in terms of shape, and another bias in which object names are linked to object kinds. The function of learning would be just to connect up the two pairs of representations; the sharpening could reflect either a decrease in noise with expanded computational resources (see Landau 1994 for discussion) or an enhancement due to input that reinforces the importance of shape. A second possibility is that both vocabulary growth and the sharpening of the shape bias are a consequence of a third factor, such as the ability to detect which words are count nouns (hence object names). Syntactic growth (with which the child could determine which words are count nouns) has long been thought to be a possible cause of the so-called vocabulary explosion (for discussion, see Landau and Gleitman 1985). A third possibility is that the sharpening of the shape bias is a genuine reflection of the child’s learning that shape matters for object names. These possibilities are currently being tested.

7. I thank Misha Becker for helping collect data on these distinctions.
8. The characterization of Tzeltal as especially "visual" seems unmotivated; most of the shape distinctions it carries can also be represented by other spatial systems, most notably, haptics. I thank Paul Bloom for reminding me of this fact.
9. I thank Manish Singh for illuminating discussion of this issue.

References

- Allan, K. (1977). Classifiers. *Language*, 53(2), 285–311.
- Becker, A. H., and Ward, T. B. (1991). Children's use of shape in extending novel labels to animate objects: Identity versus postural change. *Cognitive Development*, 6, 3–16.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Binford, O. (1971). Visual perception by computer. Paper presented at IEEE Systems, Science, and Cybernetics Conference, Miami.
- Bloom, P. (1994). Possible names: The role of syntax-semantics mappings in the acquisition of nominals. In L. R. Gleitman and B. Landau (Eds.), *Lexical acquisition*. Special volume. *Lingua*, 92, 297–332.
- Bornstein, M. (1985). Color-name versus shape-name learning in young children. *Journal of Child Language*, 12, 387–393.
- Bowerman, M. (1991). The origins of children's spatial semantic categories: Cognitive vs. linguistic determinants. In J. J. Gumperz and S. C. Levinson (Eds.), *Rethinking linguistic relativity*. Cambridge, MA: Cambridge University Press.
- Breedin, S., and Saffran, E. M. (in preparation). Sentence processing in the face of semantic loss: A case study. Manuscript, Temple University.
- Breedin, S., Saffran, E. M., and Coslett, H. B. (1994). Reversal of the concreteness effect with semantic dementia. *Cognitive Neuropsychology*, 11, 617–660.
- Brown, P. (1993). The role of shape in the acquisition of Tzeltal (Mayan) locatives. Paper presented at the 25th Annual Child Language Research Forum. April, Stanford University, Stanford, CA.
- Carlson-Radvansky, L. A., and Irwin, D. (1993). Frames of reference in vision and language: Where is above? *Cognition*, 46(3), 223–244.
- Choi, S., and Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41, 83–122.
- Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and acquisition of language*, 65–110. New York: Academic Press.
- Farah, M., Hammond, K., Levine, D., and Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, 20, 439–462.

- Francis, W. N., and Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Hayward, W., and Tarr, M. (1994). Spatial language and spatial representation. *Cognition*.
- Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions in English*. Cambridge: Cambridge University Press.
- Hill, C. (1975). Variation in the use of front and back in bilingual speakers. In *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*. Berkeley: University of California.
- Hoffman, D. and Richards, W. (1984). Parts of recognition. *Cognition*, 18, 65–96.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Johnston, J. R. (1985). Cognitive prerequisites: The evidence from children learning English. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition*. Vol. 2, *Theoretical issues*, 961–1004. Hillsdale, NJ: Erlbaum.
- Johnston, J. R., and Slobin, D. I. (1978). The development of locative expressions in English, Serbo-Croatian, and Turkish. *Journal of Child Language*, 6, 529–545.
- Jones, S., Smith, L., Landau, B., and Gershkoff-Stowe, L. (1992). On the origins of the shape bias in young children's novel word extensions. Paper presented at the Boston University Language Development Conference, Boston, October.
- Kuczaj, S. and Maratsos, M. (1975). On the acquisition of front, back, and side. *Child Development*, 46, 202–210.
- Landau, B. (1994). Object shape, object name, and object kind. In D. Medin (Ed.), Vol. 31, *Psychology of learning and motivation*, 253–304. New York: Academic Press.
- Landau, B., and Gleitman, L. (1985). *Language and experience*. Cambridge, MA: Harvard University Press.
- Landau, B., and Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–238, 255–265.
- Landau, B., Leyton, M., Lynch, E., and Moore, C. (1992). Rigidity, malleability, object kind, and object naming. Paper presented at the Psychonomics Society, St. Louis, Mo.
- Landau, B., Smith, L., and Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3, 299–321.
- Landau, B., Smith, L., and Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language*, 31, 807–825.
- Landau, B., and Stecker, D. (1990). Objects and places: Syntactic and geometric representations in early lexical learning. *Cognitive Development*, 5, 287–312.
- Levine, D., Warach, J., and Farah, M. (1985). Two visual systems in mental imagery: Dissociation of "what" and "where" in imagery disorders due to bilateral posterior cerebral lesions. *Neurology*, 35, 1010–1018.

- Levine, S. C., and Carey, S. (1982). Up front: The acquisition of a concept and a word. *Journal of Child Language*, 9, 645–657.
- Levinson, S. (1992). Vision, shape, and linguistic description: Tzeltal body-part terminology and object description. Working paper no. 12, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Leyton, M. (1992). *Symmetry, causality, mind*. Cambridge, MA: MIT Press.
- Livingstone, M., and Hubel, D. (1989). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, 240, 740–749.
- Lowe, D. (1985). *Perceptual organization and visual recognition*. Dordrecht: Kluwer.
- Marr, D. (1982). *Vision*. New York: Freeman.
- Marr, D., and Vaina, L. (1982). Representation and recognition of the movement of shapes. *Proceedings of the Royal Society, London*, 214, 501–524.
- Michotte, A. (1963). *The perception of causality*. London: Methuen.
- Miller, G., and Johnson-Laird, P. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Narissiman, B. (1993). The lexical semantics of “length,” “width,” and “height.” Unpublished manuscript. Boston University.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Piaget, J., and Inhelder, B. (1948). *The child's conception of space*. Reprint, New York: Norton, 1967.
- Piaget, J., Inhelder, B., and Szeminska, A. (1960). *The child's conception of geometry*. Reprint, New York: Norton, 1981.
- Rosch, E., Mervis, C., Gray, W., Johnson, D., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.
- Schneider, G. E. (1969). Two visual systems. *Science*, 163, 895–902.
- Schyns, P., and Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science*, 5, 195–200.
- Shiffrar, M. (1994). When what meets where. *Current Directions in Psychological Science*, 3, 96–100.
- Sinha, C., Thorseng, L., Hayashi, M., and Plunkett, K. (1993). Comparative spatial semantics and language acquisition: Evidence from Danish, English, and Japanese. Paper presented at the International Conference on the Psychology of Language and Communication, Glasgow.
- Soja, N., Carey, S., and Spelke, E. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, 38, 179–211.
- Smith, L., Jones, S., and Landau, B. (1992). Count nouns, adjectives, and perceptual properties in children's novel word interpretations. *Developmental Psychology*, 28, 273–286.

- Subrahmanyam, K. (1993). Perceptual processes and syntactic context in the learning of count and mass nouns. Ph.D. Diss., University of California, Los Angeles.
- Talmy, L. (1983). How language structures space. In H. Pick and L. Acredolo (Eds.), *Spatial orientation: Theory, research and application*, 225–282. New York: Plenum Press.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description*. Vol. 3, *Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.
- Tanz, C. (1980). *Studies in the acquisition of deictic terms*. Cambridge: Cambridge University Press.
- Ungerleider, L. G., and Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Eds.), *Analysis of visual behavior*, 549–586. Cambridge, MA: MIT Press.
- Van Essen, D., Anderson, C., and Felleman, D. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, 255, 419–423.
- Williams, P. (1991). Children's and adults' understanding of *across*. Honors thesis, Columbia University.

Chapter 9

Preverbal Representation and Language

Jean M. Mandler

Although my interests lie in the character of the preverbal conceptual system rather than of language itself, the preverbal system forms the foundation on which language rests, and it constrains what is learnable. I shall argue that preverbal conceptual representation is largely spatial in nature and that the relationship between space and language is therefore far-reaching and pervasive. It is not just that spatial terms tell us something about spatial meanings, or that spatial meanings place constraints on spatial terms. It is that many of the most basic meanings that language expresses—both semantic and syntactic—are based on spatial representations. Such a point of view will hardly be news to cognitive linguists such as Ronald Langacker or Leonard Talmy. What I hope to contribute are a few suggestions as to *why* language should be so structured. I will suggest that language is structured in spatially relevant ways because the meaning system of the preverbal language learner is spatially structured. So with apologies to Leonard Talmy for twisting his words, the subtitle of this chapter should read: “How Space Structures Language.”

One further introductory comment. To say that the preverbal meaning system is spatially structured is not to say that it is the same as spatial perception. Rather, spatial information has been redescribed into “meaning packages,” and these meaning packages retain some spatial characteristics. I will argue that some of the categorical or packaging characteristics often ascribed to language itself are actually due to the prepackaging that is accomplished during the preverbal period. Babies do not wait until the onset of language to start thinking; the problem of packaging meanings into workable units is thus a prelinguistic one.

9.1 Sensorimotor Schemas Are Not Concepts

The more I delve into cognition in the first year of life the more it becomes apparent that many of the most basic foundations on which adult concepts rest are laid down during this period. Pace Piaget, the first year of life is far from being an exclusively

sensorimotor stage. Instead, the higher cognitive functions that (among other things) will support language acquisition are being formed in parallel with the sensorimotor learning that is going on. The research that Laraine McDonough and I have been conducting indicates that the foundations of the major conceptual domains are being laid down during this period (Mandler and McDonough 1993). Fundamental concepts of animals and vehicles are learned by around six to seven months (perhaps resting on an even earlier conceptual distinction between animate and inanimate things), and the domains of plants, furniture, and utensils follow soon after. These conceptual domains in turn are used to control inferential reasoning processes (Mandler and McDonough in press). In addition, the episodic memory system has become operational and long-term recall processes have begun (Mandler and McDonough in press). All this is happening before children learn how to speak.

Such findings should give us pause. Where is the familiar sensorimotor infant we are used to hearing about, the creature who has not yet achieved conceptual representation? It seems to have disappeared. In its place we find a baby that has already developed a rich conceptual life. For many people working in language acquisition this will come as no surprise, if for no other reason than the need to account for the complexity of the concepts that newly verbal children express in language. But the current research does make evident a tension that has been lurking in the literature for many years. According to Piaget (1951), babies are not supposed to have a conceptual representational system, yet according to linguists, to learn language requires mapping onto a conceptual base. As a result, we pay lip service to the idea that to learn language requires a preexisting conceptual system, but have avoided specifying what that system is like.

The neglect seems to be due in part to a conflict within Piagetian theory. On the one hand, Piaget (1967) said that conceptual thought is not created by language, but instead thought precedes language, which then transforms it in various ways. On the other hand, because language begins before the sensorimotor period ends, Piaget tended to characterize early verbalizations as just another kind of sensorimotor schema. He did devote a good deal of effort to describing how sensorimotor schemas might be transformed into conceptual (symbolic) representation, but he said little about how the new type of representation differed from the old. The result is a gap in his theory. Sensorimotor schemas are said to be transformed into concepts and concepts are mapped into language, but little is said about what the concepts themselves are like.

As best as I can tell, this dilemma was handled in different ways by people studying language acquisition and those studying cognitive development. Workers in language acquisition attempted to specify the various notions necessary for learning language and then, reasonably enough, left it to the developmental psychologists to explicate

the representational status of these notions. For example (with the exception of the nativist position that grammatical categories are innately given) there seems to be widespread agreement that the underlying concepts needed to learn grammatical categories are notions such as "actionality," "objecthood," "agent," "location," and "possession" (Maratsos 1983). But where the developmental psychologists were to take over, until the recent work on objects and agency began to appear (Baillargeon 1993; Leslie 1984; Spelke et al. 1992), there was largely a blank. Because Piagetian theory was silent about conceptual representation at the end of the sensorimotor period, it seems to have been assumed by default that the relevant conceptual categories were the same as the sensorimotor schemas themselves. Thus in many accounts the sensorimotor achievements were assumed to be the base onto which language is mapped. Typical examples of this approach were the various attempts to relate language acquisition to stage 6 sensorimotor accomplishments, such as object permanence, but these were not very successful (see Nelson and Lucariello 1985 for discussion).

For the most part, sensorimotor schemas are not the right sort of representation for learning language. Piaget provided some of the reasons why procedural forms of representation such as sensorimotor schemas cannot in themselves serve a semiotic function. A sensorimotor schema provides something like meaning in that it enables recognition of previously seen objects to take place, and thus for the world to seem familiar. It also allows each component of a familiar event to signal the next component to come. This kind of reaction is indexical; a conditioned stimulus predicts or "means" that some other event will follow. But a sensorimotor schema does not allow independent access to its parts for purposes of denotation or to enable the baby to think independently of the activation of the schema itself (Karmiloff-Smith 1986). In short, sensorimotor schemas are neither concepts nor symbols, which Piaget considered to be the *sine qua non* for both the development of the higher cognitive functions and language acquisition.

There are other ways in which sensorimotor knowledge also appears to be the wrong sort of base for learning language. Sensorimotor schemas structure perception and control action. These schemas consist of a large number of parameters that monitor continuously varying movements and rapidly shifting perceptual views. How are such schemas to be mapped into a discrete propositional system? Some kind of interface between perception (or action) and language is needed, something that will allow an analog-digital transformation. For example, consider putting a spoon into a bowl. This requires an intricate sequence of movements, but the conceptual system greatly simplifies it, forming a summary of the event that constitutes its meaning. In this case, the meaning might be a representation of one object containing another. It is this conceptual simplification onto which propositional language is mapped, rather than onto the sensorimotor schemas themselves.

9.2 Differences between Perceptual and Conceptual Categories

In addition to Piaget's view that at the end of infancy concepts are constructed out of sensorimotor schemas, there is an even older view of the onset of concept formation, namely, the traditional doctrine of the British empiricists, espoused in modern times by philosophers such as Quine (1977). In this view, which Keil (1991) has called the doctrine of "original sim," before children develop abstract concepts about the world they categorize objects on the basis of their physical appearance according to the laws of perceptual similarity. Once these perceptual categories are formed, various types of information become associated with them, and in so doing these perceptual categories become conceptual in nature.

This associative doctrine of the creation of concepts is exemplified in current theory by the view that the first concepts to be formed are at the basic level (Mervis and Rosch 1981). In this view, babies first form concepts such as dog and cat on the basis of the similarity of the exemplars to each other, and only much later generalize from these concepts to form a superordinate concept of animal. The details of this process have never been worked out, but it would seem to be a process along the lines of the doctrine of original sim. This view is given support by the recent findings of Eimas and Quinn and their colleagues (Eimas and Quinn 1994; Quinn, Eimas, and Rosenkrantz 1993) showing that as young as three months, babies form perceptual categories of animals after a very few exposures to pictures of contrasting classes. For example, both three-olds and six-month-olds quickly learn to distinguish horses from zebras, dogs from cats, and cats from both dogs and lions. It is agreed that these are purely perceptual accomplishments, but Quinn and Eimas (in press) believe, as I assume do many others, that these perceptual categories form the kernel around which the first concepts will develop.

Nevertheless, there are both theoretical and empirical difficulties with this view that have never been resolved. Theoretically, it does not specify in what form the information to be associated with the perceptual categories is itself couched. A property such as barking might be a perceptual category in its own right, and one could imagine how it might become associated with the perceptual category of dog. But it is difficult to understand how properties that are less clearly perceptual are represented, such as "animate" or "interacts with me." More importantly, in my opinion, this approach does not explain how the transition from perceptually based categorization to more abstract or theory-laden concept formation takes place. Indeed, Quinn and Eimas (1986), among others (e.g., Keil 1991), have pointed out that no one taking the traditional empiricist view has ever satisfactorily explained how abstract or superordinate concepts are derived from the perceptual concepts of infants, or how theory-based associations begin to supplant perceptually based ones (see also Fodor 1981).

As long as it was assumed that superordinate concepts such as animal, vehicle, and plant were late acquisitions, this difficulty might be finessed. For example, perhaps language acquisition itself contributes to superordinate concept formation (e.g., Nelson 1985). However, research in our laboratory has shown that infants have formed concepts of animal and vehicle as early as seven months of age (Mandler and McDonough 1993), and other global concepts such as plant are in place at least by eleven months (we have not yet tested younger children on this concept). This research shows that on some tasks infants distinguish global categories *before* they distinguish the basic-level categories nested within the animal class.¹ For example, on our tasks infants differentiate animals and vehicles from seven months onward. But even by eleven months, they do not differentiate dogs and rabbits or dogs and fish.² Furthermore, differentiation among various basic-level classes of mammals, such as dogs and rabbits (and also basic-level classes of land vehicles, such as cars and trucks) is still not well established at eighteen months (Mandler, Bauer, and McDonough 1991).

The details of the development on these conceptual domains is not my main concern here. Rather, I want to emphasize that the development of perceptual categories (which are sensorimotor accomplishments) does not look like the development of conceptual ones. Because most aspects of these two developments have not yet been investigated, specifying the differences between them is still problematic. Nevertheless, several reasons to make the distinction are already known. First, if there were only perceptually based categories in infancy, it would be difficult to explain how infants could manage on any kind of task to categorize two superordinate domains, whose exemplars do not look alike, while failing to categorize the basic-level classes within them, whose exemplars do look alike. The quintessential example of this dilemma is shown by infants in our experiments distinguishing between little models of birds and airplanes, all of which have outstretched wings and therefore very similar overall shapes, while at the same time not distinguishing between dogs and fish or dogs and rabbits, whose shapes are quite different (Mandler and McDonough 1993).³

Second, a purely perceptual account of categorization cannot explain why three- to six-month-old infants are apparently so much more advanced than seven- to eleven-month-olds, in particular, why the younger infants make fine discriminations among basic-level classes that the older infants do not. McDonough and I have suggested that the infants at these different ages are actually engaged in different kinds of processing, even though superficially there seem to be similar task demands in the various experiments that have been conducted. The experiments for both age ranges have used a habituation-dishabituation paradigm. However, the studies of categorization in young infants have measured times to look at pictures, whereas in our work we have measured times to manually explore objects. Apparently, the traditional

looking-time habituation-dishabituation experiments do not engage infants very deeply (Mandler and McDonough 1993); for example, there is often high subject loss in these experiments even when the infants are given something to suck on to keep them awake and happy. On the other hand, when infants are given objects to explore, they show intense interest and concentration and subject loss is virtually nil. Although this issue needs further study, our findings suggest that very young infants begin to perceptually categorize the world in the absence of meaning, but that when they are older and are given a task that engages their interest, a different process is brought to bear. This different process consists of treating objects as kinds of things, that is, as having meaning, not just as things of differing appearance.

This early conceptual processing is crude in comparison to the fine perceptual discriminations that infants make. They appear not yet to have divided the world into very many different kinds, although the kinds they have conceptualized are fundamental cuts that give meaning to the perceptual categories they are also making. That is, the primary meaning to accrue to a basic-level category such as dog is that it is an animal; it is secondary (not only for infants, but adults as well) that dogs are four-legged or bark, or are man's best friend.⁴

I am suggesting that the babies in our experiments can see that dogs look different from fish or rabbits, but do not find these differences important enough to treat them differentially. This situation is essentially the same as when an older child or adult sees the differences in the appearance of poodles and collies, but for most purposes treats them as the same kind of thing, namely, dogs. Babies see the differences in the appearance of dogs and rabbits, but having constructed fewer concepts about the world, for most purposes treat them as the same kind of thing, namely, animals. The question then becomes, exactly what does this initial concept of animal consist of and how is it learned?

Unless one wants to posit that the concept of animal consists of a set of innate ideas, the meanings that make up this concept need to be derived from information that babies can learn from observation alone. By seven months, babies are not yet independently locomoting; they have just begun to handle objects and are still unskilled at doing so. It is also unlikely that most seven-month-olds have held any kind of real animal in their hands. So what kind of information is at their disposal? The first that seems likely to be relevant is biological motion. Bertenthal (1993) has shown that three-month-olds already differentiate biological from nonbiological motion, insofar as the parameters of people's motion are concerned. It seems likely that they do the same for other animals as well because the parameters governing animate motion are quite general. Thus perception of biological versus nonbiological motion is one early source of knowledge that could be used to divide the world into classes of things that move in animate and inanimate ways.

Once these categories of motion are formed they must be characterized in some way, if the difference is not just to remain a sensorimotor distinction but to represent a meaning. One of the ways to do this is to notice that the things that move in the biological way start up on their own, whereas the things that move in the mechanical way start only when another object contacts them. Another characteristic to be noticed is that the things that move in the biological way and start on their own also interact with other objects from a distance, whereas those things that move mechanically and get pushed never interact from a distance. Notice that each of these properties is available even to very young babies. Indeed, these are some of the major properties that babies can pick up when their acuity is still not well developed. Responsivity to these characteristics of motion can explain why babies as young as two months of age respond differentially to people and to dolls (Legerstee 1992). People interact with them; dolls do not. Similarly, it can explain why, by four months, babies differentiate caused motion from self-motion (Leslie 1984).

There are, of course, many other properties of objects that babies observe as well. By four months, babies know that objects are solid, that other objects cannot pass through them, and that objects still exist when they move out of sight (Baillargeon 1993). By six months, babies have learned something about containment; they know that containers must have bottoms if they are to hold things (Kolstad 1991). As young as three months, infants have begun to learn about the properties of object support. They expect an object that loses contact with a surface to fall, unless it is supported by a hand (Baillargeon in press). Slightly older infants expect that any contact implies support, so that various insubstantial objects such as a horizontal finger touching a large box are expected to be sufficient to provide support. By seven months, babies have learned enough about contact and support to predict that something seen to overlap its supporting surface by only about 15% of its base will fall. There are undoubtedly other properties babies learn about before they begin to handle objects themselves, but these are some of the main ones that have been studied to date.

9.3 How Meanings Are Created

Self-starting, biologically moving, mechanically moving, interactive, causing-to-move, caused-to-be-moved, contacting a surface, containing—these are all observable spatial and/or kinetic properties. This is one of the reasons why I have proposed that it is spatial properties (including motion) babies analyze and abstract from perceptual displays to form meanings. I have suggested that as infants are learning to parse the world into objects, a process of perceptual analysis begins to take place (Mandler 1988, 1992). This is an attentive process that occurs when an object is being

thoroughly examined and/or is being compared with something else, unlike the usual sensorimotor processing, which occurs automatically and is typically not under the attentive control of the perceiver. This attentive analysis results in a redescription of the perceptual information being processed. Thus babies have a mechanism that enables them to abstract spatial regularities and to use these abstractions to form the beginnings of a conceptual system. The contents of this new conceptual system are sets of simplified spatial invariants. It is these invariants that form the earliest represented meanings. I claim that these spatial abstractions are sufficient in themselves to represent the initial meanings of such concepts as animate thing, inanimate thing, cause, agent, support, and container. It is not necessary to interact with objects (pick them up, hold them, move them around, or move around them) for meaning to begin to be created, although as infants mature these newfound skills will provide different kinds of information than they received before. But to begin the process, it may take no more than an intelligent eye and a mechanism to transform what the eye observes.⁵

I want to add an aside here, which I hope will clarify the position I have taken with respect to the creation of meaning (Mandler 1992). It is not a nativist position; on the contrary, it is a constructivist account. The mechanism of perceptual analysis I have described makes it unnecessary to posit innate ideas or concepts; perceptual analysis alone can build up meanings and can do so continuously throughout infancy (and for that matter, throughout life). The mechanism itself must be innate, and presumably also the basic aspects of the spatial representations that result from the analysis, but the concepts our minds conceive do not have to be carried on our genes. Thus babies can create a beginning concept of animal even though it is crude compared to the biological theory they will eventually espouse (Carey 1985). New analyses can provide new information at any time, and of course, with the onset of language, a whole new source of accumulating conceptual information arrives on the scene.

Even if we agree that the earliest meanings, such as animal or container, are derived from spatial information, their representational format need not be spatial. After all, I have just described them using language. On the other hand, because the meanings themselves result from spatial analyses, there does not seem to be any good reason to translate them into propositional form. Language will be coming along shortly and babies may not need propositional representations in the interim. Once language is learned, they will be in the advantageous position of having two kinds of representation, one of which is useful for representing continuous and dynamic analog information and the other which provides a way of representing information in a discrete compositional system. Is there any advantage in the meantime to translate spatial representations of something starting up on its own or interacting with

something else from a distance into a list of propositions such as [selfmove (thing)] or [afar (thing1, thing2) + interact (thing1, thing2)]? And how would this be accomplished? Is there a list of empty slots waiting in the mind to be appointed to each successive spatial analysis, so that, say, slot 32a becomes a symbol meaning self-moving, and slot 32b becomes a symbol meaning distant interaction? This is what Harnad (1990) called the symbol-grounding problem. People usually try to solve this problem by saying that the external world provides the meaning for symbols. But neither the external world nor perception of it can provide meaning in and of themselves. The three-month-old who categorizes dog patterns or horse patterns can do so in the absence of meaning, just as an industrial robot can categorize nuts and bolts on the assembly line without meaning entering into its programs at all. Substituting perception for meaning is no different from substituting sensorimotor schemas for concepts. Instead, meaning must come from an analysis of what is perceived. Nothing about such analysis suggests it need consist of propositions composed of discrete symbols.

One reason to translate spatial representations into another format would be if it were needed to learn language. If existing spatial representations were themselves adequate for this purpose then a preverbal propositional representational system would be superfluous. At first glance, spatial representations seem unlikely candidates for the base on which to construct language. Their continuous analog character appears to be subject to some of the same difficulties I described for sensorimotor schemas. How do they get broken down into components that allow language to be mapped onto them? Here is where image-schemas come in. These are the type of spatial representations that I have described as resulting from perceptual analysis (Mandler 1992). They are spatial abstractions of a special kind (Lakoff 1987; Mandler 1992). Image-schemas retain their continuous analog character while at the same time providing some of the desirable characteristics of propositional representations. Although they are not unitary symbols, image-schemas form discrete meaning packages. In addition, they can be combined both sequentially and recursively with other image-schemas. Thus they provide an excellent medium to bridge the transition from prelinguistic to linguistic representation.

9.4 Spatial Representation in the Form of Image-Schemas

Because of the attention that babies give to moving objects, the first image-schemas they form are apt to be those involving movement. The simplest meaning that can be taken from such movement is the image-schema path. This schema represents any object moving on any trajectory through space, without regard to detail either of the object or type of movement. But paths can themselves be analyzed, and as I discussed

earlier, these analyses lead to the concept of animal. For example, focus on the shape of the path itself leads to schemas of animate and inanimate motion. Focus on ways that trajectories begin leads to image-schemas of self-motion and caused-motion, associated with animate and inanimate objects respectively. (This is an example of the embedding nature of image-schemas: beginning-of-path and end-of-path are embedded in path itself). Although I originally called these image-schemas "dynamic" because they can represent continuous change in location, it would have been more accurate to call them "kinetic." As I have defined them, path and its parts are spatial, rather than forceful.

Other types of paths that attract babies' attention are those that go into or out of things, and onto or off surfaces, leading to image-schemas of containment, contact, and support. I have also suggested that perception of contingent motion, or interactions among objects at a distance, can be represented by the notion of coupled paths, or a family of link image-schemas. The link schemas are interesting, not only because they capture one of the ways in which animate objects behave but also because they illustrate how what at first glance seems to be a nonspatial meaning (if A, then B) has an underlying spatial base. In Mandler 1992, I discussed how the link schema that represents the meaning of one animal following another can, by a slight change in its structure, also represent two objects taking turns. This is an example of how a spatial representation can also represent time. It requires mentally following a path, which takes time but which does not require an independent representation of time. It is known, of course, that languages tend to represent time by borrowing spatial terms (e.g., Fillmore 1982; Traugott 1978). I think the reason is that it is easier to think about objects moving along paths than to think about time without any spatial aids.

Because babies are slow information processors and because they probably need a lot of comparisons to carry out any single piece of perceptual analysis, analyzing spatial relations should be easier for them than analyzing temporal relations. One can look back and forth at the various parts of an object or look back at the place where an object began to move. Temporal information is evanescent, and it may be difficult to analyze without the help of previously acquired meanings. If the infant's initial conceptual vocabulary is spatial, the easiest way to handle more difficult conceptualizations would be to use the spatial conceptions that have already been formed. In this view the *concept* of time is not a primitive notion but derived. Of course, to say that conceptualizing time is more difficult than conceptualizing space does not imply that babies are not sensitive to temporal relations; they obviously are. This discussion, however, is concerned with the ability to *think* about time and space and the representations we use to do so. All organisms are sensitive to temporal relations, but most get by without conceptualizing them. When we do think about time, we may

always do so in terms of following a path. Part of path following may include some ineffable sense of duration, but that in itself does not seem to qualify as conceptual.

It is not just time that is more difficult to analyze than space; so are dynamics and internal feelings. Talmy (1985) has suggested that image-schemas are derived from analyzing the forces acting on objects, and Johnson (1987) claims that they are derived from one's bodily experiences. For developmental reasons, however, I have stressed spatial analyses as their source. If image-schemas are to represent preverbal meanings, they must reflect the processing limitations of very young infants. Babies begin their perceptual analyses before they have yet learned to pick up and examine objects; thus many of the action schemas that might be used for purposes of image-schematic analysis have not yet been formed. The processes of image-schema analysis must be already well advanced by the time babies have become adept at manipulating the world, and long before they can move around in it.

In addition, humans are strongly visual creatures, and it should be easier for babies to analyze visual displays (or even for blind babies to analyze displays via touch) than to analyze their internal sensations. There is no evidence on this issue, but it may be noted that we are notoriously bad at introspection even as adults. It is not that babies are unaware of feelings of force or happenings within the container that is their body. But in terms of analysis, one can see the movements of objects, whereas one must typically infer the forces operating on them—and of course one cannot see internal activity at all. It simply has to be more likely that a baby will learn about containers from watching objects go in and out of other objects than from introspecting about the act of eating. This point of view is supported by the widespread phenomenon that the vocabularies of internal states are derived from the vocabularies used to describe external phenomena (e.g., Sweetser 1990). It may be that even as adults the concepts we call "internal states" are at heart spatial analyses, given their internal "flavor" by the gut sensations associated with them. Again, I am talking about *conceptions* of internal states, not the states themselves.

9.5 What is the Evidence That Spatial Analyses Structure Language Learning?

The spatial analyses I have been discussing are particularly important in learning the relational aspects of language, such as the meaning of verbs and grammatical relations. Object labels can and do get mapped ostensively onto the shapes of things, although that does not in itself give them meaning. But young children do have the global preverbal meanings of animal, plant, vehicle, furniture, kitchen utensils (and perhaps many more) at the time they begin to learn object names (Mandler, Bauer, and McDonough 1991). A good deal of what parents teach young children by the

way they name things is to carve these domains into smaller meaning packages. For example, children have the preverbal meaning of animal, and as discussed earlier, they also see the perceptual difference between dogs and cats. Now they hear that this-shaped animal has a different name from that-shaped animal, and, at least in our culture, much is made of the fact that the two kinds of animals make different sounds as well. All this must suggest to children that the difference between cats and dogs may matter. In this way language can help the process of subdividing the initially global concept of animal into subclasses that carry meaning above and beyond their animalness. It is interesting in this regard that in the initial stages of noun learning, children do not particularly rely on shape. But as differential labeling increases over the next few months, they increasingly rely on shape to determine the reference of new nouns (Jones and Smith 1993). Such a finding suggests that children are making the connection between nouns and the perceptual-shape categories they have learned over the course of the first two years.

On the other hand, shape-based perceptual categories such as “dog” and “cat” cannot be used for learning grammar because relations cannot be pointed to in the way that objects can. But the global domain-level concepts such as animal and vehicle that were used to give meaning to these perceptual categories can be used instead. Thus the image-schemas that give the meaning “animate thing” to dog and cat can also be used to frame language overall, to provide the relational notions that allow propositions to be built up. For example, once the meanings are formed for animate objects as things that move themselves and cause other things to move, one has arrived at a simple concept of agent (Mandler 1991). Similarly, once the meanings are formed for inanimate objects as things that do not move by themselves but are caused to move, one has arrived at a simple concept of patient. It may be because the earliest meanings are themselves abstract and relational that abstract and relational notions such as agent and patient can be formed so easily.

Verb acquisition provides concrete examples of this kind of image-schematic underpinning. Golinkoff et al. (1995) discuss in detail how the kinds of image-schemas I have outlined underlie verb learning. The first verbs that children learn all describe paths of various sorts rather than states. The “shapes” of these paths are represented by image-schemas. These specific path schemas are more particular than the paths that differentiate animate from inanimate motion, but are otherwise similar in kind. A typical example is the verb *to fall*, which specifies the direction of the path of motion, but leaves other details aside. This kind of image-schema allows children to ignore the details of a given event and so to generalize from one instance to the next—in short to categorize types of motion.

At a more general level, notions such as animate object, cause-to-move, agent, inanimate object, and caused-to-be-moved are exactly the kind of meanings needed

to master the distinction between transitive and intransitive verb phrases. As Slobin (1985) has discussed, this distinction, abstract though it may be and marked in a variety of ways in different languages, is universally one of the earliest grammatical forms to be acquired. The reason for this is that the ideas expressed in the distinction are among those which preverbal children have universally mastered by the time language begins. English does not mark this distinction with grammatical morphemes, but many languages do and these should be easy for children to learn. For example, Choi and Bowerman (1992) point out that Korean uses different forms for intransitive verbs of self-motion and transitive verbs of caused motion (for example, a causative inflection must be added to "roll" in "He rolled the ball into the box," whereas it is not needed in "The ball rolled into the box"). Korean children respect this distinction as soon as they begin to use these verbs and do not make cross-category errors.

When errors are made in these kinds of grammatical morphemes, they often consist of underextensions. For example, Slobin (1985) found that children first use the morphemes marking transitive verb clauses in the prototypical transitive situation in which an animate agent physically acts on an inanimate object. Only later do they extend the marking to the less prototypical cases in which the agent is inanimate or the patient is animate. This kind of underextension suggests that children may try a fairly direct mapping of the language they hear onto their already-formed conceptualizations. Of course, languages do not always cooperate, and some distinctions seem likely to give language learners trouble.

This raises the old Whorfian issue of the extent to which language is mapped onto preexisting concepts or by its own structure leads children to create new ones. I will illustrate this issue with the case of learning spatial prepositions. Let me say at the outset that because we all agree that language is to some degree mapped onto existing concepts, we are only haggling over the details. But one of those important details is the following. Have preverbal children learned all the major spatial relations that various languages express? Or have they learned only a subset and do languages teach them to attend to new ones they have not analyzed on their own?

Melissa Bowerman and I have discussed this issue quite a bit, although I am not sure whether we have agreed, or merely agreed to disagree. The particular issue involves the notions of containment, contact, and support. As Bowerman (1989) has discussed, the languages of the world divide up these relations in various ways, and furthermore do so by a variety of constructions. English, for example, makes a single general distinction between containment and support by means of the prepositions *in* and *on*, with contact being ignored. I have claimed that containment and support are among the first image-schemas to be formed; because they match the English prepositional system in a straightforward fashion, it is not surprising they are

the earliest grammatical morphemes to be learned, and are learned virtually without error (Mandler 1992).⁶ These morphemes are very frequent in adult speech, they capture a well-understood conceptual distinction, they are easy to say, and so forth.

Although containment and support sound like universal spatial primitive, Bowerman (1989) suggests that this may be a somewhat provincial view. Some languages make no distinction at all (as in Spanish *en*), and others make a three-way distinction. Furthermore, various languages make the distinctions they do make by cutting the spatial pie up in different ways. For example, German divides support relations into two, depending on whether the support is horizontal or vertical. Dutch makes a similar split but apparently uses the method of attachment to categorize the support relation, rather than the horizontal and vertical. In either language, difficult cases can appear, such as how to express that a fly is on the ceiling. Upside-down support is an unusual support relation, and one might predict that it would give young language learners trouble.⁷

Developmental psychologists have only recently begun to explore in depth the development of concepts of containment, contact, and support in preverbal infants, but the work of Baillargeon and her colleagues described earlier (e.g., Baillargeon, 1995) tells us that a great deal of detailed knowledge is accumulating in the first year. Babies apparently start with quite simple image-schemas but rapidly learn conceptual variations on these, including containment with and without contact, horizontal versus vertical support, and so forth. The data suggest that a wide variety of these conceptual notions are well established before language begins. What remains to be done is to repackage these meanings *linguistically*. Perhaps because the conceptual notions are meanings and cannot be pointed to, or perhaps just because of their abstractness, different languages repackage them in various ways (Gentner 1982), ways babies must learn by listening to their native tongue.

If the native tongue is a prepositional one, it will express a quite limited subset of spatial distinctions (Landau and Jackendoff 1993), typically making binary or trinary distinctions in relations such as containment, contact, and support. The distinctions are few enough that they should pose few problems to the language learner who comes equipped with many such preverbal meanings. There are ways to express space that are limited by other principles, however. One way is to use body parts, as in Mixtec; for example, instead of saying, "The cat is under the table," in Mixtec one would say, "The cat be-located belly-table" (Brugman 1988). The system is still spatial but ignores one set of relationships (such as containment) and instead expresses a different set (relative locations vis à vis a human or animal body). Of course, body parts are well known to the young language learner; indeed, naming body parts is a common game among parents and newly verbal children, at least in our culture. This method of linguistically partitioning space should therefore not give children trouble.

Other languages use verbs to express some of the relationships that English describes by means of prepositions. In Korean, for example, entirely different morphemes are used to express relationships of *put into*, *take out*, *put onto*, and *take off*. Furthermore, the morphemes are different for *put into tightly* versus *put into loosely*, and for putting clothes on the trunk, putting clothes on limbs, and so forth. Essentially what Korean does is to distinguish between containment and support when these relations involve loose contact, but override containment and support when tight-fitting contact occurs. It is as if the language says that if the relationship is tight-fitting both containment and support apply in equal measure so that only the type of contact will be specified.

This set of semantic categories, combined with their expression in separate verb forms means that Korean children cannot get by in the early stages of communication by widespread use of a few all-purpose prepositions such as *in* or *out* to express these relations. On the other hand, they learn the morphemes just described early and effortlessly, just as English-speaking children learn a small set of prepositions to express similar meanings. English-speaking children, of course, do not say *fit together tightly* or *put in loosely* because those ideas are not expressed by single morphemes in English. The question is whether English-speaking children already understand these particular spatial distinctions and are silent about them because of a lack in their language, or whether they do not form the relevant image-schematic meanings until the language directs them to do so.

We are back to our Whorfian issue, but we have turned it into a manageable empirical question, and Bowerman, Choi, McDonough, and I are engaged in an experimental attempt to answer it. I am not sure if we have different predictions or not. I believe that babies have had ample experience of clothes fitting tightly or of the difficulty of separating pop beads to have formed a concept of tight-fitting. Therefore, I predict we will be able to show this distinction in preverbal children. The fact that Korean children sometimes overgeneralize the tight-fitting relation to the case of clothing (Korean uses a different word for putting on clothing), indicates to me the presence of a preverbal notion (as does the more general fact that the common errors children make in learning one language are often the correct expressions of another).

We still know relatively little about the age at which these various spatial analyses begin to be made. In addition, we do not yet have good estimates of the amount of language-specific learning that takes place before word production begins. If these two factors interact, it may be difficult to disentangle their relative importance. Nevertheless, a few simple principles can be surmised. First, if a language does not make a given distinction that a preverbal baby has conceptualized, this will not cause a language-learning problem. Babies will be willing to overlook this lack of sensitivity. Second, if the language makes a distinction that the baby has already learned, that

will also not cause a problem, whether the distinction is expressed by a preposition or verb (given equal salience in the speech stream). Third, difficulty will occur only when the language makes a distinction that the baby has *not* made prelinguistically. If the baby has no conception at all of the meaning of such a morpheme or construction, it should be a very late acquisition indeed. A more common situation is likely to be one in which the morpheme excludes one of the possible and likely meanings in question. A possible example is an error Korean children sometimes make in expressing the tight-fittingness of a flat magnet on a refrigerator door (the verb for *fitting tightly* has to do with three-dimensional objects, and the status of a flat magnet is not entirely clear). The presence of such errors does not necessarily mean that the language is teaching a new relationship, only that the situations described are unusual or atypical vis à vis the particular semantic cut that the language has made.

One of the points I have made about image-schema representations of space is that they have already been simplified and schematized; they have already filtered out a great deal of the information the perceptual system takes in. Language may do some of this kind of work, as Landau and Jackendoff (1993) have hypothesized, but it seems likely to me that much of it has already been done before language is learned. Infants have been analyzing spatial relations for many months. If these spatial relations are represented in terms of image-schemas a lot of the analog-to-digital transformation needed for language learning has already been done. The result is a set of meaning packages that language can put together in a variety of ways, ignoring some, emphasizing others. At the same time, no matter what the language, the number of distinctions needed to learn the spatial pronouns and/or verbs children acquire in their first year of language is quite small, involving such notions as inside-outside, contact–no contact, horizontal-vertical, up-down, tight-loose. The language itself can help children learn the more complex relationships they master at later stages by directing perceptual analysis to aspects of stimuli they may not yet have noticed.

I will close by reiterating the importance of the conceptual level of representation to understanding language acquisition. I worry that in too many accounts language is talked about as if it were mapped onto actions or onto perception. This is a common approach in connectionist paradigms, for example. Instead, language is mapped onto a meaning system that forms an interface between analog and digital forms. This interface, which shares some of the properties of both forms, is what enables a propositional representational system to be added to the baby's repertoire.

Acknowledgment

Preparation of this chapter was supported in part by National Science Foundation research grant DBS92-21867.

Notes

1. We use the term *global* for these concepts because it does not seem correct to speak of a superordinate concept if it is not yet differentiated into subconcepts (Mandler, Bauer, and McDonough 1991).
2. Infants in our experiments do make more distinctions within the vehicle domain during this age range.
3. Domain-level categorization raises the issue of how infants identify as animals little models they have never seen before, such as a model elephant. We do not yet know which features seven-month-olds are using to identify the correct domain. We have suggested that once infants have begun to analyze object movement, it directs their attention to the parts associated with motion (Mandler and McDonough 1993). This may be why infants are sensitive to what seem (to us) like very small differences between the outstretched wings of the birds and airplanes in our experiments. They do not appear to be using face information because some of our planes are Flying Tigers with faces painted on them, and some of the bird faces do not show eyes. They might be using textural information, although texture cues are minimized in our plastic models. Whether shape or texture, however, a solely perceptual account has difficulty in explaining the shifts in use of one kind of perceptual cue to another when categorizing at the basic or global level.
4. It may be of interest that in various forms of meaning breakdown (semantic dementia), the most resilient aspect of knowledge about an object such as a dog is that it is an animal. Even when patients can no longer recognize the word *dog* or a picture of a dog or say anything specific about dog, they can often still say that it is an animal (Saffron and Schwartz 1994).
5. In the case of blind infants, an exploring hand is required instead (Landau 1988).
6. Only the present progressive, *-ing*, which expresses another preverbal image-schema, traversal of a path, is learned earlier; see Brown (1973).
7. We also must not forget the arbitrary aspects of language that arise from historical accident or for other reasons. These are more frequent than we sometimes realize. For example, in London one sees signs in the Underground saying "No Smoking Anywhere on This Station," which sounds distinctly odd to American ears, but of course perfectly fine to the British. I assume that the British expression can be traced to the fact that railway stations originally consisted of raised platforms, but the example is typical of the many arbitrary aspects of language that children must learn.

References

- Baillargeon, R. (1993). The object concept revisited: New directions in the investigation of infants' physical knowledge. In C. Granrud (Ed.), *Visual perception and cognition in infancy*, 265–315. Hillsdale, NJ: Erlbaum.
- Baillargeon, R. (1995). A model of physical reasoning in infancy. In C. Rovee-Collier and L. Lipsitt (Eds.), *Advances in infancy research*, vol. 9. Norwood, NJ: Ablex.

- Bertenthal, B. (1993). Infants' perception of biomechanical motions: Intrinsic image and knowledge-based constraints. In C. Granrud (Ed.), *Visual perception and cognition in infancy*, 175–214. Hillsdale, NJ: Erlbaum.
- Bowerman, M. (1989). Learning a semantic system: What roles do cognitive predispositions play? In M. L. Rice and R. L. Schiefelbusch (Eds.), *The teachability of language*, 133–169. Baltimore: P. H. Brookes.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Brugman, C. M. (1988). *The story of over: Polysemy, semantics, and the structure of the lexicon*. New York: Garland.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Choi, S., and Bowerman, M. (1992). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41, 83–121.
- Eimas, P. D., and Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, 65, 903–917.
- Fillmore, C. (1982). Toward a descriptive framework for spatial deixis. In R. J. Jarvella and W. Klein (Eds.), *Speech, place, and actions*. New York: Wiley.
- Fodor, J. (1981). *Representations*. Cambridge, MA: MIT Press.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), *Language development*. Vol. 2, *Language, thought, and culture*. Hillsdale, NJ: Erlbaum.
- Golinkoff, R. M., Hirsh-Pasek, K., Mervis, C. B., Frawley, W. B., and Parillo, M. (1995). Lexical principles can be extended to the acquisition of verbs. In M. Tomasello and W. Merriam (Eds.), *Beyond names for things: Young children's acquisition of verbs*, 185–221. Hillsdale, NJ: Erlbaum.
- Harnard, S. (1990). The symbol-grounding problem. *Physica D*, 42, 335–346.
- Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reasoning*. Chicago: University of Chicago Press.
- Jones, S. S., and Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development*, 8, 113–139.
- Karmiloff-Smith, A. (1986). From metaprocesses to conscious access; Evidence from children's metalinguistic and repair data. *Cognition*, 23, 95–147.
- Keil, F. C. (1991). The emergence of theoretical beliefs as constraints on concepts. In S. Carey and R. Gelman (Eds.), *The epigenesis of mind*, 237–256. Hillsdale, NJ: Erlbaum.
- Kolstad, V. T. (1991). Understanding of containment in 5.5-month-old infants. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Seattle, April.
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

- Landau, B. (1988). The construction and use of spatial knowledge in blind and sighted children. In J. Stiles-Davis, M. Kritchevsky, and U. Bellugi (Eds.), *Spatial cognition: Brain bases and development*, 343–371. Hillsdale, NJ: Erlbaum.
- Landau, B., and Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–265.
- Legerstee, M. (1992). A review of the animate-inanimate distinction in infancy: Implications for models of social and cognitive knowing. *Early Development and Parenting*, 1, 59–67.
- Leslie, A. (1984). Infant perception of a manual pick-up event. *British Journal of Developmental Psychology*, 2, 19–32.
- Mandler, J. M. (1988). How to build a baby: On the development of an accessible representational system. *Cognitive Development*, 3, 113–136.
- Mandler, J. M. (1991). Prelinguistic primitives. In L. A. Sutton and C. Johnson (Eds.), *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, 414–425. Berkeley, CA: Berkeley Linguistics Society.
- Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, 99, 587–604.
- Mandler, J. M., Bauer, P. J., and McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, 23, 263–298.
- Mandler, J. M., and McDonough, L. (1993). Concept formation in infancy. *Cognitive Development*, 8, 291–318.
- Mandler, J. M., and McDonough, L. (in press). Drinking and driving don’t mix: Inductive generalization in infancy. *Cognition*.
- Mandler, J. M., and McDonough, L. (in press). Nonverbal recall. In N. L. Stein, P. O. Ornstein, B. Tversky, and C. Brainerd (Eds.), *Memory for everyday and emotional events*, Hillsdale, NJ: Erlbaum.
- Maratsos, M. (1983). Some current issues in the study of the acquisition of grammar. In J. H. Flavell and E. M. Markman (Eds.), *Cognitive development*, Vol. 3 of P. H. Mussen (Ed.), *Handbook of child psychology*. New York: Wiley.
- Mervis, C. B., and Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32, 89–115.
- Nelson, K. (1985). *Making sense: The acquisition of shared meaning*. San Diego, CA: Academic Press.
- Nelson, K., and Lucariello, J. (1985). The development of meaning in first words. In M. Barrett (Ed.), *Children’s single-word speech*, New York: Wiley.
- Piaget, J. (1951). *Play, dreams, and imitation in childhood*. New York: Norton.
- Piaget, J. (1967). *Six psychological studies*. New York: Random House.
- Quine, W. V. (1977). Natural kinds. In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds*, 155–177. Ithaca, NY: Cornell University Press.

- Quinn, P. C., and Eimas, P. D. (1986). On categorization in early infancy. *Merrill-Palmer Quarterly*, 32, 331–363.
- Quinn, P. C., Eimas, P. D., and Rosenkrantz, S. L. (1993). Evidence for representations of perceptual similar natural categories by 3-month-old and 4-month-old infants. *Perception*, 22, 463–475.
- Quinn, P. C., and Eimas, P. D. (in press). Perceptual organization and categorization in young infants. In C. Rovee-Collier and L. Lipsitt (Eds.), *Advances in Infancy Research*, Vol. 11. Norwood, NJ: Ablex.
- Saffron, E. M., and Schwartz, M. F. (1994). Of cabbages and things: Semantic memory from a neuropsychological perspective—A tutorial review. In C. Umiltà and M. Moscovitch (Eds.), *Attention and performance XV: Conscious and unconscious information processing*, Cambridge, MA: MIT Press.
- Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition*, Vol. 2, *Theoretical issues*, 1157–1256. Hillsdale, NJ: Erlbaum.
- Spelke, E. S., Breinlinger, K., Macomber, J., and Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 99, 605–632.
- Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.
- Talmy, L. (1985). Force dynamics in language and thought. In W. H. Eilfort, P. D. Kroeber, and K. L. Peterson (Eds.), *Papers from the Parasession on Causatives and Agentivity at the Twenty-first Regional Meeting*, Chicago: Chicago Linguistic Society.
- Traugott, E. C. (1978). On the expression of spatiotemporal relations in language. In J. H. Greenberg (Ed.), *Universals of human language*. Vol. 3, *Word structure*, Stanford, CA: Stanford University Press.

Chapter 10

Learning How to Structure Space for Language: A Crosslinguistic Perspective

Melissa Bowerman

Space is an important preoccupation of young children. From birth on, infants explore the spatial properties of their environment, at first visually and proprioceptively, and then through action. With improved motor control during the second year of life, their spatial explorations become more complex, and they also begin to talk about space. Early comments on space revolve mostly around motions, with remarks about static position also beginning to appear in the second half of the second year. The following utterances from a nineteen-month-old girl learning English are typical:

- (1) a. *In*. (About to climb from the grocery compartment of a shopping cart into the child seat.)
b. Monies. *In*. (Looking under couch cushions in search of coins she has just put down the crack between the cushions.)
c. Balls. *Out*. (Trying to push round porthole pieces out of a foam boat puzzle.)
d. Books. *Out*. Books. *Back*. (Taking tiny books out of a fitted case and putting them back in.)
e. Monkey *up*. (After seeing a live monkey on TV jump up on a couch.)
f. *Down*. *Drop!* (After a toy falls off the couch where she is sitting.)
g. *On*. (Fingering a piece of cellophane tape that she finds stuck on the back of her highchair.)
h. *Off*. (Pushing her mother's hand off the paper she is drawing on.)
i. *Open* mommy. (Wants adult to straighten out a tiny flexible mommy doll whose legs are bent up).¹

Remarks like these attract little attention—the view of space they reflect is obvious to adult speakers of English. But their seeming simplicity is deceptive: on closer inspection, these little utterances raise fundamental and difficult questions about the relationship between the nonlinguistic development of spatial understanding and the acquisition of spatial language. How do children come to analyze complex events and relationships, often involving novel objects in novel configurations, into a set of

discrete spatial categories suitable for labeling? How do they decide which situations are similar enough to be referred to by the same word (e.g., the two *ins* above, and the two *outs*)? Why is their choice of spatial word occasionally odd from the adult point of view (e.g., *open* for unbending a doll)—and yet, at the same time, why is it so often appropriate?

For many years it has been widely assumed that the meanings children assign to spatial words reflect spatial concepts that arise in the infant independently of language, under the guidance of both built-in perceptual sensitivities and explorations with the spatial properties of objects (e.g., Johnston and Slobin 1979; McCune-Nicholich 1981; Slobin 1973). For example, the words *in* and *out* in the examples above might label preverbally compiled notions to do with containment, *on* and *off*, notions of contact and support; and *up* and *down*, notions of motion oriented with respect to the vertical dimension.

This view is buttressed by an impressive array of research findings with infants: for instance, toddlers clearly know a lot about spatial relationships before they begin to talk about them. It also draws support from studies that stress the existence of perceptual and environmental constraints on spatial cognition and that postulate a close correspondence between the nonlinguistic and linguistic structuring of space (e.g., Bierwisch 1967; H. H. Clark 1973; Miller and Johnson-Laird 1976; Olson and Bialystok 1983). In this view the similarity between child and adult use of spatial morphemes is not surprising: the properties of human perception and cognition mold both the meanings that languages encode and the spatial notions that speakers of all ages entertain.

I will argue that the path from a nonlinguistic understanding of spatial situations to knowledge of the meanings of spatial morphemes in any particular language is far less direct than this view suggests. The meanings spatial morphemes can express are undoubtedly constrained (e.g., Landau and Jackendoff 1993; Talmy 1983), but recent research is beginning to uncover striking differences in the way space is structured for purposes of linguistic expression (see also Levinson, chapter 4, this volume). To the extent that languages differ, nonlinguistic spatial development alone cannot be counted on to provide children with the *conceptual packaging* of space they need for their native language. Whatever form children's nonlinguistic spatial understanding may take, this understanding must be applied to the task of discovering how space is organized in the local language. Although the interaction in development between nonlinguistic and linguistic sources of spatial structuring is still poorly understood, recent crosslinguistic work suggests that the linguistic input begins to influence the child at a remarkably young age: for instance, the child whose utterances are shown above is barely more than a year and a half old, but her utterances already reflect a

profoundly language-specific spatial organization (Bowerman 1994, 1996; Choi and Bowerman 1991).

I first review studies suggesting that nonlinguistic spatial development indeed lays an important foundation for the child's acquisition of spatial words. But this is not enough: Next I discuss the problem created for learners by the existence of crosslinguistic differences in the way space is carved up into categories, and review some other aspects of spatial structuring that clearly must be *learned* on the basis of linguistic experience. After this stage setting, I describe two studies I have conducted, together with Soonja Choi, to explore how children who are learning languages that classify space in interestingly different ways arrive at the spatial categories of their language. Finally, I consider what these studies suggest about the interaction between nonlinguistic and linguistic factors in the acquisition of spatial semantic categories, and about the kinds of hypotheses children may bring to the acquisition of spatial words.

10.1 Cognitive Underpinnings of Spatial Semantic Development

If any domain has a plausible claim to strong language-independent perceptual and cognitive organization, it is space. The ability to perceive and interpret spatial relationships is clearly fundamental to human activity, and it is supported by vision and other highly structured biological systems (e.g., DeValois and DeValois 1990; von der Heydt, Peterhans, and Baumgartner 1984). Our mental representations of space are constrained not only by our biology but also by their fit to the world "out there": if we try to set an object down in midair, it falls, and if we misrepresent the location of something, we cannot find it later. Little wonder it has seemed likely to many investigators that the language of space closely mirrors the contours of nonlinguistic spatial understanding. Several kinds of empirical evidence indeed support the assumption that children know a great deal about space before they can talk about it, and that they draw on this knowledge in acquiring spatial words.

10.1.1 Prelinguistic Spatial Knowledge

10.1.1.1 Piagetian Theory: Building Spatial Representations through Action The original impetus for the modern-day hypothesis that children map spatial words onto preestablished spatial concepts came from the striking fit between Piaget's arguments about the construction of spatial knowledge in young children and the course of acquisition of spatial words.² According to Piaget and Inhelder (1956), spatial concepts do not directly reflect the perception of space but are built up on the level of

representation through the child's locomotion and actions upon objects during the first eighteen months or so of life. "The earliest spatial notions are thus closely bound to object functions such as containment or support, and to the child's concern with object permanence. Recall here the toddler's pleasure with pots and pans, towers and hiding games. In the next phase, children construct the spatial notions of proximity, separation, surrounding and order" (Johnston 1985, 969). After the emergence of these notions—often called "topological" because they do not involve perspective or measurement—projective and Euclidean spatial notions are gradually constructed.

This order is closely mirrored by the sequence in which children acquire locative morphemes such as the English prepositions. Locatives begin to come in during the second year of life, but their acquisition is a drawn-out affair. Within and across languages, they are acquired in a similar order: first come words for functional and topological notions of containment (*in*), support and contiguity (*on*), and occlusion (*under*); then for notions of proximity (*next to*, *beside*, *between*), and finally for relationships involving projective order (*in front of* and *in back of/behind*). This protracted and consistent order of acquisition of locatives, coupled with its correspondence to Piaget's claims about the course of development of spatial knowledge, has been taken as strong evidence that the learning of locatives is guided and paced by the maturation of the relevant spatial notions (Johnston 1985; Johnston and Slobin 1979; Parisi and Antinucci 1970; Slobin 1973).

10.1.1.2 Infant Spatial Perception With the explosion over the last decade of research on infant perception, the evidence for prelinguistic spatial concepts has become steadily more impressive. Challenging Piaget's emphasis on the critical role of action in the construction of spatial concepts, studies show that even very young infants are sensitive to many spatial and other physical properties of their environment. For example, habituation studies of infant perception have established that within the first few days or months of life, infants can distinguish between scenes and categorize them on the basis of spatial information such as above-below (Antell and Caron 1985; Quinn 1994), left-right (Quinn and Eimas 1986; Behl-Chadha and Eimas 1995), and different orientations of an object (Bomba 1984; Quinn and Bomba 1986; Colombo et al. 1984). Studies using the related technique of time spent looking at possible versus impossible events show that by a few months of age infants also recognize that objects continue to exist even when they are out of sight (Baillargeon 1986, 1987), that moving objects must follow a continuous trajectory and cannot pass through one another (Spelke et al. 1992), and that objects deposited in midair will fall (Needham and Baillargeon 1993).

The proper interpretation of such findings is still a matter of debate. Some researchers argue that children can represent and reason about the physical world with

core knowledge that is derived from neither action nor perception, but is inborn (e.g., Spelke et al. 1992; Spelke et al. 1994). Others argue instead for “highly constrained learning mechanisms that enable babies to quickly arrive at important generalizations about objects” (Needham and Baillargeon 1993, 145) or for powerful abilities to detect perceptual invariances in stimulus information (Gibson 1982). In any event, there can be little doubt that even babies well under a year of age command a formidable set of spatial abilities.

10.1.1.3 Temporal Priority of Nonlinguistic over Linguistic Spatial Knowledge Consistent with this, whenever children’s nonlinguistic understanding of particular aspects of space has been directly compared with their knowledge of relevant spatial words, an advantage is found for nonlinguistic understanding. For example, Levine and Carey (1982) found that children can successfully distinguish the fronts and backs of objects such as dolls, shoes, chairs, and stoves—as demonstrated, for example, by their ability to orient them appropriately to form a parade—well before they can pick out these regions in response to the words *front* and *back* (see also Johnston 1984, 1985 for a related study). Similarly, E. V. Clark (1973a) found that young children play with objects in ways that show an understanding of the notions of containment and support before they learn the words *in* and *on* (see also Freeman, Lloyd, and Sinha 1980).

10.1.2 Reliance on Nonlinguistic Spatial Knowledge in Learning New Spatial Words Not only do children show a grasp of a variety of spatial notions before they can talk about them, but they also seem to draw on this knowledge in learning new spatial words. Young children often show signs of wanting to communicate about the location of objects, and before acquiring spatial morphemes, they may do so simply by combining two nouns or a verb and a noun with what seems to be a locative intention, for example, “towel bed” for a towel *on* a bed, and “sit pool” for sitting *in* a wading pool (Bloom 1970; Bowerman 1973; Slobin 1973). The prepositions most often called for but usually missing in the speech of R. W. Brown’s (1973) three subjects were *in* and *on*. At a later stage, these were the first two prepositions to be reliably supplied. This pattern has suggested to researchers that the motor driving the acquisition of locative morphemes is the desire to communicate locative meanings that are already conceptualized (e.g., Slobin 1973).

10.1.2.1 Strategies for Interpreting Spatial Words Children’s nonlinguistic spatial notions also affect how they interpret spatial words in the speech they hear. For example, in an experiment assessing how children comply with instructions to place object A *in*, *on*, or *under* object B, E. V. Clark (1973a) found that her youngest

subjects put A 'in' if B was container-shaped, and 'on' if B had a flat, supporting surface, regardless of the preposition mentioned. This meant that they were almost always correct with *in*, correct with *on* unless B was a container, and never correct with *under*. Clark proposed that prepositions whose meanings accord with learners' nonlinguistic spatial strategies are acquired before prepositions whose meanings do not; hence, *in* is easier than *on*, which in turn is easier than *under*.

10.1.2.2 Rapid Generalization of Newly Learned Spatial Words Several authors have observed that children's earliest spatial morphemes extend rapidly over a wide range of referent situations that share an abstract spatial similarity. For example, reporting that a twelve-month-old child extended *up* on the first day of use "to all vertical movement of the child himself or of objects," Nelson (1974, 281) proposed that "there is a core representation of this action concept . . . something like Vertical Movement." Similarly, Bloom (1973, 29) concluded that the use of *up* by Leopold's (1939) daughter Hildegard in connection with objects and people, including herself, "is a function of the underlying conceptual notion itself." On the basis of data from her two subjects, Gruendel (1977) concurred that "'upness' is an early-cognized or conceptualized relation" and added that *in* also "appeared from the outset to take a readily generalizable form, suggesting that meaning relations had been articulated before production began." In studying relational words in the one-word stage speech of five children, McCune-Nicholich (1981) found that *up*, *down*, *back*, and *open*, along with several other relational words, came in abruptly, generalized rapidly, and were less likely to be imitated than other words. She concluded from this that the words encode preestablished cognitive categories—specifically, operative knowledge of the late sensorimotor period.

10.1.2.3 Underextensions and Overextensions Further evidence that children draw on their nonlinguistic spatial conceptions in acquiring spatial words is that they sometimes apply the words to a range of referents that differs systematically from the adult range. For example, English-speaking children first use *behind* and *in front of* only in connection with things located behind or in front of their own body; the intended meanings seem to be "inaccessible and/or hidden" versus "visible." Later *behind* is also used when a smaller object is next to and obscured by a larger one (*under* is also sometimes inappropriately extended to these situations). Still later, *behind* and *in front of* are also produced when an object is adjacent to the back or front of a featured object such as a doll. Finally they are also used projectively to mean "second/first in the line of sight" (Johnston 1984). According to Johnston, "when we see locative meanings change over many months in a specific, predictable fashion, we are invited to assume that new spatial knowledge is prompting growth"

(p. 421). Another example of nonadultlike usage is the common overextension of the verb *open* to actions like pulling apart paper cups or Frisbees, unlacing shoes, taking a piece out of a jigsaw puzzle, and pulling a chair out from a table (Bowerman 1978; E. V. Clark 1993; see also Griffiths and Atkinson 1978). Nonadultlike uses, whether restricted or overextended relative to adult norms, have been interpreted as strong evidence for children's reliance on their own language-independent spatial notions.

The literature just reviewed establishes that infants understand a great deal about space before they acquire spatial words, that they learn spatial words in a consistent order roughly mirroring the order in which they come to understand the relationships the words encode, and that they rely on their spatial understanding in learning new words—for example, in making predictions about what these words could mean and in extending them to novel situations. There can be little doubt, then, that non-linguistic spatial development plays an important role in children's acquisition of spatial morphemes. But does the evidence establish that children map spatial words directly onto spatial concepts that are already in place? Here there is still room for doubt.

10.2 Does Language Input Play a Role in Children's Semantic Structuring of Space?

In a dissenting view, Gopnik (1980; Gopnik and Meltzoff 1986) has argued that early spatial words do not in fact express simple spatial concepts that are already thoroughly understood, but, rather, ones that are emerging and still "problematic" for children of about eighteen months. She notes that although by about twelve to fourteen months children show an interest in how objects fall and can be balanced, and in the properties of containers, there is evidence that even fifteen- to twenty-one-month-olds do not fully understand gravity and movement into and out of containers. For instance, until seventeen months Piaget's (1954) daughter Jacqueline *threw* objects to the ground rather than dropping them, and at fifteen months she was still trying to put a larger cup into a smaller one. Gopnik (1980) suggests that language may in fact help children solve spatial puzzles during the one-word stage—for example, hearing adults say "up" and "down" in connection with their experiments with gravity "may help [children] to understand that all these preliminary actions lead to the same consequence" (p. 291).

How can we reconcile Gopnik's hypothesis that eighteen-month-olds learn words for spatial concepts that are still problematic for them with evidence that much younger babies have a relatively sophisticated perceptual understanding of space? To explain the discrepancy between what infants seem able to perceive and how they act upon objects (or do not act—cf. infants' failure to search for hidden objects despite evidence they remember the existence and location of these objects; see Baillargeon

et al. 1990), some researchers have suggested that core knowledge of the physical properties of objects and their relationships is modular, and at first somewhat inaccessible to other domains of child thought and action (Spelke et al. 1994). Others point to early limitations in problem-solving skills. In order to successfully manipulate space, children not only must have spatial knowledge but also be able to devise and execute a situation-appropriate plan, and this often appears to be difficult for reasons independent of the actor's spatial understanding (Baillargeon et al. 1990).

For some spatial notions, however, there is reason to suspect that despite evidence for some early perceptual sensitivity, understanding may still be incomplete until eighteen months of age or beyond (see also Gopnik 1988). For example, by as early as six months, babies anticipate that an opening in the surface of an object allows a second, smaller object to pass through (Sitskoorn and Smitsman 1995; see also Pieraut-Le Bonniec 1987). But it is not until about seventeen to twenty months that they seem to recognize that in order to contain something, a container must have a bottom. Only at this age do they (1) look longer at an impossible event in which a bottomless cylinder seems to contain sand than at a possible event with an intact cylinder, and (2) choose with more than chance frequency an intact cup over a bottomless cup when encouraged to imitate an action of putting cubes in a cup and rattling them (Caron, Caron, and Antell 1988; see also Bower 1982, and MacLean and Schuler 1989). Similarly, although by four to six months infants recognize that an object cannot stay in midair without any support at all (Needham and Baillargeon 1993; Sitskoorn and Smitsman 1995; Spelke et al. 1992), even toddlers as old as thirty months are not surprised when a block construction stays in place after one of its two critical supporting blocks is removed (Keil 1979).

These findings are consistent with Gopnik's proposal that toddlers talk about spatial events whose properties they are still in the process of mastering, and lend some plausibility to her suggestion that linguistic input—hearing adults use the same word across a range of situations that are in some way similar—may contribute to the process of mastery. But although Gopnik stresses that language can help children to consolidate their grasp of spatial notions, she seems to assume that the form the concepts will take is ultimately determined by nonlinguistic cognition: “the cognitive concerns of all 18-month-olds are similar enough so that they will be likely to acquire the *same sorts of meanings* by the end of the one-word period” (Gopnik and Meltzoff 1986, 219, emphasis added). So linguistic input serves primarily to reinforce natural tendencies; it does not in itself introduce novel structuring principles.

As long as we restrict our attention to children learning our own native language, we have no reason to doubt that linguistic input can at most only help to reinforce spatial concepts that children will acquire in any event. This is because the spatial categories of our language seem so “natural” to us that it is easy to imagine they are

the inevitable outcome of cognitive development. But a close look at the treatment of space in diverse languages suggests that language may play a more powerful structuring role than Gopnik suggests. For example, hearing the same word repeatedly across differing events might draw children's attention to abstract properties shared by these events that might otherwise pass unnoticed. Let us consider this possibility more closely.

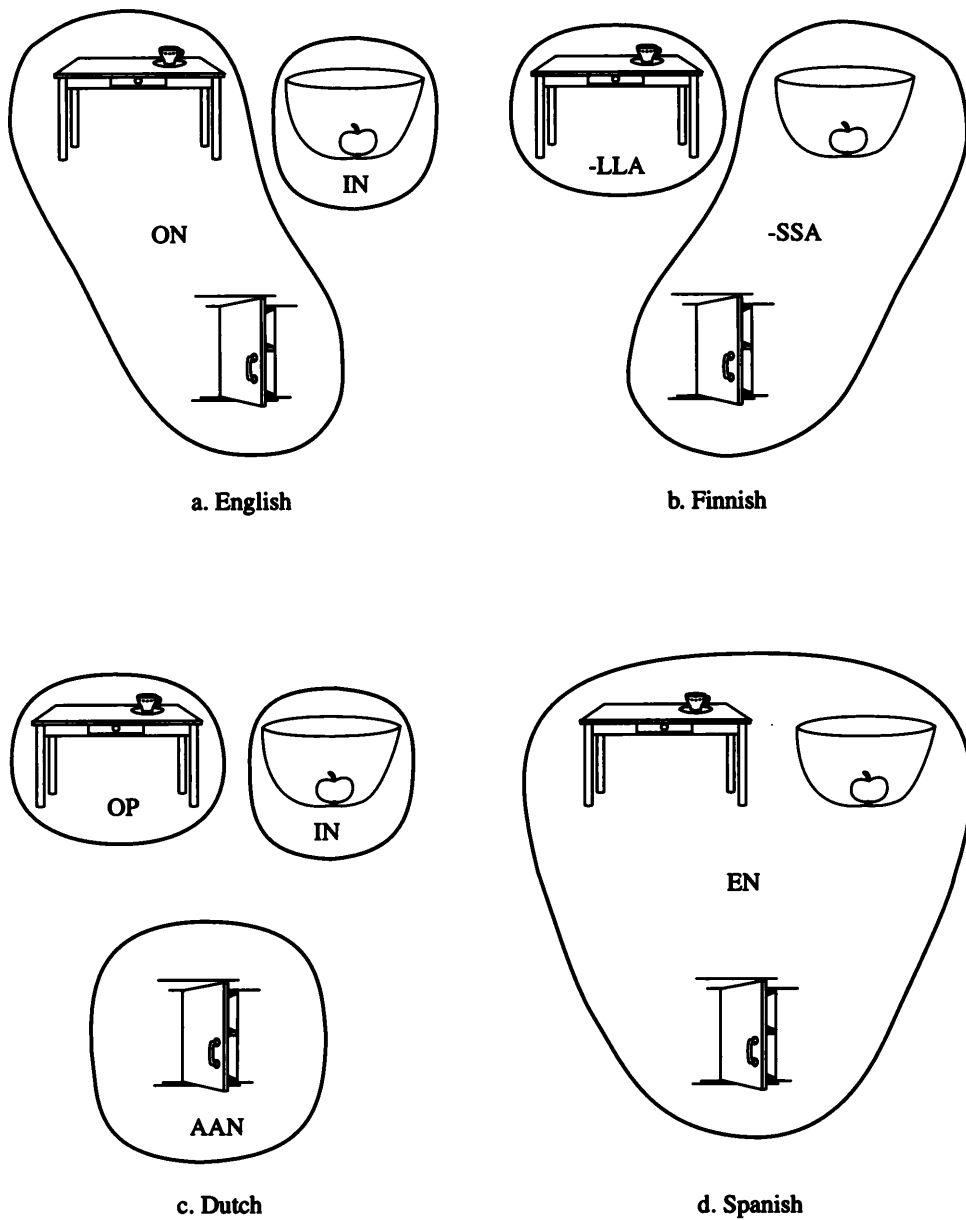
10.2.1 Crosslinguistic Perspectives on Spatial Categorization

Objectively speaking, no two objects, events, attributes, or spatial configurations are completely identical—consider two dogs, two events of falling, or two acts of kindness. But each discriminably different referent does not get its own label: one of the most basic properties of language is that it carves up the world into (often overlapping) *classes* of things that can all be referred to with the same expression, such as *dog*, *pet*, *fall*, *open*, and *kindness*. These classes, or categories, are composed of entities that can be treated as alike with respect to some equivalence metric.

Under the hypothesis that preexisting spatial concepts provide the meanings for children's spatial words, it is assumed these concepts provide the grouping principles, or, put differently, the metric along which a word will be extended to novel situations. But what principles are these? Here it is critical to realize that there is considerable variation across languages in which similarities and differences “count” in establishing whether two spatial situations belong to the same spatial semantic category—that is, can be referred to with the same spatial morpheme.

As a simple illustration, let us consider some configurations involving the often-invoked notions of contact, support, and containment: (a) “cup on table,” (b) “apple in bowl,” and (c) “handle on cupboard door” (cf. figure 10.1). In many languages, relationships involving contact with and support by a vertical surface, such as “handle on cupboard door,” are treated as similar to relationships involving contact with and support by a more-or-less horizontal surface, such as “cup on table.” In English, for example, the spatial relationships in (a) “cup on table” and (c) “handle on cupboard door” are both routinely called *on*; a different word—*in*—is needed for “containment” relations like (b) “apple in bowl.” This grouping strategy (shown in figure 10.1a) seems to make perfect sense: after all, both “cup on table” and “handle on door,” but not “apple in bowl,” involve contact with and support by an external surface.

But sensible as this strategy may seem, not all languages follow it. In Finnish, for example, situations like (c) “handle on cupboard door” are grouped linguistically with those like (b) “apple in bowl” (both are encoded with the inessive case ending *-ssa*, usually translated as “in”); for (a) “cup on table” a different case ending (the adessive, *-lla*, usually translated as “on”) is needed. The motivation for this

**Figure 10.1**

Classification of three static spatial situations in English, Finnish, Dutch, and Spanish.

grouping—shown in figure 10.1b—may be that attachment to an external surface can be seen as similar to prototypical containment, and different from horizontal support, on a dimension of “intimacy” or “incorporation” (other surface-oriented configurations that can be encoded with the case ending *-ssa*, “in,” include “Band-aid on leg,” “ring on finger,” “coat on hook,” “sticker on cupboard,” and “glue on scissors”; Bowerman, 1996).

In still a third pattern, exemplified by Dutch, situations like (c) can be collapsed together with neither (a) (*op* ‘on₁’) nor (b) (*in* ‘in’), but are characterized with a third spatial morpheme, *aan* ‘on₂’, that is somewhat specialized to relations of hanging and other projecting attachment, (e.g., “picture on wall,” “apple on twig,” “balloon on string,” “coat on hook,” “hook on door”; Bowerman 1989, 1996); this pattern is shown in figure 10.1c. And in a fourth pattern, displayed by Spanish, it is quite unnecessary to differentiate among (a), (b), and (c)—a single preposition, *en*, can comfortably be applied to all of them! (figure 10.1d). (If desired, the situations can be distinguished by use of *encima de* ‘on top of’ for (a) and *dentro de* ‘inside of’ for (b)).³ These various classification patterns, although different, all make good sense—class membership is in each case established on the basis of an abstract constancy in certain properties, while other properties are allowed to vary.

In still other languages, the familiar notions of “contact and support” and “containment” undergo much more radical deconstruction than in the examples shown so far. For example, in Tzeltal, a Mayan language of Mexico, there is no all-purpose containment word comparable to English *in* (P. Brown 1994). Different forms are needed to indicate that

- (2) a. A man is *in* a house (*ta y-util* ‘at its-inside’)
- b. An apple is *in* a bowl (*pachal* ‘be located’, of *something in a bowl-shaped container* or of the container itself)
- c. Water is *in* a bottle (*waxal* ‘be located’, of *something in a taller-than-wide rectangular or cylindrical object* or of the object itself)
- d. An apple is *in* a bucket of water (*t’umul* ‘be located’ *immersed in liquid*)
- e. A bag of coffee is *in* a pot (*xojol* ‘be located’, having been *inserted singly into a closely fitting container*)
- f. Pencils are *in* a cup (*xijil* ‘be located’, of *long/thin object, having been inserted carefully into a bounded object*)
- g. A bull is *in* a corral (*tik’il* ‘be located’, having been *inserted into container with a narrow opening*).

Similarly, in Mixtec, an Otomanguean language also spoken in Mexico, there is no all-purpose contact-and-support word comparable to English *on*. Instead, spatial relationships between two objects are indicated by invoking a “body part” of the

reference object in a conventionalized but completely productive way (Brugman 1983, 1984; Lakoff 1987). For example:

- (3) a. A man *on* a roof ([be.located] *siki-beʔe* ‘animal.back-house’)
- b. A man *on* a hill (... *šini-yúku* ‘head-hill’)
- c. A cat *on* mat (... *nũũ-yuu* ‘face-mat’)
- d. A man *on* a tree branch (... *ndaʔa-yúnu* ‘arm-tree’).

Some of these forms can also be used for an area adjacent to the named “body part” of the reference object, for example, [be.located] *šini-yúnu* ‘head-tree’ could be said of a bird either located *on* the top of a tree, or hovering *above* the tree. Comparable body part systems are also employed by Tzeltal and other Mayan languages (Levinson 1994) and many other languages of Meso-America and Africa, although details of body-part assignment vary widely (Heine 1989; MacLaury 1989).

Let us take an example from a different domain, manipulations of objects. Consider these three actions: (a) “hanging up a coat,” (b) “hanging up a mobile,” and (c) “hooking two toy train cars together.” English speakers will typically use *hang* (*up*) for both (a) and (b), conceptualizing them as similar on grounds that in both events, an entity is arranged so that it dangles downward with gravity. They will use a different expression—perhaps *hook together*—for (c), which lacks this property. This categorization pattern is shown in figure 10.2a. Korean speakers will make a different implicit grouping, using the verb *kelta* for both (a) and (c), and a different verb, *talta*, for (b). (Korean lacks the semantic category associated with English *hang*.) This pattern is shown in figure 10.2b. Why is hanging up a coat assigned to the same spatial category as hooking together two train cars? Because of the way they are attached: in both events, an entity is fixed to something by mediation of a hooking configuration (*kelta*), whereas in the “hanging a mobile” event shown in (b), the entity is attached directly (*talta*; this verb could also be used for attaching a sideways-projecting handle to a door).

Notice that both these classification strategies can achieve the same communicative effect—e.g., to call a listener’s attention to an action of hanging up a coat. But they do so in different ways. When English speakers use *hang* for hanging up a coat, they assert that the coat is arranged so that it dangles with gravity, but they say nothing about how it is attached; the listener must infer the most likely kind of attachment on the basis of his knowledge of how dangling coats are usually attached. Conversely, when speakers of Korean use *kelta* for the same action, they assert that the coat is attached by hooking, but they say nothing about dangling with gravity; again, the listener must infer on the basis of his world knowledge that when coats are hooked to something, dangling with gravity is likely to ensue. For communicative purposes, then, the expressions of the two language are equivalent: in concrete contexts, they

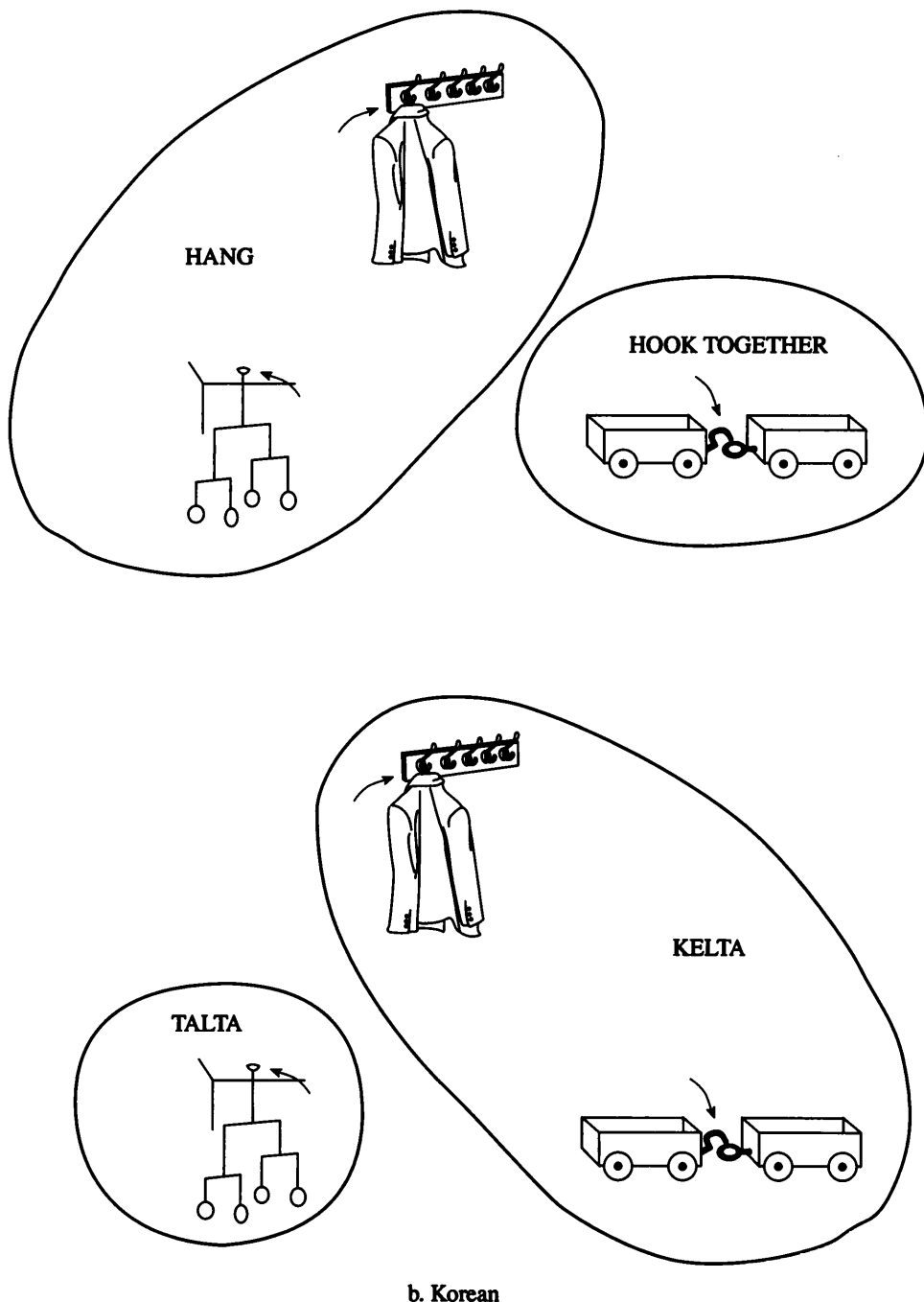


Figure 10.2
Classification of three actions in English and Korean.

can invoke the same scenes in the listener's mind. But the spatial concepts underlying the words are different, and so, consequently, are the overall sets of events they pick out.

It is clear, then, that the situations that fall together as instances of "the same spatial category" vary widely across languages in accordance with differences in the properties of situations that are conventionally used to compute similarity for purposes of selecting a word. The resulting categories cross-cut each other in complex ways. For example, the situations in (3), which are distinguished in Mixtec, all involve an object resting on a horizontal supporting surface and so are relatively prototypical for English *on*. However, Mixtec does not simply subdivide the English category of *on* more finely: recall that situations that English obligatorily distinguishes as *on* versus *above* often fall together in Mixtec—both instantiate adjacency to the named body part of the reference object.

In order to talk about space, then, it is not sufficient for children to understand that objects fall if not supported, that one object can be put above, on, below, inside, or occluding another object, and so on. A perceptual or action-based understanding of what is going on in given spatial situations is probably a necessary condition for learning to talk about space, but this knowledge alone does not buy children knowledge of how to *classify* space in their language—for example, it will not tell them whether an apple in a bowl should be seen as instantiating the same spatial relationship as a bag of coffee in a pot, or whether hanging a coat should be treated as more similar to hanging a mobile or to hooking two train cars together. To be able to make these decisions in a language-appropriate way, it is essential to discover the implicit patterning in how spatial words are distributed across contexts.⁴

10.2.2 What Else Does the Child Need to Learn?

Determining the right way to categorize spatial relations is an important problem for the language learner, but it is not the only task revealed by an examination of how different languages deal with space. A few others can be briefly summarized as follows.⁵

10.2.2.1 What Do Languages Conventionally Treat as 'Spatial Relationships' to Begin With? In the discussion of figure 10.1, I simply assumed that all the configurations shown can be construed as "spatial"—the problem was just to identify which properties languages are sensitive to in classifying them as instances of one spatial category or another. But languages in fact differ not only in how they classify spatial configurations, but also in the likelihood that they will treat certain configurations as spatial at all.

Some relationships seem to be amenable to spatial characterization perhaps in all languages—for example, a cup on a table, an apple in a bowl, and a tree adjacent

to a house. But other relationships are treated more variably. In some languages, including English, part-whole relations are readily described with the same spatial expressions used for locating independent objects with respect to each other; e.g., “the handle *on* the cupboard door (is broken)” “the muscles *in* my left calf (are sore)”, and “the lid *on* this pickle jar (has a funny picture on it).” But in many languages, analogous constructions sound odd or impossible; for example, speakers of Polish consistently use genitive constructions along the lines of “the handle *of* the cupboard door,” “the muscles *of* my left calf,” and “the lid *of* the pickle jar.”

In a second example, consider entities that do not have “good Gestalt,” such as unbounded substances like glue, butter, and mud, or bounded “negative object parts” (Herskovits 1986; Landau and Jackendoff 1993) like cracks and holes. English speakers are again relatively liberal in their willingness to treat these entities as “located objects”—e.g., “Why is there butter *on* my scissors!?” (or “Why do my scissors have butter *on* them?”) and “There’s a crack *in* my favorite cup!” But speakers of many languages resist “locating” such entities with respect to another entity, preferring instead constructions comparable to “My scissors *are buttery/have butter*” and “My cup is *cracked/has a crack*.”⁶

Differences in the applicability of spatial language to entities like butter and cracks seem to reflect pervasive crosslinguistic differences in conventions about whether constructions that are typically used for *locating* objects—for example, for narrowing the search space in response to a “where” question—can be used for describing what objects look like, or how they are configured with respect to each other (cf. Wilkins and Senft 1994). Notice that when English speakers exclaim, “Why is there butter *on* my scissors?” or “There’s a crack *in* my cup!” they are not telling their listeners “where” the butter or the crack is, but rather making an observation about the condition of the cup or the scissors. Different conventions about the use of spatial language for describing what things look like also seem to lie behind the tendency of Spanish speakers to choose constructions with *tener* ‘have’ in many contexts where English speakers would use spatial language; compare “There’s a ribbon *around* the Christmas candle” with “The Christmas candle *has (tiene)* a ribbon”.

10.2.2.2 What Should Be Located with Respect to What? The difference between directing listeners to where something is versus telling them what something looks like probably also lies at the bottom of another intriguing difference between languages. Assuming a spatial characterization of the relationship between two entities, which one will be treated as the figure (located object) and which as the ground (referent object)?

As Talmy (1983) has pointed out, it is usual for speakers to treat the smaller, more mobile object as the figure and the larger, more stable object as the ground:

- (4) a. The book is on the table.
b. ?The table is under the book.
- (5) a. The bicycle is near the church.
b. ?The church is near the bicycle.

This principle is likely to be universal when the purpose of language is to guide the listeners' search for an entity whose location is unknown to them. But when spatial language is used for a more descriptive purpose, languages may follow different conventions. For example, when one entity completely covers the surface of another, English consistently assigns the role of figure to the "coverer" and the role of ground to that which is covered (cf. sentences 6a and 7a). Dutch, however, reverses this assignment (sentences 6b and 7b):

- (6) a. There's paint *all over* my hands.
b. Mijn handen zitten helemaal *onder* de verf.
'My hands sit completely *under* the paint.'
- (7) a. There's ivy *all over* the tree.
b. De boom zit helemaal *onder* de klimop.
'The tree sits completely *under* the ivy.'

This difference between English and Dutch might be ascribable to the lack in Dutch of an equivalent to the English expression *all over*—but we can also ask whether the absence of such an expression may not be due to a conventional assignment of figure and ground that renders it unnecessary.

10.2.2.3 How Are Objects Conventionally Conceptualized for Purposes of Spatial Description? Many crosslinguistic differences in spatial organization are due, as discussed in section 10.2.1, to variation in the makeup of spatial semantic categories—that is, in the *meaning* of spatial words. But even when morphemes have roughly similar meanings in different languages, variations in encoding may arise because of systematic differences in the way objects are conventionally *conceptualized*.

Consider, for examples, *in front of* and *behind*. In section 10.1.2.3, it was pointed out that English-speaking children initially use these words only in the context of "featured" referent objects—objects that have inherent fronts and backs. But which objects are these? People and animals are clearly featured. Trees are often mentioned as examples of objects that are not. But it turns out that this is a matter of convention. For speakers of English and familiar European languages, trees indeed do not have inherent fronts and backs. But for speakers of the African language Chamus, they do!—the front of a tree is the side toward which it leans, or, if it does not lean, the side on which it has its longest branches (Heine 1989; see also Hill 1978 for some

systematic crosslinguistic differences in the assignment of front and back regions to *nonfeatured* objects). Cienki (1989) has suggested that many differences between English, Polish, and Russian in the application of prepositions meaning “in” and “on” to concrete situations are due to differences not in the meanings of the morphemes themselves, but in whether given referent objects are conceptualized as planes or containers. Children must learn, then, not only what the spatial morphemes of their language mean, but also how the objects in their environment should be construed for purposes of their “fit” to these meanings.

10.2.2.4 How Much Information Should a Spatial Description Convey? From among all the details that could be encoded in characterizing a given situation spatially, speakers make a certain selection. Within a language, the choice between a less versus more detailed characterization of a scene (e.g., “The vase is on the cupboard” versus “the vase is on top of the cupboard”) is influenced in part by pragmatic considerations like the potential for listener misunderstanding. But holding context constant, there are striking crosslinguistic differences in conventions for how much and what kind of information to give in particular situations (see also Berman and Slobin 1994; Slobin 1987).

For example, for situations in which objects are “in” or “on” objects in a canonical way (e.g., “cup on table”, “cigarette in mouth”), speakers of many languages, such as Korean, typically use a very general locative marker and let listeners infer the exact nature of the relationship on the basis of their knowledge of the objects. English, in contrast, is relatively picky, often insisting on a distinction between *in* and *on* regardless of whether there is any potential for confusion. But English speakers are more lax when it comes to relationships that canonically involve encirclement as well as contact and support: although they can say *around*, this often seems excessive (“ring *on/?around* finger,” “put your seatbelt *on/?around* you”). For most Dutch speakers, in contrast, the encoding of encirclement wherever it obtains (with *om* ‘around’) is as routine as the distinction between *in* and *on* in English. This attentiveness to encirclement may in a sense be “forced” by the lack in Dutch of an equivalent to the English all-purpose *on*: both *op* ‘on₁’ and *aan* ‘on₂’ cover a narrower range of topological relationships, and neither one seems quite appropriate for most cases of “encirclement with contact and support.”

Another kind of information that is supplied much more frequently in some languages than in others is the motion that led up to a currently static spatial situation. In English and other Germanic languages, it is common to encode a static scene without reference to this event: for example, “There’s a fly *in* my cup” and “There’s a squirrel *up in* the tree!” Although a static description of such scenes is also possible in Korean, speakers typically describe them instead with a verb that explicitly

specifies the preceding event, as suggested by the English sentences "A fly has *entered* my cup" and "A squirrel has *ascended* the tree."

There are also crosslinguistic differences in the amount of information typically provided in descriptions of motion events (Berman and Slobin 1994). Speakers of languages with rich repertoires of spatial particles, like English and German, tend to characterize motion trajectories in considerable detail (e.g., "The boy and dog fell *off* the cliff *down into* the water"), while speakers of languages that express information about trajectory primarily in the verb, such as Spanish, give less information overall about trajectory (e.g., "fell *from* the cliff"/"fell *to* the water"), and often simply imply the kind of trajectory that must have been followed by providing static descriptions of the locations of landmarks (in this case: there is a cliff above, there is water below, and the boy and dog fall).

To summarize, I have argued that different languages structure space in different ways. Most basically, they partition space into disparate and often crosscutting semantic categories by using different criteria for establishing whether two spatial situations should be considered as "the same" or "different" in kind. In addition, they differ in which classes of situations can be characterized readily in spatial terms at all, in how the roles of figure and ground are assigned in certain contexts, in how objects are conventionally conceptualized for purposes of spatial description, and in how much and what kind of information spatial descriptions routinely convey. These differences mean that there is a big discrepancy between what children know about space on a nonlinguistic basis and what they need to know in order to talk about it in a language-appropriate way.

Accounts of spatial semantic development over the last twenty-five years have neglected crosslinguistic differences like these. Among students of language acquisition there has been a strong tendency to equate "semantic structure" directly with "conceptual structure"—to view the meanings of words and other morphemes to a large extent as a direct printout of the units of human thought. But although semantic structure is certainly dependent on human conceptual and perceptual abilities, it is by no means identical: the meanings of morphemes—and often of larger constructions (Goldberg 1995)—represent a highly structured and conventionalized layer of organization, different in different languages (see Bierwisch 1981; Bowerman 1985; Lakoff 1987; Langacker 1987; Levinson, in press; Pinker 1989). In failing to fully appreciate the distinction between "conceptual" and "semantic," developmentalists have overestimated the part played in spatial semantic development by children's nonlinguistic concepts, and so underestimated the magnitude of what children must learn. In consequence, we as yet have little understanding of how nonlinguistic spatial understanding and linguistic input interact in children's construction of the spatial system of their native language.

10.3 Studying Spatial Semantic Categorization Crosslinguistically

How early in life do children arrive at language-specific spatial semantic categories? If the hypothesis is correct that the structure of spatial semantic concepts is provided—at least initially—by nonlinguistic spatial cognition, we would expect language specificity to be preceded by a period of crosslinguistic uniformity (or of individual differences that are no greater between than within languages). Hypothesizing along these lines for spatial and other meanings encoded by grammatical morphemes, Slobin (1985, 1174) proposed that “children discover principles of grammatical marking according to their own categories—categories that are not yet tuned to the distinctions that are grammaticized in the parental language”; only later are they led by the language-specific uses of particular markers to “conceive of grammaticizable notions in conformity with the speech community.” This scenario predicts extensive errors at first in the use of spatial morphemes, possibly suggestive of the guiding influence of “child-style” spatial concepts that are similar across languages.

Another possibility is that although children may perceive many properties of spatial situations, they do not start out strongly biased in favor of certain grouping principles over others. In this case they might be receptive from a very early age to semantic categories introduced by the linguistic input and quickly home in on the needed principles with relatively few errors. Of course, there are many possible gradations between the two extreme scenarios sketched here—that is, early reliance on nonlinguistic concepts versus early induction of categories strictly on the basis of the linguistic input. And some domains may be more susceptible to linguistic structuring than others. For example, Gentner (1982) has argued that the mapping between verbs and other relational words onto events is less transparent—more imposed by language—than the mapping between concrete object nouns and their referents (see also note 21 on differential transparency in another domain).

The hypothesis that language can influence the formation of children’s semantic categories from the start of lexical development played an important role in earlier views of how children learn the meanings of words. For example, Roger Brown likened the process of learning word meanings to a game (“The Original Word Game”) in which the child player makes guesses about how to classify referents on the basis of the distribution of forms in adult speech, and he suggested that “a speech invariance [e.g., hearing the same word repeatedly in different contexts] is a signal to form some hypothesis about the corresponding invariance of referent” (1958, 228). But this approach to learning word meanings has been out of fashion for a number of years.

One reason for its unpopularity is that it clashes with the contemporary stress in developmental theorizing on the need for *constraints* on word learning: “an observer

who notices *everything* can learn *nothing*, for there is no end of categories known and constructable to describe a situation" (Gleitman 1990, 12; see also Keil 1990 and Markman 1989). Another reason is that the appeal to guidance by language in the construction of semantic categories is associated with the perennially controversial Whorfian hypothesis (Whorf 1956)—the proposal that the way human beings view reality is molded by the semantic and grammatical organization of their language. The Whorfian position has seemed implausible to many, especially as infant research shows ever more clearly the richness of the mental lives of babies (although see Levinson and Brown 1994; Lucy 1992; and Gumperz and Levinson 1996 for new perspectives on the Whorfian hypothesis). But in the widespread rejection of the Whorfian hypothesis, the baby has been thrown out with the bathwater. Regardless of whether the semantic categories of our language play a role in fundamental cognitive activities like perceiving, problem solving, and remembering, we must still *learn* them in order to speak our native language fluently. But how learners home in on these categories is a topic that has been little explored.⁸

In trying to evaluate the relative strength of nonlinguistic cognitive organization and the linguistic input in guiding children's early semantic structuring of space, a useful research strategy is to compare same-age children learning languages with strikingly different spatial categories. Because we are interested in how early children can arrive at language-specific ways of structuring space, it is sensible to focus on meanings that are known in principle to be accessible to young children (thus, 'in' and 'on'-type meanings are preferable to projective 'in front of'/'behind'-type meanings). With this in mind, I have been exploring, in projects together with various colleagues (Soonja Choi, Dedre Gentner, Lourdes de León, and Eric Pederson), how children, and languages, handle topological notions of contact, separation, inclusion, and encirclement; functional and causal notions like support, containment, attachment, and adhesion; and notions to do with vertical motion and orientation (up and down).

10.3.1 Spatial Encoding in the Spontaneous Speech of Learners of Korean and English

In one study, Soonja Choi and I compared how children talk about spontaneous and caused motion in English and Korean (Choi and Bowerman 1991; Bowerman 1994). These two languages differ typologically in their expression of directed motion. English is what Talmy (1985, 1991) calls a "satellite-framed" language. These languages—which include most Indo-European languages and also, for example, Chinese and Finnish—characteristically express path notions (movement into, out of, up, down, on, off, etc.) in a constituent that is a "satellite" to the main verb, such as a prefix or (as in the case of English) a particle/preposition. Korean, in con-

trast, is a “verb-framed” language; these languages—which include, for example, Hebrew, Turkish, and Spanish—express path in the verb itself (Korean lacks a class of spatial particles or prepositions entirely).

For present purposes, the most important difference between English and Korean is that many of their semantic categories of path are different. In general, the prepositions and particles of English identify paths that are highly abstract and schematic, whereas most of the path verbs of Korean are more specific. For example, in English, a motion along a particular path is encoded in the same way regardless of whether the motion is spontaneous or caused (cf. “Go *in* the closet” versus “Put it *in* the closet”; “Get *out* of the bathtub” versus “Take it *out* of the bathtub”). In Korean, in contrast, spontaneous versus caused motions along a particular path are typically encoded with entirely different verb roots (cf. *tule* ‘enter’ versus *nehta* ‘put loosely in (or around)’; *na* ‘exit’ versus *kkenayta* ‘take out (or take from loosely around)’.⁹ Further, English path categories are relatively indifferent to variation in the shape and identity of the figure and ground objects, whereas Korean path categories are more sensitive to this, with the result that they subdivide and crosscut the English path categories in complex ways; this is illustrated in table 10.1 (see Choi and Bowerman 1991 for more detail). The overall tendency for path categories to be larger and more schematic in English than in Korean is no doubt related to the systematic difference in how they are expressed: with *closed-class* morphemes (prepositions and particles) in English and *open-class* morphemes (verbs) in Korean (see also Landau and Jackendoff 1993 and Talmy 1983).

If the meanings that children initially associate with spatial morphemes come directly from their nonlinguistic conceptions of space, these differences in the way spatial meanings are structured in English versus Korean should have no effect on learners’ early use of spatial words—children should extend the words on the basis of their own spatial concepts, not the categories of the input language. To see whether this is so, Choi and I compared spontaneous speech samples collected longitudinally from children learning English and Korean.¹⁰

We found that both sets of children first produced spatial morphemes at about fourteen to sixteen months (particles like *up*, *down*, and *in* for the English speakers; verbs like *kkita* ‘fit tightly’ and its opposite *ppayta* ‘unfit’ for the Korean speakers; cf. table 10.1), and began to use them productively (i.e., for events involving novel configurations of objects) by sixteen to twenty months. They also talked about similar events, for example, manipulations such as putting on and taking off clothing; opening and closing containers, putting things in and taking them out, and attaching things like Lego pieces; position and posture changes such as climbing up and down from furniture and laps; and being picked up and put down. The spatial concerns of children learning quite different languages are, it seems, quite similar at this age,

Table 10.1

Relative Sensitivity of Path Categories to Variation in Properties of Figure and Ground Objects in English and Korean

English:

- in* (e.g., put ball *in* box, earplug *in* ear, flower *in* vase, cherries *in* basket)
on (e.g., put box *on* table, sticker/magnet *on* refrigerator, hat/coat/shoes/bracelet *on*)
up (e.g., put a cup *up* high, pick a child *up*, sit *up*, stand *up*)

Korean:

- neh̥ta* 'put loosely in (or around)'
 (e.g., ball *in* box, loose ring *on* pole)
kkita 'fit tightly; put tightly in/on/together/around'
 (e.g., earplug *in* ear, top *on* pen, two Lego pieces *together*, tight ring *on* pole)
kkoc̥ta 'put elongated object to base'
 (e.g., flower *in* vase, hairpin *in* hair, book upright *on* shelf)
tḁm̥ta 'put multiple object in container'
 (e.g., cherries *in* basket)
no̥hta 'put on horizontal surface'
 (e.g., box *on* table)
pwuch̥ita 'stick, juxtapose surfaces that are flat, or can be conceptualized as if flat' (e.g., sticker/magnet *on* refrigerator, two Lego pieces *together*)
ssuta 'put clothing *on* head' (e.g., hat, scarf, mask, glasses)
ipta 'put clothing *on* trunk' (e.g., shirt, coat, pants)
sinta 'put clothing *on* feet' (e.g., socks, shoes)
chata 'put clothing *on/at* waist or wrist' (e.g., belt, diaper, dagger, bracelet)
ollita 'cause to ascend' (e.g., lift a cup *up*)
anta 'pick up/hold in arms' (e.g., pick a child *up*)
ancta 'assume a sitting posture' (e.g., sit *up*, sit *down*)
(ile)seta 'assume a standing posture' (e.g., stand *up*)

revolving primarily around topological notions and motion up and down (see also section 10.1, and Sinha et al. 1994). But were the children's spatial *semantic categories* similar, as inferred from the range of referent events to which they extended their words? They were not. By twenty months of age, the path semantic categories of the two sets of children were quite different from each other and clearly aligned with the categories of the input language. For example:

1. The English learners used their spatial particles indiscriminately for both spontaneous and caused motion into and out of containment, up and down, and so on. In contrast, the Korean children used strictly different verbs (intransitive vs. transitive) for spontaneous and caused motion along a path. For instance, English learners said *in* both when they climbed *into* the bathtub and put magnetic letters *into* a small box; in comparable situations the Korean learners used the verbs *tule* 'enter' versus *nehta* 'put loosely in (or around)'.
2. The English learners used *up* and *down* for a wide range of events involving vertical motion, including climbing on and off furniture, posture changes (sitting and standing up, sitting and lying down), raising and lowering things, and wanting to be picked up or put down. Recall that, as reviewed in section 10.1.2.2, the rapid generalization of *up* and *down* has been interpreted as evidence that these words are coupled to non-linguistic spatial concepts. But the Korean children used no words for a comparable range of motion up or down: as is appropriate in their language, they used different words for posture changes, climbing up or down, being picked up and put down, and so forth.
3. The English learners distinguished systematically between putting things into containers of all sorts (*in*) and putting them onto surfaces (*on*), but were indifferent to whether the figure fit the container tightly or loosely, or whether it was set loosely on a horizontal surface or attached tightly to a surface in any orientation, or—in the case of clothing items—what part of the body it went onto. The Korean learners, in contrast, distinguished between tight and loose containment (*kkita* 'fit tightly' versus *nehta* 'put loosely in (or around)'), between attaching things to a surface (*kkita* again) and setting things on a surface (*nohta* 'put on horizontal surface'), and between putting clothing on the head (*ssuta*), trunk (*ipta*), and feet (*sinta*). Some examples of these differences are given in table 10.2.

Although the children had clearly discovered many language-specific features of spatial encoding in their input language, their command of the adult path categories was by no means perfect—there were also errors suggesting difficulties in identifying the boundaries of the adult categories, such as the use of *open* for unbending a doll (cf. last example in (1) of introduction), or the use of *kkita* 'fit tightly' for flat surface attachments involving stickers and magnets (e.g., entry 6 in table 10.2; this

Table 10.2

The Treatment of Containment and Surface Contact Relations in the Spontaneous Speech of Children Learning English and Korean

	Age (in months)	Utterance	Situation	Relation
English				
1.	18	<i>In 'gain.</i>	Trying to shove toy chair through narrow door of doll house.	Tight containment (Korean <i>kkita</i>)
2.	19	<i>In.</i>	When mother dips her foot into the washtub of water.	Loose containment (Korean <i>nehta</i>)
3.	17	<i>On.</i> <i>Horsie on.</i>	Looking for rein of rocking horse; it has come off and she wants to attach it back on to the edge of the horse's mouth.	Tight surface contact (Korean <i>kkita</i>)
4.	23	Can't wow-wow <i>on.</i>	Frustrated trying to put toy dog on a moving phonograph record.	Loose surface contact (Korean <i>nohta</i>)
Korean				
5.	17	<i>Kkita.</i>	Putting peg doll into perfectly fitting niche-seat on small horse that investigator has brought.	Tight containment (English <i>in</i>)
6.	27	<i>Kkita.</i>	Attaching a magnetic fish to magnetic beak of duck.	Tight surface contact (English <i>on</i>)
7.	20	<i>Nehta.</i>	Putting blocks into a pan.	Loose containment (English <i>in</i>)
8.	28	<i>Nohta.</i>	Putting one block on top of another.	Loose surface contact (English <i>on</i>)

The Korean examples show only citation form of the verb, not whole utterances.

should be *pwuchita* 'stick, juxtapose flat surfaces'; cf. table 10.1). These errors are important because they suggest that the language specificity of the learners' categories cannot be dismissed on grounds that the children perhaps were simply mimicking what they had heard people say in particular situations, and had no real grasp of the underlying semantic concepts. (Appropriate usage for *novel* situations, as illustrated by most of the examples in table 10.2, also argues against this interpretation.) We will come back to errors later, because they provide invaluable clues to children's relative sensitivity to different kinds of spatial semantic distinctions.

10.3.2 Spatial Encoding in Elicited Descriptions of Actions in Children Learning English, Korean, and Dutch

The examination of spontaneous speech can give a good overview of the early stages of spatial semantic development, and this approach has the advantage that, because

the utterances are freely offered, they reflect how children are conceptualizing situations for their own purposes. But a disadvantage is that the specific spatial situations that children happen to talk about vary, so comparing the distribution of forms requires matching situations that are not identical (as is done in table 10.2).

To get more control over what subjects talked about, Choi and I decided to conduct a production study in which we elicited descriptions of a standardized set of spatial actions from all subjects (Bowerman and Choi 1994). This time we focused exclusively on caused motion involving spatial manipulations of objects. To English and Korean, we added Dutch. Recall that an interesting way in which Dutch differs from English is its breakdown of spatial relations encompassed by English *on* into two subclasses, *op* 'on₁' (e.g., "cup *op* table") and *aan* 'on₂' (e.g., "handle *aan* cupboard door"); these differences are relevant to motion as well as to static spatial configuration.

The actions we used—seventy-nine in all—were selected on grounds that they are grouped and distinguished in interestingly different ways in the three languages. They were both familiar and novel, and covered a broad range of "joining" and "separating" situations such as donning and doffing clothing of different kinds (carried out with a doll), manipulations with containers and surfaces (e.g., putting a toy boat *into* a baby bathtub and taking it *out*, laying a doll *on* a towel after her bath, taking a dirty pillowcase *off* a pillow and putting a clean one *on*), *opening* and *closing* things (e.g., a suitcase, a cardboard box with flaps), putting tight- and loose-fitting rings *on* a pole and taking them *off*, buttoning and unbuttoning, hanging and "unhanging" (towel *on/off* hook), hooking (train cars *together/apart*), sticking (Band-aid *on* hand, suction hook *on/off* wall), and otherwise attaching and detaching things (e.g., magnetic train cars, Lego pieces, Popbeads, Bristle blocks). For these last-mentioned actions, we varied whether the objects were moved laterally or vertically, and whether the motions were symmetrical (e.g., one Lego piece in each hand, both hands moving together) or asymmetrical (e.g., one hand joins a Lego piece to a stack of two Legos pieces held in the other hand). (English and Dutch, but not Korean, are sensitive to these properties—compare, for example, *put on* with *put together*, and *take off* with *take apart*.)

For each language we had 40 subjects: 10 adults, and 30 children, 10 each in the age ranges 2;0–2;5, 2;6–2;11, and 3;0–3;5 years. Subjects were tested individually. We elicited spatial descriptions by showing the objects involved in each action and indicating what kind of spatial action should be performed with them, but not quite performing it, and saying things like "What should I do? Tell me what to do."¹¹ This procedure worked quite well: even in the youngest age group, 87% of the children gave a relevant verbal response, although not necessarily the same one the adults gave. Typical responses from the children learning English and Dutch were particles,

either alone (e.g., *in*, *on*) or with verbs (e.g., *put it in*); from the children learning Korean they were verbs (e.g., *kkie*, imperative form of *kkita* ‘fit tightly’).

10.3.2.1 Action Descriptions as Similarity Data The data collected can be seen as analogous to the data obtained in a sorting study. But instead of giving subjects a set of cards with, say, pictures of stimuli, and asking them to sort these into piles of stimuli that “go together,” we take each word produced by a subject as defining a category (analogous to a pile), and look to see which actions the subject applied the word to (i.e., sorted into that pile). Actions a speaker refers to with the same expression are considered more alike for that speaker than actions referred to with different expressions.¹² Seen in this way, the data can be analyzed with any technique suitable for similarity data, such as multidimensional scaling or cluster analysis.¹³

In one analysis, the data from all the subjects were subjected to a multidimensional scaling analysis that allowed us to plot the actions in two-dimensional space on the basis of how similar each action was to each other action (as determined by how often speakers across all three languages characterized both actions with the same expression). This was done separately for the set of “joining” actions and the set of “separating” actions, after earlier analyses had showed that, with rare (child) exceptions, these were distinguished by subjects of all ages and languages. The two resulting plots—somewhat modified by hand to spread out actions that were bunched very tightly together (because they were very often described with the same expression)—then serve as grids on which we can display the categorization system of any individual, or the dominant categorization of a group of individuals, by drawing in “circles” (i.e., Venn diagrams) that encompass all the actions that were described in the same way.

To see how this works, consider figures 10.3 and 10.4. Figures 10.3a and 10.3b show the dominant classification of the “joining” actions by the English-speaking adults and youngest group of English-speaking children (2;0–2;5 years); Figures 10.4a and 10.4b give the same information for the Korean subjects. The number of subjects (out of 10) who produced a given response is indicated on the grid near the label for the action.¹⁴ A quick overview of similarities and differences in how different groups of subjects classified the actions can be obtained by an eyeball comparison of the relevant figures:

- Figures 10.3a and 10.4a: *adult speakers of English versus Korean*;
- Figures 10.3b and 10.4b: *same-age child speakers of English versus Korean*;
- Figures 10.3a and 10.3b: *adult versus child speakers of English*;
- Figures 10.4a and 10.4b: *adult versus child speakers of Korean*.

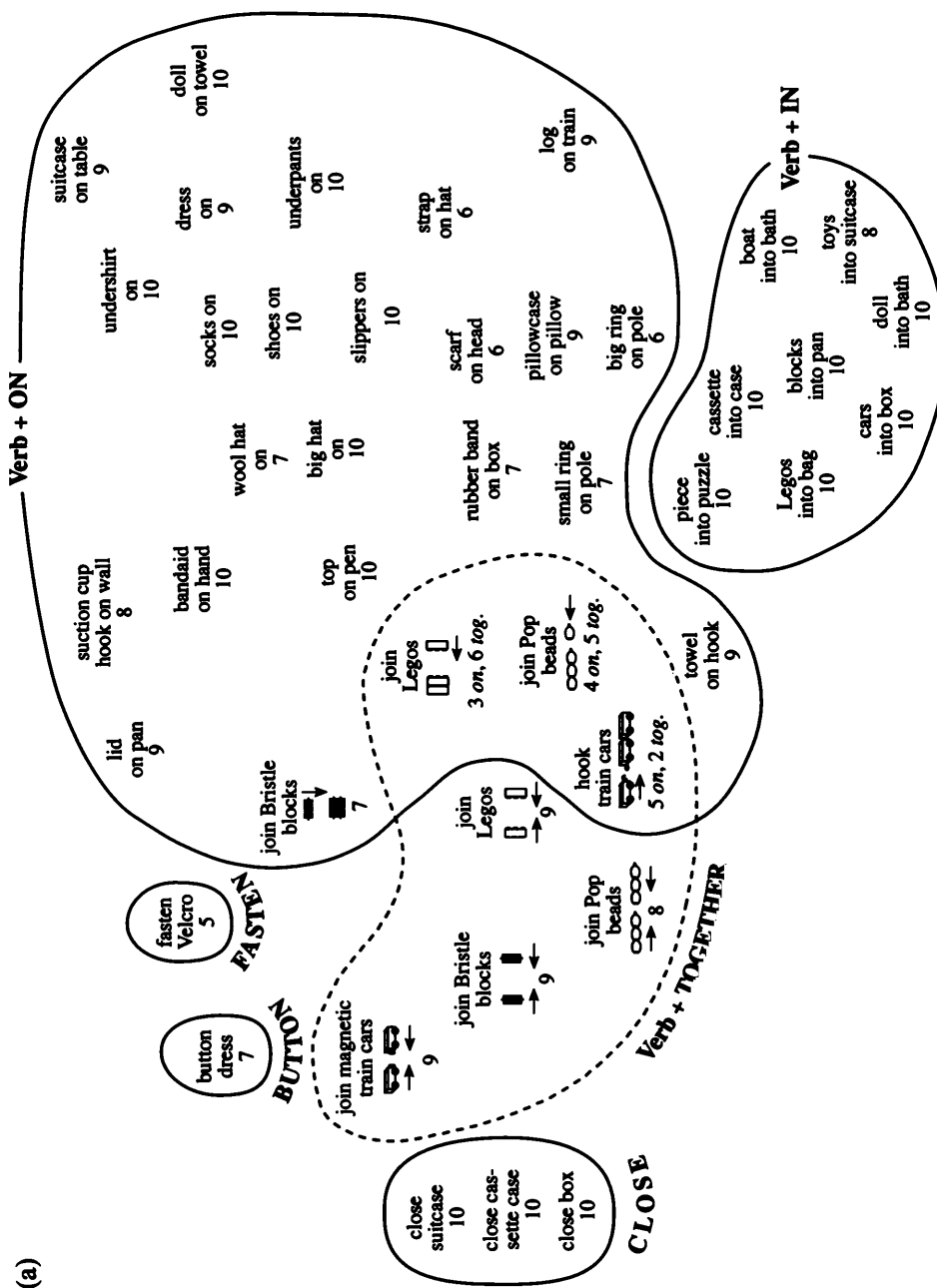
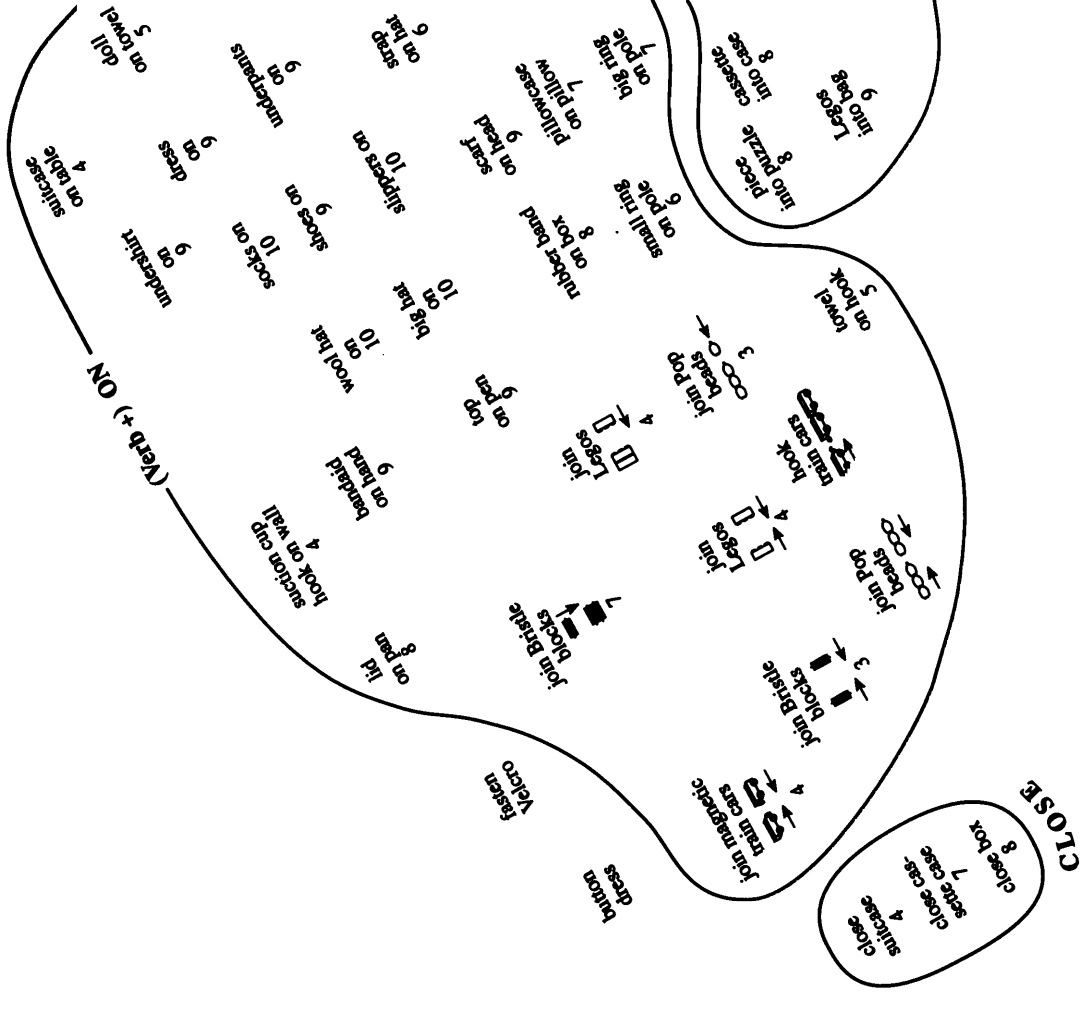


Figure 10.3
 (a) Categorization of joining actions by English-speaking adults. (b) Categorization of joining actions by English-speaking children, age 2 years to 2 years and 5 months.

Figure 10.3 (continued)



9

(a)

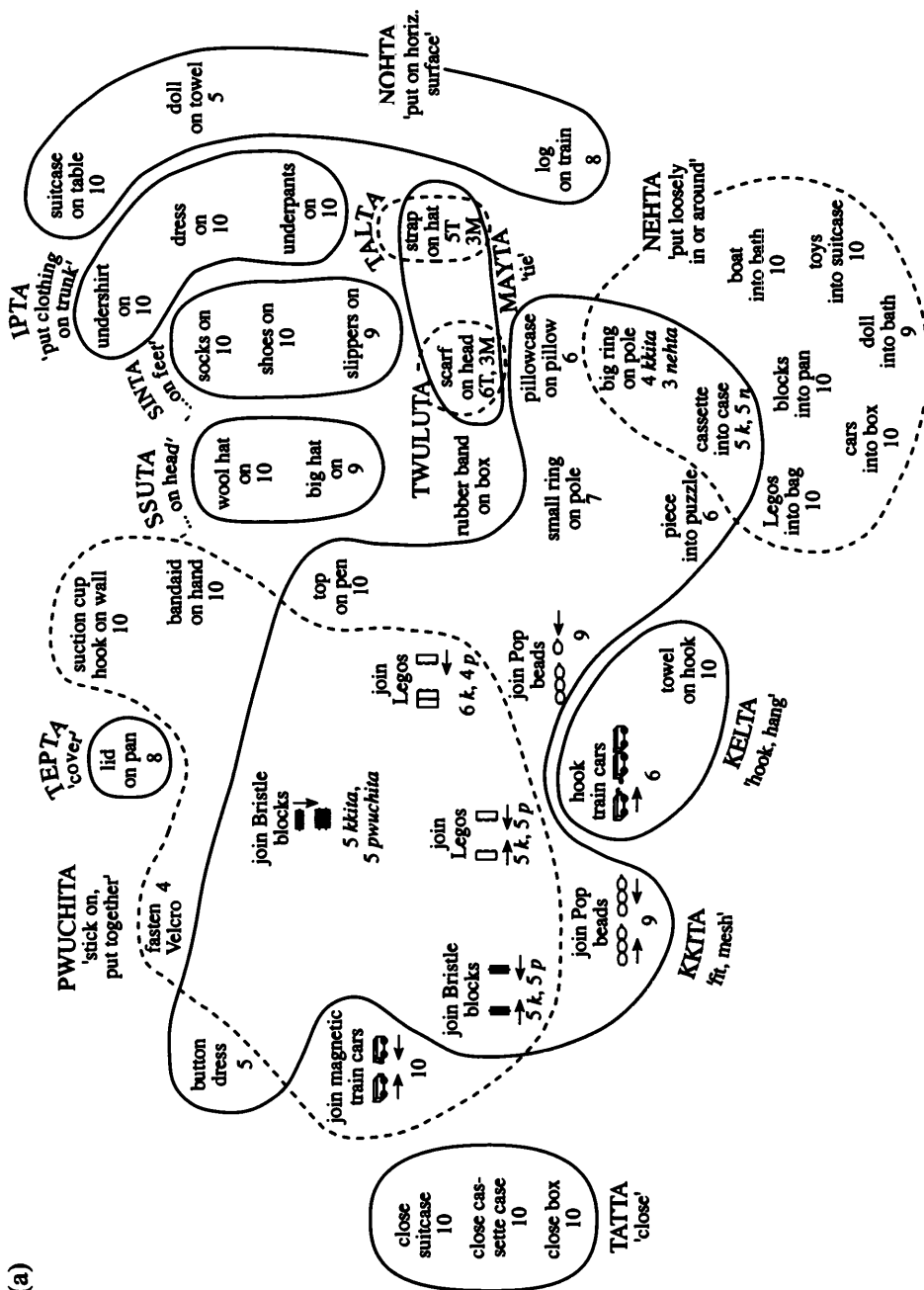


Figure 10.4

(a) Categorization of joining actions by Korean adults. (b) Categorization of joining actions by Korean children, age 2 years to 2 years and 5 months.

These comparisons reveal both similarities and differences across subject groups. For example, in addition to agreeing that joining and separating actions should be described differently, subjects of all ages and languages agree on categorizing the “closing” actions together (to far left on grid), and also the “putting into loose container” actions (lower right). But they disagree quite dramatically on the classification of actions of “putting into a tight container,” actions of encirclement, putting on clothing, and so forth.

In general outline, the children’s classification patterns are similar to those of the adult speakers of their language, but they are simpler. The children lack some words the adults use (e.g., *together* in English; *pwuchita* ‘stick or juxtapose surfaces that are flat, or can be conceptualized as if flat,’ in Korean), and they overextend certain words relative to the adult pattern—for example, many English learners overextend *on* to “together” situations; and many Korean children overextend *kkita* ‘fit tightly’ to hooking train cars together and hanging a towel on a hook, and *nehta* ‘put loosely in (or around)’ to putting a pillowcase on a pillow.

10.3.2.2 Interpreting Children’s Categorization Patterns Comparing across the three languages, these elicited production data suggest that the way children initially classify space for language is the outcome of a complex interaction between their own nonlinguistic recognition of similarities and differences among spatial situations, on the one hand, and the way space is classified by adult speakers of their language, on the other. Overall, the influence of the input language is quite strong: statistical analysis shows that in all three languages, the youngest age group of children classified the spatial actions more similarly to *adult* speakers of their *own* language than to *same-age children* learning *other* languages.¹⁵ But obedience to the adult system was by no means perfect. Patterns of conformity with and deviation from the adult target system appear to be influenced by a mix of linguistic and nonlinguistic factors. Let us consider two examples.

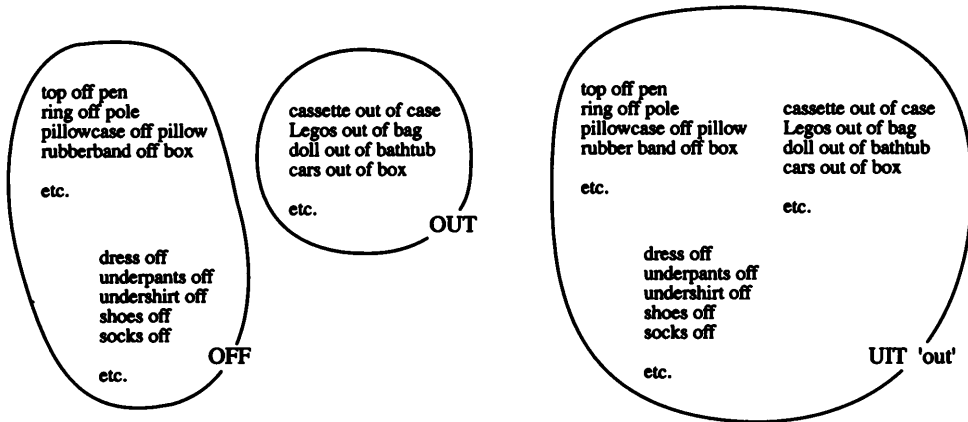
1. When children of a certain age are in principle capable of making a particular semantic distinction (as inferred from the observation that children in some language do so), the speed with which they begin to make it (if it is needed for their language) is strongly influenced by the clarity and consistency with which adult speakers mark it. For example, even the youngest age group of English speakers, like the adults, made a systematic split between “removal from containment” (*out*) and “removal from contact with an external surface” (*off*); this is illustrated in figure 10.5a with a subset of the relevant actions.¹⁶ Like English speakers, adult Dutch speakers also make a distinction between “removal from containment” (*uit* ‘out’) and “removal from contact with an external surface” (*af* ‘off’). But the youngest group of Dutch

children did not observe it—as shown in figure 10.5b, they vastly overextended *uit* ‘out’ to actions for which adults use *af* ‘off’, like taking a ring off a pole, a pillowcase off a pillow, and a rubber band off a box.

Why do the two sets of children differ in this way? Comparison of the adult systems is revealing. In English, the distribution of *out* and *off* correlates closely with removal from a concavity versus removal from a flat or convex surface (including body parts). In Dutch, the distribution of *uit* ‘out’ and *af* ‘off’ is based on the same principle, but with one important class of exceptions: whereas English uses *off* for the removal of enveloping clothing like coats, pants, shoes, and socks, Dutch uses *uit* ‘out’ (“take *out* your shoes/coat”; cf. figure 10.5c). When adult Dutch speakers are asked why they say “take out your shoes (coat, etc.),” they often seem to discover the anomaly for the first time: “It’s strange—when you take your shoe *uit* [‘out’], it’s really your foot that comes out of the shoe, isn’t it, not the shoe that comes out of your foot!” This reaction suggests that adults store this clothing use of *uit* separately from its normal use (i.e., as a separate polyseme). But this atypical use seems to be sufficiently salient to young children to obscure the distinction otherwise routinely made in Dutch between removal from surfaces and removal from containers.

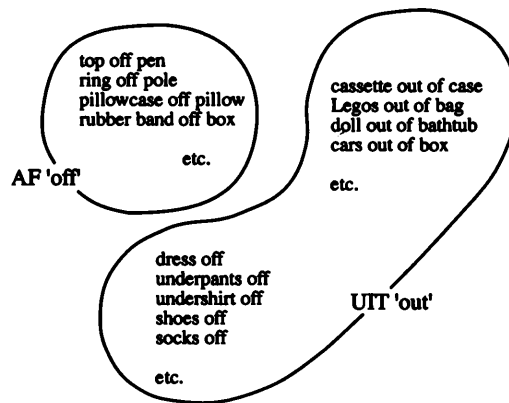
This example is intriguing because it goes squarely against a common claim about early word learning: that children at first learn and use words only in very specific contexts. According to this hypothesis, Dutch children should learn the use of *uit* for taking off clothing essentially as an independent lexical item. If so, they should proceed on the same schedule as learners of English to discover the semantic contrast between more canonical uses of *uit* ‘out’ and *af* ‘off’. But this does not happen: Dutch children appear to try to discover a coherent meaning for *uit* ‘out’ that can encompass *both* clothing- and container-oriented uses. The only meaning consistent with both uses, in that it is indifferent to the distinction between removal from a surface and removal from containment, is the notion of “removal” itself. Once children have linked this notion to *uit* ‘out’ it licenses them to use the word indiscriminately across the ‘out’/‘off’ boundary, which is exactly what they do, as shown in figure 10.5b.¹⁷

2. Children’s errors in using spatial words have often been interpreted as a direct pipeline to their nonlinguistic spatial cognition; for instance, in interpreting the somewhat different patterns of extension of the words *open* and *off* in my two daughters’ speech, I once suggested that the children had arrived at different ways of categorizing separations of various kinds on the basis of their own dealings with the physical world (Bowerman 1980). Overextensions do often seem to be conditioned by factors for which it is difficult to think of an intralinguistic explanation: for example, across all three languages in Choi’s and my study, children tended to overextend words for



a. Children learning English, age 2;0 - 2;5

b. Children learning Dutch, age 2;0 - 2;5



c. Dutch adults

Figure 10.5
Classification of actions as 'off' versus 'out' in English and Dutch.

separation more broadly than words for joining; that is, they differentiated less among actions of separation, relative to the adult pattern, than among actions of joining (and this is also true for children learning Tzotzil Mayan (Bowerman, de León, and Choi 1995). But a careful look across languages suggests that linguistic factors also play an important role in overextensions: in particular, the category structure of the input influences both *which* words get overextended and the *specific patterning* of the extensions.

If overextensions of spatial morphemes were driven purely by ways children categorize spatial events nonlinguistically, we would expect similar overextensions in different languages. And we do in fact find this to some extent: for example, similar overextensions of *open* and its translation equivalents have been reported for children learning English, French, and German (see Clark 1993 for review and sources). In Choi's and my production study, *open* (also spelled *open* in Dutch) was overextended to actions for which adults never used it about 9 times by English learners and about 21 times by Dutch learners (e.g., unbuttoning a button, taking a shoe off, separating two Lego pieces, and taking a piece out of a puzzle). But Korean children hardly make this error—it does not occur at all in the spontaneous speech data we have examined, and it occurs only once in the production study (one child used *yelda* 'open' for unhooking two train cars).

Why is there this difference in the likelihood of overgeneralizing 'open' words? A plausible explanation is that it is due to differences in the size and diversity of the 'open' categories of English and Dutch (and French and German) on the one hand, and Korean on the other. In Korean, *yelda* 'open' applies to doors, boxes, suitcases, and bags, for example, but it cannot be used for a number of other actions that are also called *open* in English and Dutch, such as opening the mouth, a clamshell, doors that slide apart (*ppellita* 'separate two parts symmetrically'), the eyes (*ttuta* 'rise'), an envelope (*ttutta* 'tear away from a base'), and a book, a hand, or a fan (*phyelchita* 'spread out a flat thing'). The breadth of the 'open' category in English and Dutch—that is, the physical diversity of the objects that can felicitously be "opened"—seems to invite children to construct a very abstract meaning; put differently, the diversity discourages children from discriminating among candidate 'opening' events on the basis of object properties that *are* in fact relevant to membership in the "open" category for adults. Conversely, the greater coherence in the physical properties of the objects to which Korean *yelda* 'open' can be applied—along with the coherence of each of the other categories encompassing events that are also called "open" in English and Dutch—may facilitate Korean children's recognition of the limits on the semantic ranges of the words.

If Korean children do not overextend *yelda* 'open', do they have another word that they overextend in the domain of separation? They do. In our production study, they

overwhelmingly used *ppayta* 'unfit' for virtually all the actions of separation—even including the actions for which adults usually used *yelda* 'open', such as opening a suitcase and a box! Like *open* in English, the category of *ppayta* 'unfit' is big and diverse in adult speech: out of the 36 "separation" actions in our study, 24 were labeled *ppayta* by at least one of the 10 Korean adults. (The word was used most heavily for events like separating Popbeads, Lego pieces, and Bristle blocks, and taking a piece out of a puzzle and the top off a pen, but it was also used occasionally for (e.g.) opening a cassette case, taking Legos out a bag, taking off a hat, and taking a towel off a hook.)

Do English, Dutch, and Korean children in fact use *open*, *open* 'open', and *ppayta* 'unfit' for the *same range* of events? If so, this would suggest the power of an underlying child-basic, language-independent notion. But the situations to which children extend *open* and *ppayta* 'unfit' differ, and the differences are related to the different meanings of the words—and their different ranges of application—in adult speech.

Korean children's *ppayta* 'unfit' category seems to have its center—as in adult speech—in the notion of "separating fitted or 'meshing' objects with a bit of force" (e.g., pulling Popbeads and Lego pieces apart, taking the top off a pen—9 out of the 10 children used *ppayta* for these actions). It is extended from this center to taking things out of containers, and overextended, relative to patterns in the adult data, to opening containers, "unsticking" adhering and magnetized objects, and taking off clothing. In contrast, English-speaking children's *open* category is centered on acts of separation as a means of *making something accessible* (e.g., opening a box to find something inside; opening a door to go into another room), and it is extended from this center only occasionally to pulling apart Popbeads and Legos and taking off clothing (both much more often called *off* in the elicited production study), and to taking things out of containers (much more often called *out*). English-speaking children also use *open* for actions in which something is made accessible without any separation at all, such as turning on TVs, radios, water faucets, and electric light switches (Bowerman 1978, 1980). Korean children do not overextend *ppayta* 'unfit' to events of this kind, probably because its use in adult speech is concentrated on acts of physical separation per se, and not on separation as a means of making something accessible.

In sum, children learning these different languages show a shared tendency, probably conditioned by nonlinguistic cognitive factors, to underdifferentiate referent events in the domain of separation—that is, they overextend words in violation of distinctions that their target language honors. But which words they "select" to overextend, and the semantic categories defined by the range of events across which they extend them, are closely related to the semantic structure of the input language.

10.4 How Do Children Construct the Spatial Semantic System of Their Language?

We have seen that language learners are influenced by the semantic categorization of space in their input language from a remarkably young age. This does not mean, of course, that they simply passively register the patterns displayed in the input—they do make errors, and these suggest that learners find some distinctions and grouping principles employed by the input language either difficult or unclear (or both). There is, then, an intricate interaction between nonlinguistic and linguistic factors in the process of spatial semantic development. In this final section, let us speculate about how this interaction takes place.

10.4.1 Is the Hypothesis That Children Map Spatial Morphemes onto Prelinguistically Compiled Spatial Concepts Still Viable?

The evidence for early language specificity in semantic categorization presented in section 10.3 might seem to argue strongly against the hypothesis that children start out by mapping spatial words onto prepackaged notions of space. But Mandler (1992 and chapter 9, this volume) suggests that the two can, after all, be reconciled.

Inspired by the work of cognitively minded linguists such as Langacker (1987), Lakoff (1987), and Talmy (1983, 1985), Mandler hypothesizes that an important step in the prelinguistic development of infants is the “redescription” of perceptual information into “image-schemas”—representations that abstract away from perceptual details to present information in a more schematic form. Preverbal image schemas would play a number of roles in infant mental life, but of special relevance for us is Mandler’s (1992, 598) suggestion that they “would seem to be particularly useful in the acquisition of various relational categories in language.” In particular, Mandler suggests that words meaning ‘in’ and ‘on’ are mapped to the image-schemas of *containment* (and the related notions of *going in* and *going out*) and *support*:

(8) Containment:  Going in:  Going out: 

(9) Support: 

In considering evidence that languages partition spatial situations in different ways, as discussed in Bowerman (1989) and Choi and Bowerman (1991), Mandler (1992, 599) suggests that “however the cuts are made, they will be interpreted [by the learner] within the framework of the underlying meanings represented by nonverbal image-schemas.” This means that children “do not have to consider countless variations in meaning suggested by the infinite variety of perceptual displays with which they are confronted; meaningful partitions have already taken place” (p. 599). Reliance on the preorganization provided by the nonverbal image-schemas for containment and support will make some distinctions harder to learn than others; for

example, Mandler suggests that children acquiring Dutch will have to learn how to break down the support schema into two subtypes of support (*op* 'on₁' and *aan* 'on₂'; cf. section 10.2.1), and this might well take some time (which is in fact true; see Bowerman 1993). On the other hand, Mandler predicts no difficulty for Spanish-speaking children in learning *en* 'in, on' (this seems also to be true) because this involves only collapsing the distinction between containment and support.

But what about the 'tight fit' category of the Korean verb *kkita*, which crosscuts the categories of both *in* and *on* in English, and, as Choi and Bowerman (1991) showed (cf. section 10.3.1), is acquired very early? Mandler (1992, 599) suggests that the early mapping of *kkita* onto the 'tight fit' meaning "is only a puzzle if one assumes that *in* and *on* are the only kinds of spatial analyses of containment and support that have been carried out." But 'tight fit' may well be an additional meaning that is prelinguistically analyzed, and thus is available for mapping to a word. Mandler acknowledges that we do not yet have independent evidence for this concept in prelinguistic infants, as we do for containment and support, and adds that "until such research is carried out it will not be possible to determine whether a given language merely tells the child how to categorize [i.e., subdivide or lump] a set of meanings the child has already analyzed or whether the language tells the child it is time to carry out new perceptual analyses" (pp. 599–600).

Mandler's hypothesis is by no means implausible, but it comes at a price. Suppose we discover that, from a very young age, toddlers learning a newly researched language, *L*, extend a word across a range of referents that breaks down or crosscuts the spatial semantic categories we already know children are sensitive to, like the categories defined by the putative image-schemas of containment, support, and tight fit. This means, by the logic of Mandler's argument, that there is yet another universal preverbal image-schema out there that we were not aware of before, and we must assume that all children everywhere have it, regardless of whether they will ever need it for the language they are learning.

This price may be acceptable as long as the putative preverbal image schemas uncovered by future research are not too numerous, and do not overlap each other in complex and subtle ways. But this seems doubtful, even on the basis of the limited data that is currently available. For example, the categories picked out by *open* and *ppayta* 'unfit' in the early speech of children learning English versus Korean overlap extensively. This might suggest that both words are mapped to the same preverbal image schema, but, as argued earlier, the overall range of the two categories in fact differs, each one being centered on core uses of the relevant word in adult speech.¹⁸

Mandler points out (personal communication) that her hypothesis is designed to account only for the very first stage of acquiring spatial morphemes; differing

extension patterns such as those just discussed may represent developments beyond this point. This is possible. But in this case the spatial image-schemas are doing little of the work that has often motivated the postulation that children map words to prelinguistically established concepts—namely, to provide a principled basis on which children can extend their morphemes beyond the situations in which they have frequently heard them. Regardless of whether image-schemas serve as the starting points, then, it seems we cannot rely on them to account for productivity in children's uses of spatial morphemes. For this, we will have to appeal to a process of learning in which children *build* spatial semantic categories in response to the distribution of spatial morphemes across contexts in the language they hear.

10.4.2 Semantic Primitives and Domain-specific Constraints

If semantic categories are constructed, they must be constructed out of something, and an important question is what this something is. Here we come squarely up against one of the oldest and most difficult problems for theorists interested in the structure of mind: identifying the ultimate stuff of which meaning is made.

Among students of language, a time-honored approach to this problem has been to invoke a set of semantic primitives—privileged meaning components that are available to speakers of all languages, but that can be combined in different ways to make up different word meanings.¹⁹ In searching for the ultimate elements from which the meanings of closed-class spatial words such as the set of English prepositions are composed, researchers have been struck by the relative sparseness of what can be important. Among the things that can play a role are notions like verticality, horizontality, place, region, inclusion, contact, support, gravity, attachment, dimensionality (point, line, plane, or volume), distance, movement, and path (cf. Bierwisch 1967; H. H. Clark 1973; Landau and Jackendoff 1993; Miller and Johnson-Laird 1976; Olson and Bialystok 1983; Talmy 1983; Wierzbicka 1972). Among things that never seem to play a role are, for example, the color, exact size or shape, or smell of the figure and ground objects (although see also Brown 1994).

10.4.2.1 Domain-specific Learning? If the meanings of closed-class spatial morphemes are so restricted—and restricted in similar ways across languages—children might take advantage of this in trying to figure out the meanings of new spatial forms. That is, they might approach the task of learning spatial morphemes with a *constrained hypothesis space*, entertaining only elements of meaning that are likely to be relevant for words in this domain.

Reasoning in this way, Landau and Stecker (1990) hypothesized that although children should be prepared to take shape into account in learning new words for objects, they should attend to shape only minimally in hypothesizing meanings for new spatial

words. To test this hypothesis, they showed three- and five-year-old learners of English a novel object on the top front corner of a box, and told them either “This is *a corp*” (count noun condition) or “This is *acorp* my box” (preposition condition). Subjects in the count noun condition generalized the new word to objects of the same shape, ignoring the object’s location, whereas subjects in the preposition condition generalized it to objects of any shape, as long as they were in approximately the same location as the original (the top region of the box).²⁰

While these findings are compatible with the claim that children’s hypotheses about the meaning of a new preposition are constrained by their obedience to domain-specific restrictions on what can be relevant to a closed-class spatial word, they are not compelling evidence. The subjects had, after all, already learned a number of English prepositions for which the shape of the figure is unimportant, so they may have been influenced by a *learned* language-specific bias to disregard shape in hypothesizing a meaning for a new preposition.²¹ Whether the claimed biases exist prior to linguistic experience is, then, still uncertain.²²

In hypothesizing about constraints on the meanings of spatial morphemes, and constraints on children in learning them, researchers have concentrated on closed-class spatial words—it is agreed that spatial *verbs*, as open-class items, can incorporate a wide range of information about the shape, properties, position, and even identity of figure and ground objects, and about the manner of motion (Landau and Jackendoff 1993, 235–236; Talmy 1983, 273). Following the logic of “constraints” argumentation, children’s hypothesis space about closed-class spatial morphemes should therefore be more constrained than their hypothesis space about spatial verbs, since spatial verbs—especially in languages that rely heavily on them, like Korean—are sensitive to the same things that spatial prepositions are sensitive to, and a lot more besides.²³ Because the advantage of built-in constraints is supposed to be that they enable learners to quickly home in on a word’s meaning without having to sift endlessly through all the things that could conceivably be relevant, it seems that children should have an easier time arriving at the meanings of closed-class spatial morphemes (more constrained) than of spatial verbs (more open).

This is an empirical question, and one that can be examined by comparing, for example, whether children acquiring English learn the meanings of spatial particles more quickly than children acquiring Korean learn the meanings of roughly comparable spatial verbs. But in Choi’s and my studies, children learning Korean were just as fast at approximating the adult meanings of common spatial verbs used to encode actions of joining and separation as children learning English were at approximating the adult meanings of English particles used to encode the same actions (cf. figures 10.3 and 10.4). And this is true even though a number of the Korean children’s early verbs incorporated shape or object-related information such

as “figure is a clothing item,” “ground is the head/the trunk/the feet” (Choi and Bowerman 1991, 116).

It was, then, apparently no harder for children to figure out the meanings of putatively less constrained spatial verbs than of more constrained closed-class spatial morphemes. This outcome casts doubt on what these domain-specific constraints are buying for the child, and whether they are really needed in our theory of acquisition.

10.4.2.2 Does Learning Spatial Words Involve Bundling Semantic Primitives? Regardless of whether children acquiring closed-class spatial morphemes are assisted by domain-specific constraints, we can still ask whether the task of formulating the meanings of spatial words is correctly seen as a process of assembling semantic primitives into the right configurations. The appeal to semantic primitives has a long history in the study of language acquisition—a particularly influential statement of this position was E. V. Clark’s (1973b) Semantic Features Hypothesis, which held that the development of a word’s meaning is a process of adding semantic components one by one until the adult meaning of the word has been reached. Clark’s approach was discarded after extensive testing and analysis, even by Clark herself (1983), and for good reason—various predictions made by the theory were simply not met (see Richards 1979 and Carey 1982 for reviews and discussions).

In an analysis of what went wrong, Carey (1982, 367) makes an important point for our purposes: many candidate semantic features are “theory-laden”—they “represent a systematization of knowledge, the linguistic community’s theory building. As such, they depend upon knowledge unavailable to the young child, and they are therefore not likely candidates for developmental primitives” (see also Gopnik 1988 and Murphy and Medin 1985 for related arguments).

Illustrating with an example from the domain of space, Carey points out that the component [tertiary (extent)]—proposed by Bierwisch (1967) as one of a set of semantic features (along with [primary] and [secondary]) needed to distinguish *long*, *tall*, *wide*, and *thick*—is highly abstract. It is implausible, she suggests, that young children start out with a notion of [tertiary] that allows them to make sense of the use of the word *thick* in such diverse contexts as the thickness of a door, the thickness of an orange peel, and the thickness of a slice of bread. More likely, they at first understand what *thick* picks out in each of these contexts independently, and only later extract what these various uses of *thick* have in common to arrive at the feature [tertiary]. A similar analysis is applied to the word *tall* by Carey (1978) and Keil and Carroll (1980): at first children learn how to use *tall* in the context of specific referents (e.g., building: ground up; person: head to toe), and only later extract the abstract features (e.g., [spatial extent] [vertical]) that unites these uses. According to this critique, then, semantic features are the *outcome* of a lengthy developmental process—

the “lexical organizers” (Carey 1978) that children extract from words to make sense of their use across contexts—not the elements in terms of which learners analyze their experience to begin with.

Carey’s criticism of semantic primitives can be seen as related to the problem of category structure that has preoccupied us throughout this chapter. Proposed primitives are usually designated with words of a particular language, often English. Although authors may insist that they do not intend their primitives to be identical with the meanings of words in any actual language, it is not clear what they *do* in fact intend them to mean. Each language offers a different idea of what some candidate primitive is, and the child must discover this view.

Consider, for example, *support*. Does this candidate primitive include support from all directions, as in English? (cf. “The pillars support the roof,” “The drunkard supported himself by leaning against the wall,” “The actor was supported by invisible wires as he flew across the stage”). Or is it restricted to support from below, like the closest equivalent to the English word *support* in German, *stützen*? Interestingly, these two notions of support are closely aligned with the meaning of ‘on’ morphemes in the two languages: English *on* is indifferent to the orientation of the supporting surface, whereas German *auf* ‘on’ is largely restricted to support from below. Figuring out what ‘support’ is, then, is not entirely a matter of analyzing the circumstances under which objects do and do not fall—it also requires discovering how ‘support’ is conceptualized in one’s language.

Invoking semantic primitives to explain the acquisition of spatial morphemes has, in the end, a lulling effect—it makes us think we understand the acquisition process better than we do. To the extent that languages differ in what counts as ‘support’, as ‘containment’ (or ‘inclusion’), as a ‘plane’, a ‘point’ or a ‘volume’, and so on, these concepts cannot serve as the ultimate building blocks out of which children construct their meanings. Still left largely unresolved, then, is one of most recalcitrant puzzles of human development: how children go beyond their processing of particular morphemes in particular contexts—for example, “(this) cup *on* (this) table”, “(this) picture *on* (this) wall”—to a more abstract understanding of what the morphemes mean.

To conclude, I have argued that the existence of crosslinguistic variation in the semantic packaging of spatial notions creates a complex learning problem for the child. Even if learners begin by mapping spatial morphemes directly onto pre-compiled concepts of space—which is not at all obvious—they cannot get far in this way; instead, they must work out the meanings of the forms by observing how they are distributed across contexts in fluent speech. Learners’ powers of observation appear to be very acute, since their spatial semantic categories show remarkable language specificity by as early as seventeen to twenty months of age. Current

theories about the acquisition of spatial words do not yet dispel the mystery surrounding this feat. In our attempts to get a better grip on the problem, evidence from children learning different languages will continue to play an invaluable role.

Acknowledgments

I am grateful to Paul Bloom, Mary Peterson, and David Wilkins for their comments on an earlier draft of this chapter, and to Soonja Choi, Lourdes de León, Dedre Gentner, Eric Pederson, Dan Slobin, Len Talmy, and David Wilkins for the many stimulating discussions I have had with them over the years about spatial semantic organization. For judgments about their languages discussed in section 10.2, I am grateful to Magdalena Smoczyńska (Polish); Susana Lopez (Castilian Spanish); Riikka Alanen, Olli Nuutinen, Saskia Stössel-Deschner, and Erling Wande (Finnish); Soonja Choi (Korean); and many colleagues at the Max Planck Institute for Psycholinguistics (Dutch).

Notes

1. These examples are taken from diary records of my daughter E (cf. Bowerman 1978, 1980; Choi and Bowerman 1991).
2. Of course, the idea that human beings apprehend space with a priori categories of mind has a much older philosophical tradition.
3. David Wilkins (personal communication) suggests that Arrernte, an Arandic language of Central Australia, may instantiate the fifth logical possibility—grouping (a) and (b) together (on grounds that both the cup and the apple are easily grasped and moved independently—both covered by a general locative morpheme) and treating (c) differently (on grounds that the handle, being tightly attached, cannot be moved without moving the whole door).
4. A similar but more general point is made by Schlesinger (1977), who argues that languages depend on many categories that are not needed and will not be constructed purely in the course of nonlinguistic cognitive development. In a related point, Olson (1970, 188–189) notes that “linguistic decisions require information...of a kind that had not previously been selected, or attended, or perceived, because there was no occasion to look for it.”
5. Some of these crosslinguistic differences were identified in the course of typological research I conducted together with Eric Pederson on how languages express static topological spatial relations (Bowerman and Pederson 1992).
6. Some analysts have considered constructions like “the scissors have butter”, “the handle of the kitchen door”, and “the scissors are buttery” to be underlyingly spatial (see Lyons 1967 on possessive constructions and Talmy 1972 on attributive adjectives like *buttery* and *muddy*). The question remains, however, why some languages permit only these descriptions of certain relationships between entities, while others also readily describe them with overtly spatial characterizations.
7. Finnish takes the same perspective as Dutch on which is figure and which is ground, but instead of locating the hands/tree “under” the paint/ivy, Finnish locates them *in* the paint/ivy (paint/ivy-*ssa*). An English alternative that at first glance might seem comparable to the

Dutch/Finnish construction is the passive, for example, "The tree is covered by/with/in ivy." This sentence does allow the "covered" entity to be the subject of the sentence, but the verb *cover* still assigns the role of figure to the coverer (the ivy) and the role of ground to the covered (the tree) (cf. "ivy covers the tree"), and the covered entity can be gotten into subject position only by passivization.

8. To decouple the patently important question of how speakers come to control the semantic categories of their language from the loaded Whorfian issue, Slobin (1987) has coined the expression "thinking for speaking."

9. Here and subsequently, the reader should keep in mind that the English glosses given for the Korean verbs serve only as rough guides to their meaning. The actual meanings do not in fact correspond to the meanings of any English words, and can only be inferred on the basis of careful analysis of the situations in which the words are used.

10. The English data came from detailed diary records of my two daughters from the start of the one-word stage, supplemented by the extensive literature on the early use of English path particles reviewed in section 10.1.2. Two sets of Korean data were used: (1) from 4 children videotaped every 3–4 weeks by Choi from 14 months old to 24–28 months old; and (2) from 4 additional children taped by Choi, Pat Clancy, and Youngjoo Kim every 2 to 4 weeks from 19–20 months old to 25–34 months old. We are grateful to Clancy and Kim for generously sharing their data.

11. We adopted this procedure rather than, for example, asking children to describe actions we had already performed because several studies have shown that children first produce change-of-state predicates, including spatial morphemes, either as requests for someone to carry out an action or when they themselves are about to perform an action—the words seem to function to announce *plans* of intended action (Gopnik 1980; Gopnik and Meltzoff 1986; Huttenlocher, Smiley, and Charney 1983). If a child failed to respond after several attempts to elicit a request/command for an about-to-be-performed action, we would go ahead and perform it and then ask the child, "What did I do?" For adults, who caught on immediately to what kind of response we were looking for, we often soon abandoned the command scenario and simply displayed the actions we wanted labeled.

12. Degrees of similarity can also be computed—for example, two actions both called "take out" can be regarded as entirely similar, two called "take out" and "pull out" are partially similar, and two called "take out" and "put on" are not at all similar. For certain kinds of analyses, it is useful to organize each subject's data as a similarity matrix showing whether, for each action paired with each other action, the subject used the same (e.g., put a 1 in the cell), similar (e.g., .5) or different (0) expressions; this allows us to disregard the fact that the expressions themselves are different across languages, as, of course, is the number of expressions used by different subjects.

13. In the quantitative analyses of the data, Choi and I have been joined in our collaboration by James Boster (see, for example, Boster 1991 for a relevant comparative analysis applied to the nonlinguistic classification of mammals by children and adults in two cultures).

14. Actions that fall outside of all the circles in a figure were responded to either very inconsistently (i.e., no "dominant response" could be identified) or (in the case of the children) received few relevant verbal responses. The use of solid versus dotted lines for the circles has no special significance—it just makes it easier to visually distinguish overlapping categories.

15. This analysis involved comparing the similarity matrices (cf. note 12) of speakers in different groups. We first constructed an aggregate matrix for the adult speakers of each language. We then correlated the similarity matrix of each child with the aggregate adult matrix for each language and with the matrices of all the other children. (The cells of the matrices, e.g., action 1 paired with action 2, action 1 paired with action 3, etc., constitute the list of variables over which the correlation is carried out.) Finally, we tested whether the children in the youngest age group for each language correlated significantly better with the adult aggregate matrix for their own language, or with same-age children speaking each of the other two languages. (We also assessed their correlation with adult speakers of each of the other two languages.)

16. The only action to which both *out* and *off* were applied (by different children) was taking a piece out of a jigsaw puzzle, and this is readily understandable: the “container” (the piece-shaped hole in the wooden base) was extremely shallow in this case, so it is probably unclear to learners whether to construe it as a “container” or a “surface” (see section 10.2.2.3 on the problem of learning the conventional conceptualization of particular objects). (For the converse action of putting the piece into the puzzle, eight children said “in” and only one said “on.”) Another action presenting a similar construal problem was “put log on train car.” The train car in question had short poles sticking up, two on a side, to keep the tiny logs from falling off. Despite the poles, 27 of the 30 adults across the three languages conceptualized this situation as one of placing a log ‘on’ a horizontal supporting surface (English *on* (*top*), Korean *nohta* ‘put on horizontal supporting surface’, Dutch (*boven*) *op* ‘on (top)’). But of the 30 children in the youngest age group across the three languages, only 5 used these words; their most typical response was *in* (English and Dutch) and *nehta* ‘put loosely in’ or *kkita* ‘fit tightly’ (Korean).

17. This pattern in Dutch also argues against a hypothesis that several people have suggested to me: that English-speaking children may learn *on* and *off* in connection with clothing as a separate, self-contained pair of meanings, so these uses should not be analyzed as part of a more general pattern of associating *on* and *off* with surface-oriented relationships. The clothing use of *uit* ‘out’ seems to interact in the course of development with other uses of *uit* in Dutch children, so this argument is incorrect for Dutch, and by extension probably also for English. (See Choi and Bowerman 1991, 110–113, for other empirical arguments against the proposal that there is extensive homonymy or polysemy in children’s early acquisition of spatial words.)

18. A similar example is provided by children learning Tzotzil Mayan (Bowerman, de León, and Choi 1995). One of the earliest spatial morphemes for “joining” actions that these children acquire is the verb *xoj*, and they seem to use it, before age 2, for a range of events that corresponds neither to the English child categories *in* or *on* nor to the Korean child category *kkita* ‘fit tightly’. In adult speech, the root *xoj* picks out a configuration of a long thing encircled by a ring-shaped thing, and can be used, for example, to describe either putting a pole through a ring or a ring over a pole. When adult Tzotzil speakers were informally tested on the same set of spatial actions Choi and I used in the elicited production described in section 10.3.2, they used *xoj* for putting tight- and loose-fitting rings on poles and occasionally for putting on clothing (the ring-and-pole configuration is instantiated by the encirclement of arms and legs by sleeves and pantlegs, feet by socks and shoes, and head by wool cap). (Adults more often described donning clothing with a verb that means “put on clothing.”) Very small Tzotzil

children also used *xoj* for putting rings on poles and (more frequently than adults) for putting on shoes, socks, and wool hat, and, beyond these manipulations with our experimental materials, they used it for other actions conforming to or approximating the ring-and-pole configuration such as threading beads, putting a coiled rope over a peg, and putting a car into a long thin box. This range overlaps the *in* and *on* categories of English-speaking children but is more restricted than either (see figure 10.3b); it also overlaps the *kkita* 'fit tightly' and *nehta* 'put loosely in (or around)' categories of the Korean children, but, again, is different from both (cf. figure 10.4b).

19. Opinions vary on whether proposed semantic primitives are irreducible units only in their role as building blocks for meaning in language, or are also perceptual or conceptual primitives on a nonlinguistic level. The remarks in this section apply either way.

20. In a different approach to whether a learner constrained by domain-specific sensitivities can acquire the meanings of spatial words across languages, Regier (1995) equipped a connectionist model with specific structural devices motivated by neurobiological and psychophysical evidence on the human visual system. Presented with frame-by-frame films instantiating the meaning of spatial words, the model was able to home in on schematized versions of several spatial categories in English, Mixtec (cf. (3) in section 10.2.1), and Russian. Whether such a model can learn to classify a more realistic set of spatial situations, including diverse objects in all their complicated functional relationships, remains to be seen.

21. A study by Imai and Gentner (1993) shows that biases in what learners think a novel word means can indeed arise through experience with the properties of a particular language. These investigators showed that English- and Japanese-speaking subjects, both child and adult, agreed in assuming that a word introduced in connection with a complex object referred to the object, and that a word introduced in the context of a gooey substance referred to the substance. But they differed in their assumptions about a word introduced in the context of a novel simple object, such as a cork pyramid. English children and adults assumed that the word referred to same-shaped objects regardless of material, whereas their Japanese counterparts assumed that it referred to entities made of the same material, regardless of shape. Imai and Gentner had predicted this outcome on the basis of Lucy's (1992) hypotheses about differences in the meanings of nouns in languages that do and do not have numeral classifiers.

22. Also uncertain is the possible cause of these biases. For example, if children are biased against detailed shape information in learning closed-class spatial words, is this because the words are *spatial*, or because they are *closed-class*? (As Talmy 1983, 1985 has argued, closed-class morphemes have highly schematic meanings across a wide range of semantic domains.)

23. Pinker (1989, 172–176) has proposed a set of meaning components particularly relevant for learning verbs, but this set is far less constrained than the set relevant for closed-class spatial morphemes. (It includes "the main event": a state or motion; path, direction, and location; causation; manner; properties of a theme or actor; and temporal distribution (aspect and phase); purpose, etc.) Nor are the components supposed to capture everything that can be important to the meaning of a verb, but only those aspects of meaning that can be relevant to a verb's syntactic behavior.

References

- Antell, S. E. G., and Caron, A. J. (1985). Neonatal perception of spatial relationships. *Infant Behavior and Development*, 8, 15–23.
- Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition*, 23, 21–41.
- Baillargeon, R. (1987). Object permanence in 3.5- and 4.5-month-old infants. *Developmental Psychology*, 23, 655–664.
- Baillargeon, R., Graber, M., DeVos, J., and Black, J. C. (1990). Why do young infants fail to search for hidden objects? *Cognition*, 36, 255–284.
- Behl-Chadha, G., and Eimas, P. D. (1995). Infant categorization of left-right spatial relations. *British Journal of Developmental Psychology*, 13, 69–79.
- Berman, R. A., and Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum.
- Bierwisch, M. (1967). Some semantic universals of German adjectivals. *Foundations of Language*, 3, 1–36.
- Bierwisch, M. (1981). Basic issues in the development of word meaning. In W. Deutch (Ed.), *The child's construction of language*, 341–387. New York: Academic Press.
- Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.
- Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax*. The Hague: Mouton.
- Bomba, P. C. (1984). The development of orientation categories between 2 and 4 months of age. *Journal of Experimental Child Psychology*, 37, 609–636.
- Boster, J. (1991). The information economy model applied to biological similarity data. In L. Resnick, J. Levine, and S. D. Teasley (Eds.), *Socially shared cognition*, 203–225. Washington, DC: American Psychological Association.
- Bower, T. G. R. (1982). *Development in infancy*, 2d ed. San Francisco: Freeman.
- Bowerman, M. (1973). *Early syntactic development: A cross-linguistic study with special reference to Finnish*. Cambridge: Cambridge University Press.
- Bowerman, M. (1978). The acquisition of word meaning: An investigation into some current conflicts. In N. Waterson and C. Snow (Eds.), *The development of communication*, 263–287. New York: Wiley.
- Bowerman, M. (1980). The structure and origin of semantic categories in the language-learning child. In M. L. Foster and S. H. Brandes (Eds.), *Symbol as sense: New approaches to the analysis of meaning*, 277–299. New York: Academic Press.
- Bowerman, M. (1985). What shapes children's grammars? In D. I. Slobin (Ed.), *The cross-linguistic study of language acquisition*. Vol. 2, *Theoretical issues*, 1257–1319. Hillsdale, NJ: Lawrence Erlbaum.

Bowerman, M. (1989). Learning a semantic system: What role do cognitive predispositions play? In M. L. Rice and R. L. Schiefelbusch (Eds.), *The teachability of language*, 133–169. Baltimore: Brooks.

Bowerman, M. (1993). Typological perspectives on language acquisition: Do crosslinguistic patterns predict development? In E. V. Clark (Ed.), *The proceedings of the Twenty-fifth Annual Child Language Research Forum*, 7–15. Stanford CA: Center for the Study of Language and Information.

Bowerman, M. (1994). From universal to language-specific in early grammatical development. *Philosophical Transactions of the Royal Society, London, B346*, 37–45.

Bowerman, M. (1996). The origins of children's spatial semantic categories: Cognitive versus linguistic determinants In J. J. Gumperz and S. C. Levinson (Eds.), *Rethinking linguistic relativity*, 145–176. Cambridge: Cambridge University Press.

Bowerman, M., and Choi, S. (1994). Linguistic and nonlinguistic determinants of spatial semantic development. Paper presented at the Boston University Conference on Language Development, January.

Bowerman, M., de León, L., and Choi, S. (1995). Verbs, particles, and spatial semantics: Learning to talk about spatial actions in typologically different languages. In E. V. Clark (Ed.), *Proceedings of the Twenty-seventh Annual Child Language Research Forum*, 101–110. Stanford, CA: Center for the Study of Language and Information.

Bowerman, M., and Pederson, E. (1992). Crosslinguistic perspectives on topological spatial relationships. Paper presented at the annual meeting of the American Anthropological Association, San Francisco, December.

Brown, P. (1994). The INs and ONs of Tzeltal locative expressions: The semantics of static descriptions of location. *Linguistics*, 32, 743–790.

Brown, R. W. (1958). *Words and things*. New York: Free Press.

Brown, R. W. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Brugman, C. (1983). *The use of body-part terms as locatives in Chalcatongo Mixtec*, 235–290. Report no. 4 of the Survey of California and Other Indian Languages. Berkeley: University of California.

Brugman, C. (1984). Metaphor in the elaboration of grammatical categories in Mixtec. Unpublished manuscript, Linguistics Department, University of California, Berkeley.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, and G. A. Miller (Eds.), *Linguistic theory and psychological reality*, 264–293. Cambridge, MA: MIT Press.

Carey, S. (1982). Semantic development: The state of the art. In E. Wanner and L. Gleitman (Eds.), *Language acquisition: The state of the art*, 347–389. Cambridge: Cambridge University Press.

Caron, A. J., Caron, R. F., and Antell, S. E. (1988). Infant understanding of containment: An affordance perceived or a relationship conceived? *Developmental Psychology*, 24, 620–627.

- Choi, S., and Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41, 83–121.
- Cienki, A. J. (1989). *Spatial cognition and the semantics of prepositions in English, Polish, and Russian*. Munich: Sagner.
- Clark, E. V. (1973a). Nonlinguistic strategies and the acquisition of word meanings. *Cognition*, 2, 161–182.
- Clark, E. V. (1973b). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*, 65–110. New York: Academic Press.
- Clark, E. V. (1983). Meanings and concepts. In J. H. Flavell and E. M. Markman (Eds.), *Mussen handbook of child psychology*. Vol. 3, *Cognitive development and the acquisition of language*, 787–840. New York: Academic Press.
- Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*, 27–63. New York: Academic Press.
- Colombo, J., Laurie, C., Martelli, T., and Hartig, B. (1984). Stimulus context and infant orientation discrimination. *Journal of Experimental Child Psychology*, 37, 576–586.
- DeValois, R., and DeValois, K. (1990). *Spatial vision*. Oxford: Oxford University Press.
- Freeman, N. H., Lloyd, S., and Sinha, C. G. (1980). Infant search tasks reveal early concepts of containment and canonical usage of objects. *Cognition*, 8, 243–262.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), *Language development*. Vol. 2, *Language, thought, and culture*, 301–334. Hillsdale, NJ: Erlbaum.
- Gibson, E. J. (1982). The concept of affordances in development: The renaissance of functionalism. In W. A. Collins (Ed.), *The concept of development*, 55–81. Minnesota Symposia on Child Psychology, vol. 15. Hillsdale, NJ: Erlbaum.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1, 3–55.
- Goldberg, A. E. (1995). *Constructions*. Chicago: University of Chicago Press.
- Gopnik, A. (1980). The development of non-nominal expressions in 12–24-month-old children. Ph.D. diss., Oxford University.
- Gopnik, A. (1988). Conceptual and semantic development as theory change. *Mind and Language*, 3, 197–216.
- Gopnik, A., and Meltzoff, A. N. (1986). Words, plans, things, and locations: Interactions between semantic and cognitive development in the one-word stage. In S. A. Kuczaj II and M. D. Barrett (Eds.), *The development of word meaning*, 199–223. Berlin: Springer.
- Griffiths, P., and Atkinson, M. (1978). A 'door' to verbs. In N. Waterson and C. Snow (Eds.), *The development of communication*, 311–331. New York: Wiley.

Gruendel, J. (1977). Locative production in the single-word utterance period: Study of "up-down," "on-off," and "in-out." Paper presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans, March.

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

Heine, B. (1989). Adpositions in African languages. *Linguistique Africaine*, 2, 77–127.

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

Hill, C. A. (1978). Linguistic representation of spatial and temporal orientation. *Berkeley Linguistics Society*, 4, 524–538.

Huttenlocher, J., Smiley, P., and Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review*, 90, 72–93.

Imai, M., and Gentner, D. (1993). Linguistic relativity vs. universal ontology: Crosslinguistic studies of the object/substance distinction. In *Proceedings of the Chicago Linguistic Society*, 29.

Johnston, J. R. (1984). Acquisition of locative meanings: *Behind* and *in front of*. *Journal of Child Language*, 11, 407–422.

Johnston, J. R. (1985). Cognitive prerequisites: The evidence from children learning English. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition*. Vol. 2, 961–1004. Hillsdale, NJ: Erlbaum.

Johnston, J. R., and Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. *Journal of Child Language*, 6, 529–545.

Keil, F. C. (1979). The development of the young child's ability to anticipate the outcomes of simple causal events. *Child Development*, 50, 455–462.

Keil, F. C. (1990). Constraints on constraints: Surveying the epigenetic landscape. *Cognitive Science*, 14, 135–168.

Keil, F. C., and Carroll, J. J. (1980). The child's acquisition of "tall": Implications for an alternative view of semantic development. *Papers and Reports on Child Language Development*, 19, 21–28.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Landau, B., and Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–238.

Landau, B., and Stecker, D. S. (1990). Objects and places: Syntactic geometric representations in early lexical learning. *Cognitive Development*, 5, 287–312.

Langacker, R. W. (1987). *Foundations of cognitive grammar*. Vol. 1, *Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Leopold, W. (1939). *Speech development of a bilingual child*. Vol. 1. Evanston, IL: Northwestern University Press.

- Levine, S. C., and Carey, S. (1982). Up front: The acquisition of a concept and a word. *Journal of Child Language*, 9, 645–657.
- Levinson, S. C. (1994). Vision, shape, and linguistic description: Tzeltal body-part terminology and object description. *Linguistics*, 32, 791–855.
- Levinson, S. C. (in press). From outer to inner space: Linguistic categories and nonlinguistic thinking. In J. Nuyts and E. Pederson (Eds.), *Linguistic and conceptual representation*. Cambridge: Cambridge University Press.
- Levinson, S. C., and Brown, P. (1994). Immanuel Kant among the Tenejapans: Anthropology as Empirical Philosophy. *Ethos*, 22, 3–41.
- Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.
- Lyons, J. (1967). A note on possessive, existential, and locative sentences. *Foundations of Language*, 3, 390–396.
- MacLaury, R. E. (1989). Zapotec body-part locatives: Prototypes and metaphoric extensions. *International Journal of American Linguistics*, 55, 119–154.
- MacLean, D. J., and Schuler, M. (1989). Conceptual development in infancy: The understanding of containment. *Child Development*, 60, 1126–1137.
- Mandler, J. (1992). How to build a baby: II Conceptual primitives. *Psychological Review*, 99, 587–604.
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
- McCune-Nicholich, L. (1981). The cognitive bases of relational words in the single-word period. *Journal of Child Language*, 8, 15–34.
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Murphy, G. L., and Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Needham, A., and Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, 47, 121–148.
- Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review*, 81, 267–285.
- Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, 77, 257–273.
- Olson, D. R., and Bialystok, E. (1983). *Spatial cognition: The structure and development of the mental representation of spatial relations*. Hillsdale, NJ: Erlbaum.
- Parisi, D., and Antinucci, F. (1970). Lexical competence. In G. B. Flores D'Arcais and W. J. M. Levelt (Eds.), *Advances in psycholinguistics*, 197–210. Amsterdam: North-Holland.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Piaget, J., and Inhelder, B. (1956). *The child's conception of space*. London: Routledge and Kegan Paul.

Pieraut-Le Bonniec, G. (1987). From visual-motor anticipation to conceptualization: Reaction to solid and hollow objects and knowledge of the function of containment. *Infant Behavior and Development*, 8, 413–424.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Quinn, P. C. (1994). The categorization of above and below spatial relations by young infants. *Child Development*, 65, 58–69.

Quinn, P. C., and Bomba, P. C. (1986). Evidence for a general category of oblique orientations in four-month-old infants. *Journal of Experimental Child Psychology*, 42, 345–354.

Quinn, P. C., and Eimas, P. D. (1986). On categorization in early infancy. *Merrill-Palmer Quarterly*, 32, 331–363.

Regier, T. (1995). A model of the human capacity for categorizing spatial relations. *Cognitive Linguistics*, 6, 63–88.

Richards, M. M. (1979). Sorting out what's in a word from what's not: Evaluating Clark's semantic features acquisition theory. *Journal of Experimental Child Psychology*, 27, 1–47.

Schlesinger, I. M. (1977). The role of cognitive development and linguistic input in language development. *Journal of Child Language*, 4, 153–169.

Sinha, C., Thorseng, L. A., Hayashi, M., and Plunkett, K. (1994). Comparative spatial semantics and language acquisition: Evidence from Danish, English, and Japanese. *Journal of Semantics*, 11, 253–287.

Sitskoorn, M. M., and Smitsman, A. W. (1995). Infants' perception of dynamic relations between objects: Passing through or support? *Developmental Psychology*, 31, 437–447.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson and D. I. Slobin (Eds.), *Studies of child language development*, 175–208. New York: Holt, Rinehart, and Winston.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition*. Vol. 2, *Theoretical issues*, 1157–1256. Hillsdale, NJ: Erlbaum.

Slobin, D. I. (1987). Thinking for speaking. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 13, 435–444.

Spelke, E. S., Breinlinger, K., Macomber, J., and Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 99, 605–632.

Spelke, E. S., Katz, G., Purcell, S. E., Ehrlich, S. M., and Breinlinger, K. (1994). Early knowledge of object motion: Continuity and inertia. *Cognition*, 51, 107–130.

Talmy, L. (1972). Semantic structures in English and Atsugewi. Ph.D. diss. University of California, Berkeley.

- Talmy, L. (1983). How language structures space. In H. Pick and L. Acredolo (Eds.), *Spatial orientation: Theory, research, and application*, 225–282. New York: Plenum.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical form. In T. Shopen (Ed.), *Language typology and syntactic description*. Vol. 3, *Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.
- Talmy, L. (1991). Path to realization: A typology of event conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, 17, 480–519. [Supplement in the *Buffalo Papers in Linguistics*, 91-01, 182–187.]
- von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224, 1260–1262.
- Whorf, B. L. (1956). *Language, thought, and reality*. Edited by J. B. Carroll. Cambridge, MA: MIT Press.
- Wierzbicka, A. (1972). *Semantic primitives*. Frankfurt: Athenäum.
- Wilkins, D., and Senft, G. (1994). A man, a tree—and forget about the pigs: Space games, spatial reference and an attempt to identify functional equivalents across languages. Paper presented at the Nineteenth International L.A.U.D. Symposium on Language and Space, Duisburg, March.

Chapter 11

Space to Think

Philip N. Johnson-Laird

11.1 Introduction

Perception is the transformation of local information at the sensorium into a mental model of the world at a distance, thinking is the manipulation of such models, and action is guided by its results. This account of human cognition goes back to the remarkable Scottish psychologist, Kenneth Craik (1943), and it has provided both a program of research for the study of human cognition and a central component of the theory of mental representations. Thus the final stage of visual perception, according to Marr (1982), delivers a three-dimensional model of the world, which the visual system has inferred from the pattern of light intensities falling on the retinas. Mental models likewise underlie one account of verbal comprehension: to understand discourse is, on this account, to construct a mental model of the situation that it describes (see, for example, Johnson-Laird 1983; Garnham 1987). The author and his colleagues have developed this account into a theory of reasoning—both inductive and deductive—in which thinkers reason by manipulating models of the world (see, for example, Johnson-Laird and Byrne 1991).

The idea of mental models as the basis for deductive thinking has its origins in the following idea:

Consider the inference

The box is on the right of the chair,
The ball is between the box and the chair,
Therefore, the ball is on the right of the chair.

The most likely way in which such an inference is made involves setting up an internal representation of the scene depicted by the premises. This representation may be a vivid image or a fleeting abstract delineation—its substance is of no concern. The crucial point is that its formal properties mirror the spatial relations of the scene so that the conclusion can be read off in almost as direct a fashion as from an actual array of objects. It may be objected, however, that such a depiction of the premises is unnecessary, that the inference can be made

by an appeal to general principles, or rules of inference, which indicate that items related by *between* must be collinear, etc. However, this view—that relational terms are tagged according to the inference schema they permit—founders on more complex inferences. An inference of the following sort, for instance, seems to be far too complicated to be handled without constructing an internal representation of the scene:

The black ball is directly beyond the cue ball. The green ball is on the right of the cue ball, and there is a red ball between them.

Therefore, if I move so that the red ball is between me and the black ball, then the cue ball is on my left.

Even if it is possible to frame inference schema that permit such inferences to be made without the construction of an internal representation, it is most unlikely that this approach is actually adopted in making the inference. (Johnson-Laird 1975, 12–13)

This passage captures the essence of the model theory of deduction, but the intuition that spatial inferences are made by imagining spatial scenes turned out *not* to be shared by all investigators.

Twenty years have passed since the argument above was first formulated, and so the aim of this chapter is, in essence, to bring the story up to date. It contrasts the model theory with an account based on formal rules of inference, and it presents evidence that spatial reasoning is indeed based on models. It then argues that spatial models may underlie other sorts of thinking—even thinking that is not about spatial relations. It presents some new results showing that individuals often reason about temporal relations by constructing quasi-spatial models. Finally, it demonstrates that one secret in using diagrams as an aid to thinking is that their spatial representations should make alternative possibilities explicit.

11.2 Propositional Representations and Mental Models

What does one mean by a mental model? The essence of the answer is that its structure corresponds to the structure of what it represents. A mental model is accordingly similar in structure to a physical model of the situation, for example, a biochemist's model of a molecule, or an architect's model of a house. The parts of the model correspond to the relevant parts of the situation, and the structural relations between the parts of the model are analogous to the structural relations in the world. Hence, individual entities in the situation will be represented as individuals in the model, their properties will be represented by their properties in the model, and the relations among them will be represented by relations among them in the model. Mental models are partial in that they represent only certain aspects of the situation, and they thus correspond to many possible states of affairs, that is, there is a many-to-one mapping from situations in the world to a model. Images, too, have these properties,

but models need not be visualizable, and unlike images, they may represent several distinct sets of possibilities. These abstract characterizations are hard to follow, but they can be clarified by contrasting mental models with so-called propositional representations.

To illustrate a propositional representation, consider the assertion:

A triangle is on the right of a circle.

Its propositional representation relies on some sort of predicate argument structure, such as the following expression in the predicate calculus:

$(\exists x)(\exists y)(\text{Triangle}(x) \ \& \ \text{Circle}(y) \ \& \ \text{Right-of}(x, y)),$

where $\exists$ denotes the existential quantifier “for some” and the variables range over individuals in the domain of discourse, i.e. the situation that is under description. The expression can accordingly be paraphrased in “Loglish”—a hybrid language spoken only by logicians—as follows:

For some x and for some y , such that x is a triangle and y is a circle, x is on the right of y .

The information in the further assertion

The circle is on the right of a line

can be integrated to form the following expression representing both assertions:

$(\exists x)(\exists y)(\exists z)(\text{Triangle}(x) \ \& \ \text{Circle}(y) \ \& \ \text{Line}(z) \ \& \ \text{Right-of}(x, y) \ \& \ \text{Right-of}(y, z)).$

A salient feature of this representation is that its structure does *not* correspond to the structure of what it represents. The key component of the propositional representation is

$\text{Right-of}(x, y) \ \& \ \text{Right-of}(y, z),$

in which there are four tokens representing variables. In contrast, the situation itself has three entities in a particular spatial relation. Hence, a mental model of the situation must have the same structure, which is depicted in the following diagram:



where the horizontal dimension corresponds to the left-to-right dimension in the situation. In what follows, such diagrams are supposed to depict mental models, and will often be referred to as though they were mental models. Each token in the present mental model has a property corresponding to the shape of the entity it represents, and the three tokens are in a spatial relation corresponding to the relation between the three entities in the situation described by the assertions. In the case of such a

spatial model, a critical feature is that elements in the model can be accessed and updated in terms of parameters corresponding to axes.

The process of inference for propositional representations calls for a system based on rules, and psychologists have proposed such systems for spatial inference based on formal rules of inference (see, for example, Hagert 1984; Ohlsson 1984). Hence, in order to infer from the premises above the valid conclusion

A triangle is on the right of a line,

it is necessary to rely on a statement of the transitivity of "on the right of":

$(\forall x)(\forall y)(\forall z)((\text{Right-of}(x, y) \ \& \ \text{Right-of}(y, z)) \rightarrow \text{Right-of}(x, z)),$

where $\forall$ denotes the universal quantifier "for any" and $\rightarrow$ denotes material implication ("if . . . , then . . ."). With this additional premise (a so-called meaning postulate) and a set of rules of inference for the predicate calculus, the conclusion can be derived in the following chain of inferential steps.

The premises are

- (1) $(\exists x)(\exists y)(\text{Triangle}(x) \ \& \ \text{Circle}(y) \ \& \ \text{Right-of}(x, y))$
- (2) $(\exists y)(\exists z)(\text{Circle}(y) \ \& \ \text{Line}(z) \ \& \ \text{Right-of}(y, z))$
- (3) $(\forall x)(\forall y)(\forall z)((\text{Right-of}(x, y) \ \& \ \text{Right-of}(y, z)) \rightarrow \text{Right-of}(x, z))$

The proof calls for the appropriate instantiations of the quantified variables, that is, one replaces the quantified variables by constants denoting particular entities:

- (4) $(\exists y)(\text{Triangle}(a) \ \& \ \text{Circle}(y) \ \& \ \text{Right-of}(a, y))$ [from (1)]
- (5) $(\text{Triangle}(a) \ \& \ \text{Circle}(b) \ \& \ \text{Right-of}(a, b))$ [from (4)]
- (6, 7) $(\text{Circle}(b) \ \& \ \text{Line}(c) \ \& \ \text{Right-of}(b, c))$ [from (2)]

There are constraints on the process of instantiating variables that are existentially quantified, but universal quantifiers range over all entities in the domain, and so the meaning postulate can be freely instantiated as follows:

- (8–10) $((\text{Right-of}(a, b) \ \& \ \text{Right-of}(b, c)) \rightarrow \text{Right-of}(a, c))$ [from (3)]

The next steps use formal rules of inference for the connectives. A rule for conjunction stipulates that given a premise of the form $(A \ \& \ B)$, where A and B can denote compound assertions of any degree of complexity, one can derive the conclusion B . Hence one can detach part of line 5 as follows:

- (11) $\text{Right-of}(a, b)$ [from (5)]

and part of line 7 as follows:

(12) Right-of(b, c) [from (7)]

Another rule allows any two assertions in separate lines to be conjoined, that is, given premises of the form A, B , one can derive the conclusion $(A \& B)$. This rule allows a conjunction to be formed from the previous two lines in the derivation:

(13) (Right-of(a, b) & Right-of(b, c)) [from (11), (12)]

This assertion matches the antecedent of line 10, and a rule known as “modus ponens” stipulates that given any premises of the form $(A \rightarrow B), A$, one can derive the conclusion B . The next step of the derivation proceeds accordingly:

(14) Right-of(a, c) [from (10), (13)]

The rules for conjunction allow the detachment of propositions from previous lines and their assembly in the following conclusion:

(15–18) ((Triangle(a) & Line(c)) & Right-of(a, c)) [from (5), (7), (14)]

Finally, this propositional representation can be translated back into English:

Therefore, the triangle is on the right of the line.

The process of inference for models is different. The theory relies on the following simple idea. A valid deduction, by definition, is one in which the conclusion must be true if the premises are true. Hence what is needed is a model-based method to test for this condition. Assertions can be true in indefinitely many different situations, and so it is out of the question to test that a conclusion holds true in all of them. But testing can be done in certain domains precisely because a mental model can stand for indefinitely many situations. Here, in principle, is how it is done for spatial inferences. Consider, again, the example above:

A triangle is on the right of a circle.

The circle is on the right of a line.

The assertions say nothing about the actual distances between the objects. Instead of trying to envisage all the different possible situations that satisfy these premises, a mental model leaves open the details and captures only the structure that all the different situations have in common:

| ○ △

where the left-to-right axis corresponds to the left-right axis in space, but the distances between the tokens have no significance. This model represents only the spatial sequence of the objects, and it is the only possible model of the premises, that is, no other model corresponding to a different left-to-right sequence of the three objects satisfies the premises. Now consider the further assertion:

The triangle is on the right of the line.

It is true in the model, and, because there are no other models of the premises, it *must* be true given that the premises are true. The deduction is valid, and because reasoners can determine that there are no other possible models of the premises, they can not only make this deduction but also *know* that it is valid (see Barwise 1993).

The same principles allow us to determine that an inference is invalid. Given, say, the inference

A triangle is on the right of a circle,
A line is on the right of the circle,
Therefore, the triangle is on the right of the line,
the first premise yields the model

○ △

but now when we try to add the information from the second premise, the relation between the triangle and the line is uncertain. One way to respond to such an indeterminacy is to build separate models for each possibility:

○ | △ ○ △ |

ignoring the possibility that the triangle and the line might be, say, one on top of the other. The first of these models shows that the putative conclusion is possible, but the second model is a counterexample to it. It follows that the triangle *may* be on the right of the line, but it does not follow that the triangle *must* be on the right of the line.

Does the model theory abandon the idea of propositional representations? Not at all. It turns out to be essential to have a representation of the meaning of an assertion independent of its particular realization in a model. The theory accordingly assumes that the first step in recovering the meaning of a premise is the construction of its propositional representation—a representation of the truth conditions of the premise. This representation is then used to update the set of models of the premises.

The use of mental models in reasoning has two considerable advantages over the use of formal rules. The first advantage is that it yields a decision procedure—at least for domains such as spatial reasoning that can have one, because the predicate calculus is provably without any possible decision procedure. An inference is valid if its conclusion holds in all the possible models of the premises, and it is invalid if it fails to hold in at least one of the possible models of the problems. Granted that problems remain within the capacity of working memory, then it is a simple matter to decide whether or not an inference is valid. One examines the models of the premises, and a conclusion is valid if, and only if, it is true in all of them. The situation is very

different in the case of formal rules. They have no decision procedure. Quine (1974, 75) commented on this point in contrasting a semantic decision procedure for the propositional calculus (akin in some ways to the mental model account of that domain) and an approach based on formal rules. Of the use of formal rules, he wrote: "It is inferior in that it affords no general way of reaching a verdict of invalidity; failure to discover a proof for a schema can mean either invalidity or mere bad luck." The same problem, as Barwise (1993) has pointed out, haunts psychological theories based on formal rules. The search space of possible derivations is vast, and thus such theories have to assume that reasoners explore it for a certain amount of time and then give up. Barwise remarks: "The 'search till you're exhausted' strategy gives one at best an educated, correct guess that something does not follow" (337). Models allow individuals to *know* that there is no valid conclusion.

The second advantage of mental models is that they extend naturally to inductive inferences and to the informal arguments of daily life to which it is so hard, if not impossible, to apply formal rules of inference (see, for example, Toulmin 1958). Such inferences and arguments nevertheless differ in their strength (Osherson, Smith, and Shafir 1986). The model theory implies that the strength of an inference—any inference—depends on the believability of its premises and on the proportion of models of the premises in which the conclusion is true (Johnson-Laird 1994). Hence the model theory provides a unified account of inference:

- If the conclusion holds in all possible models of the premises, it is *necessary* given the premises, that is, deductively valid.
- If it holds in most of the models of the premises, then it is *probable*.
- If it holds in some model of the premises, then it is *possible*.
- If it holds in only a few models of the premises, then it is *improbable*.
- If it holds in none of the models of the premises, then it is *impossible*, that is, inconsistent with the premises.

The theory forms a bridge between models and the heuristic approach to judgments of probability based on scenarios (see, for example, Tversky and Kahneman 1973). As the number of indeterminacies in premises increases, there is an exponential growth in the number of possible models. Hence the procedure is intractable for all but small numbers of indeterminacies. However, once individuals have constructed a model in which a highly believable conclusion holds, they tend not to search for alternative models that refute the conclusion. The theory according provides a mechanism for *inferential satisficing* (cf. Simon 1959). This mechanism accounts for the common failure to consider alternative lines of argument—a failure shown by studies of inference, both deductive (e.g., Johnson-Laird and Byrne 1991) and informal (e.g., Perkins, Allen, and Hafner 1983; Kuhn 1991), and by many real-life disasters, for

example, the operators at Three Mile Island inferred that a relief valve was leaking and overlooked the possibility that it was stuck open.

11.3 Algorithm for Spatial Reasoning Based on Mental Models

The machinery required for reasoning by model calls, not for formal rules of inference, but procedures for constructing models, formulating conclusions true in models, and testing whether conclusions are true in models. The present author has implemented computer programs that make inferences using such an algorithm for syllogisms, sentential connectives, doubly quantified assertions, and several other domains including spatial reasoning. The algorithm for spatial inferences works in the following way. The initial interpretation of the first premise

The triangle is on the right of the circle

yields a propositional representation, which is constructed by a “compositional semantics”:

$((1\ 0\ 0) \ \Delta \ \bigcirc).$

The parameters (1 0 0) specify which axes need to be incremented in order to relate the triangle to the circle (increment the right-left axis, i.e., keep adding 1 to it, as necessary; hold the front-back axis constant, i.e., increment it by 0; and hold the up-down axis constant, i.e., increment it by 0). There are no existing models of the discourse, because the assertion is first, and so a procedure is called that uses this propositional representation to build a minimal spatial representation:

$\bigcirc \quad \Delta.$

In the program, the spatial model is represented by an array. Likewise, the interpretation of the second premise

The circle is on the right of a line

yields the propositional representation

$((1\ 0\ 0) \ \bigcirc \ |).$

This representation contains an item in the initial model, and so a procedure is called that uses the propositional representation to update this model by adding the line in the appropriate position:

$| \quad \bigcirc \quad \Delta.$

Given the further, third assertion

The triangle is on the right of the line,

both items in its propositional representation occur in an existing model, and thus a procedure is called to verify the propositional representation. This procedure returns the value true, and with the proviso that the algorithm always constructs all possible models of the premises, the conclusion is therefore valid.

The algorithm has no need for a postulate capturing the transitivity of relations, such as “on the right of,” which are emergent properties of the meaning of the relation and of how it is used to construct models. This emergence of logical properties has the advantage of accounting for a puzzling phenomenon—the vagaries in everyday spatial inferences. The inferences modeled in the program are for the “deictic” interpretation of “on the right of,” that is, the relation as perceived from a speaker’s point of view. Other entities have an intrinsic right-hand side and left-hand side, for example, human beings (see Miller and Johnson-Laird 1976, section 6.1.3). Hence the following premises:

Matthew is on Mark’s right

Mark is on Luke’s right

can refer to the position of three individuals in relation to the intrinsic right-hand sides of Mark and Luke. To build a model of the spatial relation, the inferential system needs to locate Mark, then to establish a frame of reference around him based on his orientation, and then to use the semantics of “on *X*’s right” to add Matthew to the model in a position on the right-hand side of the lateral plane passing through Mark (see Johnson-Laird 1983, 261). The same semantics as the program uses for “on the right” can be used, but instead of applying to the axes of the spatial array, it applies to axes centered on each individual according to their orientation. Hence, if the individuals are seated in a line, as in Leonardo da Vinci’s painting of the Last Supper, then the model supports the transitive conclusion

Matthew is on Luke’s right.

On the other hand, if they are seated round a small circular table, each premise can be true, but the transitive conclusion false. Depending on the size of the table and the number of individuals seated around it, transitivity can occur over limited regions, and the same semantics for “on *X*’s right” accounts for all the vagaries in the inference.

11.4 Experiment in Spatial Reasoning

The key feature of spatial models is not that they represent spatial relations—propositional representations also do that—but rather that they are functionally organized on spatial axes and, in particular, that information in them can be accessed

by way of these axes. Does such an organization imply that when you have a spatial model of a situation, the relevant information will be laid out in your brain in a spatially isomorphic way? Not necessarily. A programming language, such as LISP, allows a program to manipulate spatial arrays by way of the coordinate values of their axes, but the data structure is only functionally an array and no corresponding physical array of data is necessarily to be found in a computer's memory as it runs the program. The same functional principle may well apply to high-level spatial models in human cognition.

The model theory makes systematically different predictions from those of theories based on formal rules. In an experiment reported by Byrne and Johnson-Laird (1989), the subjects carried out three sorts of spatial inference. The first sort were problems that could be answered by constructing just a single model of the premises, such as the following:

The knife is on the right of the plate.
 The spoon is on the left of the plate.
 The fork is in front of the spoon.
 The cup is in front of the knife.
 What's the relation between the fork and cup?

We knew from previous results that individuals tend to imagine symmetric arrangements of objects, and so these premises call for a model of this sort:

s	p	k
f		c

where *s* denotes a representation of the spoon, *p* a representation of the plate, and so on. This model yields the conclusion

The fork (*f*) is on the left of the cup (*c*).

There is no model of the premises that refutes this conclusion, and thus it follows validly from this single model of the premises. In contrast, if individuals reach this conclusion on the basis of a formal derivation, they must first derive the relation between the spoon and the knife. They need, for example, to infer from the second premise

The spoon is on the left of the plate

that the converse proposition follows:

The plate is on the right of the spoon.

They can then use the transitivity of "on the right of" to infer from this intermediate conclusion and the first premise that it follows that

The knife is on the right of the spoon.

At this point, they can use certain postulates about two-dimensional relations to derive the relation between the fork and the cup (see Hagert 1984 and Ohlsson 1984 for such formal rule systems of spatial inference).

Problems of the second sort yield multiple models because of a spatial indeterminacy, but they nevertheless support a valid answer. They were constructed by changing one word in the second premise:

The knife is on the right of the plate.

The spoon is on the left of the knife.

The fork is in front of the spoon.

The cup is in front of the knife.

What's the relation between the fork and cup?

The description yields models corresponding to two distinct layouts:

s	p	k
f		c
p	s	k
	f	c

Both these models, however, support the same conclusion:

The fork is on the left of the cup.

The model theory predicts that this problem should be harder than the previous one, because reasoners have to construct more than one model. In contrast, theories based on formal rules and propositional representations predict that this problem should be easier than the previous one because there is no need to infer the relation between the spoon and the knife: it is asserted by the second premise.

Problems of the third sort were similar but did not yield any valid relation between the two items in the question, for example:

The knife is on the right of the plate.

The spoon is on the left of the knife.

The fork is in front of the spoon.

The cup is in front of the plate.

What's the relation between the fork and cup?

In one of the experiments, eighteen subjects acted as their own controls and carried out the task with six problems of each of the three sorts presented in a random order. They drew reliably more correct conclusions to the one-model problems (70%) than to the multiple-model problems with valid answers (46%). Their correct conclusions

were also reliably faster to the one-model problems (a mean of 3.1 seconds) than to the multiple-model problems with valid answers (3.6 seconds). It might be argued that the multiple-model problems are harder because they contain an irrelevant premise that plays no part in the inference. However, in another experiment, the one-model problems contained an irrelevant premise, for example:

The knife is on the right of the plate.

The spoon is on the left of the plate.

The fork is in front of the spoon.

The cup is in front of the plate.

What's the relation between the fork and cup?

This description yields the following sort of model:

s	p	k
f	c	

and, of course, the first premise is irrelevant to the deduction. Such problems were reliably easier (61% correct) than the multiple-model problems with valid conclusions (50% correct). Thus the results of the two experiments corroborate the model theory but run counter to theories that assume that reasoning depends on formal rules of inference.

11.5 Space for Time: Models of Temporal Relations

It seems entirely natural that human reasoners would represent spatial relations by imagining a spatial arrangement, but let us push the argument one step further. Perhaps spatial models underlie reasoning in other domains, that is, inferences that hinge on nonspatial matters may be made by manipulating models that are functionally organized in the same way as those representing spatial relations (see section 11.3). A plausible extrapolation is to *temporal* reasoning. Before we examine this extension, let us see how formal rules of inference might cope.

Formal rules might be used for temporal reasoning, but there are some obstacles to them. An obvious difficulty is the large variety of linguistic expressions, at least in Indo-European languages, that convey temporal information. Consider just a handful of illustrative cases. Verbs differ strikingly in their temporal semantics (see, for example, Dowty 1979; Kenny 1963; and Ryle 1949). For instance, the assertion "He was looking out of the window" means that for some interval of time at a reference time prior to the utterance the observer's gaze was out of the window. In contrast, the assertion "He was glancing out of the window" means that for a similar interval the observer's gaze was alternately out of the window and not out of the window. Tempo-

ral adverbials can move the time of an event from the time of the utterance ("He is running now") to a time in the future ("He is running tomorrow"; see, for example, Bull 1963; Lyons 1977; and Partee 1984). General knowledge can lead to a sequential construal of sentential connectives, as in "He crashed the car and climbed out," or to a concurrent interpretation, as in "He crashed the car and damaged the fender." A theory of temporal language has to specify the semantics of these expressions, and particularly their contribution to the truth conditions of assertions. Formal rule theories of inference, in addition, must specify a set of inferential rules for temporal expressions.

In fact, no psychological theory based on formal rules of inference has so far been proposed for temporal reasoning, but logicians have proposed various analyses of temporal expressions. Quine (1974, 82) discusses the following pair of assertions:

I knew him before he lost his fortune
I knew him while he was with Sunnynurse

and suggests treating them as assertions of the form, Some F are G , where F represents "moments in which I knew him" and G represents for the first assertion, "moments before he lost his fortune," and for the second assertion, "moments in which he was with Sunnynurse." This treatment does not readily yield transitive inferences of the form

a before b ,
 b before c ,
Therefore, a before c .

Other logicians have framed temporal logics as variants of modal logic (see, for example, Prior 1967; Rescher and Urquhart 1971), but these logics depend on simple temporal operators that do not correspond to the tense systems of natural language. Their scope is thus too narrow for the various forms of everyday expressions of time. Hence a more plausible way to incorporate temporal reasoning within a psychological theory based on formal rules of inference is to specify the logical properties of temporal expressions in "meaning postulates" in a way that is analogous to the psychological theories of spatial reasoning described in section 11.2.

Temporal relations probably cannot be imagined in a single visual image. In any case, the events themselves may not be visualizable, and manipulations of this factor have no detectable effects on reasoning (see, for example, Newstead, Manktelow, and Evans 1982; Richardson 1987; and Johnson-Laird, Byrne, and Tabossi 1989). When one imagines a temporal sequence, however, it often seems to unfold in time like the original events, though not necessarily at the same speed. This sort of representation

uses time itself to represent the temporal axis (see Johnson-Laird 1983, 10). However, another possibility is to represent temporal relations in a static *spatial* model of the sequence of events in which one axis corresponds to time.

For example, the representation of the assertion

The clerk sounded the alarm after the suspect ran away
calls for a model of the form

r a

in which the time axis runs from left to right, *r* denotes a representation of the suspect running away, and *a* denotes a representation of the clerk sounding the alarm. Events can be described as momentary or as having durations, definite or indefinite. Hence the further assertion

The manager was stabbed while the alarm was ringing

means that the stabbing occurred at some time between the onset and offset of the alarm:

r a———
 s

where *s* denotes a representation of the stabbing, and the vertical dimension allows for contemporaneous events. This model corresponds to infinitely many different situations that have in common only the truth of the two premises. Thus the model contains no explicit representation of the duration for which the alarm sounded, or of the precise point at which the stabbing occurred. Yet, the conclusion

The stabbing occurred after the suspect ran away

is true in this model, and there is no model of the two premises that falsifies it.

I have implemented a computer program that carries out temporal inferences in exactly this way. It attempts to construct all the possible models of the premises. If the number grows too large, it then attempts to use the question—if there is one—to guide its construction of models so as to minimize the number it has to construct. Consider, for example, the following premises:

h happens before *b*

a happens before *b*

b happens before *c*

e happens before *d*

f happens before *d*

c happens before *d*

What's the relation between *a* and *d*?

When the program works through the premises in their stated order, it has to construct 239 models to answer the question—a number that vastly exceeds the capacity of human working memory. If the program's capacity is set more plausibly, say, to four models, it will give up working forwards and then try a depth-first search based on the question: What's the relation between *a* and *d*? It discovers the chain leading from the second premise (referring to *a*) through the third premise (referring to event *b*, which is also referred to by the second premise) to the final premise (referring to *d*), and constructs just the single model that these premises support. This model yields the conclusion that *a* happens before *d*. The advantages of this procedure are twofold. First, it ignores all irrelevant premises. Second, it deals with the premises in a coreferential order in which each premise after the first refers to an event already represented in the set of models. Of course, there are problems that defy the program's capacity for models even if it ignores irrelevant premises. In everyday life, however, individuals are unlikely to present information in an amount or in an order that overburdens human working memory; they are likely to be sensitive to the limitations of their audience (see Grice 1975). Hence it seemed appropriate in our experimental study of temporal reasoning to use similarly straightforward materials.

11.6 Experimental Study of Temporal Reasoning

Psychologists have not hitherto studied deductive reasoning based on temporal relations, and so Walter Schaeken, Gery d'Ydewalle (of the University of Leuven in Belgium), and the present author have carried out an series of experiments examining the topic.

Consider the premises of the following sort:

a before *b*

b before *c*

d while *a*

e while *c*

What's the relation between *d* and *e*?

where *a*, *b*, and so on stand for everyday events, such as "John shaves," "he drinks his coffee," and so on. These events call for the construction of a single model:

<i>a</i>	<i>b</i>	<i>c</i>
<i>d</i>		<i>e</i>

where the vertical dimension allows for events to be contemporaneous. This model supports the conclusion

d before *e*.

The model theory predicts that this one-model problem should be easier than a similar inference that contains an indeterminacy. For example, the following premises call for several models:

a before *c*

b before *c*

d while *b*

e while *c*

What's the relation between *d* and *e*?

The premises are satisfied by the following models:

a	b	c		b	a	c		a	c
				d		e		b	
	d	e						d	e

In all three models, *d* happens before *e*, and so it is a valid conclusion. The model theory also predicts that the time subjects spend reading the second premise, which creates the indeterminacy leading to multiple models, should be longer than the reading time of the second premise of the one-model problem. This multiple-model problem contains an irrelevant first premise, but the following one-model problem also contains an irrelevant first premise:

a before *b*

b before *c*

d while *b*

e while *c*

What's the relation between *d* and *e*?

In one of our experiments, we tested twenty-four university students with eight versions of each of the three sorts of problems above, and eight versions of a multiple-model problem that had no valid answer. The thirty-two problems were presented under computer control in a different random order to each subject. The two sorts of one model problem were easy and did not differ reliably (93% correct for the problems with no irrelevant premise and 89% correct for the problems with an irrelevant premise), but they were reliably easier than the multiple-model problems with valid conclusions (81% correct responses), which in turn were reliably easier than the multiple-model problems with no valid conclusions (44% correct responses). One would expect the latter problems to be difficult because it is vital to construct more than one model in order to appreciate that they have no valid conclusion, whereas the valid answer will emerge from any of the multiple models of the problems with a valid answer. Figure 11.1 shows the reading times for the four premises of the problems.

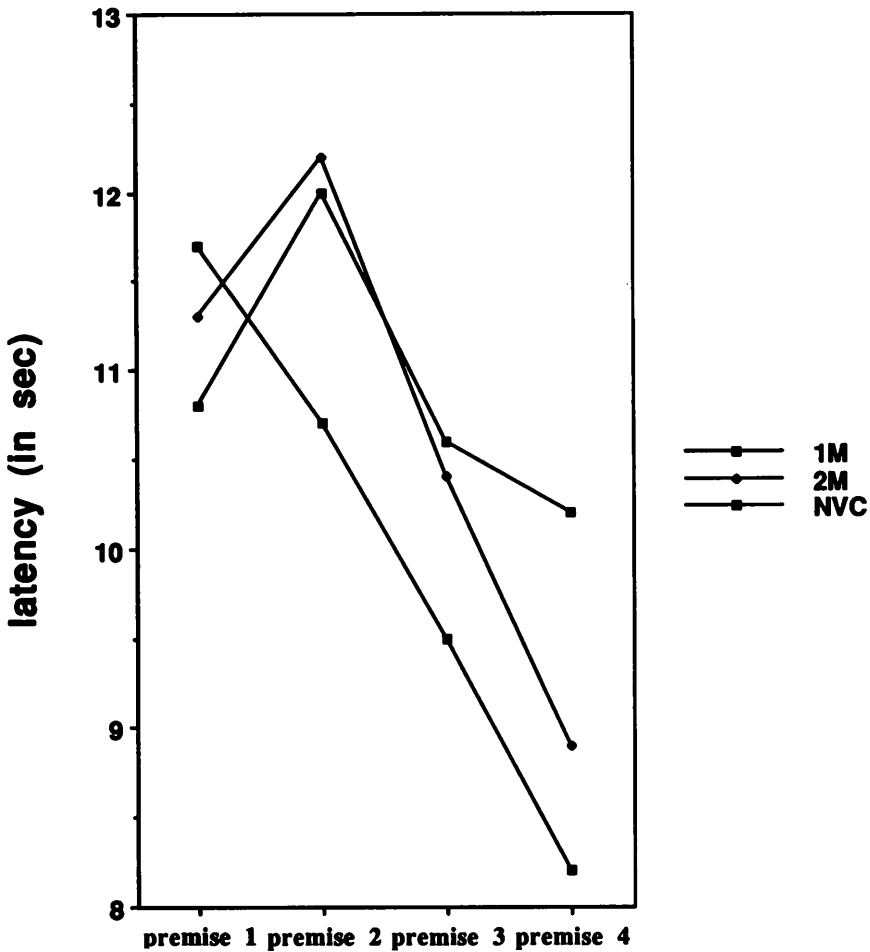


Figure 11.1
The mean latencies for reading the premises in the temporal inference experiment. The means are for one-model problems (1-M) collapsing over the two sorts, the multiple-model problems with a valid conclusion (2-M), and the multiple-model problems with no valid conclusion (NVC).

As the figure shows, subjects took reliably longer to read the second premise of the multiple-model problems—the premise that calls for the construction of more than one model—than to read the second premise of the one-model problems.

Our results, both for this experiment and others that we carried out, establish three main phenomena, and they imply that reasoning about temporal relations depends on mental models of the sequences of events. The first phenomenon concerns the number of models. When a description is consistent with just one model, the reasoning task is simple and subjects typically draw over 90% correct conclusions. When a description is consistent with more than one model, there is a reliable decline in performance. As in the earlier study of spatial reasoning, we pitted the predictions of the model theory against contrasting predictions based on formal rules of inference. The results showed that the one-model problems were reliably easier than the multiple-model problems, even though the one-model problems call for longer formal derivations than the multiple-model problems.

The second phenomenon concerns the subjects' erroneous conclusions. Formal rule theories make no specific predictions about the nature of such conclusions: subjects are said to err because they misapply a rule or fail to find a correct derivation. The model theory, however, predicts that erroneous conclusions arise because reasoners fail to consider all the models of the premises, and thus these conclusions should tend to be consistent with the premises (i.e., true in at least one model of them) rather than inconsistent with premises (i.e., not true in any model of them). The results corroborated this prediction of the model theory.

The third phenomenon concerns the time subjects took to read the premises and to respond to the questions. As we have seen, they took reliably longer to read a premise that led to multiple models than to read a corresponding premise in a one-model problem. Formal rule theories make no such prediction, and it is hard to reconcile this result with such theories because they make no use of models. The result also suggests that subjects do not construct models that represent indeterminacies within a single model. If they had done so, then they should have taken no longer to read these premises than the corresponding premises of one-model problems. And, of course, they should not have been more prone to err with indeterminate problems. The times to respond to the questions also bore out the greater difficulty of the multiple-model problems.

One final comment on our temporal experiments. Problems that depend on a transitive chain of events, as in the following one-model problem:

a	b	c
d		e

make an interesting contrast with one-model problems in which the transitive chain is not relevant to the answer:

a	b	c
	d	e

If subjects were imagining the events unfolding in time at a more or less constant rate, then presumably they ought to be able to respond slightly faster in the second case than in the first. That is to say, the actual temporal interval between *d* and *e* must be shorter in the second case than in the first. We examined this difference in the experiment described above. The mean latencies to respond were as follows: 7.0 seconds in the first case and 5.8 seconds in the second case. This difference was not too far from significance, and thus perhaps at least some of our subjects were imagining events as unfolding in time rather than simply constructing spatial models of the temporal relations.

11.7 Space for Space: How Diagrams Can Help Reasoning

Diagrams are often said to be helpful aids to thinking. They can make it easier to find relevant information—one can scan from one element to another element nearby much more rapidly than one might be able to find the equivalent information in a list of numbers or verbal assertions. Diagrams can make it easier to identify instances of a concept—an iconic representation can be recognized faster than a verbal description. Their symmetries can cut down on the number of cases that need to be examined. But can diagrams help the process of thought itself? Larkin and Simon (1987) grant that diagrams help reasoners to find information and to recognize it, but doubt whether they help the process of inference itself. According to Barwise and Etchemendy (1992, 82), who have developed a computer program, Hyperproof, that helps users to learn logic: “diagrams and pictures are extremely good at presenting a wealth of specific, conjunctive information. It is much harder to use them to present indefinite information, negative information, or disjunctive information. For these, sentences are often better.” Hyperproof accordingly captures conjunctions in diagrams, but expresses disjunctions in verbal statements. The model theory, however, makes a different prediction. A major problem in deduction is to keep track of the possible models of premises. Hence a diagram that helps to make them explicit should also help people to reason. The result of perceiving such a diagram is a model—according to Marr’s (1982) of vision—and thus one has a more direct route to a model than that provided by a verbal description. The verbal description needs to be parsed and a compositional semantics needs to be used to construct its propositional representation, which is then used in turn to construct a model. Hence it should be easier to reason from diagrams than from verbal descriptions.

We tested this prediction in two experiments based on so-called double disjunctions (Bauer and Johnson-Laird 1993). These are deductive problems, which are exemplified in verbal form by the following problem:

Julia is in Atlanta, or Raphael is in Tacoma, but not both.

Julia is in Seattle, or Paul is in Philadelphia, but not both.

What follows?

The model theory predicts that such problems based on exclusive disjunctions should be easier than those based on inclusive disjunctions:

Julia is in Atlanta, or Raphael is in Tacoma, or both.

Julia is in Seattle, or Paul is in Philadelphia, or both.

What follows?

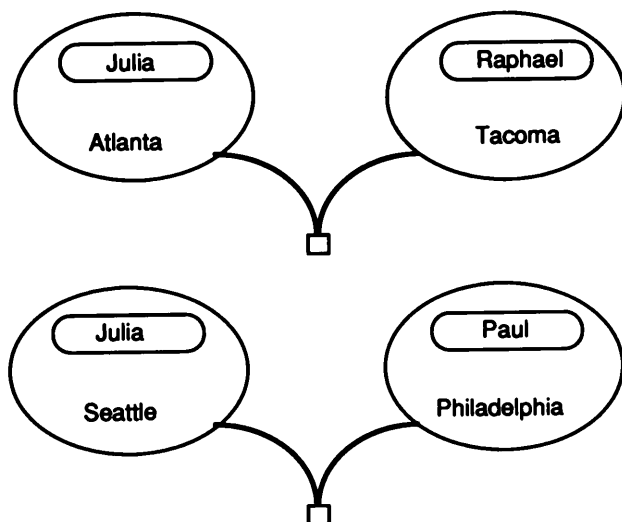
Each exclusive disjunction calls for only two models, whereas each inclusive disjunction calls for three models. Likewise, when the premises are combined, the exclusive problem yields three models:

a		p
s	t	
	t	p

Here *a* is a representation of Julia in Atlanta, *s* is a representation of Julia in Seattle, *t* is a representation of Raphael in Tacoma, and *p* is a representation of Paul in Philadelphia. In contrast, the inclusive problem yields a total of five models:

a		p
s	t	
	t	p
a	t	p
s	t	p

In our first experiment, premises of this sort were presented either verbally or else in the form of a diagram, such as figure 11.2. To represent, say, Julia in Atlanta, the diagram has a lozenge labeled "Julia" lying within the ellipse labeled "Atlanta." Inclusive disjunction, as the figure shows, is represented by a box connected by lines to the two component diagrams making up the premise as a whole. The experiment confirmed that exclusive disjunctions were easier than inclusive disjunctions (for both the percentages of correct responses and their latencies); it also confirmed that "identical" problems, in which the individual common to both premises was in the same place in both of them, were easier than "contrastive" problems such as the one above. But the experiment failed completely to detect any effect of diagrams: they yielded



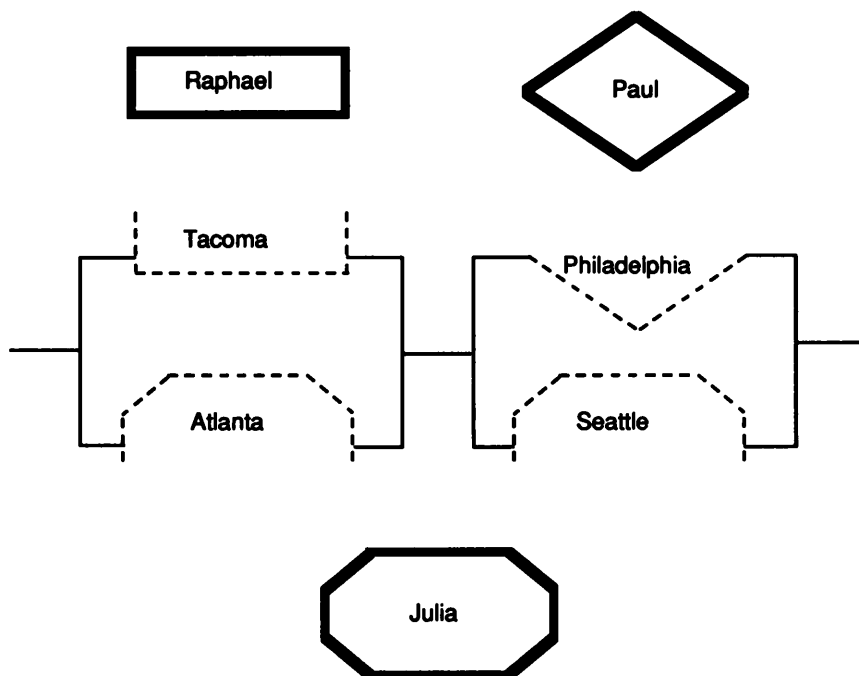
What follows?

Figure 11.2

The diagrammatic presentation of double disjunctions in the first diagram experiment.

28% correct conclusions in comparison to the 30% correct for the verbal problems. Double disjunctions remained difficult, and these diagrams were no help at all.

With hindsight, the problem with the diagrams was that they used arbitrary symbols to represent disjunction and thus failed to make the alternative possibilities explicit. In a second experiment, we therefore used a new sort of diagram, as shown in figure 11.3, which is analogous to an electrical circuit. The idea, which we explained to the subjects, was to complete a path from one side of the diagram to the other by moving the shapes corresponding to people into the slots corresponding to cities. We tested four separate groups of subjects with logically equivalent problems: one group received diagrams of people and places (as in the figure); a second group received problems in the form of circuit diagrams of electrical switches; a third group received problems in the form of verbal premises about people and places; and a fourth group received problems in the form of verbal premises about electrical switches. There was no effect of the content of the problems—whether they were about people or switches—and therefore we have pooled the results. The percentages of correct responses are presented in figure 11.4. As the figure shows, there was a striking effect of mode of presentation: 74% correct responses to the diagrammatic problems in comparison to only 46% correct responses to the verbal problems. The



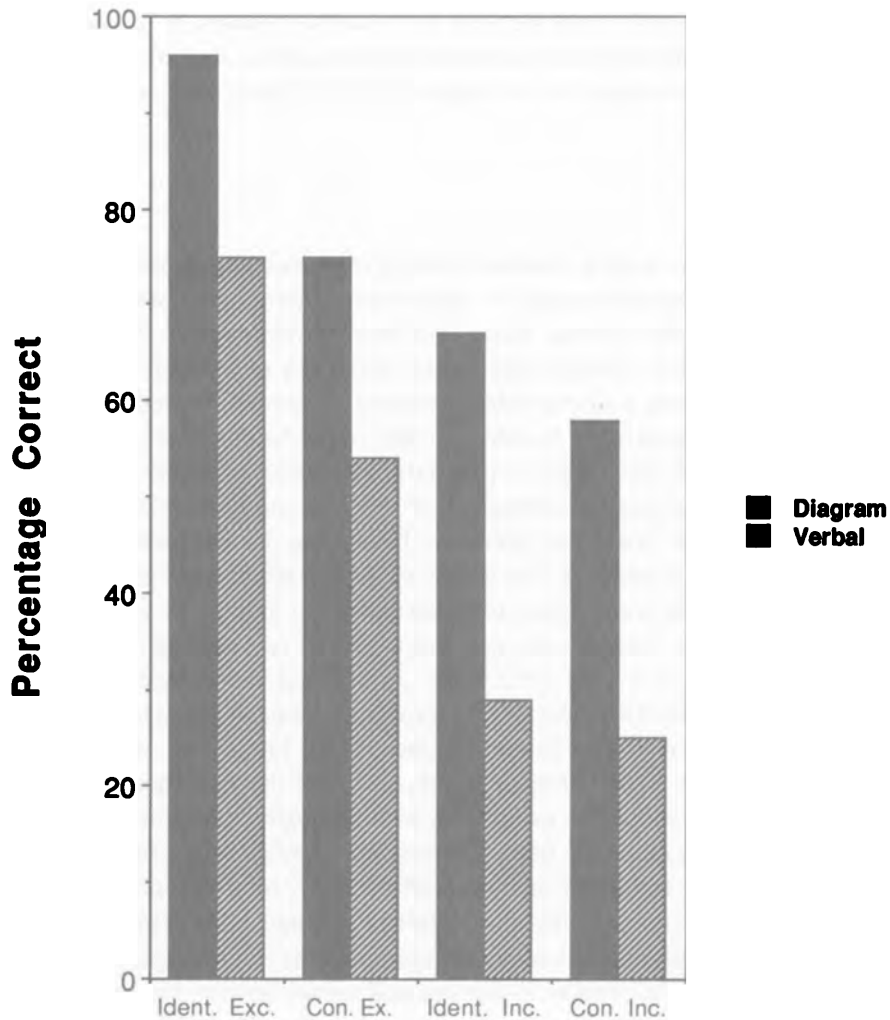
The event is occurring.
What follows?

Figure 11.3

The diagrammatic presentation of double disjunctions in the second diagram experiment.

results also corroborated the model theory's predictions that exclusive disjunctions should be easier than inclusive disjunctions, and that identical problems should be easier than contrastive problems. The latencies of the subjects' correct responses had exactly the same pattern, for example, subjects were faster to reason with exclusive disjunctions than inclusive disjunctions, and they were reliably faster to respond to the diagrammatic problems (a mean of 99 seconds) than to the verbal problems (a mean of 135 seconds).

People evidently reason by trying to construct models of the alternative possibilities, and diagrams that enable these alternatives to be made explicit can be very helpful. With a diagram of the sort we used in our second experiment, individuals perceive the layout and in their mind's eye can move people into places and out again. By manipulating the model underlying the visual image, they can construct the alter-



Form of Disjunction

Figure 11.4

The percentages of correct responses in the second diagram experiment. There are two sorts of disjunction: exclusive (exc.) and inclusive (inc.), and two sorts of relation between premises: identical (ident.) and contrastive (con.).

native possibilities more readily than they can from verbal premises. It follows that diagrams are not merely encoded in propositional representations equivalent to those constructed from verbal premises (but see Baylor 1971, Pylyshyn 1973, and Palmer 1975 for opposing views).

11.8 Conclusions

Mental models are in many ways a primitive form of representation, which may owe their origin to the selective advantage of constructing internal representations of spatial representations in the external world. The evidence reviewed in this chapter suggests that mental models underpin the spatial reasoning of logically untutored individuals and may also play a similar role in temporal reasoning. Indeed, it may be that human inference in general is founded on the ability to construct spatial, or quasi-spatial models, which also appear to play a significant part in syllogistic reasoning and reasoning with multiple quantifiers (Johnson-Laird and Byrne 1991).

Historians of science and scientists themselves have often drawn attention to the role of diagrams in scientific thinking. Our studies show that not just any diagram has a helpful role to play. It is crucial that diagrams make the alternative possibilities explicit. Theories based on formal rules and propositional representations have to postulate the extraction of logical form from an internal description of visual percepts. In contrast, the model theory allows for inferences based on visual perception, which has a mental model as its end product (Marr 1982). The two theories accordingly diverge on the matter of diagrams. Formal rule theories argue that performance with a diagram should be worse than with the logically equivalent verbal premises: with a diagram, reasoners have to construct an internal description from which they can extract a logical form. The model theory, however, predicts that performance with a diagram that makes the alternative possibilities explicit should be better than with logically equivalent verbal premises: with a diagram, reasoners do not need to engage in the process of parsing and compositional semantics. The evidence indeed suggests that human reasoners use functionally spatial models to think about space, but they also appear to use such models in order to think in general.

Acknowledgments

I am grateful to Ruth Byrne for her collaboration in developing the theory of deduction based on mental models. I am also grateful to her, to Malcolm Bauer, and to Walter Schaeken for ideas and help in carrying out the present experiments. The research was supported in part by the James S. McDonnell Foundation.

References

- Barwise, J. (1993). Everyday reasoning and logical inference. *Behavioral and Brain Sciences*, 16, 337–338. Commentary on Johnson-Laird and Byrne 1991.
- Barwise, J., and Etchemendy, J. (1992). Hyperproof: Logical reasoning with diagrams. In N. H. Narayanan (Ed.), *AAAI Spring Symposium on Reasoning with Diagrammatic Representations*, 80–84. 25–27 March, Stanford University, Stanford, CA.
- Bauer, M. I., and Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science*, 4, 372–378.
- Baylor, G. W. (1971). Programs and protocol analysis on a mental imagery task. *First International Joint Conference on Artificial Intelligence*. N. P.
- Bull, W. E. (1963). *Time, tense, and the verb*. Berkeley: University of California Press.
- Byrne, R. M. J., and Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language*, 28, 564–575.
- Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.
- Dowty, D. R. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Garnham, A. (1987). *Mental models as representations of discourse and text*. Chichester: Ellis Horwood.
- Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and semantics*. Vol. 3: *Speech acts*. New York: Seminar Press.
- Hagert, G. (1984). Modeling mental models: Experiments in cognitive modeling of spatial reasoning. In T. O'Shea (Ed.), *Advances in artificial intelligence*, Amsterdam: North-Holland.
- Johnson-Laird, P. N. (1975). Models of deduction. In R. Falmagne (Ed.), *Reasoning: Representation and process*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press; Cambridge: Cambridge University Press.
- Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition*, 189–209.
- Johnson-Laird, P. N., and Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N., Byrne, R. M. J., and Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review*, 96, 658–673.
- Kenny, A. (1963). *Action, emotion, and will*. New York: Humanities Press.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Larkin, J., and Simon, H. (1987). Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science*, 11, 65–99.
- Lyons, J. (1977). *Semantics*. Vols. 1 and 2. Cambridge: Cambridge University Press.

- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Newstead, S. E., Manktelow, K. I., and Evans, J. St. B. T. (1982). The role of imagery in the representation of linear orderings. *Current Psychological Research*, 2, 21–32.
- Ohlsson, S. (1984). Induced strategy shifts in spatial reasoning. *Acta Psychologica*, 57, 46–67.
- Osherson, D. N., Smith, E. E., and Shafir, E. B. (1986). Some origins of belief. *Cognition*, 24, 197–224.
- Palmer, S. E. (1975). Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman, D. E. Rumelhart, and the LNR Research Group (Eds.), *Explorations in cognition*, 279–307. San Francisco: Freeman.
- Partee, B. (1984). Nominal and temporal anaphora. *Linguistics and Philosophy*, 7, 243–286.
- Perkins, D. N., Allen, R., and Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell (Ed.), *Thinking*. Philadelphia: Franklin Institute Press.
- Prior, A. N. (1967). *Past, Present, and Future*. Oxford: Clarendon Press.
- Pylyshyn, Z. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 80, 1–24.
- Quine, W. V. O. (1974). *Methods of logic*. 3d ed. London: Routledge and Kegan Paul.
- Rescher, N., and Urquhart, A. (1971). *Temporal logic*. New York: Springer.
- Richardson, J. T. E. (1987). The role of mental imagery in models of transitive inference. *British Journal of Psychology*, 78, 189–203.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Schaeken, W., Johnson-Laird, P. N., and d'Ydewalle, G. (1994). Mental models and temporal reasoning. *Cognition*, in press.
- Simon, H. A. (1959). Theories of decision making in economics and behavioral science. *American Economic Review*, 49, 253–283.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Tversky, A., and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.

Chapter 12

Spatial Perspective in Descriptions

Barbara Tversky

12.1 Central Issue in Perspective

When viewing an object or a scene, people necessarily have a specific perspective on it. Yet when thinking about or describing an object or scene, people can free themselves from their own perception and their own perspective. For example, when recollecting events, people often describe their memory images as including themselves (Nigro and Neisser 1983) rather than from the perspective of experience. Or, when describing a simple scene to others, speakers often take their addressees' perspective rather than their own (Schober 1993). Given the freedom to select a perspective, what determines the perspective selected?

Spatial perspective has been a central issue to scholars with many interests, object recognition, environmental cognition, developmental psychology, neuropsychology, and language. Naturally, researchers in each area have their own concerns, and although some of these are shared, they often work in blissful ignorance of each other. What accounts for the fascination of perspective, what is it that draws researchers with such diverse interests and methods to study it? Although people cannot help but experience the world from their own necessarily limited point of view, taking other points of view is essential for a range of cognitive functions and social interactions, from recognizing an object from a novel point of view to navigating an environment to understanding someone else's position. Emerging from the restrictions of the self seems at the basis of human thought and society. Not surprisingly, each discipline has approached the problem of perspective with its own set of issues, developing its own set of distinctions.

Before examining determinants of choice of perspective in describing space and in comprehending spatial descriptions, I will first survey views on perspective in several diverse areas of cognitive science, most notably, object recognition, environmental cognition, and language, framing research in the issues relevant to each discipline. The distinctions in perspective made by each of the disciplines contain instructive

similarities and differences. Of special concern here are the bases for spatial reference frames in each discipline.

12.2 Some Perspectives on Perspective

12.2.1 Object Recognition

Viewing a three-dimensional object reveals only part of the object, yet recognizing an object can entail knowing what it looks like from other points of view. A critical issue in object recognition is the formation of mental representations that allow recognition of novel stimuli, both the same objects from different points of view and objects from the same class never before encountered. One question is the extent to which objects can be recognized solely on the basis of information from visual input, without drawing on information stored in memory, that is, from bottom-up information as opposed to top-down information (e.g., Marr 1982; Marr and Nishihara 1978). The visual input gives a *viewer-centered* representation of an object, derived from the information projected on a viewer's retina at the viewer's current perspective. It yields some depth information but, without added assumptions, no information as to how an object would look from sides not currently in the field of vision. Because it is based on experience viewing objects from many different points of view (see, for example, Tarr and Pinker 1989), and perhaps on geometric principles that allow mental transformations (e.g., Shepard 1984), memory can provide an *object-centered* representation, a more abstract representation that yields information about how an object would look from a different perspective. In many cases, recognition of an object currently under view, for example, an upside-down or tilted object, seems to depend on mental comparison to an object in memory that is canonically oriented (e.g., Jolicoeur 1985). Whereas a viewer-centered representation has a specific perspective, an object-centered representation might have a specific perspective, such as a canonical view, or it might have multiple representations each with its own perspective, or it might be perspective-free, as in a structural description (Pinker 1984). In any case, the distinction between the viewer and the object viewed as bases for perspective has been critical to thinking about mental representations of objects.

12.2.2 Environmental Cognition

A similar issue arises in the study of environmental cognition. In perceiving a scene, the viewer regards it from a specific perspective, yet more general knowledge of scenes from many perspectives is required for successful navigation. Environments are experienced from specific points of view along specific routes. Yet people are able to make spatial inferences, such as reversing routes or constructing novel ones (see, for example, Landau 1988; Levine, Jankovic, and Palij 1982 and Presson and Hazelrigg 1984).

problem for development is similar to that of acquisition. How do children come to take perspectives other than their own? Most accounts of mental representations of environments propose that as people move about an environment, they perceive the changing spatial relations of objects or landmarks to themselves, and use that information perhaps in concert with (implicit) knowledge of geometry to construct more general mental representations of the spatial relations among landmarks independent of a particular perspective. As for object recognition, the initial perspective is viewer-centered, often called *egocentric*. Later, people come to use what have been termed *allocentric* reference frames (e.g., Hart and Moore 1973; Pick and Lockman 1981). Allocentric reference frames are defined with respect to a reference system external to the environment, usually the canonical axes, north-south, east-west. However, other objects, notably landmarks, are also external to a viewer and turn out to be important in organizing environmental knowledge (e.g., Couclelis et al. 1987; Hirtle and Jonides 1985; Lynch 1960; Sadalla, Burroughs, and Staplin 1980). In environmental cognition, then, the viewer and other objects in the scene serve as bases for spatial reference frames in addition to external or extrinsic bases.

12.2.3 Neuropsychological Support

Neuropsychological evidence from different sources supports the finding by environmental psychologists that there are three bases for spatial reference systems: the viewer, landmarks, and an external reference frame. Perrett et al. (1990) have recorded responses to observed movements in the temporal lobes of monkeys, finding evidence for three bases for reference frames, namely, the viewer, the object being viewed, and the goal of the movement. In the terms of environmental cognition, both the latter categories, the object under view and the goal of the movement, can be regarded as landmarks. From recordings taken from the hippocampi of rats as they explore environments, O'Keefe and Nadel (1978; O'Keefe, chapter 7, this volume) and others have concluded that the hippocampus represents known environments with respect to an external reference frame.

12.2.4 Spatial Language

People's ability to take perspectives not currently their own is revealed in their use of language from perspectives other than the perspective under view as well as in their recognition of objects and navigation of environments. Accounts of spatial language have also found it useful to distinguish three bases for spatial reference: the viewer, other objects, and external sources (e.g., Bühler 1934; Fillmore 1975; 1982; Levelt 1984, 1989, and chapter 3, this volume; Levinson, chapter 4, this volume; Miller and Johnson-Laird 1976). These three bases at first seem to correspond to deictic, intrinsic, and extrinsic uses of language, though it will turn out not to be that simple.

Before getting into the complexities, I will review deictic, intrinsic, and extrinsic uses of language.

The term *deictic* derives from a Greek root meaning “to show” or “to point.” Deictic uses cannot be accounted for by the language alone, but require additional knowledge of “certain details of the interactional situation in which the utterances are produced,” according to Fillmore (1982, 35) or, put differently by Levelt (1989, 45), “an audio-visual scene which is more or less shared between the interlocutors, the places and orientations of the interlocutors in this scene at the moment of utterance, and the place of the utterance in the temporal flow of events.” Several kinds of deixis have been distinguished (see, for example, Fillmore 1975, 1982; Levelt 1989), notably, person, place, and time, prototypically represented in language by “I,” “here,” and “now.” For example, in person deixis, understanding the referents of “you” and “I” in a discourse depends on knowing who is speaking to whom. In place deixis, understanding the uses of “this” and “that” or “here” and “there” require knowing where the participants in a discourse are, relative to the objects in a scene. Miller and Johnson-Laird define *place deixis* as “the linguistic system for talking about space relative to a speaker’s egocentric origin and coordinate axes” (Miller and Johnson-Laird 1976, 396). It is place deixis that is of concern here. Deictic uses can be subtle, and there is not always agreement on them, as suggested by nuances in the definitions quoted above.

Some of the subtlety of deixis comes from the fact that many deictic terms can be used nondeictically, especially intrinsically, such as *front* and *left*. If I say, “The tent is in front of the boulder,” I am using the term *front* deictically. The boulder has no front side, so I must mean that the tent is located between my front side and the boulder. In that case, you must know where I am located and how I am oriented with respect to the boulder to understand what I mean. In contrast, if I say, “My pack is in front of the tent,” I can be using the term *front* either deictically, as for the boulder, or intrinsically, that is, with respect to the object’s natural sides. Unlike a boulder, but like a person, a tent has a natural front, back, top, and bottom, and a natural left and right derived from the other sides. Thus, for the intrinsic use, I mean that my pack is located near the front side of the tent. In this case, knowing where I am standing is unnecessary to understand what I mean.

The extrinsic case is the clearest. Extrinsic uses of language rely on an external reference system, such as the canonical directions, north-south, east-west. If I say, “The tent is south of the boulder,” I am using language extrinsically.

If we just stop here, it seems as though, in deictic cases, the basis for a reference frame is the viewer; in intrinsic cases, an object; and in extrinsic cases, an external reference frame. Unfortunately, things are not that simple. For one thing, speakers can refer to their own bodies intrinsically. As Fillmore puts it, “It should be clear that

it is also possible for the speaker of a sentence to regard his own body as a physical object with an orientation in space; expressions like ‘in front of me,’ ‘behind me,’ or ‘on my left side,’ are deictic by containing a first person pronoun but they are not instances of the deictic use of the orientational expressions” (Fillmore 1975, 6). Continuing this line of reasoning, Levinson (chapter 4, this volume) shows that egocentric or viewer-based uses crosscut intrinsic and extrinsic uses rather than contrasting with them. Fillmore’s examples are simultaneously egocentric and intrinsic, as in “the boulder is in front of me.” Speakers can also be simultaneously egocentric and extrinsic, as in “the boulder is south of me.”

Levinson suggests a different classification of spatial reference frames in language use: relative, intrinsic, and absolute. To illustrate the distinctions, Levinson uses the same spatial scenario for all three cases: a man is located in front of a house. The target object is the man, whose location is described relative to the referent object, the house, whose location and orientation are known. In Levinson’s analysis, the intrinsic and absolute (extrinsic) reference frames are binary, that is, they require two terms to specify the location of the target object; the target object and the referent object. Speaking intrinsically, I can say, “The man is in front of the house,” meaning close to the house’s intrinsic front. Speaking absolutely or extrinsically, I can say, “The man is north of the house.” The relative case adds the location of a viewer, and uses three terms, that is, it requires a ternary relation. If I am a viewer located away from the house’s left side, looking at the man and the house, I can say, “The man is to the left of the house,” that is, the man is left of the house with respect to me, to my left, from my current location and orientation. The relative reference frame is more complex because it requires knowing my location and orientation as well as the locations of the man and the house. According to Levinson’s analysis, what Levelt (1989) termed *primary deixis* is intrinsic, as when I say, “The tent is in front of me”, and what Levelt termed *secondary deixis* is relative, as when I say, “The tent is to the right of the boulder.”

12.2.5 Bases for Spatial Reference

For a variety of reasons, some shared and some unique, the analysis of spatial reference systems and perspective has been central to several disciplines within cognitive science, notably, object recognition, environmental cognition, and language. Each of these disciplines has regarded the viewer as an important basis for spatial reference, primarily because perception and cognition begin with the viewer’s perspective. Most have also regarded an object in the scene (or in the case of language, the self, referred to as an object) and a reference frame external or extrinsic to the scene as important bases for spatial reference systems. They provide perspectives more general than that from a particular viewpoint at a particular time. The considerations leading to the

spatial reference systems described above have been primarily cognitive. I turn now to work raising social considerations and pointing to social categories of spatial reference.

12.3 Social Categories

Spatial descriptions, like most discourse, occur in a social context; there is either a real addressee or an implicit one. Schober (1993) investigated the use of perspective with real or implicit addressees. He developed a task that required participants to take a personal perspective, either their own, or that of their addressee. In one task, pairs of subjects who could talk to each other but not see each other had diagrams with two identical circles embedded in a larger circle. The viewpoints of each of the subjects were indicated outside the larger circle. On one subject's diagram, one of the smaller circles had an *X*. That subject's task was to describe the location of the *X* so that the other subject could put an *X* in the analogous circle on the diagram. The task allowed only personal perspectives, either that of the speaker or that of the addressee. There were no other objects to anchor an intrinsic perspective and insufficient knowledge for an extrinsic one. Schober (1993) found that, on the whole, speakers took the perspective of the addressee. In a variation of the task, speakers explained which circle had an *X* to an unknown addressee, in a situation that was not interactive. When there was no active participant to the discourse, speakers were even more likely to take the addressee's perspective. Thus, what was of interest in Schober's task was whose perspective, speaker's or addressee's, speakers would adopt under what conditions.

Although Schober's task did not allow it, another possibility is to use a *neutral* perspective rather than a *personal* perspective. A neutral perspective is one that is neither the addressee's nor the speaker's. Neutral perspectives include the possibilities raised earlier, namely, using a landmark, referent object, on the extrinsic system as a basis for spatial reference. Mine, yours, or neutral are social categories, and language, more than object recognition or navigation, is social.

12.4 Determinants of Perspective Choice

Now I return to the determinants of perspective choice. After a brief review of previous analysis and research, I will describe aspects of three ongoing projects relevant to the question. As Levinson (chapter 4, this volume) has pointed out, not every language uses all three systems; thus some determinants are linguistic. Because English uses all three systems, the question of determinants of perspective choice can be

addressed in English. The experts do not agree on a dominant or default perspective. For example, Levelt (1989, 52) asserts: "Still, it is a general finding that the dominant or default system for most speakers is deictic reference, either primary or secondary." In contrast, Miller and Johnson-Laird (1976, 398) maintain: "But intrinsic interpretations usually dominate deictic ones; if a deictic interpretation is intended when an intrinsic interpretation is possible, the speaker will usually add explicitly 'from my point of view' or 'as I am looking at it.'" As it happens, the disagreeing experts all seem to be correct, but in different situations.

For extended discourse, in contrast to the single utterances that have often been analyzed, other issues arise. One of these is consistency of perspective. Many theoreticians have assumed that speakers will adopt a consistent perspective, for several reasons. Consistency of perspective is a necessary consequence of the assumption of a default perspective; anyone arguing for a single default perspective also argues for a consistent perspective. Even if the possibility of different perspectives is recognized, consistency of perspective within a discourse can provide coherence to a discourse, rendering it more comprehensible. Switching perspective carries cognitive costs, at least for comprehension (e.g., Black, Turner, and Bower 1979).

A second issue of interest for extended discourse is determining the order of presenting information, independent of perspective. As Levelt (1982a, 1989) has observed, the world is multidimensional, but speech is linear. To describe the world linearly, it makes sense to choose a natural order. Because a natural way of experiencing an environment is by moving through it, a natural way of conveying an environment is through a mental tour (Levelt 1982a, 1989).

Mental tours abound in spatial descriptions. In their classic study, Linde and Labov (1975) found that New Yorkers used tours to describe their apartments. Similarly, respondents took listeners on mental tours of simple networks (Levelt 1982a, b; Robin and Denis 1991), of the rooms where they lived (Ullmer-Ehrich 1982), and of dollhouse rooms (Ehrich and Koster 1983). Tours, though common, are by no means universal. For example, in describing locations in a complex network, a path or tour was only one of several styles adopted by subjects (Garrod and Anderson 1987). And on closer inspection, many of the room tours were "gaze tours" rather than "walking tours." Gaze tours are also natural ways of perceiving environments, from a stationary view point rather than a changing one. The discourse of a gaze tour, however, differs markedly from that of a walking tour (Ullmer-Ehrich 1982). In a gaze tour, the noun phrases are usually headed by objects and the verbs express states; for example, "the lamp is behind the table." In a walking tour, the noun phrases are headed by the addressee and the verbs express actions; for example, "you turn left at the end of the corridor and see the table on your right." Finally, the range of environments studied has been limited: single rooms, strings of rooms, and networks.

12.4.1 Pragmatic Considerations

Assertions about default and consistent perspectives notwithstanding, given that English and many other languages have all three reference systems, it makes sense that all three be used. Rather than there being a default perspective, choice of perspective is likely to be pragmatically determined. One pragmatic consideration is cognitive difficulty. Certain terms, like *left* and *right*, are more difficult for people than others, like *up* and *down* (see, for example, Clark 1973; Farrell 1979). What is easier or harder can also depend on the number or degree of mental transformation required to produce or comprehend an utterance. Some environments may lend themselves to one perspective or another, so that describing them using a different perspective may increase difficulty. It stands to reason that speakers would avoid cognitively difficult tasks, all other things being equal.

Another pragmatic consideration is the audience. Speakers tailor their speech to their addressees. In many cases, including the prototypic face-to-face conversation, the perspective of speakers and addressees differ. Because addressees have the harder job of comprehending, speakers may wish to ease the burden of addressees by using the addressees' perspective rather than their own (Schober 1993). Moreover, speakers presumably desire that their communications be understood and therefore attempt to construct their contributions to be as comprehensible as possible, given the situation (e.g., Clark 1987). Taking the addressee's perspective should make communications more likely to be understood. Finally, using the addressee's perspective is polite (Brown and Levinson 1987).

In other situations, speakers may wish to avoid taking either their own or their addressee's perspective and to adopt instead a perspective that is neutral, neither speaker's nor addressee's. Where there is some controversy between the speaker's view and the addressee's view, a neutral perspective may diffuse tension. Or more simply, the interlocutors may wish to avoid confusion over whose left and right. Whether the reasons are social or cognitive, speakers may use a neutral perspective, using landmarks as referents or an extrinsic system. Landmarks have the advantage of being visible in a scene, and an extrinsic system has the advantage of being more permanent and independent of the scene. In the remainder of the chapter I will discuss three examples, drawn from current research projects, illustrating the effects of pragmatic considerations on the selection of perspective in the comprehension or production of spatial descriptions.

A number of years ago, Nancy Franklin, Holly Taylor, and I began studying the nature of the spatial mental models engendered by language alone. We were stimulated by the research of Mani and Johnson-Laird (1982) and Johnson-Laird (1983), demonstrating the use of mental models in solving verbal syllogism, and of Glenberg, Meyer, and Lindem (1987) and Morrow, Greenspan, and Bower (1987; also Morrow,

Bower, and Greenspan 1989), demonstrating effects of distance on situation models constructed from text. Like Mani and Johnson-Laird, Franklin and I were interested in mental representations and inference of spatial relations. Franklin and I, later joined by David Bryant, began with descriptions of the immediate environment surrounding a person (Franklin and Tversky 1990; Bryant, Tversky, and Franklin 1992). Like Perrig and Kintsch (1985), Taylor and I were interested in comprehension and later production of longer discourses; we therefore focused on descriptions of larger environments (Taylor and Tversky 1992a,b). Both projects brought us to the study of perspective. Scott Mainwaring and Diane Schiano joined in a third project, investigating perspective in variations on Schober's paradigm (Mainwaring, Tversky, and Schiano 1995). Let me describe those enterprises in that order, beginning with the project on environments immediately surrounding people.

12.5 Comprehension: Nature of the Described Environment

As we turn in and move about the world, we seem to be able to keep track of the locations of objects around us without noticeable effort, updating their relative locations, even unseen locations, with every step. Franklin and I wanted to simulate that process, using language (Franklin and Tversky 1990). We wrote a series of narratives, describing "you," the subject, in various environments, some exotic like an opera house, some mundane, like a barn. In each setting, "you" were surrounded by objects, such as a bouquet of flowers or a saddle, to all six sides of your body, from your head, feet, front, back, left, and right. After studying an environment, subjects turned to a computer that repeatedly reoriented them to face one of the objects, and then probed them with direction terms, *front*, *back*, *head*, *feet*, *right*, and *left*, for the names of the objects in those directions. Subjects performed this task easily, almost without error, so the data of importance are the times to access the objects in the six directions from the body. A schematic of the situation appears in figure 12.1.

We considered three models for accessing objects around the body. According to the *equiavailability model*, no area of space is privileged over any other area, much as in scanning a picture; this model predicts equal reaction times to all directions (Levine, Jankovic, and Palij 1982). However, a three-dimensional world surrounding a subject, even an imaginary one, is different from a picture all in front of a subject. For this case, objects directly in the imaginary field of view might have an advantage relative to objects at increasing angles from the imaginary field of view. The *mental transformation model*, inspired by the classic work in imagery (see, for example, Kosslyn 1980; Shepard and Cooper 1982), takes this into account. According to this model, subjects imagine themselves in the setting, facing frontward. When given a direction and asked to identify the associated object, they imagine themselves turning

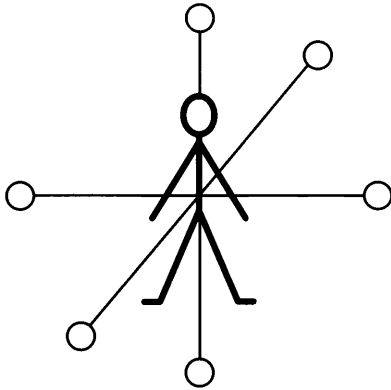


Figure 12.1

Schematic of situation where observer is upright and surrounded by objects.

to face that direction in order to access the object. In this case, times to front should be fastest, and times to back slowest, with times to head, feet, left, and right in between. The obtained pattern of data, displayed in table 12.1, contradicted both these models, but supported a third model, the *spatial framework model*.

The reaction times to access objects in the six directions from the body fit the third model, the spatial framework model. This model was inspired by analyses of Clark (1973), Fillmore (1982), Levelt (1984), Miller and Johnson-Laird (1976), and Shepard and Hurwitz (1984), but differs from each of them. According to it, subjects construct a mental spatial framework, consisting of extensions of the three body axes, and associate objects to the appropriate direction. The mental framework preserves the relative locations of the objects as the subject mentally turns to face a new object, allowing rapid updating. Accessibility of directions seems to depend on the enduring characteristics of the body and the perceptual world, rather than on the immediate imagery of the world. For an upright observer, the head/feet axis is most accessible both because it is an asymmetric axis of the body and because it coincides with the axis of gravity, the only asymmetric axis of the world. The front/back axis is next because it is also an asymmetric body axis, and the left/right axis is least accessible, having no salient asymmetries. The (upright) spatial framework pattern of reaction times, head/feet faster than front/back faster than left/right, was obtained in five experiments (Franklin and Tversky 1990) and in several replications since (e.g., Bryant and Tversky 1991; Bryant, Tversky, and Franklin 1992).

When the observer is described as reclining in the scene, the observer is described as sometimes lying on front, sometimes back, sometimes each side, so that no axis of the body coincides with gravity. Accessibility of objects, then, depends primarily on

Table 12.1
Representative Mean Reaction Time from Spatial Framework Experiments (ms)

	Head/ feet	Front	Back	Front/ back	Left/ right
Upright internal ^a	1.51	1.55	1.68	1.62	1.92
Reclining internal ^b	2.14			1.82	2.59
Upright external ^a	1.30	1.54	1.49	1.52	1.76
Two perspectives, different scenes ^c	3.50			3.99	4.48
Two perspectives, same scenes ^d	3.80			3.81	4.05

Sources:

a. Bryant, Tversky, and Franklin 1992, experiment 4.

b. Franklin and Tversky 1990, experiment 5.

c. Franklin, Tversky, and Coon 1992, experiment 4.

d. Franklin, Tversky, and Coon 1992, experiment 3.

Technique differed for Franklin, Tversky, and Coon; times are therefore not comparable to previous studies.

the relative salience of the body axes. The asymmetries of the front/back body axis are most salient because they separate the world that can be easily sensed and easily manipulated from the world that is difficult to sense or manipulate. The head/feet axis is next most salient, for its asymmetries, and the left/right axis is least salient. This pattern of data (see table 12.1), the reclining spatial framework pattern, with front/back faster than head/feet faster than left/right, appeared in two experiments (Franklin and Tversky 1990) and in subsequent replications (e.g., Bryant and Tversky 1991; Bryant, Tversky, and Franklin 1992). In this study and the previous ones, narratives addressed the subject as “you,” determining the subject’s perspective as that of the observer, surrounded by a set of objects.

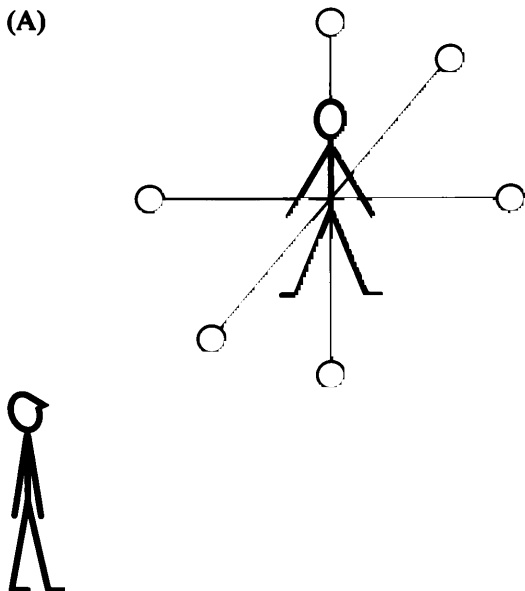
12.5.1 Central Third-Person Character and Objects

The spatial framework studies discussed thus far serve as background for the studies investigating perspective I will now describe. These studies also presented narratives describing objects surrounding observers, but subjects were free to choose a perspective among several possible ones (Bryant, Tversky, and Franklin 1992; Franklin, Tversky, and Coon 1992). In the studies described previously, narratives used the second-person “you” to draw the reader into the scene and induce the reader to take the perspective of a central character surrounded by a set of objects. Bryant, Franklin, and I (Bryant, Tversky, and Franklin 1992) wondered whether use of the second-person pronoun was necessary for perspective taking, or whether readers would take

the perspective of an observer described in the third person, or even take the perspective of an object. Because, according to literary lore, readers often identify with protagonists, we expected readers to take the perspectives of third-person observers as long as the spatial probes were from that perspective. We also expected readers to take the perspectives of objects when the spatial probes were from that perspective. Nevertheless, it was also possible that readers would take the perspective of an outside observer, looking onto the scene. We altered the narratives so that in one experiment, "you" was replaced by a proper name, half male and half female, and in another experiment, "you" was replaced by a central object. The central objects were chosen to have intrinsic sides and were turned in the scene by an outside force to face different objects. One example was a saddle in a barn, surrounded by appropriate objects. For both cases, it would be natural for subjects to take an *external* perspective, looking onto the character or object surrounded by objects rather than the *internal* perspective of the central character or object.

In order to distinguish which perspective subjects adopted in these narratives, we first needed to know the reaction time patterns for external perspective. We knew the pattern for internal perspectives, that is, the upright spatial framework pattern obtained in previous studies. We developed two types of explicitly external narratives, one where narratives described a second-person observer looking onto a scene where a character was surrounded by objects to all six sides of the character's body and one where narratives described a second-person observer looking onto a cubic array of six objects. Figure 12.2 portrays both situations schematically. The spatial framework in this case is constructed from extensions of the three body axes in front of the observer, to the scene, but because the objects are located with respect to the central character and not the observer, the relative salience of the observer's body axes is not relevant to accessibility. The characteristics of the observer's visual field are relevant to accessibility. The pattern predicted is similar to the upright internal spatial framework, but for slightly different reasons. Head/feet should be fastest because of gravity. Front/back should be next fastest because of asymmetries in the front/back visual field. In the case of external arrays, all of the objects are in front of the observer, but those described as to the front (this is English, not Hausa; cf. Hill 1982) appear larger and closer and may occlude or partially occlude those to the back. The left/right visual field has no asymmetries, and thus is predicted to be slowest. There is one difference expected between internal and external spatial frameworks. Front is expected to be faster than back for the internal case because the objects to the back cannot be seen, but not faster for the external case. The predicted patterns appeared for the two external arrays as well as for the internal arrays (see table 12.1). Thus one important factor in determining perspective in narrative comprehension is the perspective of the narrative. Subjects adopted an external point of view when narratives

(A)



(B)

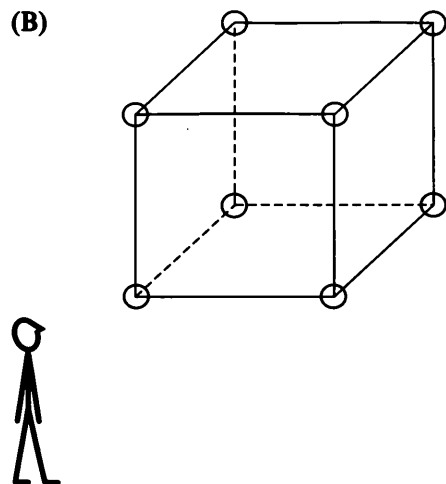


Figure 12.2

Schematic of external situations: (A) An observer looking at a central character surrounded by objects. (B) An observer looking at a cubic array of objects.

questioned them from that point of view, and an internal point of view when narratives questioned them from an internal point of view. The next step was to see what perspective subjects would adopt when narratives allowed either option.

With these findings in mind, we can return to the situation of a single central character or object surrounded by objects and described in the third person. If readers take the internal perspective of the central character or object, then times to front should be faster than times to back. If they take the external perspective of someone observing the scene, then times to front and back should not differ. In fact, times to front were faster than times to back, suggesting that readers spontaneously adopt the perspective of a central character or object, even if the character or object is described in the third person. The patterns of time to characters and objects differed in one way. For objects, the terms *head* and *feet* are not as appropriate as the terms *top* and *bottom*, so the latter terms were used. *Top*, however, can refer both to the intrinsic top of an object and the top currently upward. The converse holds for *bottom*. For objects with intrinsic sides oriented in an upright manner, these uses coincide. For objects turned on their sides, the two uses of top (and bottom) conflict, and, indeed, reaction times to judge what object was located away from the central object's top and bottom were unusually long when objects were turned on their sides. In any case, readers readily take the perspective of either a character or an object central in a scene, even when the character or object is described in the third person.

12.5.2 Two Perspectives in the Same Narrative

The second set of studies investigated perspective taking in narratives describing two different perspectives (Franklin, Tversky, and Coon 1992). The question of interest was how subjects would handle the two perspectives. Would they switch between perspectives depending on which perspective was probed, or would they take a perspective that included both but was neither? There were several different kinds of narratives, describing two people in a scene, surrounded by the same or different set of objects, or two people in two different scenes, surrounded by different sets of objects, or the same person in the same scene, surrounded by the same set of objects, but facing different objects at different times. A schematic of some of the situations appears in figure 12.3. Subjects could adopt one of two strategies for the case of two viewpoints. They could take each perspective in turn as each was probed. That would require perspective switching. Alternatively, they could adopt a single perspective, one neutral in the sense of not being the perspective of any of the characters, but one that includes both viewpoints. An oblique perspective, for example, overhead or nearly overhead, could include both viewpoints, all the relevant characters and objects. If subjects take each observer's viewpoint in turn, then the spatial framework pattern of data should be evident. If they adopt a perspective that includes both

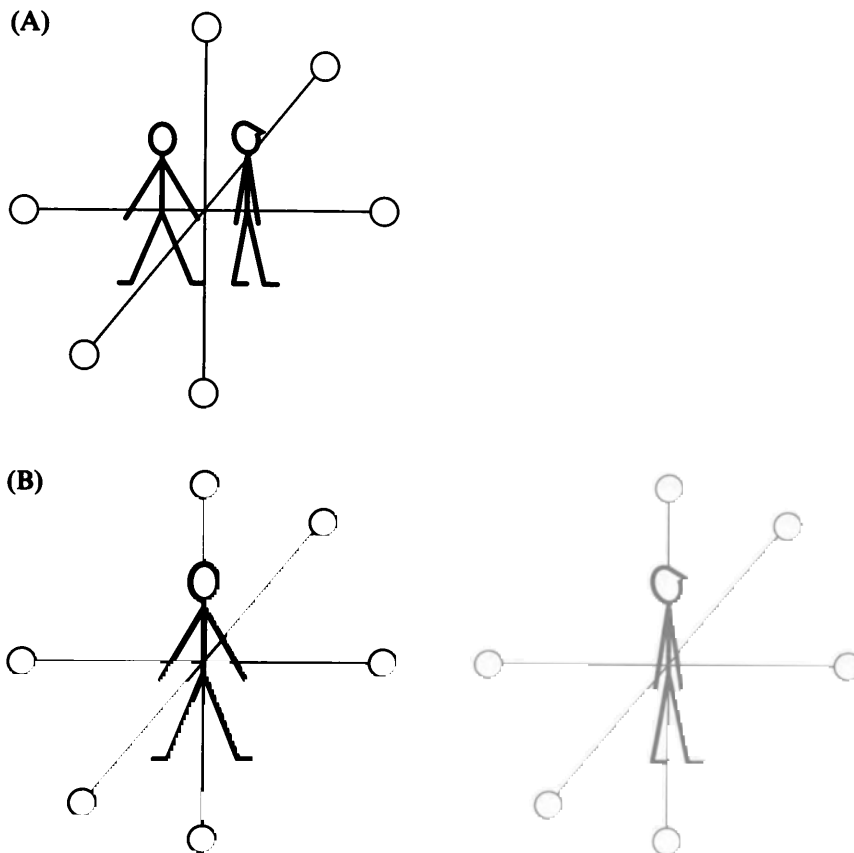


Figure 12.3

Schematic of situations with two viewpoints: (A) Two observers surrounded by different objects facing different directions in same scene. (B) Two observers surrounded by different objects, either in same scene or different scenes.

viewpoints but is not equivalent to either, then some other pattern of reaction times may emerge.

The two strategies seem to differ cognitively. To take each perspective in turn, subjects need to keep in mind a smaller set of tokens for characters and objects, only those currently associated with that perspective. However, this would require mentally changing the viewpoint and mentally changing the set of tokens each time a new viewpoint is probed. To take a neutral perspective on the entire scene would entail keeping more tokens in mind, but would not require mentally changing the set of tokens each time a new viewpoint is probed. The external spatial framework pattern

would not be expected in this case because two characters and objects need to be kept in mind. This seems to require taking an oblique viewpoint in which the bodies of the characters are not aligned with the body of the subject in the mental viewpoint.

The two strategies seem to trade off the size of the mental model with the need to switch mental models. Despite their cognitive differences, neither strategy was preferred overall. Subjects used both strategies, depending on the narrative. When narratives described two observers in the same scene, whether surrounded by the same or different objects, subjects seemed to adopt a neutral oblique perspective, rather than the viewpoints of either observer. In this case, the data did not correspond to the spatial framework pattern but rather to the equiavailability pattern, or to what we termed *weak equiavailability*. Either times were equal to all directions or times to right/left were a little slower. This pattern appeared even when one of the characters in the scene was described as "you," and the other was described in the third person. This corroborates the finding of Bryant, Tversky, and Franklin (1992) that qualities of the described scene determine perspective, not whether the central character is described in the second or third person. When narratives described two observers in difference scenes, subjects took the viewpoint of each observer in turn. In this case, the spatial framework pattern of reaction times obtained (see table 12.1).

In both the cases where narratives described a central character or object in the third person and the cases where narratives described more than one perspective, readers appeared to adopt one perspective for each scene. When there were two observers each with their own viewpoint but in the same scene, readers adopted a neutral perspective rather than that of the observers. When there were two observers in different scenes, readers took the viewpoints of the observer in each scene. Thus qualities of the scene, in this case, the described scene, determine perspective.

To summarize the results, it seems that readers prefer to take a single perspective on a single described scene. If there is a single character (or object), readers will adopt that character's perspective whether or not that perspective is explicit in the description. If there is more than one perspective explicit in the described scene, readers will adopt a neutral perspective that includes the entire scene. Would the same effects appear for scenes that are viewed, as opposed to described? We would not expect *viewers* of a scene to readily take any perspective other than their own. Without closing their eyes, viewers cannot easily get out of their own perspectives. To simultaneously hold their own view as well as the view of another or a neutral view imposes an extra cognitive burden, one that people assume on occasion, but not without effort.

12.6 Production: Nature of the Environment to be Described

Perusing a shelfful of travel guidebooks reveals two popular styles of describing a city or other tourist attraction. A *route* description takes “you,” the reader, on a mental tour; it uses a changing view from within the environment, and locates landmarks with respect to you in terms of “your” front, back, left, and right. A *survey* description, in contrast, takes a static view from above the environment and locates landmarks with respect to each other in terms of north, south, east, and west. A route description uses an intrinsic perspective, where locations are described in terms of the intrinsic sides of “you.” A survey description uses an extrinsic perspective. Thus, both perspectives are neutral because they are not the perspectives of the participants.

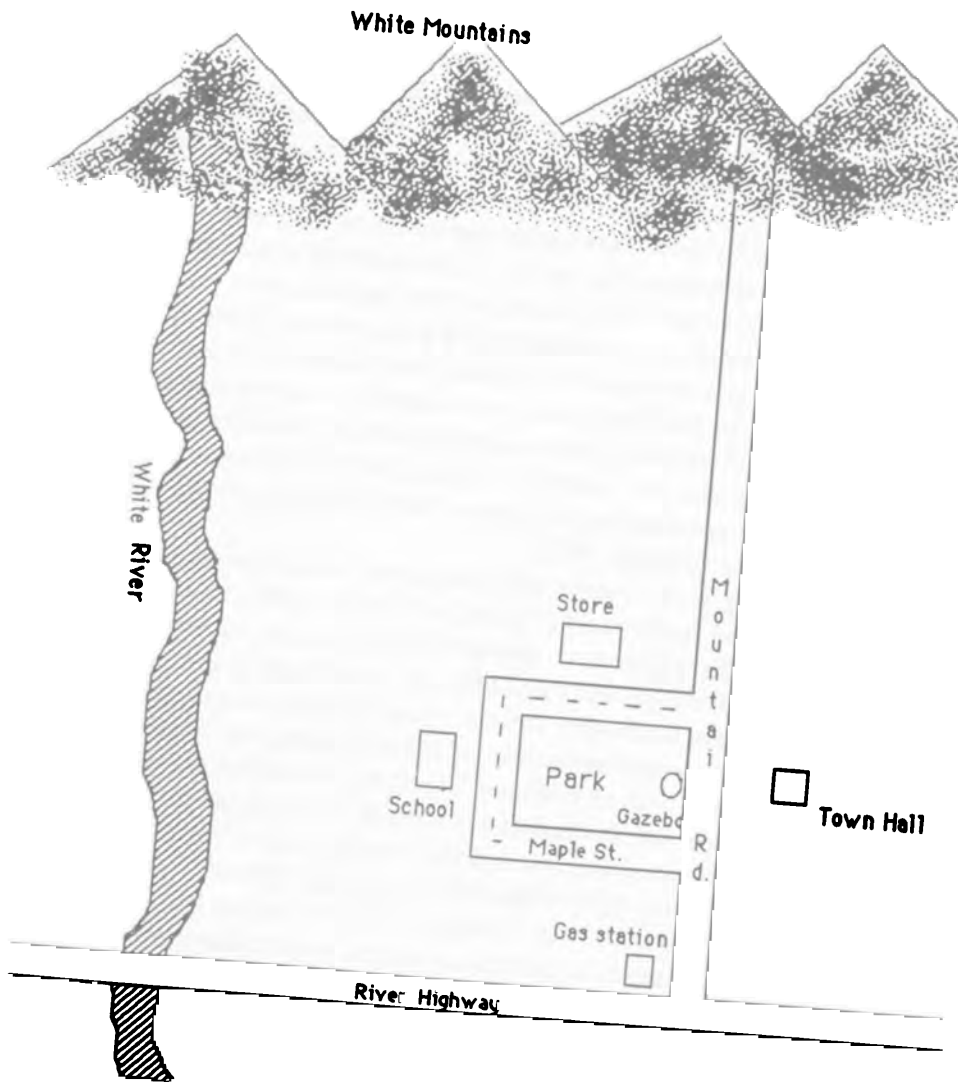
As noted previously, Levelt (1989) has argued that because a tour is a natural way of experiencing an environment, a mental tour is a natural way of describing one. A survey, too, is a natural way to experience, hence describe, an environment. A survey view can be obtained by climbing a tree or a mountain. A survey is analogous to a map in many ways, and maps have been created by cultures for millennia, even before the advent of writing (see, for example, Wilford 1981). Moreover, there is good evidence that survey knowledge can be inferred from route experience (e.g., Landau 1988; Levine, Jankovic, and Palij 1982).

In order to investigate the perspectives that people spontaneously use when describing environments, Taylor and I (Taylor and Tversky 1992a, 1996) gave subjects one of three maps to learn. The maps were of a small town, an amusement park, and a convention center. The town and the convention center maps appear in figure 12.4. Each had about a dozen landmarks. After learning the maps, subjects were asked to describe them from memory. Importantly, all subjects treated the maps as representing environments rather than as marks on paper; they described the environments, not the marks on paper (cf. Garrod and Anderson 1987). In contrast to previous research, subjects used not only route but also survey perspectives in their descriptions. Only one of the sixty-eight subjects did something different; that subject constructed a gaze tour from a stationary viewpoint. This is curious because it required X-ray vision. Also in contrast to previous research, subjects frequently mixed perspectives, nearly half of them, usually without signaling. For example, several subjects described the town by first describing the major landmarks, the mountains, river, and highways, in relation to the canonical directions, and then took readers on a tour of the park and the surrounding buildings. Often subjects combined perspectives, for example, “You turn north” or “X is on your right, north of Y.”

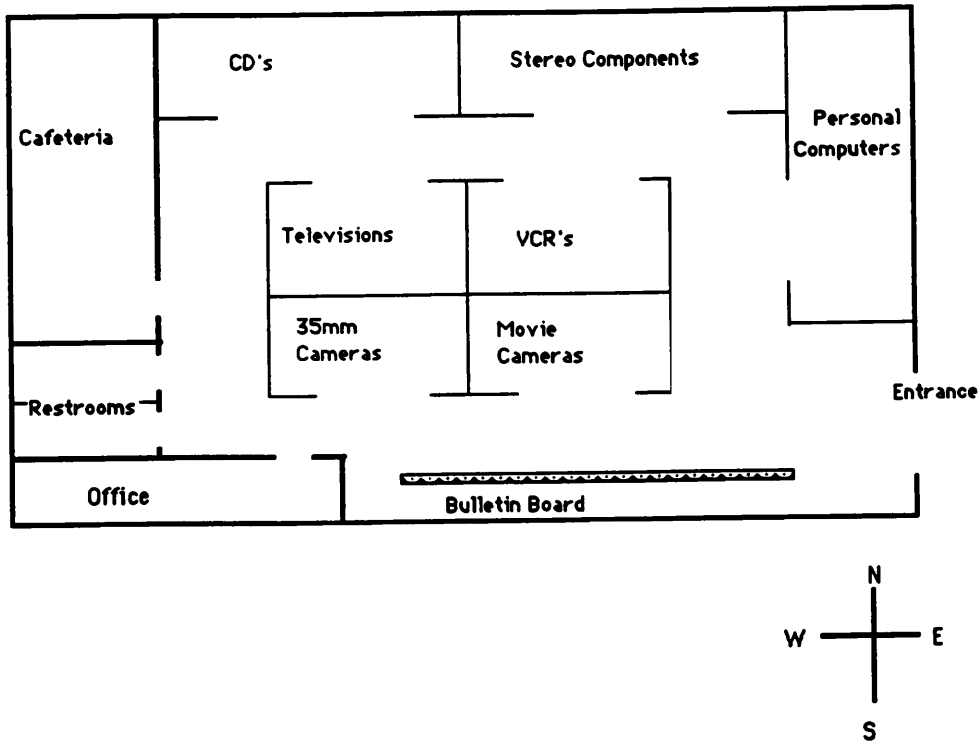
The descriptions that subjects produced were accurate and complete. They allowed other subjects to produce maps that had very few errors or omissions. By this measure, the mixed perspective descriptions were as accurate as the pure ones.

(A)

Barbara Tversky

TOWN

(B) CONVENTION CENTER

**Figure 12.4**

Maps of the town (A) and the convention center (B) from Taylor and Tversky (1992a, b). Used with permission.

We initially categorized the descriptions as route, survey, or mixed on the basis of intuitions and agreed between ourselves. Then we counted frequencies of perspective-relevant uses of language for each perspective category. Route descriptions used active verbs such as *go* or *turn* most frequently, and survey descriptions used stative verbs such as *is* most frequently, with mixed descriptions in the middle. Survey descriptions also used motion verbs statively (see Talmy, chapter 6, this volume); for example, the “road runs east and west.” Route descriptions were most likely to use viewer-centered relational terms, such as *front* and *left*, and survey descriptions were most likely to use environment-centered relational terms, such as *north* and *east*, with mixed descriptions in between. Route descriptions were most likely to use the viewer

as a referent for the location of landmarks, and survey descriptions were most likely to use landmarks as the referent for other landmarks, again with mixed descriptions in between.

With respect to the referent for the location landmarks, route descriptions resembled that of Ullmer-Ehrich's (1982) walking tour. Landmarks were described relative to "your" changing location, as in "if you turn left on Maple St., you will see the School straight ahead." Similarly, the discourse of survey descriptions resembled that of Ullmer-Ehrich's gaze tour. Landmarks were described relative to other landmarks, as in "The Town Hall is east of the Gazebo across Mountain Road," or "The lamp is behind the table." Because it is fixed and external to the scene, the viewpoint of a gaze tour functions like the cardinal directions in a survey tour. Nevertheless, gaze tours may be relative in Levinson's sense (see chapter 4, this volume); for example, "The bookcase is to the right of the lamp" is a ternary relation requiring knowledge of the speaker's location and orientation. Gaze tours, routes, and surveys, then, are ways to organize extended discourses, corresponding to relative, intrinsic, and extrinsic perspectives, respectively.

Although language was used quite differently in route and survey descriptions, the environments were organized similarly for both perspectives (Taylor and Tversky 1992a). A simple and widely used index of mental organization is the order of mentioning items in free recall (see, for example, Tulving 1962); in this case, the order of mentioning landmarks. The basic idea, an idea underlying association in memory, is that things that are related are remembered together. The landmarks in the maps could be studied and learned in any order; thus the order of mentioning them is imposed by the subject, and presumably reflects the way the subject has organized them in memory. There was a high correlation across subjects in the order of mentioning landmarks irrespective of description perspective. Organization of description and perspective of description appeared to be independent. Organization was hierarchical, with natural starting points perceptually and/or functionally determined. Environments were decomposed to regions by proximity, size, or function. Starting points were typically entrances or large landmarks.

Overall, approximately equal numbers of subjects gave route, survey, and mixed descriptions, but the proportion of each was not the same for each map. Perspective seemed to depend on the environment. For the town, there were very few pure route descriptions; the majority of descriptions were evenly split between mixed and survey. For the convention center, there were very few pure survey descriptions, and the majority of descriptions were evenly split between mixed and route. For the amusement park, no dominant perspective was evident. Both the mixing of perspectives and the priority of organization over perspective choice are consistent with Levelt's distinction between macroplanning and microplanning in speech (Levelt 1989 and

chapter 3, this volume). Overall organization of the environment would be part of macroplanning, and perspective choice part of subsequent microplanning.

The correlation of perspective with environment suggested that features of the environment determine perspective in language. The convention center and town differed in several ways. The convention center was relatively small and the town relatively large; the convention center was enclosed and the town open. In the convention center, the landmarks, in this case, the exhibition rooms, were on the same size scale. In the town, the landmarks were on different size scales, the mountains and river formed one scale, the roads and highways another, and the buildings a third. Finally, there was a single path for navigating the convention center, but several ways to navigate the town.

In a subsequent study (Taylor and Tversky 1996), we created sixteen maps to counterbalance these four factors; whether the environment was large or small, whether the environment was closed or open, whether the landmarks were on a single size scale or several size scales, and whether there was a single or several paths through the environment. Subjects studied four maps and wrote descriptions after each. The descriptions were coded as route, survey, or mixed as before. In contrast to the earlier study, where frequency of route, survey, and mixed descriptions were about equal, in this study, 22% of the descriptions were route, 36% were mixed, and 42% were survey. Neither the overall size of the environments nor whether the environments were enclosed or open—that is, neither global feature—had any effect on description perspective. Rather, it was the internal structure of the environments that affected the relative proportions of route and mixed perspectives (the proportion of survey descriptions remained constant). When landmarks were on a single size scale, there were relatively more route and relatively fewer mixed perspective descriptions than when the landmarks were on several size scales. When there was a single path through the environment, there were relatively more route and relatively fewer mixed perspective descriptions than when there were multiple paths through the environment. Of course, it is simpler to plot a route among all the landmarks where there is one and only one. The apartments that Linde and Labov's (1975) subjects described typically had landmarks, that is, rooms, on a single size scale and had a single path through the environment, and yielded primarily route descriptions.

In extended discourse, people frequently switched perspective rather than maintaining a single perspective. Perhaps because the organization of the description superseded the choice of perspective, switching perspective did seem to reduce comprehensibility of description. Choice of perspective, whether route, survey, or mixed, was affected by features of the environment. Both route and survey descriptions are analogous to natural ways of experiencing environments but seem appropriate to different situations. Route descriptions or mental tours were more likely when there

was only a single way to navigate an environment and when an environment had a uniform size scale of landmarks. Finally, gaze tours have been obtained for descriptions of single rooms (Ehrich and Koster 1983; Ullmer-Ehrich 1982) as well as for simple networks on a page (Levelt 1982a,b). Gaze tours seem more likely when the entire environment can be viewed from a single place.

12.7 Production: Cognitive and Social Determinants

The previous studies have investigated some of the cognitive factors affecting choice of perspective, the nature of the described scene, and the nature of the environment. As Schober and Hermann (cited in Schober 1993) have observed, social factors also affect perspective choice. To incorporate both, I have proposed another way of categorizing perspective, first as to whether perspective is personal or neutral. Personal perspective can be decomposed to “yours” or “mine,” that is, speaker’s or addressee’s. Neutral perspective can also be decomposed, to intrinsic or extrinsic.

To get greater clarity on determinants of perspective in simple situations, Mainwaring, Schiano, and I (Mainwaring, Tversky, and Schiano 1994) have developed several variants of the paradigm of Schober (1993) described earlier. One of these will be described here. We constructed diagrams that were structurally similar to Schober’s; in each case, there were two objects, identical except for location. The subject’s task was to describe the location of the critical object. The situation is sketched in figure 12.5, though the actual diagrams were different. Schober’s task forced subjects to use a personal reference system, either the speaker’s or that of the

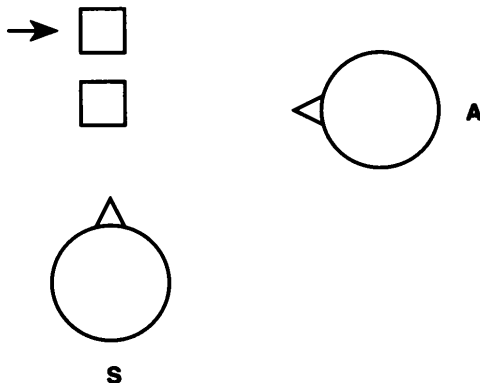


Figure 12.5

Schematic of situation where speaker and addressee are at right angles and objects are aligned with speaker.

addressee. This was the case for some of our diagrams, but for others, we added either a landmark or extrinsic directions, so that subjects had the option of using either a personal or a neutral reference system on many diagrams.

The diagrams manipulated the difficulty of the personal perspectives by varying the spatial relations between speaker and addressee and between objects and participants. The speaker was either facing the addressee or at right angles to the addressee. The two objects were either lined up with the speaker, so that from the speaker's point of view one was near and the other far, or positioned so that one object was to the speaker's left and the other to the speaker's right. When the speaker and the addressee were facing each other, then the type of relation, near/far or left/right, was the same for both, but when the speaker and addressee were at right angles, then a near/far relation for one was a right/left relation for the other. In the first case, difficulty was the same for speaker and addressee, but in the second case, where speaker and addressee were at right angles, what was easier for speaker was harder for addressee, and vice versa. Instead of communicating in pairs, subjects gave descriptions for an unknown other. With only personal reference systems possible, Schober had found that speakers tended to take the addressee's perspective. The frequency of taking the other's perspective increased when the other was unknown, rather than an active partner.

We also added a cover story. You and the other were special agents in a secret security agency. The diagrams represented dangerous missions that the two of you undertook. Each diagram portrayed a scene in which the locations of you and your partner were indicated, as well as the locations of two identical objects, bombs, treasures, or the like. In each case, you knew which object was the critical one, and when your partner gave a signal, you described the critical object briefly and directly into your secret decoder pad for your partner.

The data I am reporting are preliminary; data collection is continuing. Some effects are already apparent. From Schober's (1993) research, we expected that when only a personal perspective was possible, the speaker would take the addressee's. However, we expected cognitive difficulty to attenuate that tendency. Left/right distinctions are more difficult to produce and comprehend than near/far distinctions. When the speaker and addressee are at right angles and the objects are lined up with the speaker, the speaker needs to use *left* or *right* in order to take the addressee's perspective (see figure 12.5). If speakers realize this difficulty, they may choose to use their own perspective and the simpler terms *closer* or *farther*, sacrificing politeness to reduce difficulty. In fact, in 37% of the cases, speakers did exactly that, compared to 2% of the cases where the objects were lined up with the addressee and the speaker could use *closer* or *farther* from the addressee's perspective (could reverse the positions of speaker and addressee in figure 12.5).

We also expected the presence of a neutral perspective to attenuate the tendency of speakers to take addressees' perspectives. Selecting a neutral reference avoids the entire issue of whose perspective to take. When subjects were told which direction was north, that is, when an extrinsic reference frame was available, they took a personal perspective only 56% of the time. The presence of a landmark also reduced the frequency of taking a personal perspective, but to a lesser extent, to 64% of the time. An extrinsic system may be more likely to replace a personal system than a landmark because an extrinsic system is more global and permanent than a landmark. This is supported by the finding that subjects were more likely to describe the location redundantly, that is, to use both a personal and a neutral perspective, when the neutral perspective was a landmark than when the neutral option was the cardinal directions. Whether a landmark was used depended on the difficulty of describing it; here, difficulty translates into binary or ternary in Levinson's terms (see chapter 4). Using a landmark was more frequent when the target object could be described as closer or farther to the landmark from the addressee's perspective, that is, used intrinsically, than when the target object had to be described as left or right of the landmark from the addressee's perspective, that is, used relatively.

These results illustrate the complex interplay between social and cognitive factors in selecting a perspective. When only a personal reference system was available, there was a strong tendency, even stronger in a hypothetical rather than a real interaction (Schober 1993), for the speaker to take the addressee's perspective. In the present data, that tendency was sometimes overcome when the addressee's perspective was more difficult to produce and comprehend than the speaker's. When a neutral perspective was available in addition to a personal perspective, there was a weak tendency for the speaker to take the addressee's perspective, especially when the neutral perspective was extrinsic, rather than a landmark. An extrinsic reference is more global and permanent than a landmark, a characteristic of the environment. Cognitive difficulty also affected choice between a personal and a neutral perspective. When a landmark was easier to describe than a personal reference, it was more likely to be used. Note that these different choices of reference systems appeared in the same subjects communicating with the same hypothetical addressees. Perspective was anything but consistent. We can infer from this that the cognitive cost of switching perspective was often less than the cognitive cost of describing from certain perspectives.

12.8 Summary and Conclusion

Many disciplines in cognitive science have been intrigued with the issue of perspective. It is critical to theories of recognizing objects and navigating environments, and the development of these abilities; it has been of concern to neuropsychologists and

linguists. Despite many differences in issues, a survey of these disciplines yielded three main bases for spatial reference systems: relative (viewer-centered, egocentric, personal), intrinsic (object-centered, landmark-based), and extrinsic (external).

Perspective in language use is of particular interest because language allows us to use perspectives other than those given by perception. Although there have been many claims about perspective use in language, research on what people actually do is just beginning. Some of that research was reviewed here, along with more detailed descriptions of three current projects related to perspective choice.

Several conclusions emerge from the review of these studies on the comprehension and production of perspective in descriptions. First, there does not seem to be a default perspective. Different perspectives are adopted in different situations. Some of the influences on perspective choice are cognitive and include the viewpoint of the description, the characteristics of the described scene or scene to be described, and the relative difficulty of various perspectives. Second, perspective is not necessarily consistent. People not only spontaneously select different perspectives for different situations, they also switch perspectives, often without signaling, or use more than one perspective redundantly, even in the same discourse. Third, perspective might be better classified another way, one with distinctions at two levels. The primary distinction would be between perspectives that are personal and perspectives that are neutral. Each of these classes subdivides into two further classes. Personal perspectives are those of the participants in the discourse; they include yours and mine, that is, the speaker's and the addressee's. Neutral perspectives do not belong to the participants in the discourse; they include intrinsic or landmark-based perspectives and extrinsic or external perspectives. This classification draws attention to social influences on perspective choice, for example, attributions about the addressee. Interestingly, many of the relevant attributions about addressees are cognitive in nature, for example, what may be more or less difficult for an addressee to comprehend.

Of necessity, individuals begin with their own perspectives, yet to function in the world, to recognize objects, to find one's way in the world, to communicate to others, other perspectives must be known and used. Figuring out how we come to have perspectives other than our own has attracted scholars from many disciplines. Yet another reason researchers are drawn to the study of perspective is its social sense. Individuals have different perspectives, not just on space, but on the events that take place in space. They also have different perspectives on beliefs, attitudes, and values. For the endless discussions people have on these topics, the mine-yours-neutral distinction is essential. Reconciling my memory or beliefs or attitudes or values to yours might (or might not) best be accomplished by moving from personal to neutral ground. Going beyond personal perspective is as critical to social interaction as it is to spatial cognition.

Acknowledgments

I am indebted to my collaborators, Nancy Franklin, Holly Taylor, David Bryant, Scott Mainwaring, and Diane Schiano, for years of lively interchanges, to Mary Peterson and Lynn Nadel for valuable comments on an earlier draft, and to Eve Clark, Herb Clark, Pim Levelt, Steve Levinson, Eric Pederson, Michael Schober, and Pam Smul for ongoing discussions on deixis and perspective. Research reviewed here was supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under grant or cooperative agreement number AFOSR 89-0076 to Stanford University, and by Interval Research Corporation.

References

- Black, J. B., Turner, T. J., and Bower, G. H. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning and Verbal Behavior*, 18, 187–198.
- Brown, P., and Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Bryant, D. J., and Tversky, B. (1991). Locating objects from memory or from sight. Paper presented at Thirty-second Annual Meeting of the Psychonomic Society, San Francisco, November.
- Bryant, D. J., Tversky, B., and Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Language and Memory*, 31, 74–98.
- Bühler, K. (1934). The deictic field of language and deictic words Translated from the German and reprinted in R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action*, 9–30. New York: Wiley, 1982.
- Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*, 27–63. New York: Academic Press.
- Clark, H. H. (1987). Four dimensions of language use. In J. Vershueren and M. Bertucci-Papi (Eds.), *The pragmatic perspective*, 9–25. Amsterdam: Benjamins.
- Couclelis, H., Golledge, R. G., Gale, N. and Tobler, W. (1987). Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7, 99–122.
- Ehrich, V., and Koster, C. (1983). Discourse organization and sentence form: The structure of room descriptions in Dutch. *Discourse Processes*, 6, 169–195.
- Farrell, W. S. (1979). Coding left and right. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 42–51.
- Fillmore, C. (1975). *Santa Cruz lectures on Deixis*. Bloomington, In: Indiana University Linguistics Club.
- Fillmore, C. (1982). Toward a descriptive framework for spatial deixis. In R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action*, 31–59. London: Wiley.
- Franklin, N., and Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119, 63–76.

- Franklin, N., Tversky, B., and Coon, V. (1992). Switching points of view in spatial mental models acquired from text. *Memory and Cognition*, 20, 507–518.
- Garrod, S., and Anderson, S. (1987). Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition*, 27, 181–218.
- Glenberg, A. M., Meyer, M., and Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory Language*, 26, 69–83.
- Hart, R. A. and Moore, G. T. (1973). The development of spatial cognition. In R. M. Downs and D. Sten (Eds.), *Image and environment*, 246–288. Chicago: Aldine.
- Hill, C. (1982). Up/down, front/back, left/right: A contrastive study of Hausa and English. In J. Weissen and W. Klein (Eds.), *Here and there: Crosslinguistic studies on deixis and demonstration*, 13–42. Amsterdam: Benjamins.
- Hirtle, S. C., and Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory and Cognition*, 13, 208–217.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory and Cognition*, 13, 289–303.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Landau, B. (1988). The construction and use of spatial knowledge in blind and sighted children. In J. Stiles-Davis, M. Kritchewsky, and U. Bellugi (Eds.), *Spatial cognition: Brain bases and development*, 343–371. Hillsdale, NJ: Erlbaum.
- Levelt, W. J. M. (1982a). Cognitive styles in the use of spatial direction terms. In R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action*, 251–268. Chichester: Wiley.
- Levelt, W. J. M. (1982b). Linearization in describing spatial networks. In S. Peters and E. Saarinen (Eds.), *Processes, beliefs, and questions*, 199–220. Dordrecht: Reidel.
- Levelt, W. J. M. (1984). Some perceptual limitations on talking about space. In A. J. van Doorn, W. A. van der Grind, and J. J. Koenderink (Eds.), *Limits on perception*, 323–358. Utrecht: VNU Science Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levine, M., Jankovic, I. N., and Palij, M. (1982). Principles of spatial problem solving. *Journal of Experimental Psychology: General*, 111, 157–175.
- Linde, C., and Labov, W. (1975). Spatial structures as a site for the study of language and thought. *Language*, 51, 924–939.
- Lynch, K. (1960). *The image of the city*. Cambridge: MIT Press.
- Mainwaring, S. D., Tversky, B., and Schiano, D. (1996). Perspective choice in spatial descriptions. Technical report. Palo Alto, CA: Interval Research Corp.
- Mani, K., and Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory and Cognition*, 10, 181–187.

- Marr, D. (1982). *Vision*. New York: Freeman.
- Marr, D., and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society, London, B200*, 269–291.
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Morrow, D. G., Bower, G. H., and Greenspan, S. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language*, 28, 292–312.
- Morrow, D. G., Greenspan, S., and Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*, 26, 165–187.
- Nigro, G., and Neisser, U. (1983). Point of view in personal memories. *Cognitive Psychology*, 15, 467–482.
- O'Keefe, J., and Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Oxford University Press.
- Perrett, D., Harries, M., Mistlin, A. J., and Chitty, A. J. (1990). Three stages in the classification of body movements by visual neurons. In H. Barlow, C. Blakemore, and M. Weston-Smith (Eds.), *Images and understanding*, 94–107. Cambridge: Cambridge University Press.
- Perrig, W., and Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language*, 24, 503–518.
- Pick, H. L., Jr., and Lockman, J. J. (1981). From frames of reference to spatial representations. In L. S. Liben, A. H. Patterson, and N. Newcombe (Eds.), *Spatial representation and behavior across the lifespan: Theory and application*, 39–60. New York: Academic Press.
- Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, 18, 1–63.
- Presson, C. C., and Hazelrigg, M. D. (1984). Building spatial representations through primary and secondary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 716–722.
- Robin, F., and Denis, M. (1991). Description of perceived or imagined spatial networks. In R. H. Logie and M. Denis (Eds.), *Mental images in human cognition*, 141–152. Amsterdam: North-Holland.
- Sadalla, E. K., Burroughs, W. J., and Staplin, L. J. (1980). Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 516–528.
- Schober, M. F. (1993). Spatial perspective taking in conversation. *Cognition*, 47, 1–24.
- Shepard, R. N. (1984). Ecological constraints on internal representations: Resonant kinematics of perceiving, imaging, thinking, and dreaming. *Psychological Review*, 91, 417–447.
- Shepard, R. N., and Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Shepard, R. N., and Hurwitz, S. (1984). Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition*, 18, 161–193.

- Tarr, M., and Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology*, 21, 233–282.
- Taylor, H. A., and Tversky, B. (1992a). Descriptions and depictions of environments. *Memory and Cognition*, 20, 483–496.
- Taylor, H. A., and Tversky, B. (1992b). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, 31, 261–292.
- Taylor, H. A., and Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, 35.
- Tulving, E. (1962). Subjective organization in free recall of “unrelated” words. *Psychological Review*, 69, 344–354.
- Ullmer-Ehrich, V. (1982). The structure of living space descriptions. In R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action*, 219–249. New York: Wiley.
- Wilford, J. N. (1981). *The mapmakers*. New York: Knopf.

Chapter 13

A Computational Analysis of the Apprehension of Spatial Relations

Gordon D. Logan and Daniel D. Sadler

13.1 Introduction

Spatial relations are important in many areas of cognitive science and cognitive neuroscience, including linguistics, philosophy, anthropology, and psychology. Each area has contributed substantially to our understanding of spatial relations over the last couple of decades, as is evident in the other chapters in this volume. The psychologists' contribution is a concern for *how* spatial relations are apprehended, a concern for the interaction of representations and processes underlying an individual's apprehension of spatial relations. This chapter presents a computational analysis of the representations and processes involved in apprehending spatial relations and interprets this analysis as a psychological theory of apprehension. The chapter begins with a theory and ends with data that test the assumptions of the theory and with some comments about generality.

13.2 Three Classes of Spatial Relations

A computational theory accounts for a phenomenon in terms of the representations and processes that underlie it, specifying how the processes operate on the representations to produce the observed behavior. Important clues to the nature of the representations and processes involved in the apprehension of spatial relations can be found in the linguistic and psycholinguistic literature that addresses the semantics of spatial relations (e.g., Clark 1973; Garnham 1989; Herskovits 1986; Jackendoff and Landau 1991; Levelt 1984; Miller and Johnson-Laird 1976; Talmy 1983; and Vandaloise 1991). That literature distinguishes between three classes of spatial relations, and the discriminanda that distinguish the classes suggest the requisite representations and processes.

13.2.1 Basic Relations

Garnham (1989) distinguished basic relations from deictic and intrinsic ones. Basic relations take one argument, expressing the position of one object with respect to the viewer (e.g., the viewer thinks, "This is here" and "That is there").¹ Basic relations are essentially the same as *spatial indices*, which are discussed in the literature on human and computer vision (e.g., Pylyshyn 1984, 1989; Ullman 1984). Spatial indices establish correspondence between perceptual objects and symbols, providing the viewer's cognitive system with a way to access perceptual information about an object. Spatial indices—basic relations—individuate objects without necessarily identifying, recognizing, or categorizing them. The conceptual part of a basic relation is a symbol or a token that stands for a perceptual object. It simply says, "Something is there," without saying what the "something" is. The token may be associated with an identity or a categorization, pending the results of further processing, but it need not be identified, recognized, or categorized in order to be associated with a perceptual object. The perceptual part of a basic relation is an object that occupies a specific point or region in perceptual space.

Basic relations represent space in that they associate a conceptual token with the object in a location in perceptual space. Conceptually, the representation of space is very crude—an object is "here" and "not there." Thus two objects that are indexed separately can either be in the same location or in different locations. If they are in different locations, their relative positions are not represented explicitly in the conceptual representation. Information about their relative locations may be available implicitly in perceptual space, but it is not made explicit in basic relations. Other relations and other computational machinery are necessary to make relative position explicit.

13.2.2 Deictic Relations

Although Garnham (1989) was the first to distinguish basic relations, most linguists and psycholinguists distinguish between deictic and intrinsic relations (e.g., Herskovits 1986; Jackendoff and Landau 1991; Levelt 1984; Miller and Johnson-Laird 1976; Talmy 1983; and Vandaloise 1991). Deictic relations take two or more objects as arguments, specifying the position of one object, the *located object*, in terms of the other(s), the *reference object(s)*. The position is specified with respect to the *reference frame* of the viewer, which is projected onto the reference object. Deictic relations specify the position of the located object with respect to the viewer if the viewer were to move to the position of the reference object. Thus "The ball is left of the tree" means that if the viewer were to walk to the tree, the ball would be on his or her left side.

Deictic relations are more complex computationally than basic relations because they relate objects to each other and not simply to the viewer. They represent the relative positions of objects explicitly. The arguments of deictic relations must be individuated but they need not be identified, recognized, or categorized. Individuation is necessary because the reference object is conceptually different from the located object (i.e., “*X* is above *Y*” and “*Y* is above *X*” mean different things), but the distinction between reference and located objects can be made by simply establishing tokens that represent perceptual objects, leaving identification, recognition, and categorization to subsequent processes.

13.2.3 Intrinsic Relations

Like deictic relations, intrinsic relations take two or more arguments and specify the position of a located object with respect to a reference object. They differ from deictic relations in that the position is specified with respect to a reference frame intrinsic to the reference object rather than the viewer’s reference frame projected onto the reference object. Whereas deictic relations can apply to any reference object, intrinsic relations require reference objects that have intrinsic reference frames, that is, intrinsic tops and bottoms, fronts and backs, and left and right sides. Objects like people, houses, and cars can serve as reference objects for intrinsic relations because they have fronts, backs, tops, bottoms, and left and right sides. Objects like balls cannot serve as reference objects for intrinsic relations because they have no intrinsic tops, bottoms, and so on. Objects like trees have tops and bottoms but no fronts and backs or left and right sides, so they can support intrinsic *above* and *below* relations but not intrinsic *in front of* or *left of* relations; *in front of* and *left of* would have to be specified deictically. Objects like bullets and arrows have intrinsic fronts and backs but no intrinsic tops and bottoms or left and right sides. They can support intrinsic *in front of* and *behind* relations, but *above* and *left of* would have to be specified deictically.

Intrinsic relations are more complex computationally than deictic relations because they require the viewer to extract the reference frame from the reference object. An obvious way to extract the reference frame is to recognize the reference object or classify it as a member of some category and to impose the reference frame appropriate to that category. For example, seeing an ambiguous figure as a duck or a rabbit leads the viewer to assign *front* to different regions of the object (Peterson et al. 1992). However, it may be possible in some cases to assign an intrinsic reference frame without actually identifying the object. The main axis of the reference frame may be aligned with the object’s axis of elongation (Biederman 1987; Marr and Nishihara 1978) or with the object’s axis of symmetry (Biederman 1987; Palmer 1989).

13.2.4 Implications for Computation

The distinction between the three classes of spatial relations has at least two implications for a theory of the computation involved in apprehension. First, each class of relations describes the position of the located object in terms of a reference frame. The reference frame may coincide with the viewer's, as in basic relations, it may be projected onto the reference object, as in deictic relations, or it may be extracted from the asymmetries inherent in the reference object, as in intrinsic relations. In each case, the reference frame is a central part of the meaning of the spatial relation, and this suggests that reference frame computation is a central part of the process of apprehension.

Second, the distinction between reference objects and located objects suggests that the arguments of two- or three-place relations must be individuated somehow. "*X* is above *Y*" does not mean the same as "*Y* is above *X*." The process of spatial indexing—instantiating basic relations—is well suited for this purpose. Each object can be represented by a different token, and the tokens can be associated with the arguments that correspond to the located and reference object in the conceptual representation of the relation. The distinction between located and reference objects is also important in reference frame computation because the reference frame is projected onto or extracted from the reference object, not the located object. Spatial indexing is useful here as well. It is a central part of apprehension.

13.3 Spatial Templates as Regions of Acceptability

Reference frames and the distinction between located and reference objects suggest important parts of a computational theory of apprehension, but something is missing. They do not specify how one would decide whether a given spatial relation applied to a pair or triplet of objects. This issue has been discussed extensively in the linguistic and psycholinguistic literature. Various researchers have suggested computations involving geometric (Clark 1973; Miller and Johnson-Laird 1976), volumetric (Herskovits 1986; Talmy 1983), topological (Miller and Johnson-Laird 1976; Talmy 1983), and functional (Herskovits 1986; Vandaloise 1991) relations. We propose that people decide whether a relation applies by fitting a *spatial template* to the objects that represents *regions of acceptability* for the relation in question (see also Carlson-Radvansky and Irwin 1993; Hayward and Tarr 1995; Kosslyn et al. 1992; Logan 1994, 1995; Logan and Compton 1996).

A spatial template is a representation that is centered on the reference object and aligned with the reference frame imposed on or extracted from the reference object. It is a two- or three-dimensional field representing the degree to which objects appearing in each point in space are acceptable examples of the relation in question. The

main idea is that pairs or triplets of objects vary in the degree to which they instantiate spatial relations. Roughly speaking, there are three main regions of acceptability: one reflecting good examples, one reflecting examples that are less than good but nevertheless acceptable, and one reflecting unacceptable examples. Good and acceptable regions are not distinct with a sharp border between them. Instead, they blend into one another gradually. With the relation *above*, for example, any object that is aligned with the upward projection of the up-down axis of the reference object is a good example. Any object above a horizontal plane aligned with the top of the reference object is an acceptable example, although not a good one (the closer it is to the upward projection of the up-down axis, the better). And any object below a horizontal plane aligned with the bottom of the reference object is a bad, unacceptable example.

We propose that people use spatial templates to determine whether a spatial relation applies to a pair of objects. If the located object falls in a good or an acceptable region when the template is centered on the reference object, then the relation can apply to the pair. If two relations can apply to the same pair of objects, the preferred relation is the one whose spatial template fits best. If both spatial relations fit reasonably, the viewer may assert both relations (e.g., “above and to the right”). Spatial templates provide information about goodness of fit. Exactly how information about goodness of fit is used depends on the viewer’s goals and the viewer’s task (see below).

13.4 Computational Theory of Apprehension

At this point the representations and processes necessary to apprehend spatial relations have been described in various ways, some in detail, some briefly, and some only implicitly. Now it is time to describe them explicitly and say how they work together.

13.4.1 Representations

The theory assumes that the apprehension of spatial relations depends on four different kinds of representations: a perceptual representation consisting of objects and surfaces, a conceptual representation consisting of spatial predicates, a reference frame, and a spatial template. It may be more accurate to say there are two kinds of representation, one perceptual and one conceptual, and two “intermediate” representations that map perception onto cognition and vice versa.

13.4.1.1 Perceptual Representation The perceptual representation is a two-, two-and-a-half-, or three-dimensional analog array of objects and surfaces. It is formed automatically by local parallel processes as an obligatory consequence of opening one’s eyes (see, for example, Marr 1982; Pylyshyn 1984; and Ullman 1984). The

representation contains information about the identities of the objects and the spatial relations between them, but that information is only implicit. Further computation is necessary to make it explicit. In other words, the representation contains the perceptual information required to identify the objects or to compute spatial relations between them, but that information does not result in an explicit identification of the object as an instance of a particular category or specific relation without further computation. That “further computation” is what the other representations and processes are required for.

The current version of the theory assumes that the perceptual representation is relatively low-level, and that need not be the case. We make that assumption because it is relatively clear how low-level representations can be constructed from light impinging on the retina (e.g., Biederman 1987; Marr 1982), and we want the theory to be tractable computationally. However, the spirit of the theory would not be very different if we assumed that the perceptual representation was much more abstract; for example, if we assumed that spatial information was represented amodally, combining visual, auditory, tactual, and imaginal information. The key idea is that the perceptual representation provides an analog array of objects that can be compared to a spatial template. In principle, the objects can be highly interpreted and abstracted from the sensory systems that gave rise to them.

13.4.1.2 Conceptual Representation The conceptual representation is a one-, two-, or three-place predicate that expresses a spatial relation. The conceptual representation identifies the relation (e.g., it distinguishes *above* from *below*); it individuates the arguments of the relation, distinguishing between the reference object and the located object; it identifies the relevant reference frame (depending on the nature of the reference object); and it identifies the relevant spatial template. The conceptual representation does not identify objects and relations directly in the perceptual representation; further processing and other representations are needed for that.

An important feature of the conceptual representation is that it is addressable by language. The mapping of conceptual representations onto language may be direct in some cases and indirect in others. In English, French, Dutch, and German, for example, many conceptual (spatial) relations are lexicalized as spatial prepositions; single words represent single relations. However, there is polysemy even in the class of spatial prepositions. Lakoff (1987), for example, distinguished several different senses of *over*. Moreover, some languages may use a single word to refer to different relations that are distinguished lexically in other languages. For example, English uses one word for three senses of *on* that are distinguished in Dutch (i.e., *om*, *op*, and *aan*; see Bowerman, chapter 10, this volume). Despite these complexities, we assume that

conceptual representations may be mapped onto language and vice versa. The mapping may not always be simple, but it is possible in principle (see also Jackendoff and Landau 1991; Landau and Jackendoff 1993).

13.4.1.3 Reference Frame The reference frame is a three-dimensional coordinate system that defines an origin, orientation, direction, and scale. It serves as a map between the conceptual representation and the perceptual representation, establishing correspondence between them. The distinction between reference and located objects gives a direction to the conceptual representation; the viewer's attention should move *from* the reference object *to* the located object (Logan 1995). The reference frame gives direction to perceptual space, defining *up*, *down*, *right*, *front*, and *back*. It orients the viewer in perceptual space.

We assume that reference frames are flexible representations. The different parameters can be set at will, depending on the viewer's intentions and the nature of the objects on which the reference frame is imposed. Many investigators distinguish different kinds of reference frames—viewer-based, object-based, environment-based, deictic, and intrinsic (Carlson-Radvansky and Irwin 1993, 1994; Levelt 1984; Marr 1982; Marr and Nishihara 1978). We assume that the same representation underlies all of these different reference frames (i.e., a three-dimensional, four-parameter coordinate system). The differences between them lie in the parameter settings. Viewer-based and object-based reference frames (also known as “deictic” and “intrinsic” reference frames) differ in origin (the viewer vs. the object), orientation (major axis of viewer vs. major axis of object), direction (viewer's head up vs. object's “head” up), and scale (viewer's vs. object's).

13.4.1.4 Spatial Template As we just said, the spatial template is a representation of the regions of acceptability associated with a given relation. When the spatial template is centered on the reference object and aligned with its reference frame, it specifies the goodness with which located objects in different positions exemplify the associated relation.

We assume that different relations have different spatial templates associated with them and that similar relations have similar templates. More specifically, we assume that spatial templates are associated with conceptual representations of spatial relations. Consequently, they are addressable by language, but the addressing is mediated by linguistic access to the conceptual representation. We assume there are spatial templates for lexicalized conceptual representations, but in cases of polysemy where there is more than one conceptual representation associated with a given word (e.g., *over*; Lakoff 1987), there is a different spatial template for each conceptual

representation. Moreover, we assume that spatial templates can be combined to represent compound relations (e.g., “above right”) and decomposed to represent finer distinctions (e.g., “directly above”).

13.4.2 Processes

The theory assumes that the apprehension of spatial relations depends on four different kinds of processes: spatial indexing, reference frame adjustment, spatial template alignment, and computing goodness of fit. The first two establish correspondence between perceptual and conceptual representations; the last two establish the relevance or the validity of the relation in question.

13.4.2.1 Spatial Indexing Spatial indexing is required to bind the arguments of the relation in the conceptual representation to objects in the perceptual representation. Spatial indexing amounts to establishing correspondence between a symbol and a percept. A perceptual object is “marked” in the perceptual representation (Ullman 1984), and a symbol or a token corresponding to it is set up in the conceptual representation (Pylyshyn 1984, 1989). The correspondence between them allows conceptual processes to access the perceptual representation of the object so that perceptual information about other aspects of the object can be evaluated (e.g., its identity). Essentially, the viewer asserts two or three basic relations, one for the located object and one or two for the reference objects.

13.4.2.2 Reference Frame Adjustment The relevant reference frame must be imposed on or extracted from the reference object. The processes involved translate the origin of the reference frame, rotate its axes to the relevant orientation, choose a direction, and choose a scale. Not all of these adjustments are required for every relation. *Near* requires setting the origin and the scale, whereas *above* requires setting origin, orientation, and direction.

Different processes may be involved in setting the different parameters. The origin may be set by spatial indexing (Ullman 1984) or by a process analogous to mental curve tracing (Jolicoeur, Ullman, and MacKay 1986, 1991). Orientation may be set by a process analogous to mental rotation (Cooper and Shepard 1973; Corballis 1988). Different reference frames or different parameter settings may compete with each other, and the adjustment process must resolve the competition (Carlson-Radvansky and Irwin 1994).

13.4.2.3 Spatial Template Alignment The spatial template must be imposed on the reference object and aligned with the reference frame. In deictic relations, the spatial

template is aligned with the viewer's reference frame projected onto the reference object. In intrinsic relations, it is aligned with the intrinsic reference frame extracted from the object.

13.4.2.4 Computing Goodness of Fit Once the relevant spatial template is aligned with the reference object, goodness of fit can be computed. The position occupied by the located object is compared with the template to determine whether it falls in a good, acceptable, or bad region. We assume that the comparison is done in parallel over the whole visual (or imaginal) field. Spatial templates can be represented computationally as a matrix of weights, and the activation value of each object in the visual-imaginal field can be multiplied by the weights in its region to assess goodness of fit. Weights in the good region can be set to 1.0; weights in the bad region can be set to 0.0, and weights in acceptable but not good regions can be set to values between 0.0 and 1.0. With these assumptions, the better the example, the less the activation changes when the spatial template is applied. The activation of good examples will not change at all; the activation of bad examples will vanish (to 0.0); and the activation of acceptable examples will be somewhat diminished.

Alternatively, weights for bad regions could be set to 1.0, weights for acceptable regions could be greater than 1.0, and weights for the good region could be well above 1.0. With these assumptions, the better the example, the greater the change in activation when the spatial template is applied. The activation of bad examples will not change; the weights of acceptable but not good examples will change a little; and the weights of good examples will change substantially. In either case, the acceptability of candidate objects can be assessed and rank-ordered. Other processes and other considerations can choose among the candidates.

13.4.3 Programs and Routines

Spatial relations are apprehended for different reasons in different contexts. Sometimes apprehension itself is the main purpose, as when we want to determine which horse is ahead of which at the finish line. Other times, apprehension is subordinate to other goals, as when we want to look behind the horse that finished first to see who finished second. A computational analysis of apprehension should account for this flexibility. To this end, we interpret the representations and processes described above as elements that can be arranged in different ways and executed in different orders to fulfill different purposes, like the data structures and the instruction set in a programming language. Ordered combinations of representations and processes are interpreted as *programs* or *routines* (cf. Ullman 1984). In this section, we consider three routines that serve different purposes.

13.4.3.1 Relation Judgments Apprehension is the main purpose of relation judgments. A viewer who is asked, "Where is Gordon?" or "Where is Gordon with respect to Jane?" is expected to report the relation between Gordon and a reference object. In the first case, the reference object is not given. The viewer must (1) find the located object (Gordon); (2) find a suitable reference object (i.e., one the questioner knows about or can find easily); (3) impose a reference frame on the reference object; (4) choose a relation whose region of acceptability best represents the position of the located object; and (5) produce an answer (e.g., "Gordon is in front of the statue"). In the second case, the reference object is given (i.e., Jane). The viewer must (1) find the reference object; (2) impose a reference frame on it; (3) find the located object (i.e., Gordon); (4) choose a relation whose region of acceptability best represents the position of the located object; and (5) produce an answer (e.g., "on her left side.>").

We assume that viewers find located objects by spatially indexing objects in the perceptual representation and comparing them to a description of the specified located object (e.g., "Does that look like Gordon?"). When reference objects are specified in advance, we assume they are found in the same manner. If they are not specified in advance, as in the first case, then the most prominent objects are considered as reasonable candidates for reference objects (Clark and Chase 1974; Talmay 1983). The relation itself is chosen by iterating through a set of candidate relations—imposing the associated spatial templates on the reference object, aligning them with the reference frame, and computing goodness of fit—until one with the best fit or one with an acceptable fit is found.

Relation judgments have been studied often in the psychological literature. Subjects are told in advance what the arguments of the relation will be, but they are not told the relation between them. Their task is to find the arguments, figure out the relation between them, and report it. Thus Logan and Zbrodoff (1979) had subjects report whether a word appeared above or below the fixation point; Logan (1980) had subjects decide whether an asterisk appeared above or below a word. A common focus in relation judgments is Stroop-like interference from irrelevant spatial information (e.g., the identity of the word in the first case; the position occupied by the word-asterisk pair in the second).

13.4.3.2 Cuing Tasks In cuing tasks, apprehension is used in the service of another goal. A viewer who is asked, "Who is beside Mary?" must (1) find the reference object (i.e., Mary); (2) impose reference frame on it; (3) align the relevant spatial template with the reference frame (i.e., the one for *beside*); (4) choose as the located object the perceptual object that is the best example (or the first acceptable example) of the relation; and (5) produce an answer (e.g., "Paul").

Cuing tasks have been studied extensively in the psychological literature. Experiments on visual spatial attention require subjects to report a target that stands in a prespecified relation to a cue (e.g., Eriksen and St. James 1986). The cue is the reference object and the target is the located object. Usually, the focus is on factors other than the apprehension of spatial relations; nevertheless, apprehension is a major computational requirement in these tasks (see, for example, Logan 1995).

13.4.3.3 Verification Tasks Verification tasks present the viewer with a completely specified relation (e.g., “Is Daisy sitting next to Stella?”) and ask whether it applies to a given scene or a given display. The focus may be on one or the other of the arguments, as in “Is that *Daisy* sitting next to Stella?”; or it may be on the relation itself, as in “Is Daisy sitting *next to* Stella?” If the focus is on the arguments, verification could be done as a cuing task. The viewer could (1) find the reference object (e.g., Stella); (2) impose a reference frame on it; (3) align the relevant spatial template with the reference frame (the one for *next to*); (3) choose a located object that occupies a good or acceptable region; (4) compare that object with the one specified in the question (i.e., Is it Daisy?); and (5) report “yes” if it matches or “no” if it does not. Alternatively, if the focus is on the relation, verification could be done as a judgment task. The viewer could (1) find the located object (Daisy); (2) find the reference object (Stella); (3) impose a reference frame on it; (4) iterate through spatial templates until the best fit is found or until an acceptable fit is found; (5) compare the relation associated with that template with the one asserted in the question; and (6) report “yes” if it matches and “no” if it does not.

Verification tasks are common in the psychological literature. A host of experiments in the 1970s studied comparisons between sentences and pictures, and spatial relations figured largely in that work (e.g., Clark, Carpenter, and Just 1973). Subjects were given sentences that described spatial layouts and then pictures that depicted them. The task was to decide whether the sentence described the picture.

13.5 Evidence for the Theory

The theory claims that apprehension of spatial relations requires spatial indexing, reference frame computation, and assessment of goodness of fit of spatial templates. The psychological literature contains evidence for the first two claims, and that evidence will be reviewed briefly below. The third claim has not yet been tested. The remainder of this chapter presents four pieces of evidence that bear on its validity.

13.5.1 Apprehension Requires Spatial Indexing

Logan (1994) found evidence that apprehension of spatial relations requires spatial indexing in visual search tasks. On each trial, subjects were presented with a sentence that described the relation between a dash and a plus (e.g., “dash right of plus”), followed by a display of dash-plus pairs. Half of the time, one of the pairs matched the description in the sentence (e.g., one dash was *right of* one plus), and half of the time, no pair matched the description. All pairs except the target were arranged in the opposite spatial relation (e.g., all the other dashes were *left of* the corresponding pluses). The experiments examined the relations *above*, *below*, *left of*, and *right of*.

In one experiment, the number of dash-plus pairs was varied, and reaction time increased linearly with the number of pairs. The slope was very steep (85 ms/item when the target was present; 118 ms/item when it was absent), which suggests that the pairs were examined one at a time until a target was found (i.e., the pairs were spatially indexed element by element until a target was found). A subsequent experiment replicated these results over twelve sessions of practice (6,144 trials), suggesting that subjects could not learn to compute spatial relations without spatial indexing.

In a third experiment, the number of pairs was fixed and attention was directed to one pair in the display by coloring it differently from the rest. When the differently colored pair was the target, performance was facilitated; subjects were faster and more accurate. When the differently colored pair was not the target, performance was impaired; subjects were slower and less accurate. This suggests that apprehension of spatial relations requires the kind of attentional process that is directed by cues like discrepant colors (i.e., spatial indexing).

13.5.2 Apprehension Requires Reference Frame Computation

Logan (1995) found evidence that apprehension of spatial relations requires reference frame computation in experiments in which attention was directed from a cue to a target. The relation between the cue and the target was varied within and between experiments. Overall, six relations were investigated: *above*, *below*, *front*, *back*, *left of*, and *right of*. The operation of a reference frame was inferred from differences in reaction time with different relations: *above* and *below* were faster than *front* and *back*, and *front* and *back* were faster than *left of* and *right of*. Clark (1973) predicted these differences from an analysis of the environmental support for each relation, and Tversky and colleagues confirmed Clark's predictions in tasks that required searching imagined environments (Bryant, Tversky, and Franklin 1992; Franklin and Tversky 1990). According to Clark's (1973) analysis, *above* and *below* are easy because they are consistent with gravity, consistent over translations and rotations produced by locomotion, and supported by bodily asymmetries (heads are different from feet). *Front* and *back* are harder because they are supported by bodily asymmetries but not

by gravity and they change with locomotion through the environment. *Left* and *right* are hardest of all because they are not supported by gravity or bodily asymmetries and they change with locomotion; they are often defined with reference to other axes. Our theory would account for these differences in terms of the difficulty of aligning reference frames and computing direction.

In Logan's (1995) experiments, subjects reported targets that were defined by their spatial relation to a cue. Some experiments studied deictic relations, using an asterisk as a cue and asking subjects to project their own reference frames onto the asterisk. Subjects saw a display describing a spatial relation (*above*, *below*, *left*, or *right*) and then a picture containing several objects surrounding an asterisk cue. Their task was to report the object that stood in the relation to the asterisk cue that we specified in the first display. Subjects were faster to access objects *above* and *below* the cue than to access objects *right* and *left* of it, consistent with Clark's (1973) hypothesis and with our assumption that orienting reference frames and deciding direction take time.

Other experiments studied intrinsic relations, using a picture of a human head as a cue and asking subjects to extract the intrinsic axes of the head. Again, the first display contained a relation (*above*, *below*, *front*, *back*, *left*, or *right*) and the second contained a display in which objects surrounded a picture of the head. Subjects were faster with *above* and *below* than with *front* and *back*, and faster with *front* and *back* than with *left* and *right*.

In some experiments, the same object could be accessed via different relations. Access to the object was easy when the relation was *above* or *below* and hard when it was *left* or *right*. The cue was presented in different positions, and the regions that were easy and hard to access moved around the display with the cue. This suggests that the reference frame can be translated across space.

In other experiments, the orientation of the reference frame was varied. With deictic cues, subjects were told to imagine that the left side, the right side, or the bottom of the display was the top, and the advantage of *above* and *below* over the other relations rotated with the imagined top. With intrinsic cues, the orientation of the head cue was varied, and the advantage of *above* and *below* over the other relations rotated with the orientation of the head. These data suggest that the reference frame can be rotated at will.

13.6 Evidence for Spatial Templates

The theory assumes that spatial relations are apprehended by computing the goodness of fit between the position of the located object and a spatial template representing the relation that is centered on and aligned with the reference object. The idea that spatial templates are involved in apprehension is new and there is not much evidence

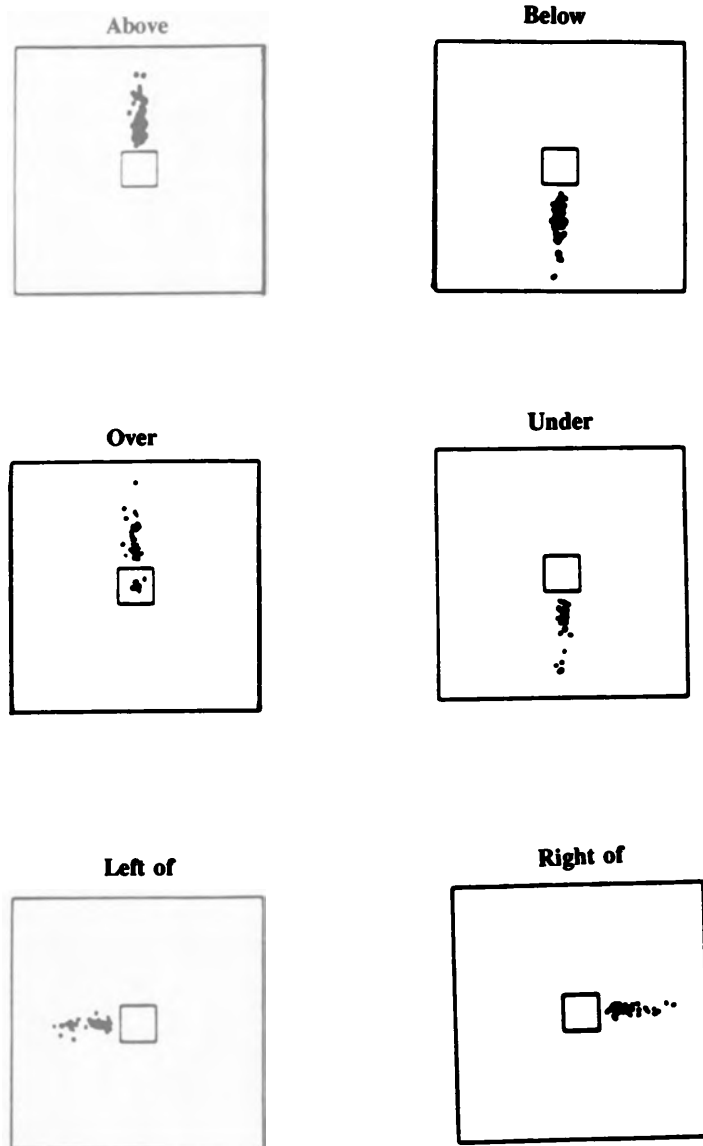
for it (but see Hayward and Tarr 1995). Sections 13.7–13.10 present four experiments that test different aspects of the idea. The first experiment assesses the parts of space that correspond to the regions of greatest acceptability, using a production task. The second assesses parts of space corresponding to good, acceptable, and bad regions, using a task in which subjects rate how well sentences describe pictures. The third assesses the importance of spatial templates in thinking about spatial relations, using a task in which subjects rate the similarities of words that describe (lexicalized) spatial relations and comparing the multidimensional similarity space underlying those ratings with one constructed from the ratings of pictures in the second experiment. The final experiment tests the idea that spatial templates are applied in parallel, using a reaction time task in which subjects verify spatial relations between objects.

13.7 Experiment 1: Production Task

The first experiment attempted to capture the regions of space corresponding to the best examples of twelve spatial relations: *above*, *below*, *left of*, *right of*, *over*, *under*, *next to*, *away from*, *near to*, *far from*, *on*, and *in*. Subjects were presented with twelve frames, with a box drawn in the center of each one; above each frame was an instruction to draw an *X* in one of the twelve relations to the box (e.g., “Draw an *X* above the box”). We assumed they would draw each *X* in the region corresponding to the best example of each relation, though we did not require them to. There were 68 subjects, who were volunteers from an introductory psychology class. The frames were drawn on three sheets of paper, four frames per sheet, and three different orders of sheets were presented.² Each frame was 5.9 cm square and the central box was 8.5 mm square.

The data were collated by making transparencies of each of the twelve frames. For each relation, we superimposed the transparency on each subject’s drawing and drew a dot on the transparency (with a felt pen) at the point corresponding to the center of the *X* that the subject drew, accumulating dots across subjects. The data for *above*, *below*, *over*, *under*, *left of*, and *right of* are presented in figure 13.1, the data for *next to*, *away from*, *near*, *far from*, *in*, and *on* are presented in figure 13.2.

The relations in figure 13.1 differ primarily in the orientation and direction of the reference frame. The patterns in each panel are similar to each other, except for rotation. The main exception is *over*, where some subjects drew *X*s that were superimposed on the box, apparently interpreting *over* as covering (which is a legitimate interpretation; see Lakoff 1987). Note that distance did not matter much. Some *X*s were placed close to the box but others were placed quite far away, near the edge of the frame. In each case, the *X*s appeared roughly centered on the axis of the reference frame extended outward from the box.

**Figure 13.1**

Data for *above*, *below*, *over*, *under*, *left of*, and *right of* from the production task in experiment 1. Each point represents the center of an *X* drawn by a different subject to stand in the relation to the central box that is specified above each frame.

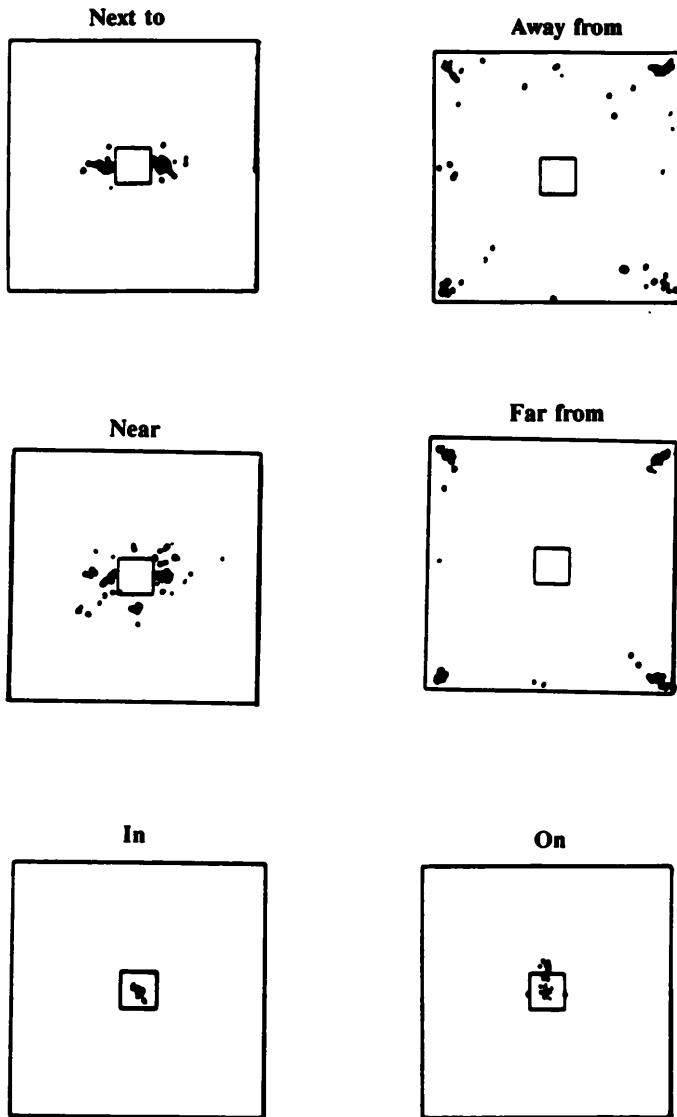


Figure 13.2

Data for *next to*, *away from*, *near*, *far from*, *in*, and *on* from the production task in experiment 1. Each point represents the center of an *X* drawn by a different subject to stand in the relation to the central box that is specified above each frame.

The relations in the top four panels of figure 13.2 depend primarily on the scale of the reference frame and not on orientation or direction. *Xs* exemplifying *next to* and *near* were placed close to the box, whereas *Xs* exemplifying *away from* and *far from* were placed some distance from it, close to the corners (especially for *far from*). One unexpected result was that *next to* was interpreted as horizontal proximity. No subject drew an *X* above or below the box for *next to*, though many did so for *near*. This unanticipated result appears again in the next experiment.

The bottom two panels of figure 13.2 represent *in* and *on*. All subjects drew their *Xs* so that their centers were within the boundaries of the box for *in*, but not all subjects did so for *on*. Some drew the *X* as if it were on top of the box, and one drew the *X* centered on each side of the box. All of these are legitimate interpretations of the relations.

13.8 Experiment 2: Goodness Rating Task

The second experiment attempted to capture the regions corresponding to good, acceptable, and bad examples of ten of the relations used in experiment 1: *above*, *below*, *left of*, *right of*, *over*, *under*, *next to*, *away from*, *near to*, and *far from*. Subjects were shown sentences, followed by pictures on computer monitors, and were asked to rate how well the sentence described the picture on a scale from 1 (bad) to 9 (good). Each sentence was of the form "The *X* is [relation] the *O*" and each picture contained an *O* in the center of a 7×7 grid and an *X* in one of the 48 surrounding positions. The grid, which was not visible to the subjects, was 8.8 cm wide and 9.3 cm high on the computer screen. Viewed at a distance of 60 cm, this corresponded to $8.3 \text{ degrees} \times 8.8 \text{ degrees}$ of visual angle. Each of the 48 positions was tested for each relation so that we could get ratings from good, acceptable, and bad regions. There were 480 trials altogether ($48 \text{ positions} \times 10 \text{ relations}$). Subjects reported their rating by pressing one of the numeric keys in the row above the standard QWERTY keyboard. There were thirty-two subjects, volunteers from an introductory psychology class. The data were collated by averaging ratings across subjects. The average ratings are plotted in figures 13.3 and 13.4 and presented in table 13.1. Subjects were very consistent; the mean standard error of the averages in figures 13.3 and 13.4 is 0.271.

Figure 13.3 presents the average ratings for *above*, *below*, *over*, *under*, *left of*, and *right of* drawn as three-dimensional graphs. Screen positions are represented in the up-down axis and the left-right axis. The up-down axis goes from upper left to lower right; the left-right axis goes from lower left to upper right. Ratings are represented in the third dimension, which is essentially vertical on the page. The central position, which was occupied by the *O*, is blank.

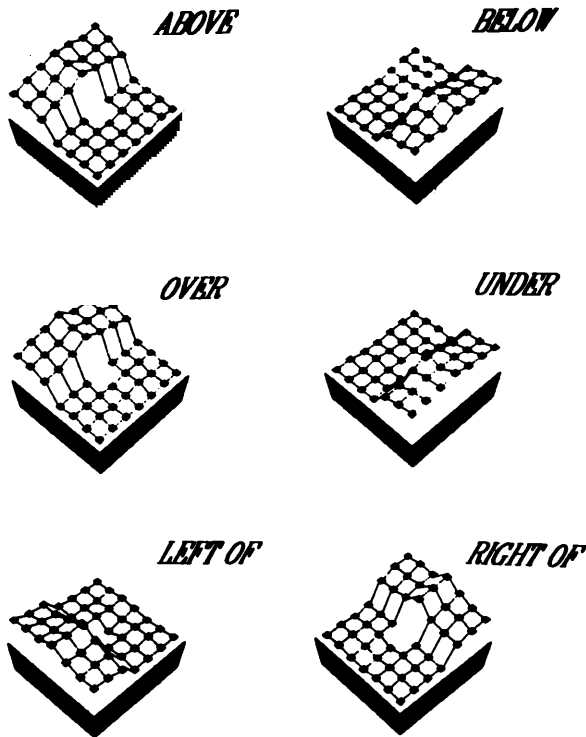


Figure 13.3

Average ratings for *above*, *below*, *over*, *under*, *left of*, and *right of* from the goodness rating task in experiment 2. Each point represents the average goodness on a scale from 1 (bad) to 9 (good) with which an *X* presented in the position of the point exemplifies the relation to an *O* presented in the central position.

As with the production task the patterns in the different panels appear to be the same except for changes in orientation and direction. The highest ratings—near 9—were given to the three points directly *above*, *below*, *over*, *under*, *left of*, or *right of* the central position, which correspond to the “best” regions that we saw in experiment 1. Note that distance did not matter much in the “best” regions; ratings were close to 9 whether the *X* was near to the *O* or far from it. Intermediate ratings were given to the 18 positions on either side of the three best positions, and the lowest ratings (near 1) were given to the remaining 27 points. There was a sharp boundary between bad and acceptable regions. The boundary between acceptable and good regions was less marked. The acceptable regions themselves were not uniform. With *above*, for example, ratings in the first position higher than the *O* tended to decrease

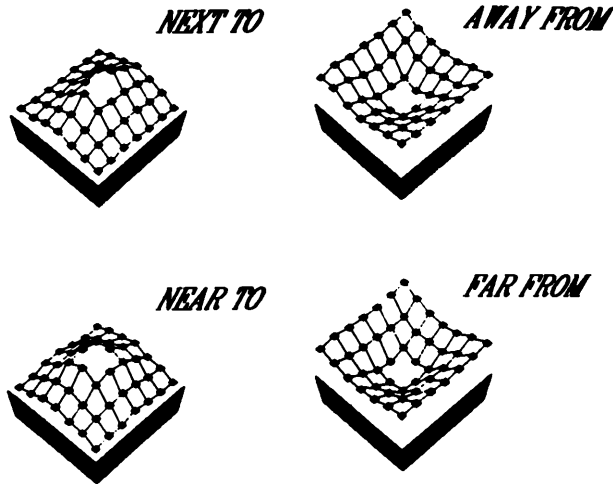


Figure 13.4

Average ratings for *next to*, *away from*, *near to*, and *far from* from the goodness rating task in experiment 2. Each point represents the average goodness on a scale from 1 (bad) to 9 (good) with which an *X* presented in the position of the point exemplifies the relation to an *O* presented in the central position.

as the position of the *X* extended farther to the left and the right, whereas ratings for the highest positions were not affected much by distance from the center, as if the region of intermediate fit were slightly U-shaped. The mean ratings for the first position higher than the *O* were 5.63, 6.41, 7.09, 8.53, 7.35, 6.74, and 5.53 from left to right. The mean ratings for positions directly above the *O* were 8.53, 8.55, and 8.61 from bottom to top. The same trends can be seen with the other relations.

The average ratings for *next to*, *away from*, *near to*, and *far from* are presented in figure 13.4 using the same three-dimensional format as figure 13.3. For *next to* and *near to*, ratings were highest in positions adjacent to the central position (occupied by the *O*) and they diminished gradually as distance increased. Consistent with experiment 1, there was a tendency to interpret *next to* horizontally; positions to the left and right of the central position were rated higher than positions the same distance away but above and below the central position. The mean ratings for the positions immediately left and right of the *O* were 8.17 and 8.39, respectively, whereas the mean ratings for the positions immediately above and below the *O* were 6.07 and 6.19, respectively.

Away from and *far from* were “mirror images” of *next to* and *near to*. Ratings were lowest in positions immediately adjacent to the central position and rose gradually as

Table 13.1

Mean Goodness Ratings for Each Relation in Experiment 2 as a Function of the Position Occupied by the *X*

<i>Above</i>						
7.00	7.66	8.10	8.61	8.19	7.32	7.66
6.69	6.56	7.66	8.55	7.13	7.16	6.88
5.63	6.41	7.09	8.53	7.35	6.74	5.53
1.94	2.16	1.88		1.97	1.88	2.00
1.94	1.78	1.66	1.13	1.63	2.41	1.66
1.81	1.94	1.42	1.03	1.50	1.84	1.58
1.44	1.38	1.34	1.19	1.34	2.08	1.44
<i>Below</i>						
1.50	1.66	1.29	1.03	1.33	1.75	1.59
1.71	2.09	1.40	1.31	1.44	1.66	1.45
1.94	2.09	1.65	1.72	1.88	2.39	2.00
2.16	2.29	2.03		2.41	1.94	2.00
5.66	6.31	6.94	8.16	6.94	6.00	5.81
6.00	7.09	7.74	8.71	7.78	7.10	6.88
7.42	5.0	6.88	8.40	7.72	7.71	7.53
<i>Over</i>						
8.84	7.65	8.10	8.90	7.59	7.38	7.10
6.75	6.94	7.19	8.29	7.45	7.32	8.41
5.69	5.97	7.07	8.42	7.19	6.38	5.58
1.91	2.19	2.09		2.13	1.94	2.25
2.28	1.91	1.71	1.28	1.97	2.09	2.00
1.69	2.00	1.28	1.45	2.19	1.69	1.66
1.52	1.59	1.52	1.20	1.28	1.66	1.66
<i>Under</i>						
1.81	1.94	1.38	1.39	1.59	1.72	1.47
1.83	1.53	2.03	1.41	1.44	1.63	1.84
1.77	1.78	1.63	1.44	1.59	1.68	2.19
2.06	2.22	1.91		2.25	2.39	2.00
5.71	5.66	6.75	8.23	6.84	5.88	5.84
6.59	7.00	7.59	7.45	7.38	6.50	6.10
7.22	7.55	7.90	8.72	7.78	7.74	7.03
<i>Left</i>						
6.56	5.65	5.28	2.56	2.13	1.88	1.66
7.00	6.06	5.39	2.25	2.16	1.53	1.80
7.13	6.52	6.34	2.31	2.47	1.94	2.10
8.35	7.83	8.06		1.10	1.59	1.94
6.84	6.39	6.65	2.16	2.03	2.10	1.41
6.03	6.23	5.63	2.48	1.90	2.22	1.59
6.16	5.77	4.94	1.90	1.94	1.94	2.03

Table 13.1 (continued)

<i>Right</i>						
1.72	1.97	1.66	2.22	5.50	6.45	6.59
1.90	2.00	2.00	2.28	5.78	6.52	7.06
1.33	1.68	2.13	2.39	6.39	6.84	7.03
1.09	1.35	1.38		8.35	8.52	8.71
1.69	1.74	2.25	2.09	6.03	6.81	7.50
1.66	1.94	1.81	2.03	5.59	6.72	6.63
1.50	1.72	1.94	1.87	5.47	6.13	6.44
<i>Next to</i>						
2.65	2.06	2.10	2.03	2.29	1.94	1.70
2.84	3.32	3.31	3.91	3.35	3.34	2.94
4.06	4.75	5.90	6.07	6.57	4.72	3.87
4.52	6.00	8.17		8.39	6.69	4.88
3.56	4.59	6.59	6.19	5.91	5.38	4.13
2.94	3.58	3.66	4.06	4.00	3.32	3.06
2.37	2.06	2.53	2.31	1.81	2.00	1.69
<i>Away from</i>						
7.38	7.94	7.45	7.74	7.72	8.10	8.44
7.41	6.84	5.74	5.16	5.69	6.72	7.22
5.90	4.75	2.94	2.91	2.78	5.13	6.47
5.35	4.38	2.13		1.88	4.58	6.25
6.32	4.81	3.09	2.50	3.44	5.41	6.45
7.28	6.09	5.34	4.97	5.41	5.75	7.66
8.10	7.50	7.58	7.63	7.44	7.83	8.26
<i>Near to</i>						
1.74	1.90	2.84	3.16	2.34	1.81	2.13
2.61	3.84	4.66	4.97	4.90	3.56	3.26
4.06	5.56	7.55	7.97	7.29	4.80	3.91
4.47	5.91	8.52		7.90	6.13	4.63
3.47	4.81	6.94	7.56	7.31	5.59	3.63
3.25	4.03	4.50	4.78	4.41	3.47	3.10
1.84	2.23	2.03	3.06	2.53	2.13	2.00
<i>Far from</i>						
7.48	7.94	7.56	7.42	7.38	7.88	8.48
6.56	5.78	5.41	5.41	5.19	5.38	7.03
5.69	4.03	2.28	1.78	2.84	4.13	6.06
5.59	3.44	1.87		1.66	4.22	5.71
6.90	4.56	2.28	1.81	2.31	4.09	6.13
7.09	6.03	4.88	5.19	5.16	6.00	7.42
7.68	7.77	7.58	7.13	7.47	7.78	8.41

distance increased. The corner positions, which were the most distant, got the highest ratings. As with figure 13.3, the ratings in figure 13.4 appear to capture the regions of best fit that were found in experiment 1. The parts of space that received the highest ratings were the parts of space in which subjects tended to draw their *Xs*.

The data in figures 13.3 and 13.4 capture our idea of spatial templates quite graphically. One can imagine centering the shape in each panel on a reference object, rotating it into alignment with a reference frame, and using it to determine whether a located object falls in a good, acceptable, or bad position.

13.9 Experiment 3: Similarity Rating Task

The data in figures 13.1–13.4 suggest a pattern of similarities among the relations. Templates corresponding to *above*, *below*, *over*, *under*, *left of*, and *right of* have similar shapes but differ from each other in orientation and direction. Templates corresponding to *next to*, *away from*, *near to*, and *far from* have different shapes from *above*, *below*, and so on, but are similar to each other except that *next to* and *near to* are reflections of *away from* and *far from*. The purpose of the third experiment was to capture these similarities in a task that did not involve external, visible relations.

Subjects were presented with all possible pairs of words describing the twelve relations, *above*, *below*, *left of*, *right of*, *over*, *under*, *next to*, *away from*, *near to*, *far from*, *in*, and *on*, and they were asked to rate their similarity on a scale of 1 (dissimilar) to 10 (similar). The words were printed in pairs with a blank beside them, in which subjects were to write their rating. The 66 pairs were presented in two single-spaced columns on a single sheet of paper. There were four groups of subjects (26, 28, 19, and 28 in each group) who received the pairs in different orders. The subjects were 101 volunteers from an introductory psychology class.

The ratings for each word pair were averaged across subjects, and the averages were subjected to a multidimensional scaling analysis, using KYST (Kruskal, Young, and Seery 1977). We tried one-, two-, and three-dimensional solutions and found that stress (a measure of goodness of fit, analogous to $1 - r^2$) was minimized with a three-dimensional fit. The stress values were .383, .191, and .077 for the one-, two-, and three-dimensional solutions, respectively. The similarity space for the three-dimensional solution is depicted in figures 13.5, 13.6, and 13.7.

Figure 13.5 shows the plot of dimension 1 against dimension 2, which appears to be a plot of an *above-below* dimension against a *near-far* dimension. *Above* and *over* appear in the bottom right, and *below* and *under* appear in the top left. *Away from* and *far* appear in the bottom left, and *next to*, *near*, *in*, and *on* appear in the top right. *Left* and *right* appear in the middle, reflecting their projection on the *above-below* × *near-far* plane.

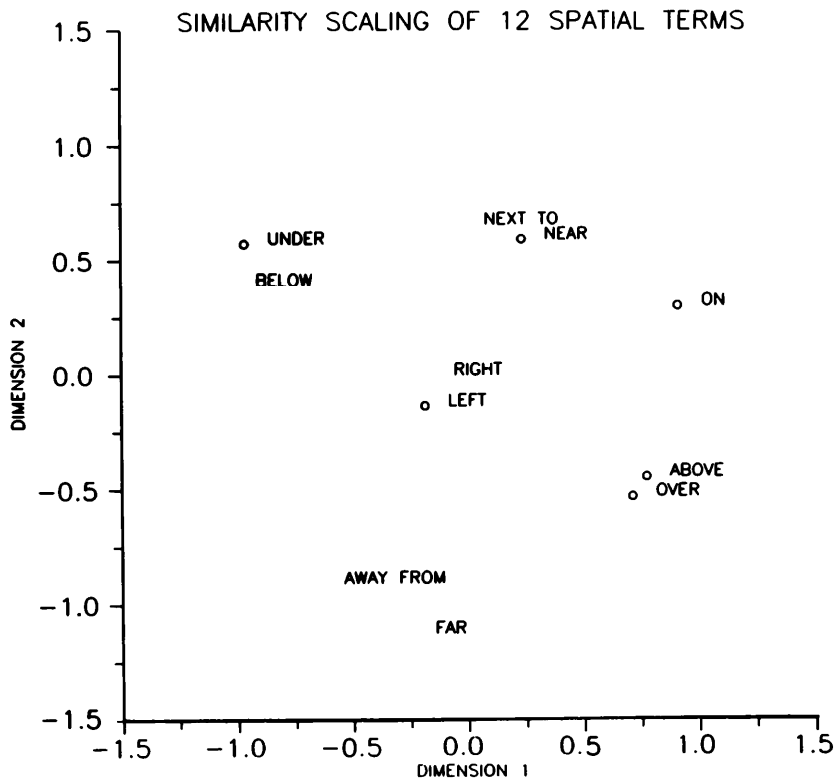


Figure 13.5

Dimension 1 $\times$ dimension 2 plotted from a similarity space constructed from a multidimensional scaling of similarity ratings of twelve spatial terms in experiment 3 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *above-below* $\times$ *near-far*.

Figure 13.6 shows the plot of dimension 1 against dimension 3, which appears to be a plot of an *above-below* dimension against a *left-right* dimension. *Above* and *over* appear on the left side, and *below* and *under* appear on the right. *Left* appears on the top, and *right* appears on the bottom. The other relations are scattered over the middle of the plot, reflecting the projection of the *near-far* axis on the *above-below* $\times$ *left-right* plane.

Figure 13.7 shows the plot of dimension 2 against dimension 3. This appears to be a plot of *near-far* against *left-right*. *In*, *on*, *next to*, and *near* appear on the top, whereas *far* and *away from* appear on the bottom. *Right* appears on the left side, while *left* appears on the right. *Above*, *over*, *below*, and *under* are scattered over the plane, reflecting the projection of the *above-below* axis on the *near-far* $\times$ *left-right* plane.

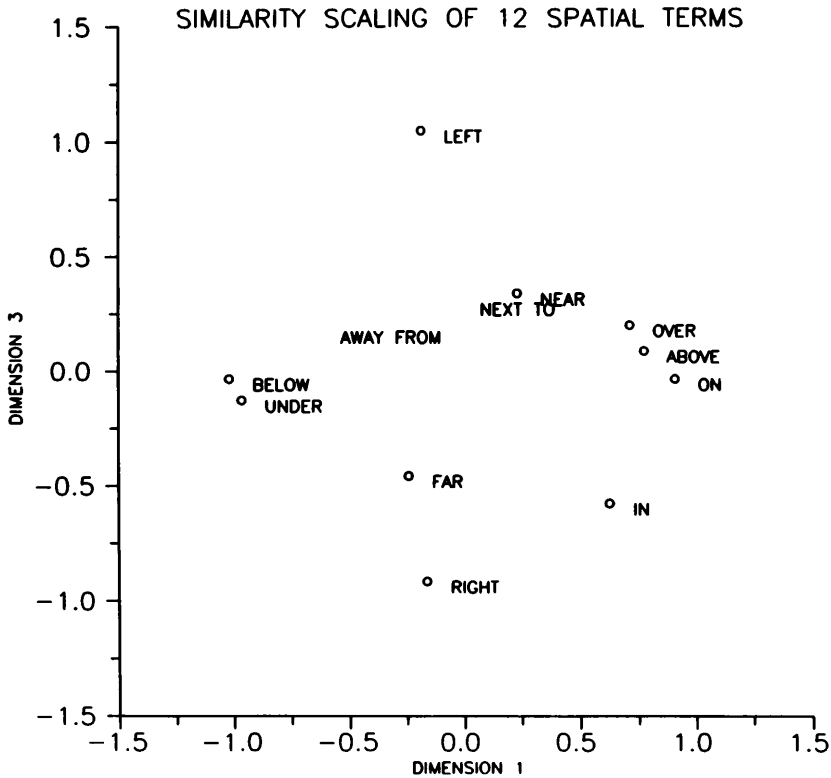


Figure 13.6

Dimension 1 $\times$ dimension 3 plotted from a similarity space constructed from a multidimensional scaling of similarity ratings of twelve spatial terms in experiment 3 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *above-below* $\times$ *left-right*.

The similarity structure in these plots resembles that seen in figures 13.1–13.4. The templates for *above* and *over* have similar shapes, opposite to those for *below* and *under*. The templates for *left* and *right* are opposite to each other and orthogonal to *above* and *below*. The templates for *far* and *away from* are similar to each other and opposite to *near* and *next to*, and all of their shapes are different from those of *above*, *below*, *left*, *right*, and so on.

In order to formalize these intuitions, we calculated similarity scores from the spatial templates in figures 13.3 and 13.4 and subjected them to multidimensional scaling, using KYST. The procedure involved several steps. We treated the forty-eight ratings for each relation as a vector and assessed similarity between relations by computing the dot product of the corresponding vectors. That is, we multiplied the

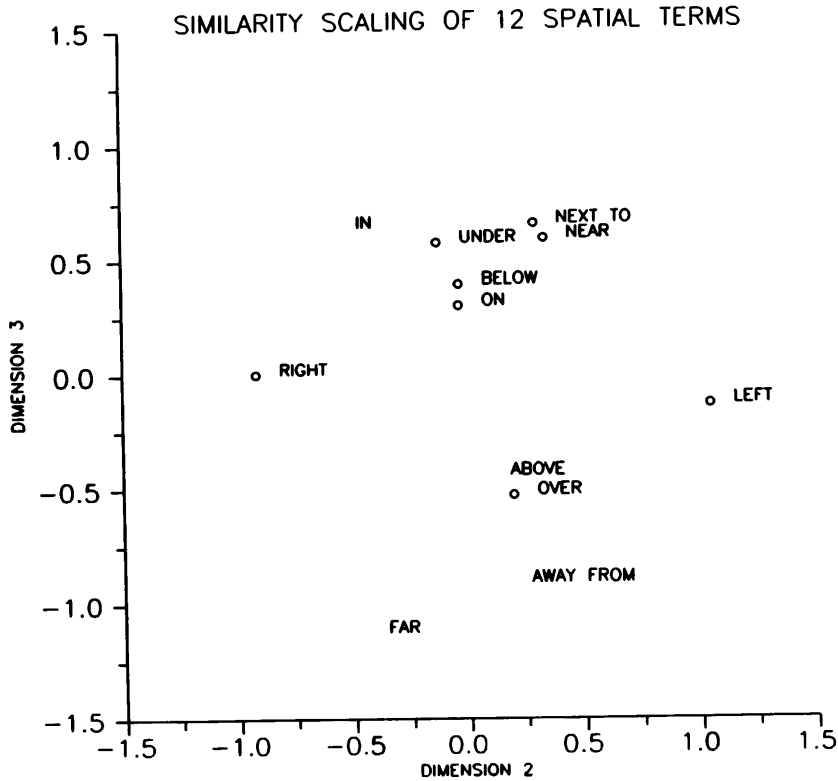


Figure 13.7

Dimension 2 $\times$ dimension 3 plotted from a similarity space constructed from a multidimensional scaling of similarity ratings of twelve spatial terms in experiment 3 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *near-far* $\times$ *left-right*.

ratings in corresponding cells and added them up to produce a similarity score analogous to a correlation coefficient. Before computing the dot product, we normalized the vectors, setting the sum of their squared values to the same value for each relation. There were forty-five dot products, reflecting all possible pairs of the ten relations examined in experiment 2. These forty-five dot products were treated as similarity ratings and ran through the KYST program. As before, we tried one-, two-, and three-dimensional solutions and found stress minimized with a three-dimensional solution. The stress values were .315, .139, and .009 for one, two, and three dimensions, respectively. The three-dimensional similarity space is plotted in figures 13.8, 13.9, and 13.10.

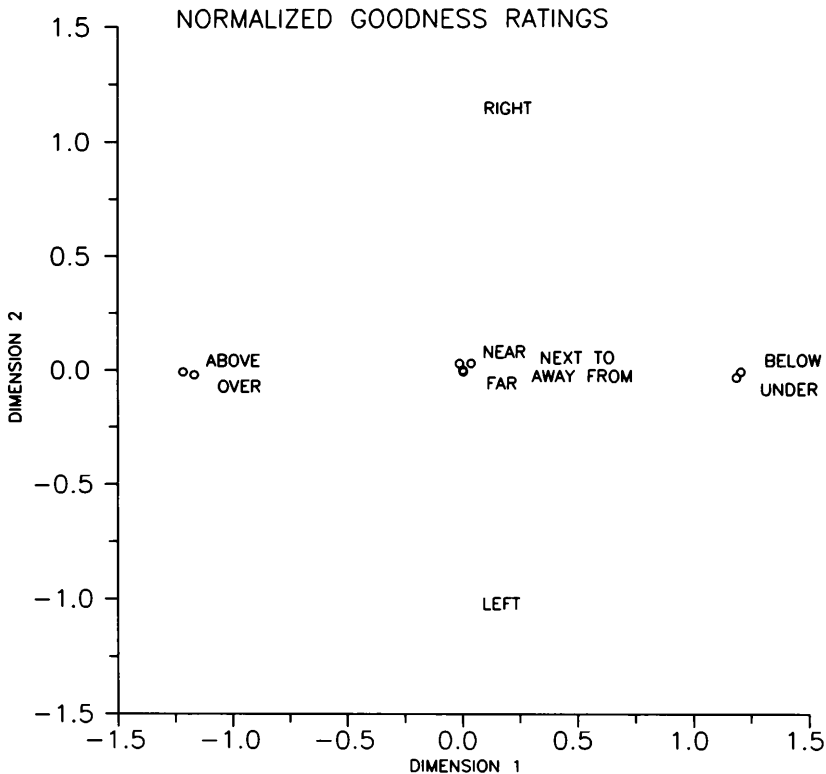


Figure 13.8

Dimension 1 $\times$ dimension 2 plotted from a similarity space constructed from a multidimensional scaling of dot products from goodness ratings of ten spatial terms from experiment 2 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *above-below* $\times$ *left-right*.

The dimensional structure that emerged from the scaling analysis of the goodness ratings was very similar to the one that emerged from the similarity ratings. The structure had three dimensions and the three dimensions could be interpreted similarly. Figure 13.8 contains the plot of dimension 1 against dimension 2, which is easily interpretable as a plot of the *above-below* axis against the *left-right* axis. Figure 13.9 contains the plot of dimension 1 against dimension 3, which appears to be a plot of the *above-below* axis against the *near-far* axis. Figure 13.10 contains the plot of dimension 2 against dimension 3, which appears to be a plot of the *left-right* axis against the *near-far* axis. We assessed the similarity of the fits quantitatively by calculating the correlation between the interpoint distances in the two solutions. Each

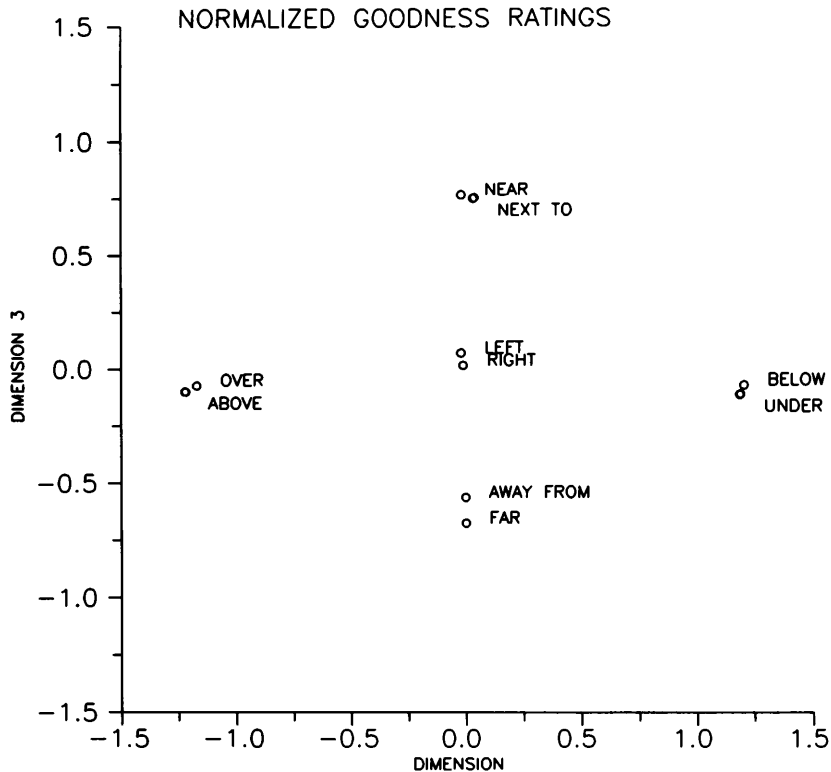


Figure 13.9

Dimension 1 $\times$ dimension 3 plotted from a similarity space constructed from a multidimensional scaling of dot products from goodness ratings of ten spatial terms from experiment 2 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *above-below* $\times$ *near-far*.

solution gives the distance between each pair of relations in multidimensional space. If the solutions are similar, then the distances between the same pairs of relations in the two spaces should be similar. The correlation was .858, indicating good agreement.

The similarity of the scaling solutions and the high correlation between distances suggests that the ratings of pictures in experiment 2 and the ratings of words in the present experiment were based on common, underlying knowledge structures. We would like to conclude that subjects used spatial templates to perform both tasks. Thus they rated pictures by aligning spatial templates with the reference object and computing the goodness of fit for the located object, and they rated words by

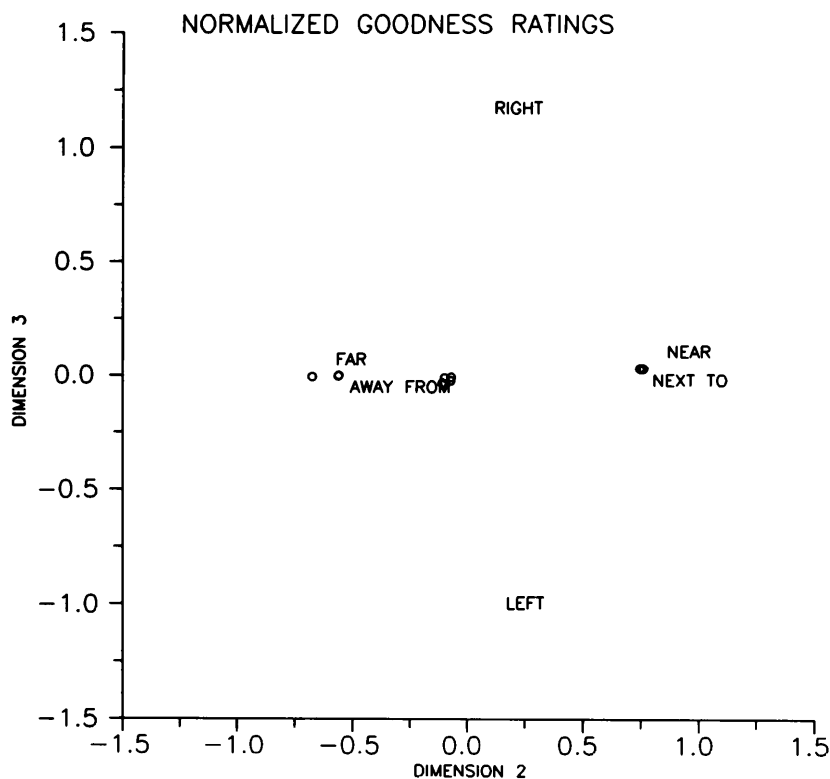


Figure 13.10

Dimension 2 $\times$ dimension 3 plotted from a similarity space constructed from a multidimensional scaling of dot products from goodness ratings of ten spatial terms from experiment 2 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *left-right* $\times$ *near-far*.

comparing the spatial templates associated with them. This conclusion is speculative, however. Although there is some evidence that subjects may compare images when given words (Shepard and Chipman 1970), other representations and processes could produce the same outcomes. The data are consistent with our conclusion, but they do not rule out competing interpretations.

13.10 Experiment 4: Relation Judgment Task

The results of experiments 1–3 are consistent with the hypothesis that spatial templates were applied in parallel to the whole perceptual representation, but they do not support that hypothesis uniquely. The same results could have been produced by

applying *serial visual routines* instead of spatial templates. Serial visual routines are processes that operate sequentially on perceptual representations to compute a number of things, including spatial relations (Ullman 1984). For example, *above* could be produced by centering a "mental cursor" on the reference object and moving upward along the up-down axis of the reference frame until the located object was found (Jolicoeur, Ullman, and MacKay 1986, 1991). If the located object was not directly above the reference object, the cursor could move from one side to the other covering the region above the top of the reference object until the located object was found. From this perspective, the spatial templates evidenced in experiments 1 and 2 may reflect preferred trajectories for serial visual routines rather than explicit representations used to compute spatial relations directly (i.e., by multiplying activation values as described earlier). The purpose of the fourth experiment was to contrast spatial templates with serial visual routines in the apprehension of spatial relations (see also Logan and Compton 1996; Sargent 1991).

The main point of contrast between spatial templates and serial visual routines is the effect of distance in judging spatial relations. Spatial templates are applied in parallel to the whole visual field, so distance between located and reference objects does not matter. The time taken to apply a spatial template should not depend on distance. By contrast, serial visual routines operate sequentially, examining the visual field bit by bit, so distance between located and reference objects should make a difference. The time taken to apply a serial visual routine should increase monotonically with distance.

Note, the evidence in experiments 1 and 2 that distance has no effect on the goodness of examples of *above*, *below*, *over*, *under*, *left of*, and *right of* does not bear on this issue because time was neither stressed nor measured. Subjects could have taken more time to rate greater distances even though they gave the same rating. The rating could have depended on the relation between the located object and the reference frame centered on the reference object, not on the time taken to compute the relation.

Experiment 4 had subjects perform a verification task in which the distance between reference and located objects was varied systematically (cf. Clark, Carpenter, and Just 1973). The range of distances used in this experiment (1–6 degrees of visual angle) was well within the range that shows monotonic increases, in reaction time in other tasks, such as mental curve tracing (Jolicoeur, Ullman, and MacKay 1986, 1991); if serial visual routines had been used to compute spatial relations in the present experiments, reaction time should therefore have increased with distance.

The experiments focused on the relations *above* and *below*. Each trial began with a fixation point exposed for 500 ms in the center of a computer screen. It was extinguished and replaced with a sentence expressing the relation between a dash and a plus (i.e., "Dash above plus?"; "Dash below plus?"; "Plus above dash?", or "Plus

below dash?") that was exposed for 1,000 ms. After the sentence was extinguished, the fixation point appeared for another 500 ms. Then a picture of a dash above or below a plus was exposed for 200 ms, too briefly to allow eye movements. Half of the time, the relation between the dash and plus matched the sentence, and half of the time, the opposite relation held. Subjects were told to respond "true" to the former case and "false" to the latter. After the 200-ms exposure of the picture, the screen went blank until the subject responded. After the response, the screen remained blank for a 1,500 ms intertrial interval. There were 384 trials in all.

The main manipulation was the distance between the dash and the plus. There were four different distances. In one version of the experiment, the dash and plus were separated by 1, 2, 3, or 4 screen lines (corresponding to .74, 1.48, 2.22, and 2.96 degrees of visual angle when viewed from a distance of 60 cm). In another version, distances were doubled. The dash and the plus were separated by 2, 4, 6, or 8 screen lines (1.48, 2.96, 4.44, or 5.92 degrees of visual angle). Stimuli separated by the different distances appeared in several different locations on the screen. In the version in which distances were 1–4 screen lines, stimuli with a distance of 1 appeared in positions 1 and 2, 2 and 3, 3 and 4, and 4 and 5; stimuli with a distance of 2 appeared in positions 1 and 3, 2 and 4, and 3 and 5; stimuli with a distance of 3 appeared in positions 1 and 4, and 3 and 5; and stimuli with a distance of 4 appeared in positions 1 and 5. The same scheme was used in the version in which distances were 2–8 screen lines, except that positions 1–5 were two lines apart. Distances, relations (*above* vs. *below*), and true and false trials occurred in random order. A different random order was constructed for each subject. The subjects were 48 volunteers from an introductory psychology class. Twenty-four served in each version of the experiment.

Mean reaction times were computed for "true" and "false" responses as a function of distance. The means across subjects are plotted in figures 13.11 and 13.12. Figure 13.11 plots reaction time as a function of absolute distance, expressed in degrees of visual angle. It shows that reaction time was longer for "false" responses than for "true" responses in both versions of the experiment, $F(1, 44) = 78.97$, $p < .01$, mean square error (MSE) = 102,274.38. Reaction time was longer in the version with the greater distances, but the difference was not significant, $F(1, 44) < 1.0$. The most important result for our present purposes is the effect of distance. Serial visual routines predict a monotonic increase in reaction time as distance increases, whereas spatial templates predict no effect. Analysis of variance showed a significant main effect of distance, $F(3, 132) = 4.33$, $p < .01$, $MSE = 57,930.55$, and the linear trend was significant, $F(1, 132) = 4.77$, $p < .01$, indicating a tendency for reaction time to *decrease* as distance increased. The observed pattern is clearly inconsistent with serial

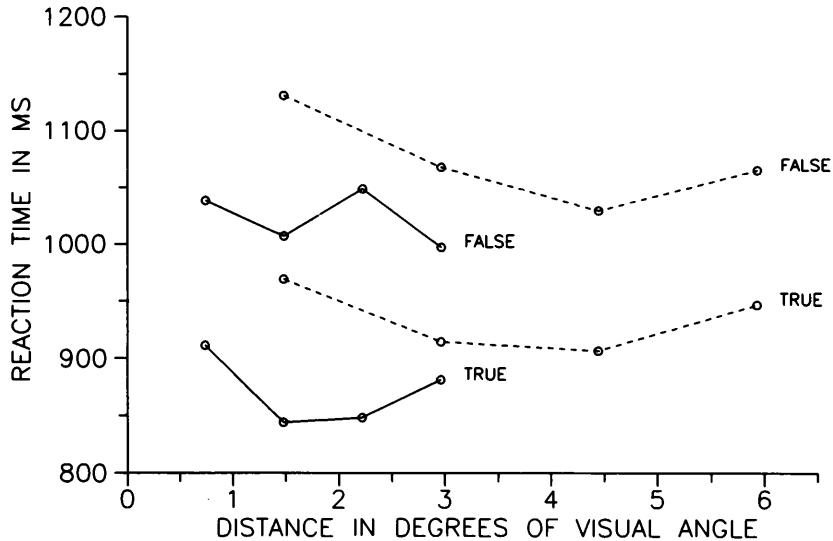


Figure 13.11

Reaction time as a function of absolute distance between reference and located objects from two versions of experiment 4 in which subjects judged *above* and *below*. “True” versus “false” response and long (dotted lines) versus short (solid lines) distances are the parameters.

visual routines. In both versions of the experiment, reaction time was longest for the shortest and longest distances and fastest for the intermediate distances.

The pattern of reaction times is not exactly what one would expect from the spatial template hypothesis, which predicted no effect of distance. However, the pattern may be consistent with theory of apprehension in which spatial templates play a part, if the slower reaction times at the longest and shortest distances can be explained. We suggest that the pattern reflects a process of reference frame adjustment. Subjects may have set the scale of their reference frames to the average distances they experienced—distances of 2 and 3 in one version and distances of 4 and 6 in the other. They may have adjusted them if the distance were longer or shorter than the average—distances 4 and 1 in one version and 8 and 2 in the other. This would produce the observed pattern of results. The effect can be seen more clearly in figure 13.12, which plots reaction time as a function of ordinal distance rather than relative distance. The patterns from the two versions of the experiment align nicely in figure 13.12. Of course, this explanation is post hoc, and must be taken with a grain or two of salt (however, no distance effects were found by Logan and Compton 1996 and by Sargent 1991).

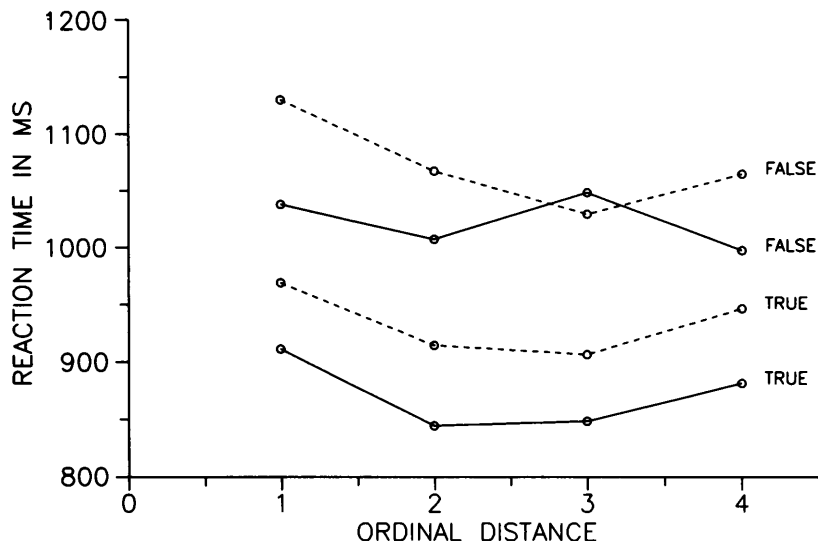


Figure 13.12

Reaction time as a function of ordinal distance between reference and located objects from two versions of experiment 4 in which subjects judged *above* and *below*. "True" versus "false" response and long (dotted lines) versus short (solid lines) distances are the parameters.

13.11 Conclusions

The data from experiments 1–4 support the idea that spatial templates underlie the apprehension of spatial relations. Experiments 1 and 2 showed that the space around a reference object is divided into regions that represent good, acceptable, and bad examples of a given relation (see also Hayward and Tarr 1995). Experiment 3 showed that similarities in the meanings of spatial terms can be accounted for in terms of similarities in the spatial templates that correspond to them. And Experiment 4 showed that distance between reference and located objects has little effect on the time required to apprehend relations, as if spatial templates were applied in to the whole visual field in simultaneously (see also Logan and Compton 1996; Sergent 1991). Together with the other data (Logan 1994, 1995), the experiments support the computational analysis of apprehension presented earlier in the chapter and argue for its viability as a psychological theory of apprehension in humans.

Several parts of the theory were taken from existing analyses of spatial relations. Reference frames and spatial indices play important roles in linguistic and psycholinguistic analyses (see Carlson-Radvansky and Irwin 1993, 1994; Clark 1973; Garnham 1989; Herskovits 1986; Jackendoff and Landau 1991; Landau and Jackendoff

1993; Levelt 1984; Logan 1995; Miller and Johnson-Laird 1976; and Talmy 1983). The novel contribution is the idea that goodness of fit is computed with spatial templates. We suggested this idea because it is computationally simple and easy to implement in software or “wetware.” It would be interesting to contrast spatial templates with other ways to compute goodness of fit in future research (e.g., geometric, volumetric, topological, or functional relations).

The theory was developed to account for the apprehension of spatial prepositions in English. As is readily apparent in the other chapters in this volume, different languages express spatial relations in different ways, so it is important to consider how the theory might generalize to other languages. What is general across languages and what is specific to English? We suspect that the theory could be adapted to most languages. Most languages express relations between objects in terms of reference frames applied to reference objects. We suspect that reference frame computation and spatial indexing (which is required to distinguish reference objects from located objects) may be common to all languages. The spatial templates applied to the reference objects may vary between languages. We suspect that spatial templates are shaped by the linguistic environment to capture the distinctions that are important in particular languages. The perceptual representation must be common to all languages because it is precognitive and thus prelinguistic. The conceptual representations clearly vary between languages. We suggest that the conceptual representations may be distinguished from each other in terms of the spatial templates with which they are associated.

The spatial templates measured in this chapter are crude approximations to the templates that people might actually use (if they use them at all). The measurements were coarse (e.g., experiment 2 used a 7×7 grid) and the reference and located objects were simple (boxes, *O*s and *X*s). We suspect that the results would generalize to finer measurements and more sophisticated objects. Indeed, Hayward and Tarr (1995) and Carlson-Radvansky and Irwin (1993) found similar results with several different reference and located objects. Certainly, the methods could be adapted to more precise measurements, different classes of objects, and even different spatial relations. Thus we do not view the experiments as the final answer, but rather, as a promising beginning to an exciting area of inquiry.

The measurements in the present experiments may not have captured all of the differences between the relations we contrasted. Experiment 1, for example, found evidence of two different senses of *over* (*above* and *covering*), whereas experiment 2 found evidence of only one of them (*above*). The displays in experiment 2 could not have picked up the second meaning because the located and reference objects were always separated. However, it should be possible to pick up the contrast with displays in which located and reference objects overlap. Subjects should rate overlapping

displays as good examples of *over* but bad examples of *above*. Thus the limitations of the present experiments lie in the specific procedures we used rather than in the general methodology. With appropriately designed displays, rating procedures should be able to capture subtle differences between relations.

Spatial templates may not capture the meanings of all spatial relations. *On*, for example, implies contact and support (Bowerman, chapter 10, this volume), neither of which can be described sufficiently in terms of occupancy of regions of space. The reference object and the located object must occupy the same region of space, but contact and support imply more than that. Contact may be assessed by examining junctions between the contours of the objects using something like templates (Biederman 1987), but support cannot be perceived so easily. *In*, as another example, implies containment (Herskovits 1986) and that is a functional relationship that cannot be described easily in terms of regions of space. Flowers *in* a vase occupy a different region of space than water *in* a vase.

Despite these limitations, spatial templates are clearly useful in describing the meanings of many spatial relations. Moreover, they are tractable computationally, and the computational analysis is readily interpretable as a psychological theory of how people actually apprehend spatial relations. The data in the present experiments and others (Carlson-Radvansky and Irwin 1993; Hayward and Tarr 1995; Logan 1994, 1995; Logan and Compton 1996) are consistent with the psychological theory, suggesting it has some validity. Competitive theories, based on assessment of geometric, topological, and functional relations, have not yet reached this stage of development.

Acknowledgments

This research was supported in part by National Science Foundation grant BNS 91-09856 to Gordon Logan. We are grateful to Jane Zbrodoff for valuable discussion. We would like to thank Paul Bloom and Mary Peterson for helpful comments on the manuscript.

Notes

1. "This is here" and "That is there" are often interpreted as deictic relations in linguistic analyses (e.g., Levelt 1984). However, in those analyses, the expressions are interpreted as sentences that one person utters to another. The listener must interpret what the speaker says in terms of two-argument relation between two external objects—the speaker as a reference object and "this" or "that" as a located object. Moreover, the listener must interpret what the speaker says in terms of the speaker's frame of reference, with "here" meaning *near* and "there" meaning *far*. Basic relations are intrapersonal rather than interpersonal. There is only one argument ("this" or "that") and there is no external frame of reference (i.e., the viewer's own frame of reference suffices). The viewer is telling himself or herself that an object exists in a location. We expressed the result of that process as a sentence to communicate the idea to the reader, but the viewer need not do so. The viewer's representation is conceptual rather than linguistic.

2. One sheet contained *under*, *near*, *in*, and *away from* in the top left, top right, bottom left, and bottom right positions, respectively. Another contained *above*, *on*, *right of*, and *next to*. The third contained *left of*, *over*, *below*, and *far from*. Roughly equal numbers of subjects received the three different orders of sheets (25, 20, and 23, respectively).

References

- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Bryant, D. J., Tversky, B., and Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language*, 31, 74–98.
- Carlson-Radvansky, L. A., and Irwin, D. E. (1993). Frames of reference in vision and language: Where is above? *Cognition*, 46, 223–244.
- Carlson-Radvansky, L. A., and Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *Journal of Memory and Language*, 33, 646–671.
- Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*, 27–63. New York: Academic Press.
- Clark, H. H., Carpenter, P. A., and Just, M. A. (1973). On the meeting of semantics and perception. In W. G. Chase (Ed.), *Visual information processing*, 311–381. New York: Academic Press.
- Clark, H. H., and Chase, W. G. (1974). Perceptual coding strategies in the formation and verification of descriptions. *Memory and Cognition*, 2, 101–111.
- Cooper, L. A., and Shepard, R. (1973). The time required to prepare for a rotated stimulus. *Memory and Cognition*, 1, 246–250.
- Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review*, 95, 115–123.
- Eriksen, C. W., and St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, 40, 225–240.
- Franklin, N., and Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119, 63–76.
- Garnham, A. (1989). A unified theory of the meaning of some spatial relational terms. *Cognition*, 31, 45–60.
- Hayward, W. G., and Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition*, 55, 39–84.
- Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions in English*. Cambridge: Cambridge University Press.
- Jackendoff, R., and Landau, B. (1991). Spatial language and spatial cognition. In D. J. Napoli and J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore festschrift for Lila Gleitman*, 145–169. Hillsdale, NJ: Erlbaum.

- Jolicoeur, P., Ullman, S., and MacKay, L. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory and Cognition*, 14, 129–140.
- Jolicoeur, P., Ullman, S., and MacKay, L. (1991). Visual curve tracing properties. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 997–1022.
- Kosslyn, S. M., Chabris, C. F., Marsolek, C. J., and Koenig, O. (1992). Categorical versus coordinate spatial relations: Computational analyses and computer simulations. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 562–577.
- Kruskal, J. B., Young, F. W., and Seery, J. B. (1977). How to use KYST-2: A very flexible program to do multidimensional scaling and unfolding. Unpublished manuscript. Bell Laboratories, Murray Hill, NJ.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Landau, B., Jackendoff, R. (1993). “What” and “where” in spatial cognition and spatial language. *Brain and Behavioral Sciences*, 16, 217–238.
- Levelt, W. J. M. (1984). Some perceptual limitations in talking about space. In A. J. van Doorn, W. A. de Grind, and J. J. Koenderink (Eds.), *Limits on perception*, 323–358. Utrecht: VNU Science Press.
- Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, 12, 523–553.
- Logan, G. D. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1015–1036.
- Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, 28, 103–174.
- Logan, G. D., and Compton, B. J. (1996). Distance and distraction effects in the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 159–172.
- Logan, G. D., and Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting trials in a Stroop-like task. *Memory and Cognition*, 7, 166–174.
- Marr, D. (1982). *Vision*. New York: Freeman.
- Marr, D., and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Philosophical Transactions of the Royal Society, London*, 200, 269–294.
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Palmer, S. E. (1989). Reference frames in the perception of shape and orientation. In B. E. Shepp and S. Ballesteros (Eds.), *Object perception: Structure and process*, 121–163. Hillsdale, NJ: Erlbaum.

- Peterson, M. A., Kihlstrom, J. F., Rose, P. M., and Glisky, M. L. (1992). Mental images can be ambiguous: Reconstruals and reference frame reversals. *Memory and Cognition*, 20, 107–123.
- Pylyshyn, Z. (1984). *Computation and cognition*. Cambridge, MA: Harvard University Press.
- Pylyshyn, Z. (1989). The role of location indices in spatial perception: A sketch of the FINST spatial index model. *Cognition*, 32, 65–97.
- Sergent, J. (1991). Judgments of relative position and distance on representations of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 762–780.
- Shepard, R. N., and Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1, 1–17.
- Talmy, L. (1983). How language structures space. In H. L. Pick and L. P. Acredolo (Eds.), *Spatial orientation: Theory, research, and application*, 225–282. New York: Plenum Press.
- Ullman, S. (1984). Visual routines. *Cognition*, 18, 97–159.
- Vandaloise, C. (1991). *Spatial preposition: A case study from French*. Chicago: University of Chicago Press.

Chapter 14

The Language-to-Object Perception Interface: Evidence from Neuropsychology

Tim Shallice

Cognitive neuropsychology has as its principal aim the elucidation of the organization of the cognitive system through the analysis of the difficulties experienced by neurological patients with selective cognitive difficulties. As far as the relation between vision and language is concerned, the area that has been most extensively investigated concerns the semantic representation of objects. By contrast, the relation between how representations of space are accessed from vision and how they are accessed from language has been little touched; spatial operations have not been subject to much cognitive neuropsychology investigation.

If we consider objects, then the Gibsonian tradition teaches us that the richness of information available in the visual field is such that many of their properties may be inferred fairly directly from the visual array. Yet there are many other aspects of the visual world that cannot be inferred from the information in the visual field alone—the structural aspects of an object that are hidden from the present viewpoint, the potential behavior of an object and of the other objects likely to be found in its vicinity or that go with it in some other way. There are also wider properties of an object that may be accessed such as the perceptual features it has when experienced through other modalities, how it is used and by whom, what its function is, what types of thought process it triggers, and what intentions it may help to create. How are the processes involved in accessing these properties of an object when it is presented visually related to the way they are accessed when it is presented verbally?

This issue has been the subject of considerable controversy in cognitive neuropsychology in recent years for two reasons. A number of striking syndromes seem to relate very directly to it. In addition, the theory that most directly reflects the surface manifestations of the disorders differs from the standard theory in other fields where the issue has been addressed.

A model widely referred to in this book and in current cognitive science is that of Jackendoff (1987). Language is viewed as involving three main types of representation—phonological structures, syntactic structures, and semantic/conceptual structures.

As far as the semantic/conceptual structures are concerned, meanings have internal organization built up from a set of primitives and principles of combination, one of the primitives being the entity "thing." However, in addition to its phonological, syntactic and conceptual structures the representation of a word may contain specifically visual structures. The visual structures involved are, however, explicitly identified with the 3-D structural description level of Marr (1982).

Although Jackendoff's theorizing was concerned specifically with words and their meanings, the issues it addresses and in particular its position on the organization of the cognitive systems mediating semantic processing are closely related to issues recently much debated by cognitive neuropsychologists. A topic on which there has been much cognitive neuropsychology research in recent years is whether the semantic systems accessed when a word is being comprehended are the same as those used in the identification of an object, given that its structural description has already been determined. Some cognitive neuropsychologists have argued that they are the same, but others have claimed that they differ at least in part.

Approaches closely related to Jackendoff's have been adopted by certain cognitive neuropsychologists (e.g., Caramazza, Berndt, and Brownell 1982; Riddoch and Humphreys 1987). The best developed current neuropsychological account of a theory of this type is the organized unitary content hypothesis (OUCH) of Caramazza et al. (1990), which utilizes a feature based theory of semantic representations. More specifically, it holds that "access to a semantic representation through an object will necessarily privilege just those perceptual predicates that are perceptually salient in an object". Thus while many elements of the semantic representation are as easily accessible from visual as from verbal input, some aspects of the semantic representation are more easily accessed from its structural description than from its phonological representation. Access properties can be asymmetrical. The authors' rationale for assuming an asymmetric relation derives from consideration of certain conditions to be discussed shortly.

There is an older tradition in neuropsychology, however, which can be traced back at least as far as Charcot (1883) and Wernicke (1886). Certain syndromes suggest that visually based knowledge may be partly separable from verbally based knowledge. This perspective has been explicitly adopted more recently by a group of neuropsychologists (e.g., Warrington 1975; Beauvois 1982; Shallice 1987; and McCarthy and Warrington 1988) using the terminology *visual semantics* and *verbal semantics*, although the conceptual basis of the two types of representation has not been clearly articulated (see Caramazza et al. 1990; Rapp, Hillis, and Caramazza 1993; and Shallice 1993).

An intermediate position has been advocated by Bub et al. (1988) and by Chertkow and Bub (1990). Following Miller and Johnson-Laird (1976), they argue that a spe-

cific stage intervenes between attaining the structural description and accessing the amodal “core concept” of an object. Accurate identification of object is held to require more than just a characterization of an object’s structure, but must involve criteria which are more functional than structural. They therefore argue for the existence of a subsystem that contains only the application of the *functional* and perceptual criteria necessary for object identification, receiving the output from the structural description system and sending output to the core amodal semantic system. Thus “visual semantics” is reduced very considerably in its scope.

We thus have one position in cognitive neuropsychology (Caramazza et al. 1990) that is entirely compatible with Jackendoff’s perspective in holding that there is a single semantic/conceptual system. In addition it, namely the Caramazza et al. perspective, holds that accessing certain aspects of the semantic representation can be easier from the structural description than from phonology. Two other positions, (Warrington 1975; Chertkow and Bub 1990) hold that Jackendoff’s view is too gross a characterization of the subdivisions of the cognitive system involved in semantic processing, and that more than one semantic/conceptual system exists. A fourth position, which has yet to be formally articulated, holds that semantic representations are processed through a connectionist network of which different regions are more specialized for different types of semantic subprocess, but neither subprocess nor region can be characterized in an all-or-none fashion (see, for example, Allport 1985; Shallice 1988a).

Two main types of syndrome have been used to argue that the semantic-conceptual system is not in fact unitary but contains a number of types of subsystem—those involving some form of category specificity, and the modality-specific aphasias, in particular, optic aphasia. I will review the evidence from each in turn and then relate them to the alternative theories. A third syndrome—selective progressive aphasia—will also be addressed.

14.1 Category Specificity

The first group of syndromes responsible for the plausibility of the position that the semantic system is not unitary but composed of a number of subsystems are those manifesting so-called category specificity. The performance of the patient for some categories of knowledge is far better than for others. Of particular relevance is the syndrome originally described in four patients with herpes simplex encephalitis (Warrington and Shallice 1984). These patients had a selective problem in identifying animals, plants, and foods, while being able to identify man-made artefacts much better. For example, one of these patients, J.B.R., could name only 6% of living things and 20% of foods but could name 54% of man-made objects. Moreover, if the

judges assessed whether a description of a line drawing of the object "grasped the core concept," the contrast was even greater (living things, 6%; foods, 20%; but man-made objects, 80%). A similar effect was found when the patient was asked to give the meaning of the object's name and this, too, was assessed as to whether the core concept was grasped (living things, 8%; foods, 30%; man-made objects, 78%).

Similar effects have now been obtained with other patients with the same etiology (Pietrini et al. 1988; Sartori and Job 1988; Silveri and Gainotti 1988; Laurent et al. 1990; Swales and Johnson 1992; Sheridan and Humphreys 1993; Sartori et al. 1993; De Renzi and Lucchelli 1994). However, in the last few years there have been a rash of claims that these dissociations are essentially a result of characteristics of the stimulus set rather than evidence for a particular type of underlying organization of the semantic system.

Funnell and Sheridan (1992) initially claimed that the dissociations might arise because words matched for word frequency as used, say, by Warrington and Shallice (1984) may not be matched for visual familiarity. Indeed, McCarthy and Shallice (see Warrington and Shallice 1984) had shown that living things were less familiar to subjects than artefacts when matched for word frequency. Warrington and Shallice (1984) had dealt with this problem by showing that the dissociations were still present when differences in familiarity were taken out as a covariate. Moreover this explanation does not account for the way that the impairment of the patients involved foods as well as living things, as McCarthy and Shallice found foods to be more familiar than artefacts when word frequency is controlled.

A stronger argument was presented by Stewart, Parkin, and Hunkin (1992), who found that the category-specific dissociation of a herpes simplex patient, H.O., disappeared when word frequency, familiarity, and visual complexity were *all* controlled simultaneously. However, the basic dissociation, while statistically significant, was much weaker in H.O. than in some of the patients described earlier. Moreover, the nonliving category included objects like swamp, geyser, volcano, and waterfall instead of being composed solely of artefacts. Most critically, Sartori, Miozzo, and Job (1993) used stimuli matched on these three variables with their patient Michaelangelo, who showed a clear and significant category-specific effect of artefacts over living things on two different stimulus sets (living things, 30% and 40%; artefacts 70% and 76%).

Yet another possible artifact has been suggested by Gaffan and Heywood (1993), who argued that a critical variable was the density of exemplars within a category, which they held to be greater for living things than for artefacts. Because living things are more similar to each other and so less discriminable than artefacts, any discriminability problem would have a greater effect in the category of living things.

Riddoch and Humphreys (1987) had made a similar point previously and shown that there was more overlap between line drawings of animals than between line drawings of artefacts.

Gaffan and Heywood buttress their position on the difficulty in discriminating between living things, as opposed to artefacts, by considering the identification performance of three groups of subjects using the Snodgrass and Vanderwart (1980) stimuli. The first group were two patients of Farah, McMullen, and Meyer (1991), who showed standard category-specific effects; the second were normal subjects, who, however, were given only a 20 ms exposure; and the third used six monkeys, who were tested on how well they could decide which of two presented items was in a previously trained set. All three groups of subjects in their very different tasks showed an advantage of man-made objects over living things.

Gaffan and Heywood (1993) argue "These results from monkeys are contrary to Warrington and Shallice's conjecture ... that a specific system for identification of man-made objects has evolved in the human brain; if Warrington and Shallice's conjecture were correct, monkeys would show relatively greater difficulty in discriminating among inanimate objects than among living things, compared to human observers." It is not apparent, however, how such a comparison can be made because the tasks carried out were so different. Moreover, for the monkeys, most of the stimuli would presumably be meaningless objects; therefore what should be critical would indeed be raw discriminability. If, however, discriminability were a key factor underlying the performance of both the monkeys and the patients, then one would expect a positive correlation within each of the living and nonliving sets of stimuli between the results of the two group of subjects. In fact, there was no correlation between the items the monkeys found difficult and those the patients found difficult in either the living or the nonliving sets.

Gaffan and Heywood's work, like that in the other critical studies, used the Snodgrass and Vanderwart (1980) stimuli, for which norms are available on a number of relevant variables. In this set of stimuli the animals, in particular, tend to be rather similar to other members of their category. Warrington and Shallice (1984), however, also used the so-called Ladybird stimuli, large clear colored pictures designed for preschool children, with three of their patients. Shallice and Cinan have obtained ratings of structural complexity, familiarity, and discriminability from normal subjects for the Ladybird stimulus set and used these to reanalyze the findings of Warrington and Shallice. With these ratings, no difference was found between all three categories of stimuli (animals, artefacts, foods) for either familiarity or discriminability, but the animals remained structurally more complex than the other two categories. Because the task the patients carried out with this stimulus set had involved

word-picture matching using a four-alternative forced-choice task, the relevant degree of discriminability on the Gaffan-Heywood hypothesis was that within each set of five; this is what the subjects of Shallice and Cinan rated. However, with these stimuli two of the three original Warrington and Shallice patients on whom the test had been used performed significantly more poorly on foods than on artefacts with the third showing a strong trend in the same direction. Moreover, on a regression analysis using the ratings obtained by Shallice and Cinan, all three patients showed a significant effect of category and no effect of the other three variables. Thus it would appear that these category specificity findings cannot just be reduced to some combination of differences in word frequency, visual familiarity, structural complexity, and within-category discriminability.

In this respect, the work of Shallice and Cinan corroborated an earlier finding of Farah, McMullen, and Meyer (1991), who used the Snodgrass and Vanderwart (1980) stimuli with two patients exhibiting the standard category-specific dissociations. In a regression analysis on picture recognition performance, Farah, McMullen, and Meyer showed that neither name frequency, name specificity, similarity to other objects, structural complexity, nor object familiarity had any significant effect. The only factor to have such an effect was category membership. The absence of a significant effect of other factors in the presence of a significant effect of category makes implausible even one final convoluted artifactual explanation put forward by Gaffan and Heywood (1993). These authors suggested that the category difference arises through performance on items differing in a way dependent upon some other dimension; following Snedecor and Cochran (1967), they pointed out that *measurement* errors on the other dimension can lead to an apparent difference in performance across categories even when the differences on the other variables are allowed for as a covariate. However, what would then be expected is that there would be a basic effect of some other dimensions; this was not in fact found in either study.

Thus it would appear that the basic category-specific effects cannot be reduced just to an artifact of some combination of differences in word frequency, visual familiarity, structural complexity, and within-category discriminability across categories. A second type of finding that supports the conclusion that all neuropsychological dissociations in this domain cannot simply be attributed to some artifact of differences in presemantic factors is the existence of the complementary phenomenon, namely a superior performance in some subjects of living things (and in two studies foods) over artefacts (Warrington and McCarthy 1983, 1987; Hillis and Caramazza 1991; Sacchett and Humphreys 1992). The first two studies involved global aphasics who could only be tested by word-picture matching using, for instance, the Ladybird stimuli discussed above. However, the subjects in the last two studies were not glob-

ally aphasic; thus naming to visual confrontation could be used (for instance, C.W. in Sacchett and Humphreys 1992 scored 19/20 on naming animals; but only 7/20 on naming artefacts). Interestingly, the location of C.W.'s lesion (left frontoparietal) differed from that characteristic of the herpes simplex encephalitis cases (for all of whom the left temporal lobe was involved).

Much the most plausible conclusion is that the category-specific effects do not arise at a presemantic level due to some difference in difficulty between the categories but reflect some qualitative difference in the semantic representations of the categories. When the herpes encephalitis syndrome was first described, it was explained in terms of a contrast between stimuli primarily differentiable in terms of their sensory qualities and those more saliently differentiable in terms of their function.

Unlike most plants and animals, man-made objects have clearly defined functions. The evolutionary development of tool using has led to finer and finer functional differentiations of artefacts for an increasing range of purposes. Individual inanimate objects have specific functions and are designed for activities appropriate to their function. Consider, for instance, chalk, crayon, and pencil; they are all used for drawing and writing, but they have subtly different functions. . . . Similarly, jar, jug, and vase are identified in terms of their function, namely, to hold a particular type of object, but the sensory features of each can vary considerably. By contrast, functional attributes contribute minimally to the identification of living things (e.g., lion, tiger, and leopard), whereas sensory attributes provide the definitive characteristics (e.g., plain, striped, or spotted). (Warrington and Shallice 1984, 849)

A closely related position was taken to explain the complementary syndrome to be discussed later (see Warrington and McCarthy 1983.)

Dector, Bub, and Chertkow (in press) take a somewhat related position based on their study of a patient, E.L.M., who suffered from bilateral temporal lobe strokes. On tests of perceptual knowledge of objects he performed normally, but he was grossly impaired at many tests involving the perceptual characteristics of animals. Dector, Bub, and Chertkow argue that the difference between the superiority of artefacts over animals arises because different tokens of the same man-made object may show a considerable variation in the shape of its parts but a consistent function that allows for a unique interpretation, thus echoing the Warrington-Shallice position. However, they then argue that artefacts "can be uniquely identified at the basic level through a functional interpretation of their parts" and this is why they are relatively preserved (see De Renzi and Lucchelli 1994 for a related position). Many artefacts with a unique function do indeed have a unique organization of distinctly functioning parts; take a lamp, for example. However, others, such as a table tennis ball, do not. As yet it remains unclear to what extent the relative sparing of artefacts depends upon their unique organization of distinctly functioning parts or on the unique functions of the whole.

14.2 Sensory Quality and Functional Aspects of Different Categories

The position just developed attributes differences in performance across different categories to the way that identification in some categories depends critically on sensory quality information but for others functional information is more critical. One can, however, consider how well different semantic aspects of the same category are understood by patients who show this category-specific pattern. When this is done, knowledge of functional aspects of biological categories tends to be much better preserved than knowledge of sensory quality aspects (Silveri and Gainotti 1988). In a related fashion, Dector, Bub, and Chertkow's (in press) patient E.L.M. was much better at answering "encyclopedic" questions about animals such as "Does a camel live in the jungle or the desert?" (85%) than visual ones such as "Does a camel have horns or no horns?" where he was at chance (55%). However, the effects are not completely clear-cut. The performance of E.L.M., say, on functional aspects of animals was still well below that of normal controls, who scored 99%. This was not due just to a general problem with carrying out semantic operations on concrete objects; when asked to identify artefacts he performed at ceiling.

A more dramatic example is given by De Renzi and Lucchelli's (1994) herpes encephalitic patient, Felicia. In explaining the perceptual difference between pairs of animals, for example, goat and sheep, or paired fruits or vegetables, for example, cherry and strawberry, she performed far worse than the worst controls (15% vs. 90%; 49% vs. 85%). However, in explaining the *visual* difference between paired objects, for example, lightning rod and TV antenna, she was somewhat better than the normal mean (90% vs. 85%). Analogous results have been reported in a number of other studies (e.g., Silveri and Gainotti 1988; Sartori and Job 1988; Farah et al. 1989) although at least one patient, Giuletta (Sartori et al. 1993), answered nonvisual questions about animals almost perfectly (see also Hart and Gordon 1992), while at the other extreme S.B. (Sheridan and Humphreys 1993) performed almost as poorly on visual as on nonvisual questions about animals (70% and 65%, respectively).

Why should the category-specific impairment generally recur if in a milder form when the patient is responding to questions about animals or foods which appear not to be based on accessing sensory qualities? Does it not undermine the explanation of category-specific effects outlined earlier, namely, that they arise from damage affecting sensory quality representations? If one articulates the theory developed thus far in a connectionist form, then the problem can be resolved. Farah and McClelland (1991) investigated a model (see figure 14.1) in which some semantic units represented the functional roles taken by an item, while others represented its visual qualities. Each of the semantic units was connected (bidirectionally) to the others, to units representing structural descriptions, and to units representing phonological word-

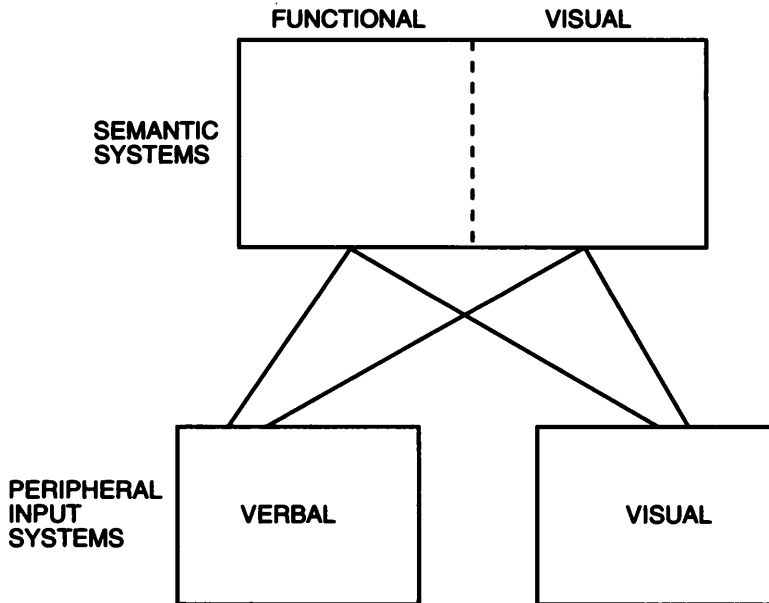


Figure 14.1

Farah and McClelland's (1991) model for explaining category-specific preservation of artefact comprehension and naming (reproduced by permission from Farah and McClelland 1991).

forms. The number of units in the two subsets of semantic representations was determined through an experiment on normal subjects. Subjects rated the description of each item in definitions of both living and nonliving things in the *American Heritage Dictionary* as to whether it described the visual appearance of the item, what the item did, or what it was for. On average there were 2.13 visual descriptions and 0.73 functional ones, but the ratio between the two types was 7.7:1 for living things and only 1.4:1 for the nonliving things. These values were then realized in the representations of living things and artefacts used for training. The network was trained using an error correction procedure based on the delta rule (Rumelhart, Hinton and McClelland 1986) applied after the network had been allowed to settle for ten cycles following presentation of each input pattern. In each of four additional variants of the basic network, one particular parameter was altered so as to establish the robustness of any effect obtained.

The most basic finding was that lesioning the "visual" semantic units led to greater impairment for living things than for artefacts with the opposite pattern shown for the lesioning of the functional semantic units. Thus the standard double dissociation was obtained due to "identification" of living things relying more on the visual

semantic units and "identification" of artefacts depending more on the functional semantic units. More interestingly, if one examines how close a match occurs over the *functional* semantic units when a lesion is made to the *visual* semantic units, then there is a difference between the two types of item. The *functional* representations of the living things were less adequately retained than those of artefacts. In the original learning process, the attainment of the representation in one of the semantic "sub-systems" helps to support the learning of the complementary representation in the other; the richer the representation is in one of the systems, the more use is made of it in learning the complementary representation. Thus the most typical relation between functional and visual impairments with living things is explained. Whether the full range of relations observed can be explained remains to be investigated.

There are two uncomfortable findings that the model would appear well designed to explain. First, the living/nonliving distinction is not absolute. Thus Y.O.T. was one of the global aphasic patients who performed very much better on word-picture matching with living things and foods than with artefacts (Warrington and McCarthy 1987). In Y.O.T.'s case, the impairment did not extend to large man-made objects such as bridges or windmills. Patient J.J. of Hillis and Caramazza (1991), who had a selective sparing of animal naming like Y.O.T., also had the naming of means of transportation spared. Complementarily, the problems of herpes encephalitic patients extended to gemstones and fabrics. The semantic representation of all these subcategories may well consist more of visual units than of functional ones, especially if function has to be linked with a specific action.

Second, the living/nonliving distinction is graded. Thus patients have been described in whom the deficit is limited, say, to animals alone (e.g., Hart, Berndt, and Caramazza 1985; Hart and Gordon 1992). The sensory quality/function contrast would seem likely to be more extreme for animals than foods, say, so that for more minor damage to sensory quality units only the least functional of the semantic categories would be affected.

Overall, this group of category-specific disorders fits with the idea that knowledge of the characteristics of objects is based on representations in more than one type of system. Realizing the different systems as sets of units in a connectionist model allows certain anomalies in the basic subsystem approach to be explained. The nature of the representations mediated by each of the systems remains unclear, however. The deficit appears not to correspond simply to damage to visual units. Thus one of the patients studied by Warrington and Shallice (1984) was unable to identify foods by taste as well as by sight. Moreover, in three of the patients where it has been assessed (Michaelangelo in Sartori and Job 1988; E.L.M. in Dector, Bub, and Chertkow, in press; and S.B. in Sheridan and Humphreys 1993), relative size judgments could be

made fairly accurately, suggesting that even the visual deficit does not extend to all visual characteristics. The issue remains open.

14.3 Optic Aphasia

A second syndrome that suggests the need to refine the conceptual/structural description contrast of Jackendoff (1987) is optic aphasia. First described by Freund (1889), optic aphasia refers to an impairment where the patient is unable to name objects presented visually but at the same time gives evidence of knowing what these objects are, for instance, by producing an appropriate mime. Moreover, the problem is not just one of naming; the patient is able to produce the name to a description or on auditory or tactile presentation. A considerable number of patients have been described who roughly fit the pattern (see Beauvois 1982; Iorio et al. 1992; Davidoff and De Bleser 1993 for reviews). If one limits consideration to patients who do not appear to have any impairment in accessing the structural description because stimulus quality does not affect naming ability, Davidoff and De Bleser (1993) list fourteen patients who have been formally described. Certain of these patients performed perfectly in gesturing the use of visually presented stimuli they could not name (Lhermitte and Beauvois 1973; Caplan and Hedley-White 1974; Gil et al. 1985).

This apparent preservation of knowledge of the visually presented object when it cannot be named has been explained most simply by assuming that the optic aphasic suffers from a disconnection between “visual semantics” and “verbal semantics,” with the name only being accessible from verbal semantics (Beauvois 1982; Shallice 1987). The distinction between subsystems at the semantic level appears to differ from the one drawn in the previous section between systems representing functional and visual or sensory quality types of information. I will address this issue in more detail later. In any case, a number of authors have contested the claim (see Riddoch and Humphreys 1987; Garrett 1992; Rapp, Hillis, and Caramazza 1993), holding that the miming could simply be based on an affordance, that is, an action characteristically induced by the shape of the object, or a cross-modal association of sensory and motor schemas, either of which might in turn be based only on an intact structural description. Alternatively, miming might require accessing only restricted parts of the semantic system, in particular those parts most strongly realized from the structural description because they are also represented in it explicitly, for example, the tines of forks; this is the privileged access theory account of Caramazza et al. (1990) and Rapp, Hillis, and Caramazza (1993). A similar explanation might also be given for the preserved drawing from memory shown in patients such as J.F. (Lhermitte and Beauvois 1973).

However, access to other types of information can be present in these patients when they cannot name. For instance, Coslett and Saffran (1992) gave their patient EM2 a task based on one devised by Warrington and Taylor (1978) in which the patient has to judge which of three items are functionally similar, for example, zipper, button, coin (see also patient C.B. in Coslett and Saffran 1989). EM2 scored at 97% on this task, with the control mean being 94%. Because the affordances of a zipper and a button are not similar, it is difficult to see how the use of affordances might be the basis for this good performance; indeed, there are no subcomponents of the two structural descriptions that are related. Rapp, Hillis, and Caramazza (1993), in confronting the argument that such a pattern of performance presents a difficulty for their privileged access position (Shallice 1993), merely respond by saying, "difficulty naming visually presented items in the face of demonstrated intact comprehension of some aspect of the visual structures, however, indicates that the full semantic description required to support naming has not been activated from a 3-D representation of the stimulus." This argument presupposes that normal performance on the function-matching test can be obtained when activation of the relevant semantic representation is reduced. This claim is merely asserted by Rapp, Hillis, and Caramazza. However, because the task is a three-alternative forced-choice test, with rather basic semantic information being required about each item—concerning its function—the assertion has some plausibility.

Similar results have, however, been obtained by Manning and Campbell (1992) on patient A.G. on semantic tasks which appear to be much more demanding. Two types of test were used with these patients. The first was the Pyramids and Palm Trees test of Howard and Patterson (1992). In a typical item of this test, the patient has to decide which tree (palm, fir) goes best with a pyramid. The stimuli can be presented either visually, verbally, or in mixed visual-verbal format. In the second test, the patient has to answer sets of questions about each item, (e.g., What is it made of?) both when the item is presented visually and when it is presented auditorily. A.G. performed at only 40%–50% in naming objects from drawings, but at 100% in naming to description and at 91% in naming tactilely presented stimuli, thus showing a specific naming defect with visual stimuli. However, A.G.'s performance on the Pyramids and Palm Trees test, while not at ceiling, was virtually identical across the visual and verbal modalities of presentation (82% vs. 84%) and in both cases was within one standard deviation of the mean of normal control subjects. A similar pattern was observed for the question-answering test (88% vs. 91%). Druks and Shallice's (1995) patient L.E.W. behaved in the same way for both types of test. That patients showed no difference and were not at ceiling on tests of auditory and verbal comprehension seems impossible to account for in Rapp, Hillis, and Caramazza's

(1993) version of the privileged access theory, which involves a unitary semantics. By contrast, these results fit well with the multiple semantic system position.

Coslett and Saffran (1992), on the other hand, present an interesting variant of the multiple store position. They agree that two semantic stores do exist and that one is disconnected from the language production mechanisms in optic aphasic patients, but they argue that the stores are primarily distinguished by hemisphere, with the right-hemisphere semantic system being disconnected from the language production systems in the left hemisphere. However, the patients described by Manning and Campbell (1992) present a difficulty for this position. In the acute condition immediately after a sudden onset lesion (e.g., vascular), the right hemisphere is supposed by right-hemisphere theorists such as Coslett and Saffran not to have access to any phonological lexicon, although they hold that over time a phonological lexicon becomes available to a semantic system in the right hemisphere (Coslett and Saffran 1989). This semantic system or the variety of output phonological word-forms that can be accessed from it is then seen to have an effective content corresponding to that of the words readable in deep dyslexia (Coltheart 1980a; Saffran et al. 1980; Coslett and Saffran 1989). In deep dyslexia, however, concrete nouns can be read reasonably well but verbs present severe problems (Coltheart 1980b). Yet while patients A.G. and L.E.W. were severely impaired in naming objects, which they could identify nonverbally, they could name actions very well. Thus A.G. was 95% correct at naming actions—the same level as controls—but worse than 50% at naming objects. This contrast in ease of accessing output phonological word-forms from an intact semantic representation is the opposite of what would be expected according to the right-hemisphere theory, where one would assume that objects should be more easily nameable than actions. The basic multiple semantic store position can perhaps explain the obtained effect by assuming the existence of another semantic subsystem—one controlling actions (Druks and Shallice 1995); being an essentially high-level output system but accessible from perceptual input, it would have connections to verbal semantics distinct from those used by the visual semantic representations of objects. This, however, remains a highly speculative account.

There remains one other counterintuitive aspect of optic aphasia. Many of the patients characterized as optic aphasic through their pattern of success and failure on naming and comprehension tests exhibit a strange set of errors when they fail to name correctly. Of the optic aphasic patients reviewed by Iorio et al. (1992), who generally correspond with Davidoff and De Bleser's (1993) group 2 optic aphasics, nearly all made both semantic and perseverative errors, with less than half also making visual errors. Moreover, in the most detailed analysis of such errors—that of Lhermitte and Beauvois (1973) of their patient J.F.—the authors consider the interaction between

Table 14.1
Errors Made by J. F. in Two Experiments

Type of error	Example	100 pictures	30 objects
<i>Horizontal errors</i>			
Semantic	<i>shoe</i> ⇒ "hat"	9	3
Visual	<i>coffee beans</i> ⇒ "hazel nuts"	2	1
Mixed visual-and-semantic	<i>orange</i> ⇒ "lemon"	6	1
<i>Vertical errors</i>			
Item and coordinate perseveration	T26 ... ⇒ "wristwatch"		
	T27 <i>scissors</i> ⇒ "wristwatch"		
	T44 ... ⇒ "newspaper"		
	T45 <i>case</i> ⇒ "two books"	8	2
<i>Mixed horizontal/vertical errors</i>			
	T43 ... ⇒ "chair"		
	T47 <i>basket</i> ⇒ "cane chair"		
	T53 <i>string</i> ⇒ "strand of weaved cane"	3	0

Source: Lhermitte and Beauvois 1973.

what they call "horizontal errors," understood strictly in terms of the processes (temporally) intervening between presentation of the stimulus and the responses, and what they call "vertical errors," where effects of preceding stimuli or responses occur. It is clear from this analysis that the perseverative and the semantic errors combine in a complex way (see table 14.1).

Why might such a strange combination of errors be characteristic of optic aphasia? Again a possible answer can be given by adding a connectionist dimension to the models. Plaut and Shallice (1993a) considered a network which had a direct pathway mapping visual representations into semantic ones. It also had a "cleanup" pathway that involved recurrent connections from the semantic units to the "cleanup" units and back (see figure 14.2). The network used an iterative version of the backpropagation learning algorithm known as backpropagation through time (Rumelhart, Hinton, and Williams 1986). Training with an algorithm of this type in such a recurrent network leads to its developing a so-called attractor structure; the effect of the operation of the cleanup pathway is to move a noisy first-pass representation at the semantic level toward one of the representations it has been trained to produce as an output, given that the initial representation is in the vicinity of the trained one.

The network contained one other major difference from other networks well known in cognitive psychology, such as Seidenberg and McClelland's (1989). In

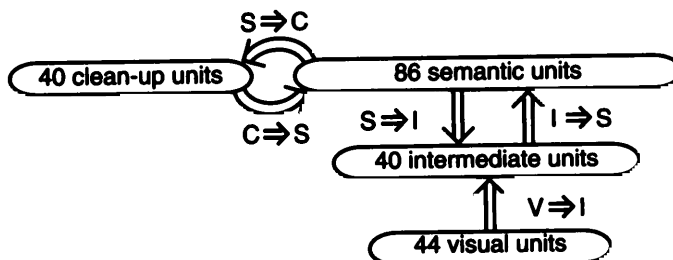


Figure 14.2

Plaut and Shallice's (1993) model for explaining the typical error pattern found in optic aphasia (reproduced from Plaut and Shallice 1993a by permission).

the nervous system, changes in synaptic efficiency at a single synapse occur at many different time scales (Kupferman 1979). The incorporation of additional connection weights that change much more rapidly in training than those standardly used in connectionist modeling is also computationally valuable; it allows for temporal binding of neighboring elements into a whole (e.g., von der Marlsburg 1988) and facilitates recursion (Hinton, personal communication described in McClelland and Kawamoto 1986). Each connection in the network therefore combined a standard, slowly changing, long-term weight with a rapidly altering, short-term weight based on the correlation between the activities of its input and output units.

A network having both types of weights tends to reflect in its behavior both its long-term reinforcement history and its most recent activity; it contains the analogue of both long-term learning and of priming. The network was trained to respond appropriately at the semantic level to the structural representations of forty different objects. Wherever the network was lesioned, it produced a few visual errors but considerably more semantic errors and typically more with both visual and semantic similarity to the stimulus. More critically, there was a strong perseverative aspect to the responses. The previous response or one of its semantic associates could well occur as an error to the present stimulus. This corresponds well to the error pattern occurring in optic aphasia.

Adding a connectionist dimension to the model therefore allows the error pattern of the syndrome to be explained. The information-processing model we used as a basis for the connectionist simulations corresponds to those of Riddoch and Humphreys (1987) and Caramazza et al. (1990), which were held to be unsatisfactory earlier in this chapter. However, the essence of the simulation is that if short- and long-term weights are combined, the errors will reflect both perseverative influences and the level of representation at which strong attractors occur.¹ Thus the obtained error pattern would also be expected if an analogous connectionist dimension were

added to the multiple semantic system models, provided that one or more of the semantic systems had analogous attractor properties.

14.4 Conclusion

In the sections 14.1 and 14.2 certain syndromes were discussed involving category-specific impairments, particularly those associated with herpes simplex encephalitis, where large differences in performance exist between identification of man-made artefacts on the one hand and of living things and foods on the other. Explanations in terms of differences between the categories on a number of potentially confounding dimensions were considered and rejected. The favored explanation assumes that partially separable systems underlie the semantic representations of the functional and of the sensory quality properties of stimuli. In section 14.3 another syndrome—optic aphasia—was considered; here it was argued that the most plausible explanation involved disconnecting “visual” and “verbal” or “lexical” semantic representations.

The evidence presented in all three sections poses difficulties for the view that a single conceptual system, together with a structural description system that can also be addressed from above, is a sufficient material base for representing semantic operations. The sensory quality component of the semantic system cannot be conflated with the structural description system because variables relevant to disorders of the latter system, for example, presentation of items from unusual views (Warrington and Taylor 1978), do not predict the stimuli that are difficult for patients with impairments to the former system (Warrington and Shallice 1984; Dector, Bub, and Chertkow in press). The issue is even clearer from the perspective of the second set of disorders. In certain optic aphasic patients much more semantic information appears to be accessible from vision than could be based on the structural description alone; yet it would appear not to be available in a generally accessible conceptual system because it cannot be used to realize naming.

By contrast, the accounts presented for these disorders fit naturally with those beginning to be developed within developmental psychology for image schemas at a level of abstraction higher than the structural description and yet not simply subsumable within verbal knowledge (see Mandler, chapter 9, this volume). However, to argue that the such visual semantic processes should be limited to what is required for visual identification alone—in Chertkow and Bub’s (1990) visual identification procedure subsystem—and that this is the only system lying between the structural description system and an amodal core semantic system does not fit well for either syndrome. In the herpes encephalitis condition what is lost are the sensory quality aspects of the item, while identification procedures, according to Miller and Johnson-Laird (1976), require primarily functional property information as well as structural

analysis. Turning to optic aphasia, one possibility to explain the syndrome might be to view it as arising from a disconnection between the visual identification procedures and the core semantic system. However, a task like Pyramids and Palm Trees involves the utilization of shared context. The Bub and Chertkow theory holds that inferred context is stored in the amodal core semantic system, so that an optic aphasic would be not expected to perform well on such tasks for words that could not be named. Patients A.G. (Manning and Campbell 1992) and L.E.W. (Druks and Shallice 1995) show the opposite pattern, namely, intact performance on this task, together with grossly impaired naming.

There are, however, certain problems in explaining the two types of syndrome in terms of the functional/sensory quality and visual/verbal dichotomies. The concepts are orthogonal. The information available in a visually or sensory quality-based semantic system, as inferred by the information lost in the herpes encephalitic patient is not the only information *accessible* from the visual modality in the optic aphasic patient. Certain optic aphasic patients, for example, A.G. and L.E.W., can access types of information from vision that would be in the functional or encyclopedic parts of the semantic system on a simple all-or-none multiple store view. Moreover, within the semantic dementia literature there are striking echoes of this visual input predominance extending outside the purely sensory quality domain in the performance of patient T.O.B. (McCarthy and Warrington 1988).² When a picture was presented to T.O.B. his identification was more than 90% accurate for both types of material, but he identified verbal input artefacts much better than living things (89% vs. 33%). Thus when the word *dolphin* was presented, the patient could say only, "A fish or a bird," but when presented with the picture, he said, "Lives in water ... they are trained to jump up and come out. ... In America during the war they started to get this particular animal to go through to look into ships." McCarthy and Warrington have argued that this patient has an impairment that affects the stored information itself rather than an input pathway because of the consistency with which particular items were or were not identifiable (see for rationale Warrington and Shallice 1979; Shallice 1987). Thus contrasting both optic aphasia and semantic dementia with herpes simplex encephalitis, it would appear that the putative lines of cleavage within the semantic system suggested by the syndromes differ.

One possibility is to postulate category-specific systems that are themselves specific to particular modalities (McCarthy and Warrington 1988). However, explanations provided for certain secondary aspects of the syndromes suggest an alternative direction in which a more economical solution might lie. A connectionist simulation of Farah and McClelland (1991) can account for certain otherwise most recalcitrant findings about category-specific disorders. For optic aphasia, the counterintuitive error pattern associated with the disorder is in turn explicable on a connectionist

simulation of Plaut and Shallice (1993a). Thus adding a connectionist dimension to the theoretical framework used to account for the characteristics of the syndromes enables a much fuller explanation of the detailed nature of the deficits to be provided. Adding such a connectionist dimension to a subsystem approach provides an account closely related to presimulation suggestions made over the last ten years or so, that the semantic system has as its material basis a large associative neural network with different concepts being represented in different combinations of its subregions, depending on the specific subset of input and output systems generally used to address them (see Allport 1985; Warrington and McCarthy 1987; Shallice 1988b; and Saffran and Schwartz 1994). How the rule-governed aspects of semantic processing would be dealt with on this type of account has not been addressed by neuropsychologists. However, the use of a connectionist network framework for explaining aspects of neuropsychological disorders does not preclude the possibility of explaining rule-governed aspects of semantic processing, provided additional elements are added to the basic network (see Touretsky and Hinton 1988; Derthick 1988; and Miikkulainen 1993). On this account the semantic/conceptual system postulated by Jackendoff would need to be realized as a complex neural network. As yet, though, no implementation adequately explains the rich and highly counterintuitive evidence that detailed study of individual neurological patients provides.

Notes

1. This is especially the case if the mapping from the visual to the semantic level is not orthogonal, as it is in language (see Plaut and Shallice 1993a); for visual presentation of objects, the visual and the semantic representations are correlated.
2. A simple peripheral explanation of the phonological word-form being damaged can also be excluded.

References

- Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. K. Newman and R. Epstein (Eds.), *Current perspectives in dysphasia*. Edinburgh: Churchill Livingstone.
- Beauvois, M. F. (1982). Optic aphasia: A process of interaction between vision and language. *Philosophical Transactions of the Royal Society, London, B298*, 33–47.
- Bub, D., Black, S., Hampson, E., and Kerkesy, A. (1988). Semantic encoding of pictures and words: Some neuropsychological observations. *Cognitive Neuropsychology*, 5, 27–66.
- Caplan, L., and Hedley-White, T. (1974). Cueing and memory dysfunction in alexia without agraphia: A case report. *Brain*, 97, 251–262.

- Caramazza, A., Berndt, R. S., and Brownell, H. H. (1982). The semantic deficit hypothesis: Perceptual parsing and object classification by aphasic patients. *Brain and Language*, 15, 161–189.
- Caramazza, A., Hillis, A. E., Rapp, B. C., Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology*, 7, 161–189.
- Charcot, J. W. (1883). Un cas de suppression brusque et isolée la vision mentale des signes et des objets (formes et couleurs). *Progrès Medical*, 11, 568–571.
- Chertkow, H., and Bub, D. (1990). Semantic memory loss in dementia of Alzheimer's type. *Brain*, 113, 397–417.
- Coltheart, M. (1980a). Deep dyslexia: A right hemisphere hypothesis. In M. Coltheart, K. E. Patterson, and J. C. Marshall (Eds.), *Deep dyslexia*. London: Routledge.
- Coltheart, M. (1980b). Deep dyslexia: A review of the syndrome. In M. Coltheart, K. E. Patterson, and J. C. Marshall (Eds.), *Deep dyslexia*. London: Routledge.
- Coslett, H. B., and Saffran, E. M. (1989). Preserved object recognition and reading comprehension in optic aphasia. *Brain*, 112, 1091–1110.
- Coslett, H. B., and Saffran, E. M. (1992). Optic aphasia and the right hemisphere: Replication and extension. *Brain and Language*, 43, 148–161.
- Davidoff, J., and De Bleser, R. (1993). Optic aphasia: A review of past studies and a reappraisal. *Aphasiology*, 7, 135–154.
- Dector, M., Bub, D., and Chertkow, H (in press). Multiple representations of object concepts: Evidence from category-specific aphasia. *Cognitive Neuropsychology*.
- De Renzi, E., and Lucchelli, F. (1994). Are semantic systems separately represented in the brain? The case of living category impairment. *Cortex*.
- Derthick, M. (1988). Mundane reasoning by parallel constraint satisfaction. Ph.D. diss., Carnegie Mellon University, Pittsburgh.
- Druks, J., and Shallice, T. (1995). Preservation of visual identification and action naming in optic aphasia. Paper presented at the Annual British Neuropsychological Society Conference, London, March.
- Farah, M. J., Hammond, K. H., Mehta, Z., and Ratcliff, G. (1989). Category specificity modality specificity in semantic memory. *Neuropsychologia*, 27, 193–200.
- Farah, M. J., and McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120, 339–357.
- Farah, M. J., McMullen, P. A., and Meyer, M. M. (1991). Can recognition of living things be selectively impaired? *Neuropsychologia*, 29, 185–194.
- Freund, D. C. (1889). Über optische Aphasie und Seelenblindheit. *Archiv für Psychiatrie und Nervenkrankheiten*, 20, 276–297.

- Funnell, E., and Sheirden, J. (1992). Categories of knowledge? Unfamiliar aspects of living and nonliving things. *Cognitive Neuropsychology*, 9, 135–153.
- Gaffan, D., and Heywood, C. A. (1993). A spurious category-specific visual agnosia for living things in normal human and nonhuman primates. *Journal of Cognitive Neuroscience*, 5, 118–128.
- Garrett, M. (1992). Disorders of lexical selection. *Cognition*, 42, 143–180.
- Gil, R., Pluchon, C., Toullat, G., Michenau, D., Rogew, R., Lefevre, J. P. (1985). Disconnexion visuo-verbale (aphasie optique) pour les objets, les images, les couleurs, et les visages avec alexie abstractive. *Neuropsychologia*, 23, 333–349.
- Hart, J., Berndt, R. S., and Caramazza, A. (1985). A category-specific naming deficit following cerebral infarction. *Nature*, 316, 439–440.
- Hart, J., and Gordon, B. (1992). Neural systems for object knowledge. *Nature*, 359, 60–64.
- Howard, F., Patterson, K. E. (1992). *Pyramids and palm trees: A test of semantic access from pictures and words*. Thames Valley.
- Iorio, L., Falango, A., Fragassi, N. A., and Grossi, D. (1992). Visual associative agnosia and optic aphasia: A single case study and a review of the syndromes. *Cortex*, 28, 23–37.
- Jackendorff, R. (1987). On beyond zebra: The relation of linguistic and visual information. *Cognition*, 26, 89–114.
- Kupferman, I. (1979). Modulatory actions of neurotransmitters. *Annual Review of Neuroscience*, 2, 447–465.
- Laurent, B., Allegri, R. F., Michel, D., Trillet, M., Naegle-Faure, B., Foyatier, N., and Pellat, J. (1990). Encéphalites herpétiques à prédominance unilatérale: Étude neuropsychologique au long cours de 9 cas. *Revue Neurologique*, 146, 671–681.
- Lhermitte, F., and Beauvois, M. F. (1973). A visual-speech disconnection syndrome: Report of a case with optic aphasia, agnosic alexia, and colour agnosia. *Brain*, 96, 695–714.
- Manning, L., and Campbell, R. (1992). Optic aphasia with spared action naming: A description and possible loci of impairment. *Neuropsychologia*, 30, 587–592.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- McCarthy, R. A., and Warrington, E. K. (1988). Evidence for modality-specific meaning systems in the brain. *Nature*, 334, 428–430.
- McClelland, J. L., and Kawamoto, A. H. (1986). Mechanisms of sentence production: Assigning roles to constituents of sentences. In J. L. McClelland and D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 2, 272–325. Cambridge, MA: MIT Press.
- Miikkulainen, R. (1993). *Subsymbolic case-role analysis of sentences with embedded clauses*. Technical report AI 93-202. Austin: University of Texas Press.
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge: Cambridge University Press.

- Pietrini, V., Nertempi, P., Vaglia, A., Revello, M. G., Pinna, V., and Ferro-Milone, F. (1988). Recovery from herpes simplex encephalitis: Selective impairment of specific semantic categories with neuroradiological correlation. *Journal of Neurology Neurosurgery and Psychiatry*, 51, 1984-1293.
- Plaut, D. C., and Shallice, T. (1993a). Perseverative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience*, 5, 89-117.
- Plaut, D. C., and Shallice, T. (1993b). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10, 377-500.
- Rapp, B. C., Hillis, A. E., and Caramazza, A. (1993). The role of representations in cognitive theory: More on multiple semantics and the agnosias. *Cognitive Neuropsychology*, 10, 235-249.
- Riddoch, M. J., and Humphreys, G. W. (1987). Visual object processing in a case of semantic access agnosia. *Cognitive Neuropsychology*, 3, 201-224.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1. Cambridge, MA: MIT Press.
- Sacchetti, C., and Humphreys, G. W. (1992). Calling a squirrel a squirrel but a canoe a wigwam: A category-specific impairment for artefactual objects and body parts. *Cognitive Neuropsychology*, 9, 73-86.
- Saffran, E. M., Bogyo, L. C., Schwartz, M. F., and Marin, O. S. M. (1980). Does deep dyslexia reflect right hemisphere reading. In M. Coltheart, K. E. Patterson, and J. C. Marshall (Eds.), *Deep dyslexia*. London: Routledge.
- Saffran, E. M., and Schwartz, M. F. (1994). Of cabbages and things: Semantic memory from a neuropsychological perspective. In C. Umiltà and M. Moscovitch (Eds.), *Attention and performance*, vol. 15, 507-536. Cambridge, MA: MIT Press.
- Sartori, G., and Job, R. (1988). The oyster with four legs: A neuropsychological study of the interaction of visual and semantic information. *Cognitive Neuropsychology*, 5, 105-132.
- Sartori, G., Job, R., Miozzo, M., Zagu, S., and Machiori, G. (1993). Category-specific form-knowledge deficit in a patient with herpes simplex encephalitis. *Journal of Clinical and Experimental Neuropsychology*, 15, 280-299.
- Sartori, M., Miozzo, M., and Job, R. (1993). Category-specific naming impairments? Yes. *Quarterly Journal of Experimental Psychology*, 46a, 489-504.
- Seidenberg, M. S., and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Shallice, T. (1987). Impairments of semantic processing: Multiple dissociations. In R. Job, G. Sartori, and M. Coltheart (Eds.), *The cognitive neuropsychology of language*. London: Erlbaum.
- Shallice, T. (1988a). *From neuropsychology to mental structure*. New York: Cambridge University Press.

- Shallice, T. (1988b). Specialization within the semantic system. *Cognitive Neuropsychology*, 5, 133–142.
- Shallice, T. (1993). Multiple semantics: Whose confusions? *Cognitive Neuropsychology*, 10, 251–261.
- Sheridan, J., and Hymphreys, G. W. (1993). A verbal-semantic category specific recognition impairment. *Cognitive Neuropsychology*, 10, 143–184.
- Silveri, M. C., and Gainotti, G. (1988). Interaction between vision and language in category-specific semantic impairment. *Cognitive Neuropsychology*, 3, 677–709.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical methods*. 6th ed. Ames: Iowa State Press.
- Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174–215.
- Stewart, F., Parkin, A. J., and Hunkin, N. M. (1992). Naming impairment following recovery from herpes simplex encephalitis: Category-specific? *Quarterly Journal of Experimental Psychology*, 44a, 261–284.
- Swales, M., and Johnson, R. (1992). Patients with semantic memory loss: Can they relearn lost concepts? *Neuropsychological Rehabilitation*, 2, 295–305.
- Von der Marlsburg, C. (1988). Pattern recognition by labeled graph matching. *Neural Networks*, 1, 141–148.
- Warrington, E. K. (1975). The selective impairments of semantic memory. *Quarterly Journal of Experimental Psychology*, 27, 635–657.
- Warrington, E. K., and McCarthy, R. (1983). Category-specific access dysphasia. *Brain*, 106, 859–878.
- Warrington, E. K., and McCarthy, R. (1987). Categories of knowledge: Further fractionation and an attempted integration. *Brain*, 110, 1273–1296.
- Warrington, E. K., and Shallice, T. (1979). Semantic access dyslexia. *Brain*, 102, 43–63.
- Warrington, E. K., and Shallice, T. (1984). Category-specific semantic impairments. *Brain*, 107, 829–854.
- Warrington, E. K., and Taylor, A. M. (1978). Two categorical stages of object recognition. *Perception*, 7, 695–705.
- Wernicke, C. (1886). Die neuen Arbeiten über Aphasie. *Fortschritte der Medizin*, 4, 371–377.
- Zingeser, L. B., and Berndt, R. S. (1988). Grammatical class and context effects in a case of pure anomia: Implications for models of language production. *Cognitive Neuropsychology*, 5, 473–516.

Chapter 14

The Language-to-Object Perception Interface: Evidence from Neuropsychology

Tim Shallice

Cognitive neuropsychology has as its principal aim the elucidation of the organization of the cognitive system through the analysis of the difficulties experienced by neurological patients with selective cognitive difficulties. As far as the relation between vision and language is concerned, the area that has been most extensively investigated concerns the semantic representation of objects. By contrast, the relation between how representations of space are accessed from vision and how they are accessed from language has been little touched; spatial operations have not been subject to much cognitive neuropsychology investigation.

If we consider objects, then the Gibsonian tradition teaches us that the richness of information available in the visual field is such that many of their properties may be inferred fairly directly from the visual array. Yet there are many other aspects of the visual world that cannot be inferred from the information in the visual field alone—the structural aspects of an object that are hidden from the present viewpoint, the potential behavior of an object and of the other objects likely to be found in its vicinity or that go with it in some other way. There are also wider properties of an object that may be accessed such as the perceptual features it has when experienced through other modalities, how it is used and by whom, what its function is, what types of thought process it triggers, and what intentions it may help to create. How are the processes involved in accessing these properties of an object when it is presented visually related to the way they are accessed when it is presented verbally?

This issue has been the subject of considerable controversy in cognitive neuropsychology in recent years for two reasons. A number of striking syndromes seem to relate very directly to it. In addition, the theory that most directly reflects the surface manifestations of the disorders differs from the standard theory in other fields where the issue has been addressed.

A model widely referred to in this book and in current cognitive science is that of Jackendoff (1987). Language is viewed as involving three main types of representation—phonological structures, syntactic structures, and semantic/conceptual structures.

As far as the semantic/conceptual structures are concerned, meanings have internal organization built up from a set of primitives and principles of combination, one of the primitives being the entity "thing." However, in addition to its phonological, syntactic and conceptual structures the representation of a word may contain specifically visual structures. The visual structures involved are, however, explicitly identified with the 3-D structural description level of Marr (1982).

Although Jackendoff's theorizing was concerned specifically with words and their meanings, the issues it addresses and in particular its position on the organization of the cognitive systems mediating semantic processing are closely related to issues recently much debated by cognitive neuropsychologists. A topic on which there has been much cognitive neuropsychology research in recent years is whether the semantic systems accessed when a word is being comprehended are the same as those used in the identification of an object, given that its structural description has already been determined. Some cognitive neuropsychologists have argued that they are the same, but others have claimed that they differ at least in part.

Approaches closely related to Jackendoff's have been adopted by certain cognitive neuropsychologists (e.g., Caramazza, Berndt, and Brownell 1982; Riddoch and Humphreys 1987). The best developed current neuropsychological account of a theory of this type is the organized unitary content hypothesis (OUCH) of Caramazza et al. (1990), which utilizes a feature based theory of semantic representations. More specifically, it holds that "access to a semantic representation through an object will necessarily privilege just those perceptual predicates that are perceptually salient in an object". Thus while many elements of the semantic representation are as easily accessible from visual as from verbal input, some aspects of the semantic representation are more easily accessed from its structural description than from its phonological representation. Access properties can be asymmetrical. The authors' rationale for assuming an asymmetric relation derives from consideration of certain conditions to be discussed shortly.

There is an older tradition in neuropsychology, however, which can be traced back at least as far as Charcot (1883) and Wernicke (1886). Certain syndromes suggest that visually based knowledge may be partly separable from verbally based knowledge. This perspective has been explicitly adopted more recently by a group of neuropsychologists (e.g., Warrington 1975; Beauvois 1982; Shallice 1987; and McCarthy and Warrington 1988) using the terminology *visual semantics* and *verbal semantics*, although the conceptual basis of the two types of representation has not been clearly articulated (see Caramazza et al. 1990; Rapp, Hillis, and Caramazza 1993; and Shallice 1993).

An intermediate position has been advocated by Bub et al. (1988) and by Chertkow and Bub (1990). Following Miller and Johnson-Laird (1976), they argue that a spe-

cific stage intervenes between attaining the structural description and accessing the amodal “core concept” of an object. Accurate identification of object is held to require more than just a characterization of an object’s structure, but must involve criteria which are more functional than structural. They therefore argue for the existence of a subsystem that contains only the application of the *functional* and perceptual criteria necessary for object identification, receiving the output from the structural description system and sending output to the core amodal semantic system. Thus “visual semantics” is reduced very considerably in its scope.

We thus have one position in cognitive neuropsychology (Caramazza et al. 1990) that is entirely compatible with Jackendoff’s perspective in holding that there is a single semantic/conceptual system. In addition it, namely the Caramazza et al. perspective, holds that accessing certain aspects of the semantic representation can be easier from the structural description than from phonology. Two other positions, (Warrington 1975; Chertkow and Bub 1990) hold that Jackendoff’s view is too gross a characterization of the subdivisions of the cognitive system involved in semantic processing, and that more than one semantic/conceptual system exists. A fourth position, which has yet to be formally articulated, holds that semantic representations are processed through a connectionist network of which different regions are more specialized for different types of semantic subprocess, but neither subprocess nor region can be characterized in an all-or-none fashion (see, for example, Allport 1985; Shallice 1988a).

Two main types of syndrome have been used to argue that the semantic-conceptual system is not in fact unitary but contains a number of types of subsystem—those involving some form of category specificity, and the modality-specific aphasias, in particular, optic aphasia. I will review the evidence from each in turn and then relate them to the alternative theories. A third syndrome—selective progressive aphasia—will also be addressed.

14.1 Category Specificity

The first group of syndromes responsible for the plausibility of the position that the semantic system is not unitary but composed of a number of subsystems are those manifesting so-called category specificity. The performance of the patient for some categories of knowledge is far better than for others. Of particular relevance is the syndrome originally described in four patients with herpes simplex encephalitis (Warrington and Shallice 1984). These patients had a selective problem in identifying animals, plants, and foods, while being able to identify man-made artefacts much better. For example, one of these patients, J.B.R., could name only 6% of living things and 20% of foods but could name 54% of man-made objects. Moreover, if the

judges assessed whether a description of a line drawing of the object "grasped the core concept," the contrast was even greater (living things, 6%; foods, 20%; but man-made objects, 80%). A similar effect was found when the patient was asked to give the meaning of the object's name and this, too, was assessed as to whether the core concept was grasped (living things, 8%; foods, 30%; man-made objects, 78%).

Similar effects have now been obtained with other patients with the same etiology (Pietrini et al. 1988; Sartori and Job 1988; Silveri and Gainotti 1988; Laurent et al. 1990; Swales and Johnson 1992; Sheridan and Humphreys 1993; Sartori et al. 1993; De Renzi and Lucchelli 1994). However, in the last few years there have been a rash of claims that these dissociations are essentially a result of characteristics of the stimulus set rather than evidence for a particular type of underlying organization of the semantic system.

Funnell and Sheridan (1992) initially claimed that the dissociations might arise because words matched for word frequency as used, say, by Warrington and Shallice (1984) may not be matched for visual familiarity. Indeed, McCarthy and Shallice (see Warrington and Shallice 1984) had shown that living things were less familiar to subjects than artefacts when matched for word frequency. Warrington and Shallice (1984) had dealt with this problem by showing that the dissociations were still present when differences in familiarity were taken out as a covariate. Moreover this explanation does not account for the way that the impairment of the patients involved foods as well as living things, as McCarthy and Shallice found foods to be more familiar than artefacts when word frequency is controlled.

A stronger argument was presented by Stewart, Parkin, and Hunkin (1992), who found that the category-specific dissociation of a herpes simplex patient, H.O., disappeared when word frequency, familiarity, and visual complexity were *all* controlled simultaneously. However, the basic dissociation, while statistically significant, was much weaker in H.O. than in some of the patients described earlier. Moreover, the nonliving category included objects like swamp, geyser, volcano, and waterfall instead of being composed solely of artefacts. Most critically, Sartori, Miozzo, and Job (1993) used stimuli matched on these three variables with their patient Michaelangelo, who showed a clear and significant category-specific effect of artefacts over living things on two different stimulus sets (living things, 30% and 40%; artefacts 70% and 76%).

Yet another possible artifact has been suggested by Gaffan and Heywood (1993), who argued that a critical variable was the density of exemplars within a category, which they held to be greater for living things than for artefacts. Because living things are more similar to each other and so less discriminable than artefacts, any discriminability problem would have a greater effect in the category of living things.

Riddoch and Humphreys (1987) had made a similar point previously and shown that there was more overlap between line drawings of animals than between line drawings of artefacts.

Gaffan and Heywood buttress their position on the difficulty in discriminating between living things, as opposed to artefacts, by considering the identification performance of three groups of subjects using the Snodgrass and Vanderwart (1980) stimuli. The first group were two patients of Farah, McMullen, and Meyer (1991), who showed standard category-specific effects; the second were normal subjects, who, however, were given only a 20 ms exposure; and the third used six monkeys, who were tested on how well they could decide which of two presented items was in a previously trained set. All three groups of subjects in their very different tasks showed an advantage of man-made objects over living things.

Gaffan and Heywood (1993) argue "These results from monkeys are contrary to Warrington and Shallice's conjecture ... that a specific system for identification of man-made objects has evolved in the human brain; if Warrington and Shallice's conjecture were correct, monkeys would show relatively greater difficulty in discriminating among inanimate objects than among living things, compared to human observers." It is not apparent, however, how such a comparison can be made because the tasks carried out were so different. Moreover, for the monkeys, most of the stimuli would presumably be meaningless objects; therefore what should be critical would indeed be raw discriminability. If, however, discriminability were a key factor underlying the performance of both the monkeys and the patients, then one would expect a positive correlation within each of the living and nonliving sets of stimuli between the results of the two group of subjects. In fact, there was no correlation between the items the monkeys found difficult and those the patients found difficult in either the living or the nonliving sets.

Gaffan and Heywood's work, like that in the other critical studies, used the Snodgrass and Vanderwart (1980) stimuli, for which norms are available on a number of relevant variables. In this set of stimuli the animals, in particular, tend to be rather similar to other members of their category. Warrington and Shallice (1984), however, also used the so-called Ladybird stimuli, large clear colored pictures designed for preschool children, with three of their patients. Shallice and Cinan have obtained ratings of structural complexity, familiarity, and discriminability from normal subjects for the Ladybird stimulus set and used these to reanalyze the findings of Warrington and Shallice. With these ratings, no difference was found between all three categories of stimuli (animals, artefacts, foods) for either familiarity or discriminability, but the animals remained structurally more complex than the other two categories. Because the task the patients carried out with this stimulus set had involved

word-picture matching using a four-alternative forced-choice task, the relevant degree of discriminability on the Gaffan-Heywood hypothesis was that within each set of five; this is what the subjects of Shallice and Cinan rated. However, with these stimuli two of the three original Warrington and Shallice patients on whom the test had been used performed significantly more poorly on foods than on artefacts with the third showing a strong trend in the same direction. Moreover, on a regression analysis using the ratings obtained by Shallice and Cinan, all three patients showed a significant effect of category and no effect of the other three variables. Thus it would appear that these category specificity findings cannot just be reduced to some combination of differences in word frequency, visual familiarity, structural complexity, and within-category discriminability.

In this respect, the work of Shallice and Cinan corroborated an earlier finding of Farah, McMullen, and Meyer (1991), who used the Snodgrass and Vanderwart (1980) stimuli with two patients exhibiting the standard category-specific dissociations. In a regression analysis on picture recognition performance, Farah, McMullen, and Meyer showed that neither name frequency, name specificity, similarity to other objects, structural complexity, nor object familiarity had any significant effect. The only factor to have such an effect was category membership. The absence of a significant effect of other factors in the presence of a significant effect of category makes implausible even one final convoluted artifactual explanation put forward by Gaffan and Heywood (1993). These authors suggested that the category difference arises through performance on items differing in a way dependent upon some other dimension; following Snedecor and Cochran (1967), they pointed out that *measurement* errors on the other dimension can lead to an apparent difference in performance across categories even when the differences on the other variables are allowed for as a covariate. However, what would then be expected is that there would be a basic effect of some other dimensions; this was not in fact found in either study.

Thus it would appear that the basic category-specific effects cannot be reduced just to an artifact of some combination of differences in word frequency, visual familiarity, structural complexity, and within-category discriminability across categories. A second type of finding that supports the conclusion that all neuropsychological dissociations in this domain cannot simply be attributed to some artifact of differences in presemantic factors is the existence of the complementary phenomenon, namely a superior performance in some subjects of living things (and in two studies foods) over artefacts (Warrington and McCarthy 1983, 1987; Hillis and Caramazza 1991; Sacchett and Humphreys 1992). The first two studies involved global aphasics who could only be tested by word-picture matching using, for instance, the Ladybird stimuli discussed above. However, the subjects in the last two studies were not glob-

ally aphasic; thus naming to visual confrontation could be used (for instance, C.W. in Sacchett and Humphreys 1992 scored 19/20 on naming animals; but only 7/20 on naming artefacts). Interestingly, the location of C.W.'s lesion (left frontoparietal) differed from that characteristic of the herpes simplex encephalitis cases (for all of whom the left temporal lobe was involved).

Much the most plausible conclusion is that the category-specific effects do not arise at a presemantic level due to some difference in difficulty between the categories but reflect some qualitative difference in the semantic representations of the categories. When the herpes encephalitis syndrome was first described, it was explained in terms of a contrast between stimuli primarily differentiable in terms of their sensory qualities and those more saliently differentiable in terms of their function.

Unlike most plants and animals, man-made objects have clearly defined functions. The evolutionary development of tool using has led to finer and finer functional differentiations of artefacts for an increasing range of purposes. Individual inanimate objects have specific functions and are designed for activities appropriate to their function. Consider, for instance, chalk, crayon, and pencil; they are all used for drawing and writing, but they have subtly different functions. . . . Similarly, jar, jug, and vase are identified in terms of their function, namely, to hold a particular type of object, but the sensory features of each can vary considerably. By contrast, functional attributes contribute minimally to the identification of living things (e.g., lion, tiger, and leopard), whereas sensory attributes provide the definitive characteristics (e.g., plain, striped, or spotted). (Warrington and Shallice 1984, 849)

A closely related position was taken to explain the complementary syndrome to be discussed later (see Warrington and McCarthy 1983.)

Dector, Bub, and Chertkow (in press) take a somewhat related position based on their study of a patient, E.L.M., who suffered from bilateral temporal lobe strokes. On tests of perceptual knowledge of objects he performed normally, but he was grossly impaired at many tests involving the perceptual characteristics of animals. Dector, Bub, and Chertkow argue that the difference between the superiority of artefacts over animals arises because different tokens of the same man-made object may show a considerable variation in the shape of its parts but a consistent function that allows for a unique interpretation, thus echoing the Warrington-Shallice position. However, they then argue that artefacts "can be uniquely identified at the basic level through a functional interpretation of their parts" and this is why they are relatively preserved (see De Renzi and Lucchelli 1994 for a related position). Many artefacts with a unique function do indeed have a unique organization of distinctly functioning parts; take a lamp, for example. However, others, such as a table tennis ball, do not. As yet it remains unclear to what extent the relative sparing of artefacts depends upon their unique organization of distinctly functioning parts or on the unique functions of the whole.

14.2 Sensory Quality and Functional Aspects of Different Categories

The position just developed attributes differences in performance across different categories to the way that identification in some categories depends critically on sensory quality information but for others functional information is more critical. One can, however, consider how well different semantic aspects of the same category are understood by patients who show this category-specific pattern. When this is done, knowledge of functional aspects of biological categories tends to be much better preserved than knowledge of sensory quality aspects (Silveri and Gainotti 1988). In a related fashion, Dector, Bub, and Chertkow's (in press) patient E.L.M. was much better at answering "encyclopedic" questions about animals such as "Does a camel live in the jungle or the desert?" (85%) than visual ones such as "Does a camel have horns or no horns?" where he was at chance (55%). However, the effects are not completely clear-cut. The performance of E.L.M., say, on functional aspects of animals was still well below that of normal controls, who scored 99%. This was not due just to a general problem with carrying out semantic operations on concrete objects; when asked to identify artefacts he performed at ceiling.

A more dramatic example is given by De Renzi and Lucchelli's (1994) herpes encephalitic patient, Felicia. In explaining the perceptual difference between pairs of animals, for example, goat and sheep, or paired fruits or vegetables, for example, cherry and strawberry, she performed far worse than the worst controls (15% vs. 90%; 49% vs. 85%). However, in explaining the *visual* difference between paired objects, for example, lightning rod and TV antenna, she was somewhat better than the normal mean (90% vs. 85%). Analogous results have been reported in a number of other studies (e.g., Silveri and Gainotti 1988; Sartori and Job 1988; Farah et al. 1989) although at least one patient, Giuletta (Sartori et al. 1993), answered nonvisual questions about animals almost perfectly (see also Hart and Gordon 1992), while at the other extreme S.B. (Sheridan and Humphreys 1993) performed almost as poorly on visual as on nonvisual questions about animals (70% and 65%, respectively).

Why should the category-specific impairment generally recur if in a milder form when the patient is responding to questions about animals or foods which appear not to be based on accessing sensory qualities? Does it not undermine the explanation of category-specific effects outlined earlier, namely, that they arise from damage affecting sensory quality representations? If one articulates the theory developed thus far in a connectionist form, then the problem can be resolved. Farah and McClelland (1991) investigated a model (see figure 14.1) in which some semantic units represented the functional roles taken by an item, while others represented its visual qualities. Each of the semantic units was connected (bidirectionally) to the others, to units representing structural descriptions, and to units representing phonological word-

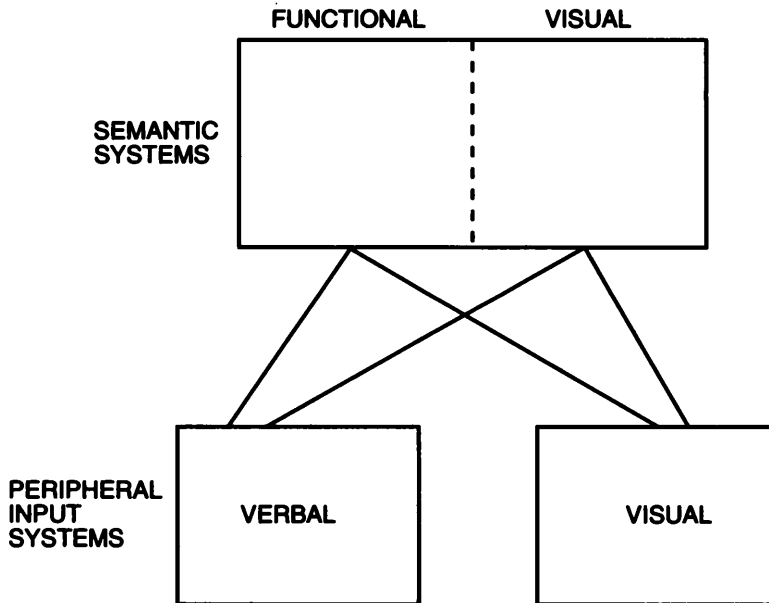


Figure 14.1

Farah and McClelland's (1991) model for explaining category-specific preservation of artefact comprehension and naming (reproduced by permission from Farah and McClelland 1991).

forms. The number of units in the two subsets of semantic representations was determined through an experiment on normal subjects. Subjects rated the description of each item in definitions of both living and nonliving things in the *American Heritage Dictionary* as to whether it described the visual appearance of the item, what the item did, or what it was for. On average there were 2.13 visual descriptions and 0.73 functional ones, but the ratio between the two types was 7.7:1 for living things and only 1.4:1 for the nonliving things. These values were then realized in the representations of living things and artefacts used for training. The network was trained using an error correction procedure based on the delta rule (Rumelhart, Hinton and McClelland 1986) applied after the network had been allowed to settle for ten cycles following presentation of each input pattern. In each of four additional variants of the basic network, one particular parameter was altered so as to establish the robustness of any effect obtained.

The most basic finding was that lesioning the "visual" semantic units led to greater impairment for living things than for artefacts with the opposite pattern shown for the lesioning of the functional semantic units. Thus the standard double dissociation was obtained due to "identification" of living things relying more on the visual

semantic units and "identification" of artefacts depending more on the functional semantic units. More interestingly, if one examines how close a match occurs over the *functional* semantic units when a lesion is made to the *visual* semantic units, then there is a difference between the two types of item. The *functional* representations of the living things were less adequately retained than those of artefacts. In the original learning process, the attainment of the representation in one of the semantic "sub-systems" helps to support the learning of the complementary representation in the other; the richer the representation is in one of the systems, the more use is made of it in learning the complementary representation. Thus the most typical relation between functional and visual impairments with living things is explained. Whether the full range of relations observed can be explained remains to be investigated.

There are two uncomfortable findings that the model would appear well designed to explain. First, the living/nonliving distinction is not absolute. Thus Y.O.T. was one of the global aphasic patients who performed very much better on word-picture matching with living things and foods than with artefacts (Warrington and McCarthy 1987). In Y.O.T.'s case, the impairment did not extend to large man-made objects such as bridges or windmills. Patient J.J. of Hillis and Caramazza (1991), who had a selective sparing of animal naming like Y.O.T., also had the naming of means of transportation spared. Complementarily, the problems of herpes encephalitic patients extended to gemstones and fabrics. The semantic representation of all these subcategories may well consist more of visual units than of functional ones, especially if function has to be linked with a specific action.

Second, the living/nonliving distinction is graded. Thus patients have been described in whom the deficit is limited, say, to animals alone (e.g., Hart, Berndt, and Caramazza 1985; Hart and Gordon 1992). The sensory quality/function contrast would seem likely to be more extreme for animals than foods, say, so that for more minor damage to sensory quality units only the least functional of the semantic categories would be affected.

Overall, this group of category-specific disorders fits with the idea that knowledge of the characteristics of objects is based on representations in more than one type of system. Realizing the different systems as sets of units in a connectionist model allows certain anomalies in the basic subsystem approach to be explained. The nature of the representations mediated by each of the systems remains unclear, however. The deficit appears not to correspond simply to damage to visual units. Thus one of the patients studied by Warrington and Shallice (1984) was unable to identify foods by taste as well as by sight. Moreover, in three of the patients where it has been assessed (Michaelangelo in Sartori and Job 1988; E.L.M. in Dector, Bub, and Chertkow, in press; and S.B. in Sheridan and Humphreys 1993), relative size judgments could be

made fairly accurately, suggesting that even the visual deficit does not extend to all visual characteristics. The issue remains open.

14.3 Optic Aphasia

A second syndrome that suggests the need to refine the conceptual/structural description contrast of Jackendoff (1987) is optic aphasia. First described by Freund (1889), optic aphasia refers to an impairment where the patient is unable to name objects presented visually but at the same time gives evidence of knowing what these objects are, for instance, by producing an appropriate mime. Moreover, the problem is not just one of naming; the patient is able to produce the name to a description or on auditory or tactile presentation. A considerable number of patients have been described who roughly fit the pattern (see Beauvois 1982; Iorio et al. 1992; Davidoff and De Bleser 1993 for reviews). If one limits consideration to patients who do not appear to have any impairment in accessing the structural description because stimulus quality does not affect naming ability, Davidoff and De Bleser (1993) list fourteen patients who have been formally described. Certain of these patients performed perfectly in gesturing the use of visually presented stimuli they could not name (Lhermitte and Beauvois 1973; Caplan and Hedley-White 1974; Gil et al. 1985).

This apparent preservation of knowledge of the visually presented object when it cannot be named has been explained most simply by assuming that the optic aphasic suffers from a disconnection between “visual semantics” and “verbal semantics,” with the name only being accessible from verbal semantics (Beauvois 1982; Shallice 1987). The distinction between subsystems at the semantic level appears to differ from the one drawn in the previous section between systems representing functional and visual or sensory quality types of information. I will address this issue in more detail later. In any case, a number of authors have contested the claim (see Riddoch and Humphreys 1987; Garrett 1992; Rapp, Hillis, and Caramazza 1993), holding that the miming could simply be based on an affordance, that is, an action characteristically induced by the shape of the object, or a cross-modal association of sensory and motor schemas, either of which might in turn be based only on an intact structural description. Alternatively, miming might require accessing only restricted parts of the semantic system, in particular those parts most strongly realized from the structural description because they are also represented in it explicitly, for example, the tines of forks; this is the privileged access theory account of Caramazza et al. (1990) and Rapp, Hillis, and Caramazza (1993). A similar explanation might also be given for the preserved drawing from memory shown in patients such as J.F. (Lhermitte and Beauvois 1973).

However, access to other types of information can be present in these patients when they cannot name. For instance, Coslett and Saffran (1992) gave their patient EM2 a task based on one devised by Warrington and Taylor (1978) in which the patient has to judge which of three items are functionally similar, for example, zipper, button, coin (see also patient C.B. in Coslett and Saffran 1989). EM2 scored at 97% on this task, with the control mean being 94%. Because the affordances of a zipper and a button are not similar, it is difficult to see how the use of affordances might be the basis for this good performance; indeed, there are no subcomponents of the two structural descriptions that are related. Rapp, Hillis, and Caramazza (1993), in confronting the argument that such a pattern of performance presents a difficulty for their privileged access position (Shallice 1993), merely respond by saying, "difficulty naming visually presented items in the face of demonstrated intact comprehension of some aspect of the visual structures, however, indicates that the full semantic description required to support naming has not been activated from a 3-D representation of the stimulus." This argument presupposes that normal performance on the function-matching test can be obtained when activation of the relevant semantic representation is reduced. This claim is merely asserted by Rapp, Hillis, and Caramazza. However, because the task is a three-alternative forced-choice test, with rather basic semantic information being required about each item—concerning its function—the assertion has some plausibility.

Similar results have, however, been obtained by Manning and Campbell (1992) on patient A.G. on semantic tasks which appear to be much more demanding. Two types of test were used with these patients. The first was the Pyramids and Palm Trees test of Howard and Patterson (1992). In a typical item of this test, the patient has to decide which tree (palm, fir) goes best with a pyramid. The stimuli can be presented either visually, verbally, or in mixed visual-verbal format. In the second test, the patient has to answer sets of questions about each item, (e.g., What is it made of?) both when the item is presented visually and when it is presented auditorily. A.G. performed at only 40%–50% in naming objects from drawings, but at 100% in naming to description and at 91% in naming tactilely presented stimuli, thus showing a specific naming defect with visual stimuli. However, A.G.'s performance on the Pyramids and Palm Trees test, while not at ceiling, was virtually identical across the visual and verbal modalities of presentation (82% vs. 84%) and in both cases was within one standard deviation of the mean of normal control subjects. A similar pattern was observed for the question-answering test (88% vs. 91%). Druks and Shallice's (1995) patient L.E.W. behaved in the same way for both types of test. That patients showed no difference and were not at ceiling on tests of auditory and verbal comprehension seems impossible to account for in Rapp, Hillis, and Caramazza's

(1993) version of the privileged access theory, which involves a unitary semantics. By contrast, these results fit well with the multiple semantic system position.

Coslett and Saffran (1992), on the other hand, present an interesting variant of the multiple store position. They agree that two semantic stores do exist and that one is disconnected from the language production mechanisms in optic aphasic patients, but they argue that the stores are primarily distinguished by hemisphere, with the right-hemisphere semantic system being disconnected from the language production systems in the left hemisphere. However, the patients described by Manning and Campbell (1992) present a difficulty for this position. In the acute condition immediately after a sudden onset lesion (e.g., vascular), the right hemisphere is supposed by right-hemisphere theorists such as Coslett and Saffran not to have access to any phonological lexicon, although they hold that over time a phonological lexicon becomes available to a semantic system in the right hemisphere (Coslett and Saffran 1989). This semantic system or the variety of output phonological word-forms that can be accessed from it is then seen to have an effective content corresponding to that of the words readable in deep dyslexia (Coltheart 1980a; Saffran et al. 1980; Coslett and Saffran 1989). In deep dyslexia, however, concrete nouns can be read reasonably well but verbs present severe problems (Coltheart 1980b). Yet while patients A.G. and L.E.W. were severely impaired in naming objects, which they could identify nonverbally, they could name actions very well. Thus A.G. was 95% correct at naming actions—the same level as controls—but worse than 50% at naming objects. This contrast in ease of accessing output phonological word-forms from an intact semantic representation is the opposite of what would be expected according to the right-hemisphere theory, where one would assume that objects should be more easily nameable than actions. The basic multiple semantic store position can perhaps explain the obtained effect by assuming the existence of another semantic subsystem—one controlling actions (Druks and Shallice 1995); being an essentially high-level output system but accessible from perceptual input, it would have connections to verbal semantics distinct from those used by the visual semantic representations of objects. This, however, remains a highly speculative account.

There remains one other counterintuitive aspect of optic aphasia. Many of the patients characterized as optic aphasic through their pattern of success and failure on naming and comprehension tests exhibit a strange set of errors when they fail to name correctly. Of the optic aphasic patients reviewed by Iorio et al. (1992), who generally correspond with Davidoff and De Bleser's (1993) group 2 optic aphasics, nearly all made both semantic and perseverative errors, with less than half also making visual errors. Moreover, in the most detailed analysis of such errors—that of Lhermitte and Beauvois (1973) of their patient J.F.—the authors consider the interaction between

Table 14.1
Errors Made by J. F. in Two Experiments

Type of error	Example	100 pictures	30 objects
<i>Horizontal errors</i>			
Semantic	<i>shoe</i> ⇒ "hat"	9	3
Visual	<i>coffee beans</i> ⇒ "hazel nuts"	2	1
Mixed visual-and-semantic	<i>orange</i> ⇒ "lemon"	6	1
<i>Vertical errors</i>			
Item and coordinate perseveration	T26 ... ⇒ "wristwatch"		
	T27 <i>scissors</i> ⇒ "wristwatch"		
	T44 ... ⇒ "newspaper"		
	T45 <i>case</i> ⇒ "two books"	8	2
<i>Mixed horizontal/vertical errors</i>			
	T43 ... ⇒ "chair"		
	T47 <i>basket</i> ⇒ "cane chair"		
	T53 <i>string</i> ⇒ "strand of weaved cane"	3	0

Source: Lhermitte and Beauvois 1973.

what they call "horizontal errors," understood strictly in terms of the processes (temporally) intervening between presentation of the stimulus and the responses, and what they call "vertical errors," where effects of preceding stimuli or responses occur. It is clear from this analysis that the perseverative and the semantic errors combine in a complex way (see table 14.1).

Why might such a strange combination of errors be characteristic of optic aphasia? Again a possible answer can be given by adding a connectionist dimension to the models. Plaut and Shallice (1993a) considered a network which had a direct pathway mapping visual representations into semantic ones. It also had a "cleanup" pathway that involved recurrent connections from the semantic units to the "cleanup" units and back (see figure 14.2). The network used an iterative version of the backpropagation learning algorithm known as backpropagation through time (Rumelhart, Hinton, and Williams 1986). Training with an algorithm of this type in such a recurrent network leads to its developing a so-called attractor structure; the effect of the operation of the cleanup pathway is to move a noisy first-pass representation at the semantic level toward one of the representations it has been trained to produce as an output, given that the initial representation is in the vicinity of the trained one.

The network contained one other major difference from other networks well known in cognitive psychology, such as Seidenberg and McClelland's (1989). In

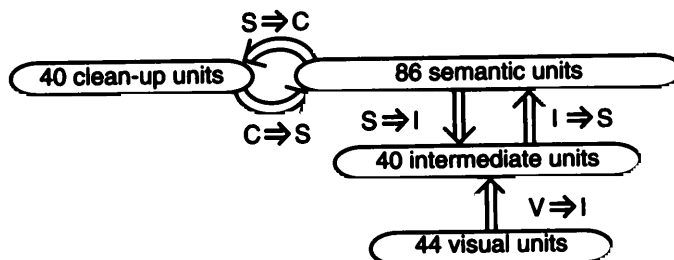


Figure 14.2

Plaut and Shallice's (1993) model for explaining the typical error pattern found in optic aphasia (reproduced from Plaut and Shallice 1993a by permission).

the nervous system, changes in synaptic efficiency at a single synapse occur at many different time scales (Kupferman 1979). The incorporation of additional connection weights that change much more rapidly in training than those standardly used in connectionist modeling is also computationally valuable; it allows for temporal binding of neighboring elements into a whole (e.g., von der Marlsburg 1988) and facilitates recursion (Hinton, personal communication described in McClelland and Kawamoto 1986). Each connection in the network therefore combined a standard, slowly changing, long-term weight with a rapidly altering, short-term weight based on the correlation between the activities of its input and output units.

A network having both types of weights tends to reflect in its behavior both its long-term reinforcement history and its most recent activity; it contains the analogue of both long-term learning and of priming. The network was trained to respond appropriately at the semantic level to the structural representations of forty different objects. Wherever the network was lesioned, it produced a few visual errors but considerably more semantic errors and typically more with both visual and semantic similarity to the stimulus. More critically, there was a strong perseverative aspect to the responses. The previous response or one of its semantic associates could well occur as an error to the present stimulus. This corresponds well to the error pattern occurring in optic aphasia.

Adding a connectionist dimension to the model therefore allows the error pattern of the syndrome to be explained. The information-processing model we used as a basis for the connectionist simulations corresponds to those of Riddoch and Humphreys (1987) and Caramazza et al. (1990), which were held to be unsatisfactory earlier in this chapter. However, the essence of the simulation is that if short- and long-term weights are combined, the errors will reflect both perseverative influences and the level of representation at which strong attractors occur.¹ Thus the obtained error pattern would also be expected if an analogous connectionist dimension were

added to the multiple semantic system models, provided that one or more of the semantic systems had analogous attractor properties.

14.4 Conclusion

In the sections 14.1 and 14.2 certain syndromes were discussed involving category-specific impairments, particularly those associated with herpes simplex encephalitis, where large differences in performance exist between identification of man-made artefacts on the one hand and of living things and foods on the other. Explanations in terms of differences between the categories on a number of potentially confounding dimensions were considered and rejected. The favored explanation assumes that partially separable systems underlie the semantic representations of the functional and of the sensory quality properties of stimuli. In section 14.3 another syndrome—optic aphasia—was considered; here it was argued that the most plausible explanation involved disconnecting “visual” and “verbal” or “lexical” semantic representations.

The evidence presented in all three sections poses difficulties for the view that a single conceptual system, together with a structural description system that can also be addressed from above, is a sufficient material base for representing semantic operations. The sensory quality component of the semantic system cannot be conflated with the structural description system because variables relevant to disorders of the latter system, for example, presentation of items from unusual views (Warrington and Taylor 1978), do not predict the stimuli that are difficult for patients with impairments to the former system (Warrington and Shallice 1984; Dector, Bub, and Chertkow in press). The issue is even clearer from the perspective of the second set of disorders. In certain optic aphasic patients much more semantic information appears to be accessible from vision than could be based on the structural description alone; yet it would appear not to be available in a generally accessible conceptual system because it cannot be used to realize naming.

By contrast, the accounts presented for these disorders fit naturally with those beginning to be developed within developmental psychology for image schemas at a level of abstraction higher than the structural description and yet not simply subsumable within verbal knowledge (see Mandler, chapter 9, this volume). However, to argue that the such visual semantic processes should be limited to what is required for visual identification alone—in Chertkow and Bub’s (1990) visual identification procedure subsystem—and that this is the only system lying between the structural description system and an amodal core semantic system does not fit well for either syndrome. In the herpes encephalitis condition what is lost are the sensory quality aspects of the item, while identification procedures, according to Miller and Johnson-Laird (1976), require primarily functional property information as well as structural

analysis. Turning to optic aphasia, one possibility to explain the syndrome might be to view it as arising from a disconnection between the visual identification procedures and the core semantic system. However, a task like Pyramids and Palm Trees involves the utilization of shared context. The Bub and Chertkow theory holds that inferred context is stored in the amodal core semantic system, so that an optic aphasic would be not expected to perform well on such tasks for words that could not be named. Patients A.G. (Manning and Campbell 1992) and L.E.W. (Druks and Shallice 1995) show the opposite pattern, namely, intact performance on this task, together with grossly impaired naming.

There are, however, certain problems in explaining the two types of syndrome in terms of the functional/sensory quality and visual/verbal dichotomies. The concepts are orthogonal. The information available in a visually or sensory quality-based semantic system, as inferred by the information lost in the herpes encephalitic patient is not the only information *accessible* from the visual modality in the optic aphasic patient. Certain optic aphasic patients, for example, A.G. and L.E.W., can access types of information from vision that would be in the functional or encyclopedic parts of the semantic system on a simple all-or-none multiple store view. Moreover, within the semantic dementia literature there are striking echoes of this visual input predominance extending outside the purely sensory quality domain in the performance of patient T.O.B. (McCarthy and Warrington 1988).² When a picture was presented to T.O.B. his identification was more than 90% accurate for both types of material, but he identified verbal input artefacts much better than living things (89% vs. 33%). Thus when the word *dolphin* was presented, the patient could say only, "A fish or a bird," but when presented with the picture, he said, "Lives in water ... they are trained to jump up and come out. ... In America during the war they started to get this particular animal to go through to look into ships." McCarthy and Warrington have argued that this patient has an impairment that affects the stored information itself rather than an input pathway because of the consistency with which particular items were or were not identifiable (see for rationale Warrington and Shallice 1979; Shallice 1987). Thus contrasting both optic aphasia and semantic dementia with herpes simplex encephalitis, it would appear that the putative lines of cleavage within the semantic system suggested by the syndromes differ.

One possibility is to postulate category-specific systems that are themselves specific to particular modalities (McCarthy and Warrington 1988). However, explanations provided for certain secondary aspects of the syndromes suggest an alternative direction in which a more economical solution might lie. A connectionist simulation of Farah and McClelland (1991) can account for certain otherwise most recalcitrant findings about category-specific disorders. For optic aphasia, the counterintuitive error pattern associated with the disorder is in turn explicable on a connectionist

simulation of Plaut and Shallice (1993a). Thus adding a connectionist dimension to the theoretical framework used to account for the characteristics of the syndromes enables a much fuller explanation of the detailed nature of the deficits to be provided. Adding such a connectionist dimension to a subsystem approach provides an account closely related to presimulation suggestions made over the last ten years or so, that the semantic system has as its material basis a large associative neural network with different concepts being represented in different combinations of its subregions, depending on the specific subset of input and output systems generally used to address them (see Allport 1985; Warrington and McCarthy 1987; Shallice 1988b; and Saffran and Schwartz 1994). How the rule-governed aspects of semantic processing would be dealt with on this type of account has not been addressed by neuropsychologists. However, the use of a connectionist network framework for explaining aspects of neuropsychological disorders does not preclude the possibility of explaining rule-governed aspects of semantic processing, provided additional elements are added to the basic network (see Touretsky and Hinton 1988; Derthick 1988; and Miikkulainen 1993). On this account the semantic/conceptual system postulated by Jackendoff would need to be realized as a complex neural network. As yet, though, no implementation adequately explains the rich and highly counterintuitive evidence that detailed study of individual neurological patients provides.

Notes

1. This is especially the case if the mapping from the visual to the semantic level is not orthogonal, as it is in language (see Plaut and Shallice 1993a); for visual presentation of objects, the visual and the semantic representations are correlated.
2. A simple peripheral explanation of the phonological word-form being damaged can also be excluded.

References

- Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. K. Newman and R. Epstein (Eds.), *Current perspectives in dysphasia*. Edinburgh: Churchill Livingstone.
- Beauvois, M. F. (1982). Optic aphasia: A process of interaction between vision and language. *Philosophical Transactions of the Royal Society, London*, B298, 33–47.
- Bub, D., Black, S., Hampson, E., and Kerkesy, A. (1988). Semantic encoding of pictures and words: Some neuropsychological observations. *Cognitive Neuropsychology*, 5, 27–66.
- Caplan, L., and Hedley-White, T. (1974). Cueing and memory dysfunction in alexia without agraphia: A case report. *Brain*, 97, 251–262.

- Caramazza, A., Berndt, R. S., and Brownell, H. H. (1982). The semantic deficit hypothesis: Perceptual parsing and object classification by aphasic patients. *Brain and Language*, 15, 161–189.
- Caramazza, A., Hillis, A. E., Rapp, B. C., Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology*, 7, 161–189.
- Charcot, J. W. (1883). Un cas de suppression brusque et isolée la vision mentale des signes et des objets (formes et couleurs). *Progrès Medical*, 11, 568–571.
- Chertkow, H., and Bub, D. (1990). Semantic memory loss in dementia of Alzheimer's type. *Brain*, 113, 397–417.
- Coltheart, M. (1980a). Deep dyslexia: A right hemisphere hypothesis. In M. Coltheart, K. E. Patterson, and J. C. Marshall (Eds.), *Deep dyslexia*. London: Routledge.
- Coltheart, M. (1980b). Deep dyslexia: A review of the syndrome. In M. Coltheart, K. E. Patterson, and J. C. Marshall (Eds.), *Deep dyslexia*. London: Routledge.
- Coslett, H. B., and Saffran, E. M. (1989). Preserved object recognition and reading comprehension in optic aphasia. *Brain*, 112, 1091–1110.
- Coslett, H. B., and Saffran, E. M. (1992). Optic aphasia and the right hemisphere: Replication and extension. *Brain and Language*, 43, 148–161.
- Davidoff, J., and De Bleser, R. (1993). Optic aphasia: A review of past studies and a reappraisal. *Aphasiology*, 7, 135–154.
- Dector, M., Bub, D., and Chertkow, H (in press). Multiple representations of object concepts: Evidence from category-specific aphasia. *Cognitive Neuropsychology*.
- De Renzi, E., and Lucchelli, F. (1994). Are semantic systems separately represented in the brain? The case of living category impairment. *Cortex*.
- Derthick, M. (1988). Mundane reasoning by parallel constraint satisfaction. Ph.D. diss., Carnegie Mellon University, Pittsburgh.
- Druks, J., and Shallice, T. (1995). Preservation of visual identification and action naming in optic aphasia. Paper presented at the Annual British Neuropsychological Society Conference, London, March.
- Farah, M. J., Hammond, K. H., Mehta, Z., and Ratcliff, G. (1989). Category specificity modality specificity in semantic memory. *Neuropsychologia*, 27, 193–200.
- Farah, M. J., and McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120, 339–357.
- Farah, M. J., McMullen, P. A., and Meyer, M. M. (1991). Can recognition of living things be selectively impaired? *Neuropsychologia*, 29, 185–194.
- Freund, D. C. (1889). Über optische Aphasie und Seelenblindheit. *Archiv für Psychiatrie und Nervenkrankheiten*, 20, 276–297.

- Funnell, E., and Sheirden, J. (1992). Categories of knowledge? Unfamiliar aspects of living and nonliving things. *Cognitive Neuropsychology*, 9, 135–153.
- Gaffan, D., and Heywood, C. A. (1993). A spurious category-specific visual agnosia for living things in normal human and nonhuman primates. *Journal of Cognitive Neuroscience*, 5, 118–128.
- Garrett, M. (1992). Disorders of lexical selection. *Cognition*, 42, 143–180.
- Gil, R., Pluchon, C., Toullat, G., Michenau, D., Rogew, R., Lefevre, J. P. (1985). Disconnexion visuo-verbale (aphasie optique) pour les objets, les images, les couleurs, et les visages avec alexie abstractive. *Neuropsychologia*, 23, 333–349.
- Hart, J., Berndt, R. S., and Caramazza, A. (1985). A category-specific naming deficit following cerebral infarction. *Nature*, 316, 439–440.
- Hart, J., and Gordon, B. (1992). Neural systems for object knowledge. *Nature*, 359, 60–64.
- Howard, F., Patterson, K. E. (1992). *Pyramids and palm trees: A test of semantic access from pictures and words*. Thames Valley.
- Iorio, L., Falango, A., Fragassi, N. A., and Grossi, D. (1992). Visual associative agnosia and optic aphasia: A single case study and a review of the syndromes. *Cortex*, 28, 23–37.
- Jackendorff, R. (1987). On beyond zebra: The relation of linguistic and visual information. *Cognition*, 26, 89–114.
- Kupferman, I. (1979). Modulatory actions of neurotransmitters. *Annual Review of Neuroscience*, 2, 447–465.
- Laurent, B., Allegri, R. F., Michel, D., Trillet, M., Naegle-Faure, B., Foyatier, N., and Pellat, J. (1990). Encéphalites herpétiques à prédominance unilatérale: Étude neuropsychologique au long cours de 9 cas. *Revue Neurologique*, 146, 671–681.
- Lhermitte, F., and Beauvois, M. F. (1973). A visual-speech disconnection syndrome: Report of a case with optic aphasia, agnosic alexia, and colour agnosia. *Brain*, 96, 695–714.
- Manning, L., and Campbell, R. (1992). Optic aphasia with spared action naming: A description and possible loci of impairment. *Neuropsychologia*, 30, 587–592.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- McCarthy, R. A., and Warrington, E. K. (1988). Evidence for modality-specific meaning systems in the brain. *Nature*, 334, 428–430.
- McClelland, J. L., and Kawamoto, A. H. (1986). Mechanisms of sentence production: Assigning roles to constituents of sentences. In J. L. McClelland and D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 2, 272–325. Cambridge, MA: MIT Press.
- Miikkulainen, R. (1993). *Subsymbolic case-role analysis of sentences with embedded clauses*. Technical report AI 93-202. Austin: University of Texas Press.
- Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge: Cambridge University Press.

- Pietrini, V., Nertempi, P., Vaglia, A., Revello, M. G., Pinna, V., and Ferro-Milone, F. (1988). Recovery from herpes simplex encephalitis: Selective impairment of specific semantic categories with neuroradiological correlation. *Journal of Neurology Neurosurgery and Psychiatry*, 51, 1984-1293.
- Plaut, D. C., and Shallice, T. (1993a). Perseverative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience*, 5, 89-117.
- Plaut, D. C., and Shallice, T. (1993b). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10, 377-500.
- Rapp, B. C., Hillis, A. E., and Caramazza, A. (1993). The role of representations in cognitive theory: More on multiple semantics and the agnosias. *Cognitive Neuropsychology*, 10, 235-249.
- Riddoch, M. J., and Humphreys, G. W. (1987). Visual object processing in a case of semantic access agnosia. *Cognitive Neuropsychology*, 3, 201-224.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1. Cambridge, MA: MIT Press.
- Sacchetti, C., and Humphreys, G. W. (1992). Calling a squirrel a squirrel but a canoe a wigwam: A category-specific impairment for artefactual objects and body parts. *Cognitive Neuropsychology*, 9, 73-86.
- Saffran, E. M., Bogyo, L. C., Schwartz, M. F., and Marin, O. S. M. (1980). Does deep dyslexia reflect right hemisphere reading. In M. Coltheart, K. E. Patterson, and J. C. Marshall (Eds.), *Deep dyslexia*. London: Routledge.
- Saffran, E. M., and Schwartz, M. F. (1994). Of cabbages and things: Semantic memory from a neuropsychological perspective. In C. Umiltà and M. Moscovitch (Eds.), *Attention and performance*, vol. 15, 507-536. Cambridge, MA: MIT Press.
- Sartori, G., and Job, R. (1988). The oyster with four legs: A neuropsychological study of the interaction of visual and semantic information. *Cognitive Neuropsychology*, 5, 105-132.
- Sartori, G., Job, R., Miozzo, M., Zagu, S., and Machiori, G. (1993). Category-specific form-knowledge deficit in a patient with herpes simplex encephalitis. *Journal of Clinical and Experimental Neuropsychology*, 15, 280-299.
- Sartori, M., Miozzo, M., and Job, R. (1993). Category-specific naming impairments? Yes. *Quarterly Journal of Experimental Psychology*, 46a, 489-504.
- Seidenberg, M. S., and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Shallice, T. (1987). Impairments of semantic processing: Multiple dissociations. In R. Job, G. Sartori, and M. Coltheart (Eds.), *The cognitive neuropsychology of language*. London: Erlbaum.
- Shallice, T. (1988a). *From neuropsychology to mental structure*. New York: Cambridge University Press.

- Shallice, T. (1988b). Specialization within the semantic system. *Cognitive Neuropsychology*, 5, 133–142.
- Shallice, T. (1993). Multiple semantics: Whose confusions? *Cognitive Neuropsychology*, 10, 251–261.
- Sheridan, J., and Hymphreys, G. W. (1993). A verbal-semantic category specific recognition impairment. *Cognitive Neuropsychology*, 10, 143–184.
- Silveri, M. C., and Gainotti, G. (1988). Interaction between vision and language in category-specific semantic impairment. *Cognitive Neuropsychology*, 3, 677–709.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical methods*. 6th ed. Ames: Iowa State Press.
- Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174–215.
- Stewart, F., Parkin, A. J., and Hunkin, N. M. (1992). Naming impairment following recovery from herpes simplex encephalitis: Category-specific? *Quarterly Journal of Experimental Psychology*, 44a, 261–284.
- Swales, M., and Johnson, R. (1992). Patients with semantic memory loss: Can they relearn lost concepts? *Neuropsychological Rehabilitation*, 2, 295–305.
- Von der Marlsburg, C. (1988). Pattern recognition by labeled graph matching. *Neural Networks*, 1, 141–148.
- Warrington, E. K. (1975). The selective impairments of semantic memory. *Quarterly Journal of Experimental Psychology*, 27, 635–657.
- Warrington, E. K., and McCarthy, R. (1983). Category-specific access dysphasia. *Brain*, 106, 859–878.
- Warrington, E. K., and McCarthy, R. (1987). Categories of knowledge: Further fractionation and an attempted integration. *Brain*, 110, 1273–1296.
- Warrington, E. K., and Shallice, T. (1979). Semantic access dyslexia. *Brain*, 102, 43–63.
- Warrington, E. K., and Shallice, T. (1984). Category-specific semantic impairments. *Brain*, 107, 829–854.
- Warrington, E. K., and Taylor, A. M. (1978). Two categorical stages of object recognition. *Perception*, 7, 695–705.
- Wernicke, C. (1886). Die neuen Arbeiten über Aphasie. *Fortschritte der Medizin*, 4, 371–377.
- Zingeser, L. B., and Berndt, R. S. (1988). Grammatical class and context effects in a case of pure anomia: Implications for models of language production. *Cognitive Neuropsychology*, 5, 473–516.

Chapter 15

Space and Language

Mary A. Peterson, Lynn Nadel, Paul Bloom, and Merrill F. Garrett

15.1 Introduction

Functioning effectively in space is essential to survival, and sophisticated spatial cognitive systems are evident in a wide range of species. In humans, the emergence of language adds another level of complexity to the organization of spatial cognition. We use language for many purposes, not least of which is the conveying of information about where important things are located (food, safety, enemies) and how to get to and from these places (for discussion of these evolutionary issues, see O'Keefe and Nadel 1978, Pinker and Bloom 1990). Given the fundamental nature and importance of spatial cognition, it is of considerable interest to determine the ways in which it connects to language. The hope that study of such connections might shed light on both the spatial cognitive faculty and on the language faculty has generated considerable interest in the domain of "language and space."

We are interested in how people talk about space and what they can and do choose to say about it. By exploring the boundaries of these cognitive domains, we hope to uncover their structure and to elucidate the ways in which they can relate to one another. By considering the role of development and culture in shaping the language-space interaction, we hope to discover the extent to which fundamental aspects of spatial cognition are given *a priori* and the extent to which spatial cognition can be altered by experience. And by analyzing the ways in which neural systems organize spatial and linguistic knowledge, we hope to shed light on how these two capacities relate to one another.

In the present chapter we analyze what we take to be the consensually accepted framework within which the relations between language and space have been considered. Based on this framework, we critically discuss some influential proposals as to the precise nature of this relationship. Finally, we return to the set of issues and questions with which we began, and reach some tentative conclusions.

15.2 Framework

The framework we adopt here for how we talk about space is based on the proposal by Jackendoff (1983, 1987), who took Fodor's (1975) "language of thought" hypothesis as a starting assumption. Fodor argued convincingly that one cannot learn a language unless one already has an original language to structure the learning process; he referred to this original language as the "language of thought." The language of thought includes the building blocks from which our concepts are constructed.

Extending Fodor's analysis, Jackendoff has argued for something along the lines of the situation represented in figure 15.1. There exist language representations (LRs), spatial representations (SRs), and conceptual representations (CRs). LRs include all aspects of language structure, including the lexicon and the grammar; SRs include all aspects of spatial structure as it is represented in the brain; and CRs are primitives that form the components of meaning, both linguistic (CR_L) and spatial (CR_S) mean-

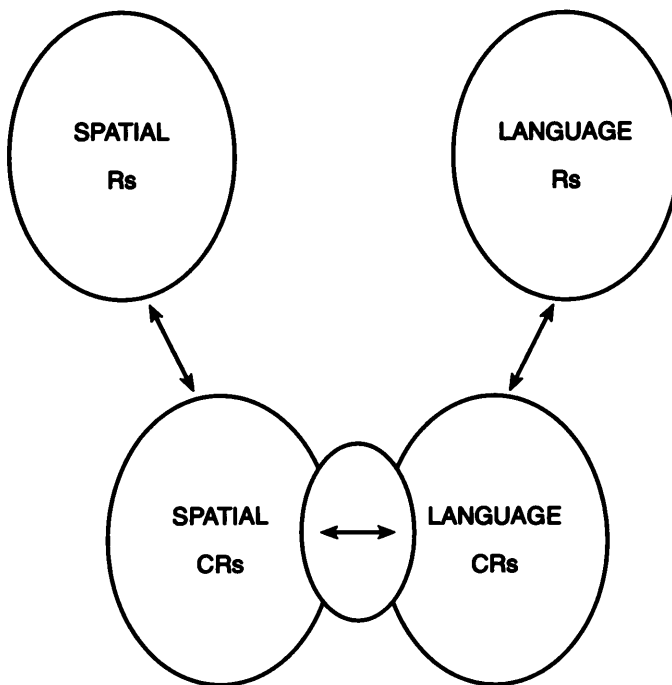


Figure 15.1

A schematic depiction of Jackendoff's analysis of the relationship between language representations (Language Rs), spatial representations (Spatial Rs), and conceptual representations (CRs) of both language and space.

ing. It is by virtue of some interface between CR_L and CR_S that we can talk (using LR) about space (SR). We can leave open whether this interface corresponds to a set of mappings between two distinct systems or to an actual shared conceptual representation. In either case, as shown in figure 15.1, it is likely that only certain aspects of CR_L and CR_S can participate in this interface. That is, it is likely that some aspects of spatial meaning cannot be expressed linguistically, just as some aspects of language do not correspond to spatial notions. If we accept this view, then it follows that the study of the “language of space” cannot fully illuminate our faculty of spatial cognition, some aspects of which we would not, according to this view, be able to talk about. Nevertheless, given the necessity of some interface, Jackendoff had the insight that one might gain substantial knowledge about the nature of at least some of the spatial CRs by analyzing linguistic spatial terms.

Jackendoff’s framework thus makes obvious the following four questions: (1) Which aspects of space can we talk about and which not, and why might this be so? (2) Which aspects of language reflect particular spatial attributes and which do not? (3) Are spatial CRs changed by linguistic experience? (4) What light can the study of space and language shed on the nature of conceptual representations? The answers to these questions invite analyses from a number of different perspectives, many of which are represented by the authors in this volume.

15.3 Space and Spatial Information

What do we mean by “space”? It is clear that space can contain objects and events, but it need not. Empty space, and unoccupied places, exist. O’Keefe and Nadel (1978, 86) note that “the notions of place and space are logical and conceptual primitives which cannot be reduced to, or defined in terms of, other entities. Specifically, place and spaces are not, in our view, defined in terms of objects or the relations between objects.” On the other hand, objects and events both occupy spatial locations and have intrinsic spatial properties. Objects are partly defined by the spatial relations among their parts, and events are partly defined by the spatial relations among the various entities (e.g., objects, people, actions) that compose them. Are the logical distinctions among objects, events, and spaces maintained in language, and are they paralleled by dissociable neural representations? In order to determine which aspects of spatial representation are transparent to language, it would be useful to know something about the various ways in which space is represented in the mind/brain. Is there a single multimodal or amodal representation of space per se? Or are there a number of independent modules for spatial representation? Can the study of these representations help us understand space, language, and the interface between the two, and can it shed light on the nature of the “spatial” primitives in CR?

15.3.1 Evidence from Studies of the Brain

Available neurobiological evidence suggests that a relatively large number of distinct representations or “maps” of space and spatial information exist. For example, various investigators have discussed maps of motor space, auditory space, visual space, and haptic space; maps of body space, near space, far space; maps of egocentric space and allocentric space; and, maps of categorical space and coordinate space. Quite separate representations for the spatial features of objects and their locations have also been investigated. Neuroscientists have linked a variety of brain structures and systems to one or another of these spatial representations, providing converging evidence for some degree of independence of many of these forms of spatial information.

Neurobiological evidence also suggests the existence of multimodal spatial representations. It is known, for example, that the mammalian superior colliculus integrates sensory spatial maps and motor maps, such that nearby neurons are activated either by sensory inputs from, or motor outputs directed to, a particular part of egocentric space (cf. Gaithier and Stein 1979; Meredith and Stein 1983). Indeed, this integrative system is quite primitive phylogenetically; comparable integration of sensory spatial maps occurs in snakes, where visual and thermal “spatial” maps are brought into register in the superior colliculus (Newman and Hartline 1981)

There is also evidence demonstrating the presence of multimodal sensory spatial maps in various areas of the mammalian neocortex, including especially parts of the parietal cortex (Pohl 1973; Kolb et al. 1994). All of these multimodal maps seem to share the critical feature of representing space egocentrically, that is, with reference to the organism, or some part of the organism (e.g., hand, eye, head, torso). The “maps” in these multimodal brain association areas, as well as those in unimodal regions such as the visual and somatosensory cortex, are all laid out in topographic fashion, such that neighboring regions of neural space represent neighboring regions of the ego-centered world.

In addition to these data demonstrating the existence of various unimodal and multimodal ego-centered spatial maps, there is evidence for a superordinate allocentric amodal or multimodal spatial representation that subsumes or somehow integrates the spatial representations (SRs) provided by each of the various spatial maps. It is now well established that the vertebrate hippocampus subserves a spatial mapping function that is both multimodal and allocentric; that is, external space is represented independent of the momentary position of the organism, in terms of the relations between objects and the places they occupy in what appears to be an objective, absolute framework (O’Keefe and Nadel 1978; and see O’Keefe, chapter 7, this volume). This system contains information about place, distance, and direction. The “place” cells first identified by O’Keefe (e.g., O’Keefe and Dostrovsky 1971; O’Keefe

1976; O'Keefe and Nadel 1978) are active when an animal is in a given location in space, as defined by the relationship between that spatial location and other places in the environment. O'Keefe and Nadel (1978) have postulated that information about distance is provided to the hippocampus via the septal region, and a pattern of activation termed *theta*, driven by inputs from brain stem movement systems. More recently, Taube and others (Taube 1992; Taube, Muller, and Ranck 1990a,b) have described a population of "head direction" neurons in the dorsal subiculum and thalamus of the rat; these cells are active when an animal faces in a particular direction in the environment, whatever its specific location. Existing data show that the place cells and the head direction cells are tightly linked together (Knierim et al. 1993). The representation of allocentric space created in the hippocampus uses multimodal information from cortical systems including the parietal and temporal regions, as well as inputs conveying egocentric information about directions and distances.

Interestingly, this allocentric spatial representation is not neurally topographic in the way the egocentric representations are. As far as we can tell, neighboring regions of hippocampal neural space do not represent neighboring regions of the external world. While this hippocampal system has been postulated to provide the basis for certain spatial primitives (e.g., *places*), it does not appear to be necessary for a wide range of nonallocentric spatial representations, such as those subserved by the superior colliculus and neocortical regions already noted.

The spatial maps in the superior colliculus, parietal cortex, and hippocampus are thought to represent space without regard for the exact nature of the objects occupying any part of the represented space. The spatial (and other) properties of objects appear to be captured in separate neural systems. Thus considerable neurobiological evidence suggests the existence of two streams of visual processing about objects: a ventral pathway, incorporating regions of the temporal lobe that is concerned with what an object is, and a dorsal pathway, incorporating regions of the parietal cortex that is concerned with where an object is located with respect to the organism (e.g., Ungerleider and Mishkin 1982). Neuropsychological investigations of brain-damaged individuals have been taken as support for this "what" versus "where" distinction in the object representation system, as in the well-known cases of "blindsight," where subjects express a lack of awareness for the presence and nature of an object, all the while demonstrating by their behavior that they know "where" the object is located (e.g., Weiskrantz 1986). However, recent evidence indicates that the ventral and dorsal processing streams are not nearly as isolated as originally conjectured (Van Essen, Anderson, and Fellman 1992). Nor is the neuropsychological evidence for independent streams completely convincing; some neuropsychologists now argue for notions like degrees of modularity (for review, see Shallice 1988). Whatever the status of these visual processing streams, they both provide inputs to

the hippocampal system, presumably contributing to its ability to construct the allocentric representation discussed above.

The evidence from studies of the brain thus suggests that (1) there are a variety of spatial maps in the brain, which makes it unlikely that there is a single amodal spatial representation that gives rise to the entire set of spatial primitives; (2) at least some neural representations of space do not include detailed representations of objects, reflecting the logical distinction between environmental space and the spatial aspects of objects discussed above; and (3) there is some, but not total, separation within the systems representing objects between those representing what an object is, and those representing where it is located.

15.3.2 Evidence from Studies of Perception

Behavioral evidence is consistent with the idea of a variety of spatial maps. For example, consider the elegant study conducted by Loomis et al. (1992), who showed observers two targets located at different distances from the observer in an open field. The observers performed two tasks with respect to these targets. In one task, observers used a matching response to report about the perceived distance between the two targets at different distances (i.e., they adjusted the apparent horizontal distance between two objects located at a standard distance to match the apparent depth interval between the two test targets). In the second task, observers viewed the display from the same stationary vantage point used for the first task and, closing their eyes, walked first to one distal target and then to the other (the targets were removed once the observers closed their eyes and began walking). Thus the distance between two distal objects that had been visually apprehended from the same vantage point as used in the first task was motorically expressed in the second task, once the observers had walked to the location of the first object.

Loomis et al. (1992) found a dissociation between these two different estimates of the distance between the two objects, with the walking responses suggesting a more veridical perception of distance than the matching responses. The latter reflected the operation of organizing factors; the error in the perceived distance between the two distal objects appeared to increase systematically as a function of the distance from the observer's vantage point to the objects (this effect may be an instance of Gogel's 1977 equidistance tendency). On the other hand, no such increase as a function of distance from the stationary vantage point was observed when the interobject distance was assessed via walking responses. The experiment by Loomis et al. demonstrates that although certain distance representations are veridical, in that they can support accurate navigating to the two targets in turn, other representations of the distance between the two objects are systematically distorted by the operation of perceptual organizing factors.

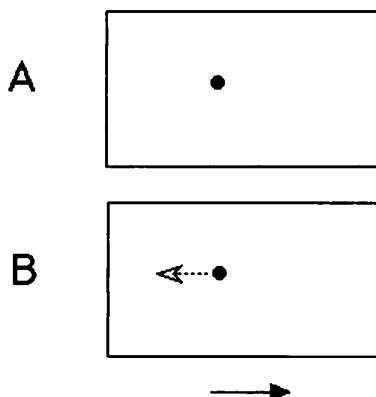


Figure 15.2

An illustration of the Duncker induced-motion display in which a small dot is enclosed in a large rectangle. A stationary display is shown in (A). A moving display is shown in (B), where the actual motion of the rectangle is indicated by a solid arrow pointing to the right, and the perceived motion of the dot is indicated by a dashed arrow pointing to the left.

Similarly, Bridgeman, Kirch, and Sperling (1981) demonstrated a striking dissociation between the location of a small target relative to an enclosing frame and its location relative to the observer. These investigators showed their observers a Duncker display, containing a small target enclosed within a larger rectangular frame, like that shown in figure 15.2. In displays such as this one, an “induced motion” illusion occurs when the frame is displaced abruptly in one direction, say, to the right. In a number of situations, abrupt displacements have been shown to mimic motion signals to the visual system, and consequently, to result in apparent motion. The unique characteristic of the Duncker induced-motion illusion is that observers perceive motion (or displacement) of the stationary dot, rather than of the displaced frame. For example, in the display shown in figure 15.2, rather than perceiving the displaced rectangular frame as moving to the right, observers perceive the stationary target as moving to the left, that is, in the direction opposite to the direction in which the frame was displaced.

The induced illusory motion of the small target is very compelling visually, and it can be canceled (i.e., the small target can be made to appear stationary when the frame is displaced to the right) by the addition of a real displacement of the small target in the same direction as the frame displacement (in this example, to the right). Thus, as with many other kinds of illusory motion (e.g., Gogel 1982; Peterson and Shyi 1988), induced motion and real motion add perceptually and may be indistinguishable.

Bridgeman, Kirch, and Sperling gathered two kinds of responses about the location of the small target in the Duncker display. One response was a cancellation response, as described. By this measure, and by self-report about what they saw, the observers in their experiment indicated perceiving the (actually stationary) target as having moved from its original location. The magnitude of the change in location inferred from the cancellation responses was about half the distance through which the frame had been displaced. On another block of trials, Bridgeman, Kirch, and Sperling asked the same observers to point, with an unseen hand, to the final perceived location of the target after viewing the induced-motion display, which disappeared from sight before they made the pointing response. Surprisingly, at least with respect to the cancellation responses given by the same subjects, the magnitude of the illusion measured by the pointing response was negligible. Under these conditions, observers pointed much closer to the actual location of the target than to the perceived location of the target (as inferred from the cancellation responses).

Thus the experiment revealed a distinction between the spatial representations mediating the cancellation response (and presumably, visually perceived location) and the spatial representations mediating the pointing response (and perhaps, motoric responses in general). It is clear that the visually perceived location reflects visual organizing factors—in this case, an organization that depends on the enclosing relationship between the frame and the target—whereas the representation of location accessed by the motor response seems relatively free of such effects. (We will return to this point below.) Although it may not be clear how best to characterize this distinction (see Loomis et al. 1992 for a lucid discussion), the results of Bridgeman, Kirch, and Sperling and those of Loomis et al. strongly imply that the maps mediating visually perceived spatial relationships differ from those mediating movement-expressed distances, locations, and/or directions.

In addition to the behavioral evidence for differential encoding of egocentric versus allocentric spaces, and for locations versus directions, there is evidence that spatial experience can reflect the combination of inputs from different modalities. For example, Lackner and his colleagues have shown that auditory, visual, and kinesthetic inputs regarding an observer's orientation in space are combined to yield a perceived space that does not correspond to the space signaled by any one input (for review, see Lackner 1985). The behavioral evidence is therefore consistent with the idea of multi-modal spatial representations, as well as with the idea of various unimodal spatial maps.

What do perceptual studies indicate about the independence of object and spatial representations? To a certain extent, the independence of these two systems has simply been assumed (see Marr 1982 and Wallach 1949 for explicit statements of this

assumption). Consistent with this assumption, object recognition does appear to exhibit location invariance (see Biederman and Cooper 1991 for recent evidence), and accurate distance perception is clearly possible for novel objects.

On the other hand, behavioral evidence has occasionally suggested that these two systems may influence one another. For example, Carlson and Tassone (1971) found that the perceived egocentric distance to objects in naturalistic settings is influenced by the familiarity of the objects. The initial experiments could not rule out a number of alternative explanations based on response tendencies or differences in the complexity of familiar and unfamiliar objects, but subsequent work excluded these possibilities (Predebon 1990). Similarly, object recognition may not be completely independent of location: recognition accuracy for individual objects located within contextually appropriate scenes is reduced when the objects are presented in inappropriate locations (e.g., a fire hydrant in a street scene is less likely to be recognized when it is located inappropriately on top of a mailbox than when it is located appropriately at street level; Biederman 1981).

Recent evidence indicates that figure-ground organization, which entails the perception of the relative distance between two adjacent regions in the visual field, is influenced by the familiarity (or recognizability) of the regions (Peterson 1994; Peterson and Gibson 1993, 1994). These findings have led Peterson and her colleagues to propose that a rapid object recognition process (a "what" process) operates before the determination of depth segregation (a classic "where" process) and that the former exerts an influence on the latter in combination with more traditional depth cues, such as binocular disparity (see Peterson 1994; Peterson and Gibson 1994). Similarly, Shiffrar and Freyd (1990; Shiffrar 1994) have shown that perceived direction of motion through space ("where") is constrained by the types of movements that are possible given the nature of the objects in motion ("what").

In addition to these effects of object identity on perceived spatial organization, there is evidence that another type of spatial information is fundamental for object identity. Object identification fails when the parts of the object are spatially rearranged (see, for example, Biederman 1987; Hummel and Biederman 1992; and Peterson, Harvey, and Weidenbacher 1991), and is delayed when a picture depicts an object misoriented with respect to its typical upright orientation (Gibson and Peterson 1994; Jolicoeur 1988; Tarr and Pinker 1989).

In sum, and consistent with conclusions drawn from neurobiological analysis, the study of perception shows that (1) a variety of independent modules for spatial representation exist; and (2) some representations deal with objects, some with spaces, and some with the interaction between the two. What can be said about how, if at all, language hooks up with each of these proposed modules?

15.3.3 Talking about Space and Spatial Relations

Taking the existence of independent spatial modules—some dealing with spaces, some dealing with objects, and some dealing with the interactions between the two—as a starting assumption, one can pose the following question. Does language express the information available in all, or only some, of these spatial modules? And are particular parts of language used to express specific forms of spatial information?

Landau and Jackendoff (1993; Jackendoff, chapter 1, this volume; Landau, chapter 8, this volume) have recently taken a modular position on the question of how language relates to representations of objects versus spatial relations between objects. Their position rests on a linguistic analysis that emphasizes the differences between the manner in which languages code for spatial relations and for objects. For example, in English, objects are described by nouns, which are open-class linguistic elements, whereas spatial relationships are described by prepositions, which are closed-class linguistic elements. The prepositions of English—and members of the corresponding grammatical class in other languages—may be specialized for speaking about space, as opposed to speaking about objects, in that they may be applied with few constraints to objects of different sizes and different sorts (e.g., see Talmy 1983). In addition, spatial relationships between objects tend to be coarsely coded by languages, at least in comparison to object identities. That is, spatial prepositions such as *near* and *far* would not support accurate motor behavior of the type studied by Bridgeman, Kirch, and Sperling (1981) and by Loomis et al. (1992). Landau and Jackendoff (1993) also stress the fact that the number of spatial prepositions in English (around 75) is quite small relative to the number of object names (30,000 or so, according to a count by Biederman 1987).

Landau and Jackendoff (1993) took these differences between prepositions and nouns as evidence that prepositions and nouns mapped onto different sorts of spatial representations. In particular, their proposal suggests that closed-class linguistic spatial terms might map to a subset of CR_ss that are about the spatial relations between objects, or between an observer and objects (i.e., relations that represent the locations of objects without regard for the specific properties of the objects occupying those locations), whereas nouns might enjoy a privileged mapping to a subset of CR_ss that are specialized for object representation (e.g., the 3-D object models of Marr 1982 or Biederman 1987). Landau and Jackendoff pointed out that the linguistic distinctions in the meanings of nouns and prepositions fit nicely with neurobiological and computational evidence indicating that “what” and “where” are represented independently (e.g., Ungerleider and Mishkin 1982; Rueckl, Cave, and Kosslyn 1989).

By incorporating modern research and theory about “what” and “where” systems, Landau and Jackendoff’s (1993) approach usefully builds on Jackendoff’s (1983) insight that we can learn about spatial conceptual representations by studying how

we talk about space. However, in our view, research programs that attempt simply to identify subdivisions of language with neural spatial systems will not be able to fully elucidate the nature of spatial conceptual representations (see also Bierwisch, Chapter 2, this volume). This follows from the fact that words express abstract conceptual notions that do not appear to be captured in any one-to-one fashion by sensory, perceptual, or neural representations. The literature on word learning illustrates this point.

In much of the discussion of the relationship between language and space, particularly by developmental psychologists, it is assumed that nouns are equivalent to object names and that objects are equivalent to entities generalized on the basis of shape. But neither of these equivalencies exists. For adults, only a minority of nouns refer to material objects; most nouns are like *day*, *family*, *joke*, *factors*, *information*, and so on. Nouns that do not refer to objects are also used, with appropriate syntax and meaning, by two- and three-year-old children, and there is considerable experimental evidence showing that children have no special problems learning such words (see Bloom 1994, in press). Even infants appear to possess CRs that do not correspond to objects: infants six months of age are capable of mentally representing, and counting, discrete sounds (Starkey, Spelke, and Gelman 1990) and individual actions, such as the jump of a puppet (Wynn 1995).

What about the claim that “just for nouns that do name objects,” a property such as shape, which can be derived from sensory/perceptual inputs, is criterial? Even this is too strong. Children learn superordinates, like *animal* and *furniture*, that refer to categories that share no common shape, relationship terms like *doctor* and *sister*, and functional terms like *clock* and *switch* (see Soja, Carey, and Spelke 1992 for further discussion). In fact, even for those objects that *do* have characteristic shapes, children know that shape is not criterial. If you alter a porcupine so that it has the shape of a cactus, three- and four-year-old children insist it remains a porcupine; they view ontological boundaries (e.g., you cannot transform an animal into a plant) as more significant than shape (Keil 1989).

Rather than assuming that nouns map directly onto a “what” system that encodes objects in terms of shape, an alternative outlined in Bloom (1994, in press) assumes that nouns map onto CRs that are nonspatial (and thus can include notions like *joke* and *day*). This is not to deny that shape is important for learning object names, as demonstrated by Landau and her colleagues (see Landau, chapter 8, this volume). In some cases, considerations of shape are relevant for determining category membership, implying an interface between CRs and the shape of an object. For instance, there is evidence that children and adults have an essentialist notion of natural kind concepts, so that the CR for porcupine is, roughly, “everything that has the same internal ‘stuff’ as previously encountered porcupines” (e.g., Putnam 1975; Keil 1989).

But since internal stuff—essence—is unobservable, we normally use an observable property that is highly correlated with essence—shape—to determine whether something is a porcupine. Shape also correlates with membership in certain functional kinds; given the purpose for which chairs are designed, they are likely to have a certain configuration (i.e., they are likely to have shapes that afford sitting).

As noted above, however, there are many words and semantic categories for which no such correlation with shape exists. Jokes and days have no shapes, doctors and fundamentalists do have typical shapes but not ones that distinguish them from lawyers and agnostics, and although categories like animals and furniture refer to entities with shapes, the entities forming the category all have different shapes. These considerations show that although there is a relationship between the category of nouns and the notion of object shape, it is not direct. Rather, it is mediated through a more abstract conceptual system of CRs. As a result, the link between language and the shapes of objects is, at least for open-class categories such as nouns, nowhere near as direct as many researchers assume it to be.

Similarly, it is clear that *spatial* terms cannot be derived simply from an interface between language and a set of sensory/perceptual maps. Consider what is actually conveyed by the spatial representation used to describe the relationship between the butterfly and the jar in figure 15.3A, captured by the following sentence:

The butterfly is in the jar.

In this description, the relationship described by the spatial preposition *in* cannot be reduced to one of mere surroundedness in the visual display. The butterfly in figure 15.3B is not *in* the tabletop, although the contours of the tabletop surround it, as the contours of the jar surround the butterfly in figure 15.3A; whereas the butterfly in figure 15.3C is correctly described as *in* the canyon, although the contours of the canyon do not surround the butterfly. Clearly, the meaning of the spatial preposition *in* cannot be defined by appealing to attributes of a sensory spatial representation only. Instead, one must appeal to some abstract relationship, such as a capacity for *containment* that jars and canyons share, but tabletops do not. The abstract notion of containment may be one of the conceptual representations linking space and language, and it may be by virtue of this CR that canyons and jars can be categorized similarly by language, but the notion of containment simply cannot be accounted for by complexes of sensed information (for discussion, see Bowerman, chapter 10, this volume, and Mandler, chapter 9, this volume).

Even though CRs will not map to complexes of sensory information, can similarities in the way linguistic terms and spatial maps are characterized help us identify an underlying isomorphism between a linguistic category such as prepositions and a particular spatial representation? For example, might the “categorizing” role played

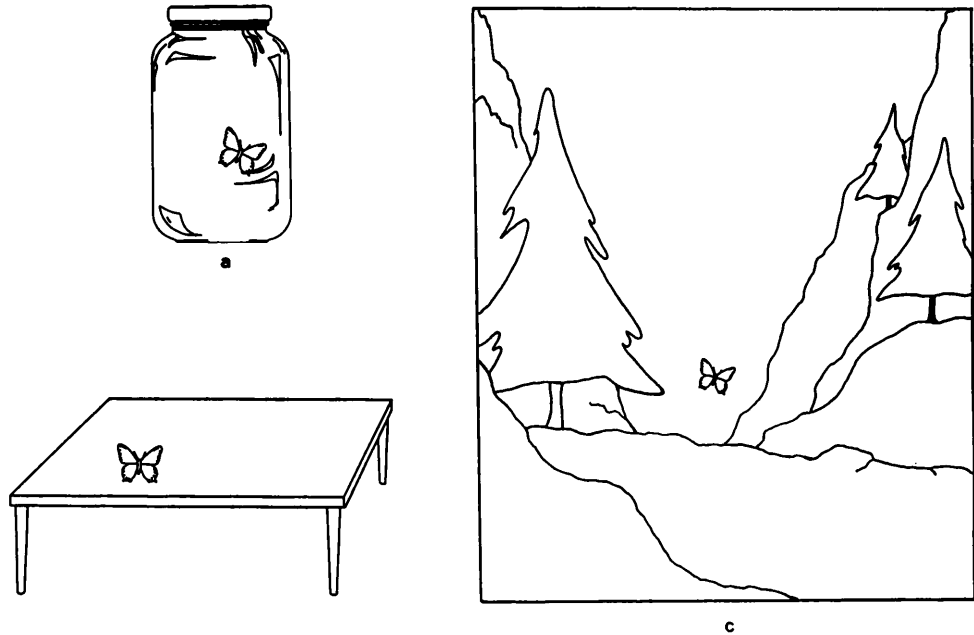


Figure 15.3

Demonstrations that the meaning of the spatial preposition *in* does not map simply to surroundedness in the visual display. The image of the butterfly is surrounded by the contours of both the jar in (a) and the table in (b), but not by the contours of the canyon in (c). Yet the spatial term *in* correctly describes the spatial location of the butterfly relative to the jar in (a) and the canyon in (c), whereas the spatial term *on* (rather than *in*) applies to the spatial location of the butterfly relative to the table in (b).

by the term *in* in the situations illustrated in figures 15.3A and 15.3C imply that spatial terms have a privileged mapping to “categorical” as opposed to “coordinate” spatial representations within the “where” system, such as those postulated by Kosslyn and his colleagues (Kosslyn et al. 1989; Kosslyn et al. 1992)? Such a solution would appeal to the surface similarity implied by the use of the term *categorical* in these two cases, but the similarity may not go beyond the surface. Categorical maps treat a set of spatial locations as equivalent; that is, these maps have low resolution. Yet terms such as *in* could as easily reflect high-resolution as low-resolution spatial representations.

Those situations in which spatial prepositional usage *can* depend upon the entities being related may be more revealing about the nature of conceptual representations than those situations that fit within a “what” versus “where” dichotomy. There are a number of examples in addition to the one illustrated in figure 15.3, showing that

nonspatial factors govern the use of spatial terms in English. For example, the considerable variability in applying the terms *front* and *back* to churches implicates functional factors as relevant to axial descriptions, given that the front of a church can be defined by functional factors (e.g., the direction people attending a service face) as well as by structural factors coded in an object representation. (See Vandeloise 1991 for a discussion of functional meanings of French prepositions). Current neuropsychological evidence, supported by computational evidence, indicates that functional representations may be critically important attributes of object meaning (see Shallice, chapter 14, this volume). Thus the nonspatial semantics of an object may govern prepositional usage. Other nonspatial semantic factors such as salience are relevant when we say whether something is “near” or “far”; these semantic factors are evident in some spatial memories as well (for review, see McNamara 1991).

A last example that language does not divide up in ways that map directly onto particular neural spatial representations can be found in linguistic directional terms. Directions are coded and lexicalized by various languages in deictic (egocentric) terms; in intrinsic (object-centered) terms; or in absolute (cardinal) terms. Does this variability imply that language interfaces directly with spatial representations of all these types, at least with regard to direction? Would this count as a case of direct linkage between particular linguistic elements and specific spatial representations?

Consider the fact that the terms *right/left* can be used to denote both the speaker's right/left (e.g., egocentric use), or the right and left of some other object or person. Yet spaces are apprehended egocentrically by an exploring animal, either by the act of moving through them or by the behavior of visually scanning them. In either case, the inputs are initially coded in terms of egocentric relations between the observer and entities such as places and objects in the environment. Spatial relations of distance and direction, for instance, and even of the arrangement of parts within an entity, are “computed” by the organism from its various egocentric inputs. Human factors work has demonstrated the primacy or importance of egocentric coding as well. For example, if one has to discriminate which of two adjacent objects on a display screen is brighter (larger, more familiar), the best response mapping is one in which a choice of the object on the right is indicated by a right-hand key press, and a choice of the object on the left by a left-hand key press. There is much evidence from developmental work showing that egocentric knowledge about spatial location precedes allocentric knowledge (e.g., Mangan and Nadel 1992; Wilcox, Rosser, and Nadel 1994). In addition, the neuropsychological deficit of neglect points to the primacy of egocentric right/left coding. Individuals who sustain damage to the right parietal lobe often “neglect” left hemispace. For example, when there are objects in their right hemispace, neglect patients ignore objects in their left hemispace (Heilman

1979; Volpe, LeDoux, and Gazzaniga 1979). Likewise, when neglect patients are asked to imagine a scene, well known to them before their brain damage, they are unable to imagine those objects that lie to their left from their imagined vantage point, and the objects omitted change as the imagined vantage point changes (Bisiach and Luzzatti 1978). Thus right/left egocentric relationships appear to be coded early in processing and appear to be critically important in spatial understanding.

Notwithstanding the evidence attesting to the importance of egocentrically based spatial information, many normal individuals have severe difficulties in mapping linguistic terms onto egocentric relationships. Part of the problem in using the terms *right/left* may arise because egocentric spaces depend on the direction the individual is facing; that is, the regions of space that lie to the right and left are interchanged by a 180° rotation. For a speaker and a listener who face each other, the space to the right of the speaker lies to the left of the listener. Even for speakers who evidence no overt difficulty in using the terms *right* and *left*, considerable effort is required to translate the frame of the speaker into that of the listener (see Tversky, chapter 12, this volume).

Those same individuals who have difficulties mapping linguistic terms onto egocentric relationships have no trouble in reaching to the right or left to catch an object falling off a table. This dissociation suggests that right/left terms do not map directly to motoric egocentric neural representations. Consistent with the possibility that the perspective taking evident in language use does not simply reflect the use of a perspectivized spatial map, Levelt (chapter 3, this volume) demonstrates that the spatial representations accessed for speaking about arrays of dots affording either an egocentric or an intrinsic description are not already coded for the egocentric or intrinsic directions speakers choose to express. In our view, the underlying spatial representations may instead be allocentric spatial representations such as those found in the hippocampus (see O'Keefe and Nadel 1978).

The discussion of linguistic evidence has thus far focused on objects and spatial relations, including distances and directions. We have yet to touch upon a critical aspect of space and that is *place*. How does language treat places? In contrast to the spatial relations of distance and direction, places are described by open-class elements rather than by closed-class elements. Some place names are count nouns, like *center*, *basement*, and *border*, and these can be extended to novel instances in much the same way as names for kinds of entities like ball, house, and country. Others are proper names, like Paris, Times Square, or the Equator, which behave much like the individual names Bill, Joan, and the Salvation Army, and are certainly as informative. We know of no count of the number of places that can be named in English to stack up against the number of spatial prepositions counted by Landau and Jackendoff (1993; $n = 75$) or against the number of object names estimated by Biederman (1987;

$n = 30,000$), but certainly the number of places that can be named is several orders of magnitude larger than 75.

This fact suggests to us a rather different way of imagining the relation between aspects of language such as open- and closed-class elements and aspects of spatial (and object) representations. Systems concerned with both space per se and with objects contain information about *entities* and about *relations between entities*. The linguistic evidence could be taken to suggest that nouns are used in the case of entities (be they places, objects, or other things—see below), and that prepositions are used in the case of relations (be they about places, objects, or other entities). In this view, prepositions are not limited to describing the spatial relationships between objects. Even putting aside more abstract usages of prepositions (as in “John went from rich to poor”; see, for example, Jackendoff 1990), sentences such as the following are perfectly acceptable:

The mist hovered *over* the sea.

John put the poison *into* the soup.

A swarm of bees flew *into* the forest.

There was an explosion *next to* my house.

Boston is *near* New York.

He swept the space *in front of* the fireplace.

In these examples, there is no problem whatsoever using prepositions to describe spatial relationships between and among substances, collections of objects, events, locations, and even empty space itself.

In sum, studies of language provide no reason to go beyond the basic framework spelled out at the outset, in particular, to propose privileged one-to-one mappings between parts of language and particular spatial representation systems. While it may be the case that the information carried by some spatial maps can be talked about, and the information carried by others cannot, we believe this is *not* a result of connections between specific types of language elements and specific spatial maps. This does not mean that we reject the idea that neurobiological and perceptual/cognitive research can shed light on the nature of the spatial conceptual representations. Rather, we suspect that investigations of the ways in which CRs interact with the maps identified by neurobiological and behavioral research will be fruitful in elucidating the nature of spatial conceptual representations. This leads us to the following modest proposal.

Some, but not all, of the spatial maps identified by neurobiological and behavioral research impose a structure that goes beyond, and in consequence alters, our interpre-

tation of the information available in the input alone. For example, the hippocampus appears to impose a Euclidean framework onto non-Euclidean inputs (O'Keefe and Nadel 1978, who see in this process the instantiation of a Kantian a priori notion of absolute space). Other examples are revealed by the organizing factors that structure some behavioral representations—factors like the equidistance tendency (Gogel, 1977; Gogel and Teitz 1977) and the constraints due to gravity identified by Shepard (for summary, see Shepard 1994). We propose that in “distorting” the sensory inputs, these spatial maps may impose an order and a structure that our spatial conceptual representations require. If this were the case, studies of language use and other spatial behaviors that revealed the operation of these organizing factors might lead to some understanding of the CRs themselves.

Before engaging in this sort of analysis it is important to first look at the ways in which the mappings between language, behavior and space vary across cultures. If there are internally imposed structures that reflect primitive spatial CRs, one would expect to see these structures preserved across cultural and linguistic boundaries. This follows from the assumption, central to our guiding framework, that the CRs are part of a universal “language of thought” that makes understanding of the world possible. If, on the other hand, spatial frameworks and perception itself can be shown to vary across culture, their utility as stable indicators of the nature of spatial CRs is questionable.

15.4 Effects of Experience

Speakers of Tzeltal code spatial relations with respect to absolute directions; they simply do not use egocentric terms to speak about space or objects (see Levinson, chapter 4, this volume). In this respect they differ from speakers of Dutch and English, for example. The critical feature of absolute directions is that they remain invariant as vantage point changes. In Tzeltal, the absolute directions that are used originate in a feature of the environment—uphill/downhill—are applied even when the geographical feature is out of sight. A tremendous amount of effort is required to keep track of the absolute directions; nevertheless, these directions seem to be well preserved in the memories of events and scenes experienced by the speakers of Tzeltal (Brown and Levinson 1993; Levinson, chapter 4, this volume). This certainly raises the question that led us to consider the effects of experience in the first place. Are there differences in the CR_s between speakers of Tzeltal and speakers of English, Dutch, or other languages that lexicalize egocentric relations rather than cardinal directions?

This possibility is difficult to address, but Levinson and his collaborators have shown that speakers of Tzeltal and speakers of Dutch behave differently in old/new

perceptual recognition tasks, problem-solving tasks, and memory tasks (Levinson, chapter 4, this volume). Furthermore, gestures employed by speakers recounting remembered scenes and events are different. The gestures employed by speakers of Tzeltal indicate absolute directions, and the gestures employed by speakers of Dutch and English indicate relative directions (Haviland 1993). Does this mean that the language one speaks, or the culture in which one lives, can change the *nature* of the underlying CR_s? Or does it support the less radical claim that the culture in which one lives, and the language one speaks, affects the *availability* of different CR_s because of differential degrees of practice utilizing them? And in either case, do such findings imply that the conceptual representations at the interface between language and thought are themselves different?

It seems clear that different languages and/or cultures can utilize different cognitive skills to different degrees. For example, Emmorey and her colleagues (see Emmorey, chapter 5, this volume) have shown that sign language may engage mental rotation skills, and consequently, may improve these skills due to practice. Of course, differential prowess at mental rotation does not imply that the CR_s are different. Nor is this implied by differences in performance on memory tasks and problem-solving tasks, such as those discussed by Levinson (chapter 4, this volume), although the existence of such differences provides evidence relevant to theories of perception, memory, and problem solving. For example, Levinson's finding (chapter 4, this volume) that differential encoding of absolute versus egocentric directions by speakers of Tzeltal and of Dutch, respectively, is evident in problem-solving tasks is consistent with psychological evidence that world knowledge (which differs from speaker to speaker) influences problem-solving behavior (Murphy and Medin 1985).

What about Levinson's claim that differential encoding of absolute directions versus egocentric directions is also evident in performance on old/new perceptual recognition memory tests? If the perceptual representations accessed, for example, by speakers of two languages *were* different, by virtue of the differential attention each had paid to particular aspects of the situation at encoding, that would be consistent with recent evidence that knowledge influences perception more than traditionally assumed (for summaries, see Peterson 1994; and Shiffrar 1994). Note that the effects of knowledge on perceptual organization may be highly constrained in that the relevant structural, semantic, or functional representations mediating such knowledge must be accessed within the normal time course of perceptual processing (see Carpenter and Grossberg 1987; Gibson and Peterson 1994; and Peterson, Harvey, and Weidenbacher 1991). Thus, if Levinson's findings are shown to reflect differences in perceptual representations *per se*, they might justify a search for correlates of absolute directions in perceptual input.¹

Alternatively, the tasks employed by Levinson may reflect differences in semantic representations between speakers of different languages and/or members of different cultures. It has always been supposed that different languages or different cultures might combine primitive CRs differently so that certain meanings are more or less salient to speakers of a given language (Bowerman 1989; Slobin 1995).² Neuropsychological findings discussed by Shallice (chapter 14, this volume), suggest that qualitatively different semantic representations may be accessed in the course of identifying artifacts and living things (see also Farah and McClelland 1991). Similarly, semantic representations of remembered scenes and events could vary in their emphasis on absolute or egocentric directions, depending upon one's culture and experience. Thus, while Levinson's results might mean that semantic representations are different for speakers of different languages (see also Bowerman, chapter 10, this volume), they would not entail that the primitive CRs themselves, from which the semantic representations are constructed, have been changed by language (or culture).

It is important not to overemphasize the differences between speakers of different languages: it is clear that spatial cognition is not necessarily constrained by the language that one knows. For example, speakers of languages that do not habitually employ absolute direction terms can do so when these terms are suited to the task (see Tversky, chapter 12, this volume). In addition, speakers of Tzeltal may use egocentric (or deictic) relations, especially when these are not overshadowed by absolute direction relations (Brown and Levinson 1993). More generally, it follows as a matter of logic that some understanding of spatial relationships must be available prior to the acquisition of spatial language; otherwise, it would be impossible for spatial language to be acquired in the first place. While this leaves open the possibility that exposure to different languages can engage certain aspects of spatial cognition to a greater extent than others, it does not support the strongest Whorfian hypothesis that one's manner of thinking about space is entirely determined by the language one learns. (For more general discussion of this point, see Fodor 1975.)

The preceding discussion, and much of the evidence in this volume, implies that the exact mappings between CR_S and CR_L are plastic in that they can look rather different in different languages. Need this have implications for the structure of CR_S and CR_L themselves? The short answer is no. One need not assume that how one talks about one's spatial concepts necessarily influences those concepts in some fundamental way. What we can talk about is certainly less than what we know, and what we know consciously, whether we can state it precisely or not, is certainly less than what we know in toto. Cultures may influence how we choose to refer to spatial attributes, and even which spatial attributes we choose to refer to, but there is little support for the view that they, or the languages they use, fundamentally alter our spatial understanding of the world. Under the assumption that experience does not fundamentally

alter perceptual and cognitive processes, the study of intrinsic organizing factors should offer a window upon the underlying conceptual representations. By comparing how we use language to refer to space and spatial relations with how we behave in space, we can gain insight into some of these “distortions.”

15.5 Conclusions

At the outset we posed four central questions in the study of space and language: (1) Which aspects of space can we talk about? (2) Which aspects of language reflect particular spatial attributes? (3) Are spatial CRs changed by experience? (4) What light can the study of space and language shed on the nature of conceptual representations?

There are aspects of spatial knowledge we cannot naturally talk about (for example, absolute distance between two objects or between an observer and an object), and aspects of spatial knowledge we can talk about (for example, spatial relations), but we cannot at present provide a satisfying distinction between these two classes of spatial knowledge. Although the distinctions can be described by terms like *precise* and *coarse* spatial representations, we do not believe that those terms accurately express the CRs that might underlie the distinction. The suggestion by Landau and Jackendoff (1993) that nouns and prepositions describe objects and spatial relations, respectively, is an important start to the project of understanding how language maps to space, but we suspect that a broader view, namely, that nouns describe entities (including, but not limited to, places and objects) and prepositions describe relations, is likely to be closer to the truth. There is evidence that different cultures refer to space in different ways, but there is no reason to suppose this involves a change in the underlying conceptual representations, as long as the distinction between CRs and semantic representations is kept clear.

We pointed out the importance of a careful analysis of the intrinsic “organizing factors” that interact with environmental information to structure our knowledge of the spatial world. These organizing factors act as a kind of “syntax” in accord with which inputs to spatial systems are ordered, and in so doing they contribute meaning to the spatial representations themselves. This is perhaps clearest in the allocentric map observed in the hippocampus, but it is also observable in other cases. It is our view that careful study of the way language reflects these organizations or “distortions” should help illuminate the CRs. By itself, however, such a study will not accomplish the entire task. The relationship between spatial language and other aspects of cognitive processing, such as our intuitive understanding of motion (Talmy, chapter 6, this volume), the on-line recognition of spatial relationships (Logan and

Sadler, chapter 13, this volume), and our deductive inferences about these relationships (Johnson-Laird, chapter 11, this volume), must also be carefully unpacked if we are to derive maximum benefit from the study of language and space. Progress in these areas should improve our understanding of the relations between space and language, which in turn could illuminate the nature of conceptual representation.

Acknowledgments

We thank the McDonnell-Pew Cognitive Neuroscience Program for fostering our collaboration on this chapter, which was written while Mary Peterson and Lynn Nadel were on sabbatical from, and supported by, the University of Arizona.

Notes

1. Levinson's claims rest on the assumption that performance on old/new recognition memory tests predominantly reflects differences in perceptual organization or processing. See Hochberg and Peterson (1987); Peterson and Hochberg (1983) for criticism of this assumption.
2. Both Fodor (1975, 85–86) and Jackendoff (1983, 17) allow that, by combining the primitive CRs differently, different languages may render a given idea more or less salient to the speakers of those languages.

References

- Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy and J. R. Pomerantz (Eds.), *Perceptual organization*, 213–253. Hillsdale, NJ: Erlbaum.
- Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Biederman, I., and Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, 20, 585–593.
- Bisiach, F., and Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, 14, 129–133.
- Bloom, P. (1994). Possible names: The role of syntax-semantics mappings in the acquisition of nominals. *Lingua*, 92, 297–329.
- Bloom, P. (in press). Theories of word learning: Rationalist alternatives to associationism. In T. K. Bhatia and W. C. Ritchie (Eds.), *Handbook of language acquisition*. New York: Academic Press.
- Bowerman, M. (1989). Learning a semantic system: What role do cognitive predispositions play? In M. L. Rice and R. L. Schiefellbusch (Eds.), *The teachability of language*, 133–169. Baltimore: Brooks.
- Bridgeman, B., Kirch, M., and Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Perception and Psychophysics*, 29, 336–342.

- Brown, P., and Levinson, S. (1993). Linguistic and nonlinguistic coding of spatial arrays: Explorations in Mayan cognition. Working paper no. 24, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Carlson, V. R., and Tassone, E. P. (1971). Familiar versus unfamiliar size: A theoretical derivation and test. *Journal of Experimental Psychology*, 87, 109–115.
- Carpenter, G. A., and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37, 54–115.
- Farah, M. J., and McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120, 339–357.
- Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
- Gaithier, N. S., and Stein, B. E. (1979). Reptiles and mammals use similar sensory organizations in the midbrain. *Science*, 205, 595–597.
- Gibson, B. S., and Peterson, M. A. (1994). Does orientation-independent object recognition precede orientation-dependent recognition? Evidence from a cueing paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 299–316.
- Gogel, W. C. (1977). The metric of visual space. In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and processes*, 129–181. New York: Wiley Interscience.
- Gogel, W. C. (1982). Analysis of the perception of motion concomitant with a lateral motion of the head. *Perception and Psychophysics*, 32, 241–250.
- Gogel, W. C., and Tietz, J. D. (1977). Eye fixation and attention as modifiers of perceived distance. *Perceptual and Motor Skills*, 45, 343–362.
- Haviland, J. (1993). Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology*, 3, 3–45.
- Heilman, K. M. (1979). Neglect and related disorders. In K. M. Heilman and E. Valenstein (Eds.), *Clinical neuropsychology*, 268–307. New York: Oxford University Press.
- Hochberg, J., and Peterson, M. A. (1987). Piecemeal organization and cognitive components in object perception: Perceptually coupled responses to moving objects. *Journal of Experimental Psychology: General*, 116, 370–380.
- Hummel, J. E., and Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99, 480–517.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1987). On beyond zebra: The relation of linguistic and visual information. *Cognition*, 26, 89–114.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology*, 42, 461–478.

- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Knierim, J. J., McNaughton, B. L., Duffield, C., and Bliss, J. (1993). On the binding of hippocampal place fields to the inertial orientation system. *Society for Neuroscience, Abstracts*, 19, 795.
- Kolb, B., Buhrman, K., McDonald, R., and Sutherland, R. J. (1994). Dissociation of the medial prefrontal, posterior parietal, and posterior temporal cortex for spatial navigation and recognition memory in the rat. *Cerebral Cortex*, 46, 664–680.
- Kosslyn, S. M., Chabris, C. F., Marsolek, C. J., and Koenig, O. (1992). Categorical versus coordinate spatial relations: Computational analyses and computer simulations. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 562–577.
- Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J., and Gabrieli, J. D. E. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 723–735.
- Lackner, J. R. (1985). Human sensory-motor adaptation to the terrestrial force environment. In D. J. Ingle, M. Jeannerod, and D. N. Lee (Eds.), *Brain mechanisms and spatial vision*, 175–209. Dordrecht: Nijoff.
- Landau, B., and Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–265.
- Loomis, J. M., DaSilva, J. A., Fujita, N., and Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 906–921.
- Mangan, P., and Nadel, L. (1992). Spatial memory development and development of the hippocampal formation in Down syndrome. Paper presented at the Twenty-fifth International Congress of Psychology, Brussels, July.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- McNamara, T. (1991). Memory’s view of space. In G. H. Bower (Ed.), *The psychology of learning and motivation*, vol. 27, 147–186. New York: Academic Press.
- Meredith, M. A., and Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221, 389–391.
- Murphy, G. L., and Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Newman, E. A., and Hartline, P. H. (1981). Integration of visual and infrared information in bimodal neurons of the rattlesnake optic tectum. *Science*, 213, 789–791.
- O’Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, 51, 78–109.
- O’Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a cognitive map: Preliminary evidence from unit activity in the freely moving rat. *Brain Research*, 34, 171–175.

- O'Keefe, J., and Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Peterson, M. A. (1994). Shape recognition can and does occur before figure-ground organization. *Current Directions in Psychological Science*, 3, 105–110.
- Peterson, M. A., and Gibson, B. S. (1993). Shape recognition contributions to figure-ground organization in three-dimensional display. *Cognitive Psychology*, 25, 383–429.
- Peterson, M. A., and Gibson, B. S. (1994). Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science*, 5, 253–259.
- Peterson, M. A., Harvey, E. H., and Weidenbacher, H. L. (1991). Shape recognition inputs to figure-ground organization: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance*, 17, 1075–1089.
- Peterson, M. A., and Hochberg, J. (1983). Opposed-set measurement procedure: A quantitative analysis of the role of local cues and intention in form perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 183–193.
- Peterson, M. A., and Shyi, G. C.-W. (1988). The perception of real and illusory concomitant rotation in a three-dimensional cube. *Perception and Psychophysics*, 44, 31–42.
- Pinker, S., and Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 585–642.
- Pohl, W. (1973). Dissociation of spatial discrimination deficits following frontal and parietal lesions in monkeys. *Journal of Comparative and Physiological Psychology*, 82, 227–239.
- Predebon, J. (1990). Relative distance judgments of familiar and unfamiliar objects viewed under representatively natural conditions. *Perception and Psychophysics*, 47, 342–348.
- Putnam, H. (1975). The meaning of “meaning.” In H. Putnam (Ed.), *Mind, language, and reality: Philosophical papers*, vol. 2, 215–271. Cambridge: Cambridge University Press.
- Rueckl, J. G., Cave, K. R., and Kosslyn, S. M. (1989). Why are “what” and “where” processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience*, 1, 171–186.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.
- Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin and Review*, 1, 2–28.
- Shiffrar, M. (1994). When what meets where. *Current Directions in Psychological Research*, 3, 96–100.
- Shiffrar, M., and Freyd, J. J. (1990). Apparent motion of the human body. *Psychological Science*, 1, 257–265.
- Slobin, D. (1995). Learning to think for speaking: Native language, cognition, and theoretical style. In J. J. Gumperz and S. C. Levinson (Eds.), *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.

- Soja, N. N., Carey, S., and Spelke, E. S. (1992). Perception, ontology, and word meaning. *Cognition*, 45, 101–107.
- Starkey, P., Spelke, E. S., and Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, 36, 97–127.
- Talmy, L. (1983). How language structures space. In H. Pick and L. Acredolo (Eds.), *Spatial orientation: Theory, research, and application*, 225–282. New York: Plenum Press.
- Tarr, M. J., and Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology*, 21, 233–282.
- Taube, J. S. (1992). Qualitative analysis of head direction cells recorded in the rat anterior thalamus. *Society for Neuroscience*, 18, 708.
- Taube, J. S., Muller, R. U., and Ranck, J. B., Jr. (1990a). Head direction cells recorded from the postsubiculum in freely moving rats. 1. Description and quantitative analysis. *Journal of Neuroscience*, 10, 420–435.
- Taube, J. S., Muller, R. U., and Ranck, J. B., Jr. (1990b). Head direction cells recorded from the postsubiculum in freely moving rats. 2. Effects of environmental manipulations. *Journal of Neuroscience*, 10, 436–447.
- Ungerleider, L. G., and Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Eds.), *Analysis of visual behavior*, 549–586. Cambridge, MA: MIT Press.
- Van Essen, D., Anderson, C., and Felleman, D. (1992). Information processing in the primate visual cortex: An integrated systems perspective. *Science*, 255, 419–423.
- Vandeloise, C. (1991). *Spatial prepositions: A case study from French*. Chicago: University of Chicago Press.
- Volpe, B. T., LeDoux, J. E., and Gazzaniga, M. S. (1979). Information processing in an “extinguished” visual field. *Nature*, 282, 722–724.
- Wallach, H. (1949). Some considerations concerning the relation between perception and cognition. *Journal of Personality*, 18, 6–13.
- Weiskrantz, L. (1986). *Blindsight*. Oxford: Clarendon Press.
- Wilcox, T., Rosser, R., and Nadel, L. (1994). Representation of object location in 6.5-month-old infants. *Cognitive Development*, 9, 193–209.
- Wynn, K. (1995). Origins of numerical knowledge. *Mathematical Cognition*, 1, 35–60.

Name Index

- Acredolo, L., 129, 130
Allan, K., 358
Allen, R., 443
Allport, D., 533
Anderson, C., 355, 557
Anderson, J., 280
Anderson, S., 172, 469, 479
Antell, S., 388, 392
Antinucci, F., 388
Atkinson, M., 391
- Baayen, H., 113, 125
Babcock, M., 244
Baillargeon, R., 367, 371, 378, 388, 389, 391, 392
Barwise, J., 443, 455
Battison, R., 172
Bauer, M., 456
Bauer, P., 369, 375
Baumgartner, G., 387
Baylor, G., 460
Beauvois, M., 532, 541, 543
Becher, A., 349
Behl-Chadha, G., 388
Bellugi, U., 171, 175, 194, 195, 196, 200, 201
Bennett, D., 280
Berlin, B., 56
Berman, R., 401, 402
Berndt, R., 532, 540
Bertenthal, B., 370
Berthoz, A., 152
Bettger, 196
Bialystok, E., 15, 386, 422
Biederman, I., 8, 9, 13, 24, 46, 317, 347, 495, 498, 561, 562, 567
Bierwisch, M., 8, 15, 19, 31–73, 32, 37, 55, 58, 65, 134, 358, 386, 402, 422, 424, 563
Binford, O., 317
Bisiach, F., 567
Black, J., 469
Bloom, P., 14, 357, 389, 390, 426, 553–573, 563
- Bomba, P., 388
Bornstein, M., 344
Bower, G., 469, 470, 471
Bower, T., 392
Bowerman, M., 7, 192, 203, 379, 385–426, 387, 389, 391, 402, 404, 405, 409, 416, 418, 419, 420, 421, 424, 428n, 498, 526, 564, 571
Boyer, P., 234
Breedin, S., 355, 358
Brewer, B., 126
Bridgeman, B., 559, 560, 562
Brown, P., 70, 80, 101, 102, 103, 104, 110, 111, 114, 122, 148,
Brown, R., 389, 403, 422
Brownell, H., 532
Brugman, C., 378
Bryant, D., 471, 473, 478, 503
Bub, D., 532, 533, 537, 538, 540, 546, 547
Bucher, N., 231
Bühler, K., 147, 465
Bull, W., 449
Burroughs, W., 465
Byrne, R., 46, 80, 437, 443, 446, 449, 460
- Campbell, J., 129, 130
Campbell, R., 542, 543, 547
Caplan, L., 541
Caramazza, A., 532, 533, 536, 540, 542, 545
Carey, S., 235, 336, 346, 372, 389, 424, 425, 563
Carlson, V., 561
Carlson-Radvansky, L., 89, 93, 94, 132, 133, 325, 496, 499, 500, 524, 525, 526
Caron, A., 388, 392
Caron, R., 392
Carpenter, G., 570
Carpenter, P., 131, 503, 521
Carramazza, A., 532
Carroll, J., 424
Cave, K., 562
Charcot, J., 532

- Chase, W., 502
 Chertkow, H., 532, 533, 537, 538, 540, 546, 547
 Chipman, S., 519
 Choi, S., 351, 377, 379, 404, 405, 409, 418, 420, 421, 423
 Chomsky, N., 33, 35, 41
 Cienki, A., 401
 Cinan, S., 535, 536
 Clark, E., 344, 389, 391, 418, 424
 Clark, H., 110, 128, 131, 134, 135, 144, 386, 422, 470, 493, 496, 502, 503, 521, 524
 Cochran, W., 536
 Cohen, D., 131
 Coltheart, M., 543
 Compton, B., 496, 521, 523, 524, 526
 Comrie, B., 301, 302
 Cook, W., 280
 Coon, V., 184, 473, 476
 Cooper, E., 561
 Cooper, L., 471, 500
 Corballis, M., 500
 Corina, D., 172, 175, 194
 Coslett, H., 355, 358, 542, 543
 Couclelis, H., 465
 Coulter, G., 172
 Craik, K., 437
 Cramazza, A., 541
 Culicover, P., 8

 Danziger, E., 113, 125, 131, 148
 Davidoff, J., 541, 543
 De Bleser, R., 541, 543
 Dector, M., 537, 538, 540, 546
 de León, L., 404, 418
 Denis, M., 469
 DeRenzi, E., 534, 537, 538
 Derthick, M., 548
 DeValois, K., 387
 DeValois, R., 387
 Dölling, J., 37
 Dorina, D., 201
 Dostrovsky, J., 556
 Dowty, D., 14, 448
 Druks, J., 543, 547
 Dunker, K., 559

 Ehrich, V., 89, 469, 484
 Eilan, N., 152
 Eimas, P., 368, 388
 Emmorey, K., 124, 171–204, 175, 185, 194, 195, 196, 201, 203, 570
 Engberg-Pedersen, E., 185
 Engel, S., 256
 Eriksen, C., 503
 Etchemendy, J., 455
 Evans, J., 449

 Fara, M., 355
 Farah, M., 10, 355, 535, 536, 538, 547, 571
 Farrell, W., 470
 Felleman, D., 355
 Fellman, D., 557
 Fillmore, C., 15, 131, 374, 465, 466, 467, 471
 Fodor, J., 41, 50, 231, 368, 554, 571
 Francis, W., 336
 Franklin, N., 184, 470, 471, 473, 476, 478, 503
 Frederick, R., 128
 Freeman, N., 389
 Freyd, J., 232, 244, 561
 Friederici, A., 88, 89, 133
 Frisk, V., 278
 Funnell, E., 534

 Gaffan, D., 534, 535, 536
 Gainotti, G., 534, 538
 Gaithier, N., 556
 Garnham, A., 89, 437, 493, 494, 524
 Garrett, M., 541, 553–573
 Garrod, S., 469, 479
 Gazzaniga, M., 567
 Gee, J., 185
 Gelman, R., 563
 Gentner, D., 378, 403, 404
 Gibson, B., 561, 570
 Gibson, E., 389
 Gleitman, L., 404
 Glenberg, A., 470
 Gogel, W., 559, 569
 Golinkoff, R., 376
 Goodhart, W., 185
 Gopnik, A., 391, 392, 393, 424
 Gordon, B., 538, 540
 Grabowski, J., 80, 88
 Greenspan, S., 470, 471
 Gregory, R., 152
 Grice, H., 451
 Griffiths, P., 391
 Grossberg, S., 570
 Gruber, J., 13, 22, 48, 278, 305
 Gruendel, J., 390
 Gumperz, J., 404

 Hafner, J., 443
 Hagert, G., 447
 Hale, K., 35
 Hankamer, J., 95
 Harnard, S., 72, 373
 Hart, J., 538, 540
 Hart, R., 465
 Hartline, P., 556
 Harvey, E., 561, 570
 Haude, R., 198
 Haviland, J., 113, 125, 134

- Hayward, W., 343, 496, 506, 524, 525, 526
 Hazelfigg, M., 464
 Hedley-White, T., 541
 Heilman, K., 566
 Heine, B., 396, 400
 Herrmann, T., 80, 88
 Herskovits, A., 128, 134, 142, 191, 265, 322, 399, 493, 494, 496, 524, 526
 Heywood, C., 534, 535, 536
 Hickok, G., 201
 Hill, A., 89
 Hill, C., 143, 339, 400, 474
 Hillis, A., 532, 536, 540, 541, 542
 Hinrichs, E., 14
 Hinton, G., 539, 544, 548
 Hirtle, S., 465
 Hjelmlev, L., 61, 62
 Hockett, C., 147
 Hoffman, D., 347
 Hoiting, N., 185
 Howard, F., 542
 Hsieh, 331
 Hubel, D., 355
 Hummel, J., 561
 Humphreys, G., 532, 534, 535, 536, 537, 538, 540, 541, 545
 Hunkin, N., 534
 Hurwitz, S., 83, 471

 Ilan, A., 196
 Inhelder, B., 129, 319, 326, 335, 387
 Iorio, L., 541
 Irwin, D., 89, 93, 94, 132, 133, 325, 496, 499, 500, 524, 525, 526

 Jackendoff, R., 1–27, 6, 7, 8, 9, 10, 11, 13, 14, 15, 22, 26n, 35, 39, 40, 44, 47, 48, 49, 63, 74n, 77, 104, 130, 176, 200, 215, 268, 272n, 278, 281, 322, 353, 354, 355, 378, 386, 399, 405, 422, 423, 493, 494, 499, 524, 531, 532, 533, 541, 554, 555, 562, 567, 568, 572
 Jammer, M., 128
 Janis, W., 174
 Jankovic, I., 464, 471, 479
 Jeannerod, M., 10
 Jepson, A., 256
 Job, R., 534, 538, 540
 Johnson, M., 268, 269, 375
 Johnson, R., 534
 Johnson-Laird, P., 5, 6, 15, 47, 81, 87, 110, 128, 132, 134, 137, 140, 144, 159n, 162, 203, 322, 386, 422, 437–460, 438, 443, 445, 449, 456, 460, 465, 466, 469, 470, 471, 472, 493, 494, 496, 524, 532, 546, 573
 Johnston, J., 326, 386, 388, 389, 390
 Jolicoeur, P., 500, 521, 561

 Jones, S., 344, 346, 376
 Jonides, J., 465
 Just, M., 131, 503, 521

 Kahneman, D., 443
 Kamp, H., 50
 Kant, E., 128
 Karmiloff-Smith, A., 367
 Katz, J., 35
 Kawamoto, A., 545
 Kay, P., 56
 Keil, F., 235, 368, 392, 404, 424, 563
 Kenny, A., 448
 Keyser, S., 35
 Kintsch, W., 471
 Kirch, M., 559, 560, 562
 Kittay, E., 6
 Klatsky, R., 152, 153
 Klima, E., 171, 175, 194, 200, 201
 Knierim, J., 557
 Kolb, B., 556
 Kolstad, V., 371
 Kosslyn, S., 45, 47, 130, 195, 196, 198, 199, 471, 496, 562, 565
 Koster, C., 469, 484
 Kruskal, J., 514
 Kubovy, M., 131
 Kucera, H., 336
 Kuczaj, S., 336, 339
 Kuhn, D., 443
 Kupferman, I., 545

 Labov, W., 270, 271, 469, 483
 Lackner, J., 560
 Lakoff, G., 13, 22, 268, 269, 373, 402, 420, 498, 499, 506
 Landau, B., 10, 15, 130, 176, 268, 281, 317–358, 327, 329, 333, 346, 347, 353, 359n, 378, 386, 399, 405, 422, 423, 464, 478, 493, 494, 499, 524, 562, 567
 Lang, E., 19, 21, 54, 55, 73
 Langacker, R., 10, 13, 22, 62, 215, 365, 402, 420
 Larkin, J., 455
 Laurent, B., 534
 Lederman, S., 152, 153
 LeDoux, J., 567
 Leech, G., 131
 Legerstee, M., 371
 Lehrer, A., 6
 Leslie, A., 367, 371
 Levelt, W., 18, 19, 77–104, 78, 81, 83, 88, 89, 93, 101, 103, 117, 132, 133, 134, 144, 146, 160, 285, 325, 335, 465, 466, 469, 472, 478, 482, 484, 493, 494, 524, 567
 Levin, B., 6
 Levine, D., 355

- Levine, M., 464, 471, 479
 Levine, S., 336, 389
 Levinson, S., 7, 80, 88, 101, 102, 103, 104, 104n, 106n, 109–157, 111, 113, 114, 117, 122, 125, 131, 132, 134, 145, 148, 158n, 160, 176, 179, 184, 254, 285, 321, 324, 325, 335, 352, 386, 402, 404, 465, 466, 468, 470, 482, 486, 569, 570, 571, 573n
 Lewis, D., 144
 Leyton, M., 242, 335, 347, 353
 Lhermitte, F., 541, 543
 Liddell, S., 174, 175, 198
 Lillo-Martin, D., 171, 174, 175, 185
 Linde, C., 270, 271, 469, 483
 Lindem, K., 470
 Livingstone, M., 355
 Lloyd, S., 389
 Locke, J., 152
 Lockman, J., 465
 Loew, R., 185
 Logan, G., 133, 493–526, 496, 499, 502, 521, 523, 524, 526, 572
 Loomis, J., 558, 560, 562
 Lowe, D., 317
 Lucariello, J., 367
 Lucchelli, F., 534, 537, 538
 Luzzatti, C., 567
 Lynch, K., 465
 Lyons, J., 134, 135, 449

 MacKay, L., 500, 521
 MacLaury, R., 396
 MacLean, D., 392
 Mainwaring, S., 471, 484
 Mandler, J., 5, 26n, 203, 365–380, 366, 369, 370, 371, 372, 373, 374, 375, 376, 381, 420, 421, 546, 564
 Mangan, P., 566
 Mani, K., 470, 471
 Manktelow, K., 449
 Manning, L., 542, 543, 547
 Maratsos, M., 336, 339, 367
 Markman, E., 404
 Marlsburg, C., 545
 Marr, D., 8, 9, 13, 21, 24, 46, 130, 141, 151, 252, 317, 335, 347, 437, 455, 460, 464, 495, 497, 498, 499, 532, 560, 562
 Matsumoto, Y., 215
 McCarthy, R., 532, 534, 536, 537, 547, 548
 McClelland, J., 538, 539, 544, 545, 547, 571
 McCullough, K., 124
 McCune-Nicholich, L., 390
 McDonough, L., 366, 369, 375, 379
 McIntire, M., 190, 191
 McMullen, P., 535, 536

 McNamara, T., 566
 Medin, D., 424, 570
 Meier, R., 171
 Meltzoff, A., 152, 391, 392
 Meredith, M., 556
 Mervis, C., 11, 368
 Metzler, J., 130, 131, 196
 Meyer, M., 470, 535, 536
 Michotte, A., 11, 343
 Miikkulainen, R., 548
 Miller, C., 185
 Miller, G., 15, 87, 110, 128, 132, 134, 140, 144, 322, 386, 422, 445, 465, 466, 469, 471, 493, 496, 532, 546
 Miller, J., 196
 Milner, B., 278
 Miozzo, M., 534
 Mishkin, M., 10, 268, 354, 557, 562
 Moore, G., 465
 Moravcsik, J., 51
 Morrow, O., 470
 Muller, R., 280, 557
 Murphy, G., 424, 570

 Nadel, L., 10, 128, 129, 159n, 268, 277, 278, 281, 294, 305, 308, 313, 465, 553–573, 555, 556, 557, 566, 567, 569
 Narasimhan, B., 19–21, 29, 335
 Needham, A., 388, 389, 392
 Neisser, U., 463
 Nelson, K., 367, 369, 390
 Newman, E., 556
 Newport, E., 171
 Newstead, S., 449
 Nicholich, L., 386
 Nigro, G., 463
 Nishihara, H., 464, 495, 499

 Ohlsson, S., 447
 O'Keefe, J., 10, 128, 129, 144, 159n, 268, 277–315, 278, 280, 281, 294, 296, 301, 305, 306, 313, 465, 553, 555, 556, 557, 567, 569
 Oliva, A., 355, 356
 Olson, D., 15, 386, 422
 Osherson, D., 443

 Padden, C., 174, 185
 Paillard, J., 129
 Palij, M., 464, 471, 479
 Palmer, S., 231, 460, 495
 Parisi, D., 388
 Parkin, A., 534
 Partee, B., 8, 449
 Patterson, K., 542
 Pears, J., 126

- Pederson, E., 117, 135, 148, 254, 404
 Perkins, D., 443
 Perrett, D., 465
 Perrig, W., 471
 Peterhans, E., 387
 Peterson, M., 495, 553–573, 559, 561, 570
 Piaget, J., 129, 319, 326, 335, 365, 366, 367, 387, 391
 Pick, H., 132, 465
 Pieraut-Le Bonniec, G., 392
 Pietrini, V., 534
 Pinker, S., 6, 8, 131, 402, 464, 553, 561
 Plaut, D., 544, 547
 Pohl, W., 556
 Poizner, H., 200, 201
 Poulin, C., 185
 Predebon, J., 561
 Presson, C., 464
 Prior, A., 449
 Pustejovsky, J., 6, 7, 14, 33, 51, 53
 Putnam, H., 11, 563
 Pylyshyn, Z., 460, 494, 497, 500

 Quine, W., 368, 443, 449
 Quinn, P., 368, 388

 Ranck, J., 280, 557
 Rapoport, T., 6
 Rapp, B., 532, 541, 542
 Rescher, N., 449
 Reyle, U., 50
 Richards, M., 424
 Richards, W., 256, 347
 Richardson, J., 449
 Riddoch, M., 532, 535, 541, 545
 Robin, F., 469
 Rock, I., 126
 Rosch, E., 11, 344, 353, 368
 Rosenkrantz, S., 368
 Rosser, R., 566
 Roth, J., 199
 Rubin, J., 256
 Rueckl, J., 562
 Rumelhart, D., 539, 544
 Ryle, G., 448

 Sacchetti, C., 536, 537
 Sadalla, G., 465
 Sadler, D., 133, 343, 493–526, 573
 Saffran, E., 355, 358, 542, 543, 548
 Sag, I., 95
 Sandler, W., 172
 Sartori, G., 534, 538, 540
 Sartori, M., 534
 Schaeken, W., 451

 Schiano, D., 471, 484
 Schmitt, B., 117
 Schneider, G., 354
 Schober, M., 183, 463, 468, 470, 484, 485, 486
 Schuler, M., 392
 Schwartz, M., 548
 Schyns, P., 355, 356
 Scoville, W., 277
 Seery, J., 514
 Seidenberg, M., 544
 Senft, G., 399
 Sergeant, J., 521, 523, 524
 Shafir, E., 443
 Shallice, T., 531–548, 533, 534, 542, 544, 546, 547, 548, 566, 571
 Shepard, R., 83, 130, 131, 196, 464, 471, 500, 519, 569
 Shepard-Kegl, J., 190, 191
 Sheridan, J., 534, 538, 540
 Shiffrar, M., 355, 561, 570
 Shyi, G., 559
 Silveri, M., 534, 538
 Simon, H., 443, 455
 Sinha, C., 358, 389, 407
 Sitskoorn, M., 392
 Slobin, D., 78, 132, 185, 377, 386, 388, 389, 401, 402, 403, 571
 Smith, L., 344, 345, 346, 376
 Smith, M., 278
 Smith, S., 443
 Smitsman, A., 392
 Snedecor, G., 536
 Snodgrass, J., 535, 536
 Soja, N., 346, 563
 Sorabji, R., 126
 Spelke, E., 346, 367, 388, 392, 563
 Sperling, A., 559, 560, 562
 St. James, J., 503
 St. John, M., 185
 Staplin, L., 465
 Starkey, P., 563
 Stecker, D., 327, 329, 422
 Stein, B., 556
 Stein, J., 129, 152
 Stewart, F., 534
 Subrahmanyam, K., 346
 Supalla, S., 185
 Supalla, T., 203
 Svorou, S., 128
 Swales, M., 534
 Sweetser, E., 375
 Szeminska, A., 326, 335

 Tabossi, P., 449
 Takano, Y., 131, 144

- Talbot, K., 198
 Talmy, L., 10, 11, 13, 22, 26n, 78, 177, 178, 179,
 211–271, 214, 215, 218, 228, 254, 256, 265, 270,
 272n, 273n, 317, 318, 319, 320, 322, 323, 324,
 326, 352, 359n, 365, 375, 386, 399, 404, 405,
 420, 422, 423, 426, 429n, 493, 494, 496, 502,
 524, 562, 572
 Tanz, C., 336
 Tarr, M., 131, 343, 464, 496, 506, 524, 525, 526,
 561
 Tassone, E., 561
 Taube, J., 280, 557
 Taylor, A., 542
 Taylor, H., 88, 184, 470, 479, 482, 483
 Teitz, J., 569
 Teller, P., 15
 Tolman, E., 129
 Toulmin, S., 443
 Touretsky, D., 548
 Traugott, E., 374
 Tulving, E., 482
 Turner, T., 469
 Tversky, B., 5, 18, 19, 80, 81, 88, 105n, 133, 160,
 184, 270, 271, 335, 343, 443, 463–487, 471, 472,
 473, 476, 478, 482, 483, 484, 504, 567, 571
 Tye, M., 130

 Ullman, S., 494, 497, 500, 501, 520, 521
 Ullmer-Ehrich, V., 469, 482, 484
 Ungerleider, L., 10, 268, 354, 557, 562
 Urquhart, A., 449

 Vaina, L., 347
 Valvo, A., 152
 Van Cleve, J., 128
 Vandalloise, C., 134, 493, 494
 Vanderwart, M., 535, 536
 Van Essen, D., 355, 557
 Verkuyl, H., 14
 Volpe, B., 567
 von der Heydt, R., 387

 Wallach, H., 560
 Warach, J., 355
 Ward, T., 349
 Warrington, E., 532, 533, 534, 535, 537, 540, 542,
 546, 547, 548
 Weidenbacher, H., 561, 570
 Weiskrantz, L., 557
 Wernicke, C., 532
 Wierzbicka, A., 422
 Wilcox, T., 566
 Wilford, J., 479
 Wilkins, D., 399
 Williams, R., 544
 Wunderlich, D., 65, 68, 69

 d'Ydewalle, G., 451
 Young, F., 514

 Zbrodoff, N., 502

Subject Index

- Aboriginal language, frame of reference in, 125, 184
About, 299–300, 304
Above, 89–90, 132, 335, 388, 495, 504
 effect on figure object, 323
 to express social status, 308
 template for, 497, 505–526
Absolute direction coding, 116, 120, 122, 566–567.
 See also Fixed heading system
 liabilities of, 146
 in map recall, 481
 and perspective discourse, 486
 preemption by, 94
 in Tzeltal, 569–570
Absolute frame of reference, 145–147
 in American Sign Language, 183
 in horizontal spatial patterns, 135–138
 in Tzeltal, 111–113
Absolute perspective, 80, 88, 89–90, 101
Absolute-relative space, theories, 128
Absolute tense, 302
Abstraction
 in fictivity, 257–260
 and I-language, 40
 in lexical systems, 36–37
 and meaning formation, 376
 neural representation of, 563
 in perspective taking, 78
 structural, 267
Abstract motion, 215. *See also* Caused motion;
 Fictive motion theory; Motion
Acceptability, in spatial cognition, 501
Accessibility to consciousness, in fictivity, 248
Access paths, 214, 242–243
Across, 178, 301
 concept of, 258
 effect on figure object, 323
 and object axis, 327
 properties of, 266
 in Tzeltal, 111
Action, 410–411, 543
Actionability, 248, 258
Active-determinative principle, 226
Addressee, 259, 470
Adjectives
 dimensional, 15, 71–72
 spatial, 321
Advent paths, 214, 241–242
Affixation, 173
African language, 396, 400
After, 304
Against, 190, 301
Agency, 227, 228–230, 376
Agentive sensory paths, in fictivity, 227
Alignment, in multiple perspectives, 89–92
Alignment paths, in fictivity, 217
Allocentricity theory, 129–130, 159n16
 frame, 285
 maps, 556
 perspective, 465
 space, 560
Along, 178, 301, 323, 327
American Sign Language, 171–204, 183
Amid, 300, 321, 324
Among, 178, 300, 321, 324
Analog-digital, and CS, 5
Anaphora, 95, 174
Anchor point, 142, 148
Angles. *See entries under* Axial
Animacy, 343
 judgment of, 342–343, 349
 and meaning formation, 372–373
 in prelinguistic cognition, 366
Animals, 12, 368–371
Aphasia, 200–203, 533–538
Apparent motion, 211, 237
Apprehension, 502–503
Apprehension, 493–526, 524
Aristotle, moored boat case, 126
Around, 55, 299–300, 304

- Arrernte language, 426n3
 Articulation, 43
Art of the Fugue, 42
 Aspect, 301
 Associative ception form, 260–261
At, 299–300, 304, 324
 Atsugewi language, 7, 324
 Attachment, 319–320, 351
 Auditory-phonological interface, 4–5
 Auditory system, temporal perception in, 173
Auf-an, 320, 321
Away from, properties of, 266
 Axial based learning, 317–360
 Axial perspective, 14–19, 474–476
 vertical, 135
 Axial representation, 317–319, 320, 325–326, 358
 for environmental comprehension, 472–473
 in mental models, 441
 schematizing reference object, 334–344
 Axial system theory, 14–15, 19, 63, 329
- Bach, J. S., musical example, 42
Back, 87–88, 123, 131–132, 133, 134, 163n51, 566
 in American Sign Language, 179–183
 in apprehension theory, 504–505
 as non-linguistic concept, 389
 relative frame of reference, 142–143
 in Tzeltal, 352
 Backpropagation, 544
 Bali language, axes system of, 146
 Basic relations, as spatial indices, 494
 Basic spatial terms, 56–57
 Bearings, in absolute frame of reference, 145–147
Before, 299, 304
Behind, 298–299, 320
 cross-linguistic perspective, 400
 overextension, 390–391
 schematization, 335–336
Behither, 298
 Belhare language, 27n19
Below, 287–289, 335, 388, 495
 causal relation use, 311–313
 effect on figure object, 323
 to express social status, 306–311
 in spatial apprehension, 504
 template for, 505–526
 in vector notation, 282–287
Beneath, 293–294, 306–311
Beside, 299
Between, 178, 300, 305, 324
 in American Sign Language, 190
 concept of, 437–438
Beyond, 296–298, 305
 Bifurcation, and representation tasks, 355
 Binding, multimodal representations, 12
 Binocular disparity, 561. *See also* Stereopsis
- Biological motion, concept formation by, 370
 Blindsight, 557
 Body-centered frame of reference. 130. *See also*
 Egocentric; *entries under* Deictic
 Body-part system, 352, 378, 395
 in Tzeltal, 352–353
 Body-tour perspective, 133
Bottom, 352
 Brain. *See also* Cortex; Hippocampus; Lateraliza-
 tion; Parietal cortex; Thalamus
 and space representation, 555–558
 Brain damage, and *left-right* representation, 566–
 567
 Brain stem, and distance information, 557
By, 299, 304
- Canonical position, 92–95
 Canonical region, determination of, 341–343
 Canonical reference frame
 encounter frame for, 16
 orientation, 16, 89–95, 561
 representation for, 45
 Cardinal direction, in perspective, 466
 Cardinal direction coding, 109, 110, 128, 566–567,
 569–570. *See also* Absolute direction coding
 in Guugu Yimithirr, 7
 Cardinal direction system. *See* Cardinal direction
 coding
 Cardinal direction system. *See* Absolute direction
 coding
 Cardinal locations, 336, 339
 Cartesian coordinate system, 187, 193
 Case systems, 61
 Case theory, 60–62, 280
 Categorization, 48, 368–370, 390. *See also*
 Relation judgment
 child learning of, 563
 crosslinguistic, 393–398
 and multiple store theory, 546
 patterns for, 411–419
 for recall, 482
 semantic, 407
 and shape, 563
 spatial, 393–398, 402, 403–420
 of visual objects, 9
 Category-specific system, 533–538, 547–548
 Causal relations, in semantic mapping, 311–313
 Caused motion, 371, 374. *See also* Motion; Fictive
 motion theory
 abstract, 215
 encoding for, 404–420, 409–410
 in verb phrases, 377
 Ception theory, 214, 245, 260–261
 Certainty, in fictivity, 248
 Chamus language, 400
 Chiapas, Mexico, 110

- Children. *See* Learners
- Chinese language, 8, 61–62, 73
- Chomsky theory, 33, 41
- Clarity, in fictivity, 246
- Classification. *See also* Categorization
of spatial knowledge, 395–398
- Classifier systems, 61
- Closed-class forms, 259, 267–268, 321, 405. *See also* Open-class forms
in fictivity, 264–265
- Coarse representation, schematizing for, 327–329
- Coercion principle, 7
- Cognition
definition, 437
developmental, 317–359
with diagrammatic icons, 455–460
multidimensional, 24
and perspective taking, 464–465
prelinguistic, 365, 386
spatial, 43–50, 195, 353–354, 387, 553
thinking for speaking, 77–78, 104, 427n8
- Cognitive map theory, 277–314
- Cognitive organization, 211–213, 231
- Color, 44, 52, 354
- Complementary phenomenon, and semantic
theory, 536–538
- Compositional semantics, construction of,
444
- Computational theory, 493–526
- Computer logic, 444, 450
- Conceptual development. *See* Cognition theory
- Conceptual-intentional system, 50
- Conceptualization
in fictivity, 245–246
opposition theory in, 129–130
prelinguistic, 369
of space, 43–50
- Conceptual lexicon, 47
- Conceptual representation, 120, 122–123, 554
in computational theory, 498
effect of linguistic coding, 117
and frame of reference, 125
and organizing factors, 572
preverbal, 365–380
- Conceptual structure, 5–8
and axial specification, 21
color, 52
and I-language, 40
time, 52
- Conceptual system, interface of, 44
- Congflation, by young learners, 334
- Conjunctive forms (Takano), 131
- Connectionist theory, 547–548
- Consciousness, 248, 268
- Constructional fictive motion, 215
- Constructivism, 372
- Contact concept, 386
crosslinguistic, 377–378, 393–398
neonate image-schema for, 374
and spatial templates, 526
- Containment, 388, 392, 564
crosslinguistic, 377–378, 393–398
image-schema for, 374
of *in-on*, 190–193
in Korean, 322
neonatal concept of, 371, 386
and spatial templates, 526
- Content-structure, in fictivity, 247–248
- Contextual frame, 17
- Contiguity, of objects, 111
- Control, 311
- Conversion
across modalities, 153–157
in perspective taking, 81–82
- Coordinate system
in American Sign Language, 179
in Gestalt theory, 126
language differences of, 138–141
ternary, 137
- Coordination, and perspective sharing, 83–88
- Core concept theory, 533
- Coreference, in American Sign Language, 174
- Cortex, 556. *See also* Brain; Brain stem; Brain
damage; Hemisphere function; Parietal cortex
spatial function, 354
- Count-mass distinction, 13–14
- Count nouns, 344, 345
- Coverage paths, 214
in fictivity, 243–244
- Criterion of economy, 12–13, 21
- Criterion of grammatical effect, 13, 22
- Criterion of interfacing, 13. *See also* Interfacing
- Crosslinguistic analysis, 7, 134, 385–428
- Cross-modality
definition, 158n1
and spatial representation, 152–157
- CS-SR interface. *See* Interfacing
- Cues, for perspective taking, 88–89
- Cuing tasks, in spatial apprehension, 502–503
- Cultural determinants, 102, 569
- Deafness, 171
- Decision procedure, 443
- Deductive reasoning, 437–438, 441
with temporal relations, 444–455
- Deep ellipsis. *See* Ellipsis
- Deictic center, in coordinate systems, 147
- Deictic direction coding, 79, 133
- Deictic frame of reference, 15
contradictions of, 135–138
developmental appearance, 131–132
redefinition of, 138

- Deictic perspective
 definition, 465–466
 direction coding, 566–567
 in narrative, 185
- Deictic perspective, 77, 96, 571
 preemption by, 94
- Deictic relation, 494–495, 526n1
- Deixis am Phantasma, 105n3
- Demonstrative paths, in fictivity, 219, 227
- Demonstratives, in perspective, 185
- Descriptions, 83–86, 401–402, 487
 and perspective, 463–487
- Descriptors, as locatives, 399
- Diagrams, as reasoning tools, 455–460
- Digital-analog, and CS, 5
- Dimensional adjectives, 15, 57–59, 71–72
- Dimensionality, and I-space, 49
- Directional terms, 99–101, 272n8, 566–567
 in cognitive mapping theory, 279
 ellipsis of, 95–102
- Direction system. *See* Perspective
- Discourse, 104, 468, 562–568
- Discrepancy, cognitive, 211–213
- Disjunctions, in deductive problem-solving, 455–460
- Displacement. *See* Motion
- Dissociation, in fictivity, 261–262
- Distance, 558. *See also* entries for each preposition
 prepositions, 294–295
- Domain-specific learning, 422–424
- Dorsal subiculum (rat), 557
- Down, 142, 287–289, 301
 in cardinal direction system, 111
 nonphysical use of, 308
 in Tzeltal, 111
- Driving tour perspective, 133. *See also* Gaze tour
- Dual coding hypothesis (Paivio), 5
- Duration, 375, 450
- During, 304
- Dutch language, 351, 394–398
 bipartite proximity, 77
 concept of support, 378
 conceptual representation, 498
 direction term selection, 102
 early spatial concepts, 351
 gestures, 570
 image-schemas, 421
 recall memory, 114
 relative coding, 115–116
 spatial encoding, 408–410
 transitive inference, 120
 vision bound perspective, 88
- Dynamism, cognitive bias for, 213, 270–271
- Dyslexia, 543
- Earth-based frame, 254. *See also* Absolute
 direction coding; Cardinal direction coding
- Egocentric direction coding, 129–130, 560, 566–567. *See also* Deictic direction coding
- distance perception in, 561
- maps, 556
- perspective, 465
- Elementary forms (Takano), 131
- Ellipsis, 95–99
 deep, 99–103
 mapping of, 164n64
- Emanation, 214
 basis of, 226–231
 evil eye concept of, 235
 and perception, 231–232
 properties of, 216–217
- Embedding, 26n7
- Encephalitis, 533, 546
- Encirclement, concept of, 401
- Encoding, in language learners, 404–420
- Energy, in emanation theory, 235
- Enter, I-space concept of, 50
- Entities, 278, 568
- Environmental cognition, and perspective taking, 464–465
- Environmental frames, 17–18, 92, 130, 132, 471–473
- Episodic information, 277
- Equiavailability model, 471, 478
- Equidistance tendency (Gogel), 558, 569
- Essence, and functional kind memberships, 563–564
- Event-process distinction, 14
- Evidentials, 259
- Evil eye, in emanation theory, 235
- Experience, effects of, 569–572
- Explicit transfer, 49
- Explicit visual perception, 268
- Exploration, neonate stimulation by, 370
- Extensions, 215, 390–391, 416–418
 under-, 377
- External region, determination of, 341
- Extrinsic frame of reference, 132
- Extrinsic perspective, 466–467, 479
- Factive representation, defined, 211
- Fictive motion theory, 211–273
 basis of, 226–231
 and ception theory, 260–262
 and dynamism, 270–273
 fire in, 221–224
 linguistic, 214–217
 in map recall, 481
 and metaphor, 268–269
 path types, 217–226, 236–244
 representations in, 244–260
- Fictivity. *See* fictive motion theory
- Figure-ground, 93, 110, 141, 321. *See also*
 Referent; Relatum

- asymmetry in language, 177–178
- definition, 159n9
- in fictivity, 217, 220
- organization of, 561
- in spatial relationships, 317
- in spatial semantics, 402
- theory, 357
- visual, 93
- Figure object, 322, 326
 - axial information on, 329–334
 - geometry of, 323–324
- Fine-grained representation, 320, 344–350
- Finnish language, 8, 393–398
- Fire, in fictivity, 221–224
- Fixed bearing system, 128, 145–147. *See also*
 - Absolute direction coding; Cardinal direction coding
 - use of gesture, 124–125
- Fodorian hypothesis, 1–2, 25n6, 41, 554
- Folk iconography, 267
- For*, 281, 294–295
- Force dynamics
 - CS-SR concept of, 11
 - in mapping, 307
 - neonate image-schema for, 375
 - sensing of, 256
- Frame-relative motion, 214, 237–240
- Frames of reference, 127table, 159n17, 554. *See also*
 - Deictic frame of reference; Egocentric direction coding; Object-centered frame; Perspective
 - absolute, 556
 - allocentric, 285
 - in American Sign Language, 179–184
 - and axial vocabulary, 14–19
 - binary-ternary, 496–497
 - computation for, 504–505
 - conceptual representation for, 125
 - cross-discipline theory, 127–134
 - crosslinguistic theory, 109–164
 - egocentric, 21
 - hierarchy of, 93–94
 - orientation of, 92
 - spatial, 126–127
 - switching, 21
 - variability in, 463
- Framework vertical constraint principle, 94
- French language, conceptual representation in, 498
- From*, 288, 303, 304, 305
- Front*, 18, 87–88, 123, 133, 134, 142–143, 163n51, 352, 389
 - in American Sign Language, 179–183
 - in apprehension theory, 504–505
 - concept variability, 566
 - frame of reference theory, 131–132
 - in map recall, 481
 - in Nilotic cultures, 162n43
 - properties of, 217
- Functionality
 - and conceptual organization, 52
 - recall of, 542
 - representation for, 535–541
 - and SR hypothesis, 23
- Gaze tour, 133, 484
 - relative perspective of, 481
- Generalization, 331–333, 392–393
- Geographical frame, 17
- Geomagnetic sense, prepositional presence of, 315n3
- Geometric frame, 15
- Geometric pattern, in fictivity, 248
- Geometric representation, 317–359
- Geonic construction (Biederman), 8, 46
- German language, 62, 63, 320, 321
 - concept of attachment, 351
 - concept of support, 378
 - dimensional adjectives in, 73
 - direction equivalents, 324
 - early spatial concepts, 351
 - representation in, 498
- Gestalt theory, 104, 242, 243. *See also* Figure-ground
 - dominant visual, 93
 - in fictivity, 269
 - frame of reference in, 126
 - influence on perspective, 102
- Gesture, 156
 - and absolute direction coding, 124–125, 570
 - in emanation theory, 234
- Ghost physics, 234–235, 388
- Gibsonian tradition, 531
- Global directional system, 296
- Global framing, 238
- Gloss, 204n1
- Gogel's equidistance tendency, 558, 569
- Goodness of fit, 501
- Goodness rating, 509–514
- Grammar
 - acquisition of, 375
 - categories for, 367
 - representation of, 554
 - spatial distinctions of, 59, 63
- Gravity
 - determination of, 89
 - early concept of, 371, 386, 388, 391, 396
 - in spatial representation, 17, 132, 281, 472
 - in vector notation, 285
- Ground, 137
- Guidebooks, perspective in, 479

- Guugu Yimithirr
 and absolute perspective system, 80
 coordinate system of, 147
 spatial expressions in, 103
 syntactic variations, 7
- Handedness, 131. *See also Left-right*
- Haptic information, 10, 44, 78
- Hausa language, 17
- Head direction neurons, 557
- Hebrew language, 405
- Hematoma, right parietal-occipital, 201–203
- Hemisphere function, 200–203, 543
- Hermann grid, and fictivity, 249–250
- Herpes simplex encephalitis, 533, 546
- Hippocampus, 129, 557, 567
 allocentric map in, 572
 in cognitive mapping theory, 279
 Euclidean framework in, 569
 mapping function, 277, 556
 role in perspective, 465
- Horizontal patterns, 135–138, 544
- Horizontal prepositions, 296–299. *See also entries for each preposition*
- Hungarian language, 61
 syntactic variations, 8
- Hyperproof, 455
- Iconography, in fictivity, 230, 233–234
- Identifiability, in fictivity, 247
- I-language, 33, 64, 73. *See also I-space*
 abstraction, 40
 grammar of, 59–62
- Image-schemas, 9–10, 420, 422. *See also Inference*
 high level, 546
 neonatal representation, 380
 spatial representation, 373–375
- Imaging, 47. *See also Representation*
- Imagistic encoding
 ception forms, 260
 representation, 101
 and SR, 9–10
- Implicit transfer, 49
- Implicit visual perception, 268
- In*, 304, 335, 351, 357, 377, 385, 401
 abstraction of, 564–566
 American Sign Language, 190–193
 crosslinguistic perspective, 134, 393–398
 and I-space, 50, 55
 mapping of, 565
 properties of, 264–268
 schematization, 322–324
 template for, 505–526
- Indexing
 in relation judgement, 502
 spatial, 504, 524
 and spatial cognition, 500
- Indexing function, in American Sign Language, 174
- Induced motion, 211, 240, 559
- Infants. *See Learners; Neonates*
- Inference
 basis of, 460
 schema for, 438 (*see also Image-schemas*)
 transitive, 117–123, 144
- Inferential potential, 81–83
- Inferential satisficing, 443
- Infixation, 173
- Influence, metaphoric expression of, 306–311
- In front of*, 92, 161n31, 161n34, 320, 335–336, 400, 466, 495
 crosslinguistic perspective, 137
 overextension, 390–391
- Inner language, 109. *See also Language of thought*
- Inside*, 324
- Integration, of maps in mammals, 556
- Interfacing, 2, 3–5, 10–13
 of I-language, 43
- Internal states, relationship to external space, 375
- Into*, 351
- Intransitive verbs, acquisition of, 377
- Intrinsic direction coding, 79, 566–567
- Intrinsic frame of reference, 15, 96, 110, 140–141
 in fictivity, 261
 in horizontal spatial patterns, 135–138
 in young learners, 131–132
- Intrinsic perspective, in guidebooks, 479
- Intrinsic relations, definition, 454
- Introspection, 375
- Invalidity, knowledge of, 443
- I-space. *See also I-language*
 definition of, 44–48
 domain, 67
enter, 50
- Italian language, 351
- Jackendoff single store theory, 44, 532, 547
- Japanese language, 77, 358
- Judgment, speed of spatial, 94
- Kantian theory, 151, 569
 absolute space theory, 128
- Kinesthetic information, 44. *See also Haptic information*
- Kinship, linguistic forms of, 259
- Knowledge, spatial, 8
- Korean language, 321, 357, 396, 401
 dimensional adjectives in, 73
 early spatial concepts, 351
 image-schemas in, 421
in-on morphemes of, 379
 locative constructions in, 69–70

- spatial encoding in, 404–420, 408–410
 - through* equivalents, 324
 - verb phrases in, 377
- KYST system, in computation theory, 517
- Ladybird stimulus set, 535
- Lak language, 61
- Language of thought (Fodor), 6, 554. *See also*
 - Fodorian hypothesis; Language of the mind
 - cross-cultural understanding of, 569
- Language of the mind, 1. *See also* Language of thought
- Lateralization, in human brain, 277
- Layout, and SR, 9
- Learners, 343. *See also* Neonates
 - axial understanding in, 317–360
 - categorization by, 563
 - early spatial concepts in, 385–393
 - object representation in, 317, 327–344
 - prepositional mastery in, 335
 - schematizing by, 344–348
 - structural abstraction in, 267–268
 - universals in, 356–357
- Leave, I-space concept of, 50
- Left-right, 92, 567
 - in American Sign Language, 172
 - apprehension theory, 504–505
 - axis for, 111, 123, 134, 162n42, 162n47, 163n51, 335, 566
 - concept of, 196, 566–567
 - ellipsis, 95
 - map recall of, 481
 - in perspective taking, 470, 485
 - recall, 114–115
 - relative frame of reference, 142
 - in spatial apprehension, 504
 - spatial template for, 505–526
- Leibnizian theory, 151
- Lexical systems
 - choice, 95–96
 - concepts, 19–24, 77, 103
 - items, 5, 12
 - representation, 532–548, 554
 - semantics, 32, 33–73, 34–36
 - spatial concepts, 56–57
 - in Tzeltal, 111
- Light radiation, 221–224
- Line, intangible, 217–218
- Linearization, 101
 - and CS, 6
 - in spoken languages, 185
- Line of sight, 220
- Linguistic-spatial interface, 1–27
- Link schemas, 374
- LISP, 446
- Local framing, 238
- Localists. *See* Locationist theory
- Localizability, in fictivity, 247
- Location, 328–329, 561
- Locationist theory, 62, 280, 281
- Locatives, 65, 389
 - acquisition of, 388
 - in American Sign Language, 175
 - crosslinguistic differences in, 399
 - in German, 32
 - in spatial representation, 44
 - in Tzeltal, 352
- Loglish, 439
- Macroplanning, 77, 103, 482
- Magic beams, depiction in emanation theory, 233
- Magno system, 355
- Mapping, 3, 104, 270
 - of axial information, 343–344
 - differences in, 560
 - of early language, 377
 - in fictivity, 230
 - in frame of reference theory, 152
 - framework of, 555
 - in hippocampus, 277
 - in I-language, 33
 - interface, 3–5, 10–13
 - of linguistic elements, 562–563
 - motoric, 567
 - perceptual, 133
 - perspective in, 80–95, 102, 479
 - topography in, 556
- Marr-Biederman theory, 10–11. *See also* 2-D Sketch; 2-1/2 D Sketch; 3-D Sketch (Marr)
- Marrian abstractions, 252
- Mass nouns, 14
- Mayan language, 110, 396
 - body-part descriptions, 162n38
- Meaning, 386. *See also* Meaning packages
 - encoding requirements, 6
 - formation of, 371–373
 - in lexical systems, 34–35
 - and mental model construction, 438–444
 - preverbal, 365–380
 - and primitives, 554–555
 - and SR, 12
- Meaning packages, 365, 375–376, 380
- Memory, 113, 114–117, 570
 - in map recall, 482
- Mental image, generation of, 195, 198–200
- Mental maps, 129
- Mental models, 46, 437
 - of environment comprehension, 471
 - language-based, 470
 - propositional representations for, 438–444
- Mental rotation skills, 570
- Mental tours, 469

- Message, preparation for, 77
 Metaphor
 in fictivity, 268–269
 linguistic, 214
 preposition function as, 281
 in vertical prepositions, 306–311
 Microplanning, 77, 103, 482
 Miming, to identify objects, 541
 Mind, 555–558. *See also* Brain
 Mirror reversals, 196
 Mixtec language, 378, 395, 398
 Mnemonics, controlling for, 123–124
 Modalities. *See also* Multimodal representation
 cross-discipline understanding, 125–134
 and CS, 6
 and frame of reference, 109, 149–152
 logic in, 449
 in spatial representation, 9, 46–47
 transmodality, 45
 Model, 437–444, 460
 Modularity, 43, 125–126
 Chomsky theory of, 41
 definition of, 158n1
 representational, 1–3
 and spatial cognition, 392, 438
 specialization of, 12–13
 Molyneux's question, 110, 152, 157
 Monkeys, discrimination ability of, 535
 Mood, 258
 Mopan language, 80, 102
 More, 290–293
 Morphemes
 closed-class, 267–268
 early production of, 405–408
 mapping of, 420
 and object relationship, 401
 Morphological categories, in I-language, 64
 Morphology, 31–32, 173
 Motion, 6, 388
 aftereffect, 240
 biological, 370
 categorization by neonates, 376
 cognitive bias for, 270–271
 CS-SR concept of, 11
 early understanding of, 391
 encoding of, 401–402, 404–420
 fictive, 211–273
 frame for, 15
 illusory, 559–560
 and magno system, 355
 self-motion, 371, 373, 377
 and spatial representation, 44–45
 Motoric maps, and *left-right* representation, 567
 Movement. *See* Fictive motion; Motion
 Multimodal representation, 556
 Multiple store theory, 532–548
 Musical formulation, 78
 Naming, 344
 impairment of, 358, 533–546, 547
 Narasimhan figures, 19–21
 Narrative, 184, 196, 278, 472–479
 Nativism, 372, 389
 Natural kind concepts, in children, 563–564
 Near, 294–295, 324, 505–526
 in American Sign Language, 190
 Negative object parts, 399
 Neocortex, map integration in, 556
 Neonates, 152, 388–389, 563. *See also* Learners
 Network theory, 544–546
 Neurophysiology, 152
 Neuropsychology, 531–548
 Neutral perspective, 468, 470
 Newtonian theory, 128, 151
 Next to, 505–526
 Nonagentive paths, 226–227
 Nonlinguistic coding, 386, 389
 in Tzeltal, 113
 Nonspatial factors, 13, 22–23, 374, 564–566
 Nouns, 567–568. *See also* Count nouns
 and abstraction, 563
 acquisition, 375–376
 role of, 562, 572
 spatial, 321
 Object, 10–13, 345
 conceptualization, 400–402
 contiguity, 111
 early representation, 317–359
 encoding, 8, 563–564
 functionality, 388
 identification, 354
 localization, 354
 and meaning formation, 371–373
 naming deficit for, 358
 and perspective, 464
 properties, 371
 recognition, 561
 schemata for, 54–55
 spatial representation, 47, 176, 344–350, 533–546, 557–558
 structure, 252–256
 support, 371
 Object-centered direction coding, 566–567
 Object-centered frame, 15, 26n9, 130–131, 132, 254. *See also* Absolute frame; Cardinal direction coding; Deictic frame
 Objectivity, 247, 273n11
 Observation, and meaning formation, 372
 Observer-centered frame, 15, 17–18. *See also* Deictic frame of reference; Deictic perspective

- Off*, 428n16
 semantic classification of, 411
- Ornidirectional prepositions, 299–300
- On*, 321, 335, 351, 377, 394–398, 398, 401
 in American Sign Language, 190–193
 effect on figure object, 323
 spatial template for, 505–526
- Ontology, 48, 51–52
- Open*, 385–386
- Open-class forms. *See also* Closed-class forms
 in fictivity, 264
 morphological, 405
- Opposite*, 301
- Opposition theory, 129–130
- Optic aphasia, 541–546, 547
- Optic flow, 238–239
- Order preservation, 5
- Organized unitary content hypothesis (OUCH), 532
- Orientation path, in fictivity, 217, 230
- Orientation theory
 frame of reference, 16, 131, 155–157
 and I-space, 49
 and spatial representation, 45
 specification of, 189
- Original sin, 368
- Original Word Game, 403
- Ostension, in fictivity, 247
- OUCH theory, 532
- Out*, 385, 428n16
 semantic classification of, 411
- Over*, 351, 505–526, 525
- Overextensions, 215, 390–391, 416–418. *See also*
 Extensions; Underextensions
- Paillard frames of reference, 159n17
- Palpability-related parameters, 213–214, 245–248, 249–251
- Parietal cortex, 354, 556, 557
- Part articulation, 347
- Particularity, linguistic forms of, 259
- Parvo system, 354
- Past*, 305
- Path, 26n14, 185
 in cognitive mapping theory, 280
 concept of, 10–13
 encoding descriptions of, 405
 in fictivity, 217–226
 as higher level concept, 26n14
 neonate image-schema for, 373–374
 sensing of, 253, 255–256
- Patency, acquisition of concept, 376
- Pattern path, 214
 in fictivity, 236–237
- Perception, 570
 definition, 437
 and distance representation, 558
 and fictive emanation, 231–232
 in fictivity, 245–246
 and prelinguistic categorization, 368
 visual, 44, 211
- Perceptual encoding, and SR, 10
- Perceptual interface, 531–548
- Perceptual organization, 573n1
- Permanence, in Piagetian theory, 317–320
- Perseveration, in naming recall, 544–546
- Person deixis, 466
- Perspective, 77–106, 88. *See also* Absolute direction coding; Cardinal direction coding; Deictic direction coding; Egocentric direction coding; Frames of reference; Intrinsic direction coding
 in American Sign Language, 179–184
 choice of, 468–469
 in computational theory, 497–498
 descriptions of, 463–487
 driving tour, 133
 multiple systems for, 95
 neutral, 484–486
 switching, 473–478, 482–484
 vision bound, 88
- Phonetic features, 33–34
- Phonology, 4, 172, 543
- Phonology-syntax interface, 2
- Phosphene effect, in fictivity, 250
- Piagetian theory, 129–130, 317–320, 335, 343, 365–367, 387–388, 391
- Pictorial representation. *See* 2-D sketch; 2-1/2 D sketch; 3-D sketch (Marr)
- Place, 26n14
 in cognitive mapping theory, 278
 CS-SR concept of, 10–13
 as higher level concept, 26n14
 linguistic representation of, 567
 and objects, 321
- Place cells, 556
- Place deixis, 466
- Pointing function, in American Sign Language, 174
- Point of view. *See also* Frames of reference; Orientation
 in American Sign Language, 179–183, 184
- Polish language, 401
- Politeness, in perspective choice, 485
- Polysemy, 498, 499
- Possession, 48, 259
- Postsubiculum, 280, 296
- Pragmatics, 80–95, 103–104, 470–471
- Precognition, 254–255
- Predications, 95
- Preemption, in multiple perspectives, 89–92
- Prelinguistic cognition, 366, 387–390
 and spatial templates, 525

- Prepositions, 65. *See also entries under each preposition*
 acquisition of, 377–378, 388
 axial occurrences, 22
 in conceptual representations, 498
 as figure-ground indicators, 178
 lexical representations of, 68–69
 linguistic function of, 562
 and object representations, 321, 565
 path-specifying, 220, 222
 role of, 567–568, 572
 as shape encoders, 176
 spatial, 15, 277–314, 321–323
 and spatial cognition, 525
 Preverbal learning, 365–380
 Primes. *See also primitives*
 conceptual, 54
 in I-language, 59, 73
 semantic, 66, 72, 125
 in UG, 34
 Primitives, 138–140, 374–375, 378, 554–555, 557, 573n2
 of conceptual-intentional system, 50
 and CS, 5
 in I-space, 64
 in modularity model, 1
 semantic role of, 422–425, 532
 Probability judgments, 443
 Problem-solving, 392, 570
 Process, in spatial cognition, 500–501
 Programs, in spatial apprehension, 501–503
 Projective relators, 134
 Pronominal deixis. *See* Deictic perspective
 Propositional formulation, 78–79
 Prosody, 34
 Prospect paths, in fictivity, 217
 Prototype image, shortcomings for mapping, 11–12
 Proximity, early concept of, 388
 Pseudointrinsic facets, 144
 Psycholinguistics, and frame of reference, 133
 Psychology of language, and frame of reference, 132
 Qualia structure (Pustejovsky), 53
 Quantification, and CS, 6
 Radiation path, in fictivity, 221, 230
 Reaction time, 89, 520–524
 Reality status, in fictivity, 258
 Reasoning, with mental models, 438–444
 Recall. *See also* Memory
 frame of reference for, 125
 Recognition memory, frame of reference for, 125
 Reduplication, 173
 Reference frame, 524
 in computational theory, 499
 sensing of, 253–254
 in spatial apprehension, 504–505
 and spatial cognition, 500
 Reference object, 32, 321, 322
 asymmetry in language, 177–178
 in coordinate system, 126
 in fictivity, 217, 220
 geometry of, 324–326
 Reference points, with temporal prepositions, 305
 Referent, 78, 92, 104, 175. *See also* Figure-ground; Relatum
 Regions of acceptability, and spatial cognition, 496–497
 Relation judgment, 502
 Relationships, 398–400, 493–526
 Relative-absolute space, theories, 127–129
 Relative frame of reference, 110, 135–138, 142–144, 179. *See also* Deictic frame of reference
 Relative perspective, 467
 Relative system, 79
 Relative tense, 302
 Relatum, 78, 104, 126, 137. *See also* Figure-ground; Referent
 Representation, 8–10, 437, 554–555. *See also* Representation system; Templates
 for action, 543
 in American Sign Language, 186
 axial specification in, 21
 criteria, 8–9
 cross-cultural, 571
 differences in, 353–357
 of distance, 558
 for environmental comprehension, 470–473
 fine-grained, 344–350
 for *in front of*, 336
 as interface, 44
 of I-space, 64
 kinds of, 497–500
 language independence of, 81
 multimodal, 560
 in neonates, 563
 and object function, 541
 and object recognition, 463
 perception, 133
 preverbal, 365–380, 372–373
 propositional, 438–444
 and reference frame, 499
 retinotopic, 10
 semantic, 39, 533
 of sensory quality, 538–541
 shape bias, 345–347
 in signed languages, 193–195
 spatial, 5, 554
 stability across cultures, 571–572

- for temporal relations, 448–455
- three types of, 531
- visual, 101
- in young learners, 317
- Representational modularity, 1–3, 24, 39–40
- Representation system. *See also* Representation
 - for axial information, 335–337
 - and frame of reference, 109, 156
 - and Molyneux's question, 152–157
 - multiple functions of, 42
- Reverse-direction conceptualization, 223–224
- Rich concepts, and conceptual organization, 53–54
- Rotation, 47, 148, 163n51, 567
 - ability in deafness, 195–198
 - and goodness of fit, 514
 - for shape recognition, 131
- Route, 18–19, 479, 482
- Routines, in spatial apprehension, 501–503
- Rules of inference, in mental models, 439–442, 446–448
- Rule theories, and diagram use, 460
- Russian language, 72–73, 351, 401
- Satellite-framed languages, 404
- Scales, in vector notation, 286–287
- Schematization, 266–268, 318, 323. *See also* 2-D, 2-1/2 D, 3-D sketch (Marr)
 - of reference object, 334–344
- Scotoma, in factivity, 249
- Self-motion, 371, 373, 377. *See also* Caused motion; Fictive Motion Theory; Motion
- Semantic Features Hypothesis, 424
- Semantics
 - breakdown in, 381n4
 - components of, 32
 - crosslinguistic categorization, 403–420
 - forms in, 36–39
 - function after stroke, 538–541
 - map for, 278
 - pathologies of, 533–548
 - and primitives, 422–425
 - processing, 531–532
 - representations for, 101, 365–380
 - and sensory quality, 538–541
 - spatial, 387, 389–391, 391–393
 - structure in, 402
- Semitic languages, 173
- Sensorimotor period, 365–367, 387–388, 390
- Sensory encoding, 6, 12, 252–257
- Sensory paths, in factivity, 224–226
- Sensory quality, 538–541, 546
- Sequential perspectival mode, 270
- Serial visual routines, in computation theory, 520–524
- Shadow paths, 224, 226
- Shape
 - ignoring, 327–329, 330
 - and object names, 563
 - recognition of, 131
 - representations for, 345–347, 358
 - and SR, 8–9
- Shape bias, 345–350
- Shepard and Metzler paradigm, 131
- Sight, 152, 224–225. *See also* Visual system
- Signed languages, 171–205. *See also* American Sign Language
- Similarity theory, 368
 - crosslinguistic study of, 393–398
 - rating, 514–520
 - testing, 542
- Since, 304, 305
- Site arrival, 241–242
- Skeletal-based schematization. *See* Axial-based learning
- 2-1/2 D Sketch (Marr), 2, 9–10, 130, 134, 142, 160n20
- 2-D Sketch (Marr), 133, 156, 323
- 3-D Sketch (Marr), 2, 8–9, 12, 46, 48, 133, 134, 156, 437, 532
 - in concept, 323
 - and object schemata, 55
 - and theory of vision, 130, 151, 455
- Snakes, map integration in, 556
- Social categories, 48, 468
- Social cognition, 6, 23
- Social determinants, 484, 487
- Social relation, 8, 306–311
- Spanish language, 68, 394–398, 399, 405
 - concept of containment, 378
 - en, 321
 - image-schemas in, 421
 - lexical concepts, 77
- Spare concepts, and conceptual organization, 53–54
- Speaker-centered system, 79
- Speaker's knowledge status, linguistic forms of, 259
- Speech, speed of articulation, 173
- SR-CS interface. *See* Interfacing
- Standard average European (SAE), 94, 105n8
- State of change, 211
- State of motion, 211
- State of occurrence, 211
- Staticism, and cognitive dynamism, 270–271
- Stationariness, 211, 238–239, 270–271
- Stereopsis, 10. *See also* Binocular disparity
- Stick figure drawings, in structure sensing, 267
- Stimulus dependence, in factivity, 248
- Strength, in factivity, 246
- Stroke, semantic function after, 537
- Stroop-like interference, 502

- Structurality, in fictivity, 250
- Structure, 40, 264–268, 386
- Subjective frame of reference, 130
- Subjective motion, 215
- Sunlight, 221–224, 233
- Superior colliculus (mammalian), 556
- Superman, depiction in emanation theory, 233
- Superordinates, 368, 563
- Support concept, 386, 388, 392
 - crosslinguistic factor, 377–378, 393–398
 - neonatal image-schema for, 374
 - in neonates, 371
 - and spatial templates, 526
- Surface ellipsis, 95–103
- Survey, 479, 482
- Symbol-grounding (Harnad), 372
- Synoptic perspectival mode, 270
- Syntax, 26n7
 - in American Sign Language, 174
 - and conceptual structure, 7
 - in I-language, 64
 - in perspective, 185
 - in phonology, 4
 - semantic component of, 31–32
 - and shape bias, 345
 - spatial representation base for, 365–380
 - in vector grammar, 313–314
 - in vector notation, 282
- Tabassarian language, 61
- Ta in Tzeltal language, 321
- Targeting paths, in fictivity, 220, 227
- Taste, impairment of, 540
- Taxonomy, and CS, 6
- Templates. *See also* Representation
 - in computational theory, 499
 - spatial, 496–497, 498, 505–526
 - for spatial cognition, 500
- Temporal cortex, 354
- Temporality, 14, 173
 - in error patterns, 544–546
 - mapping of, 277
 - in neonates, 374–375
 - reasoning, 438, 444–455
 - and rules of inference, 448
 - in scene identification, 356
 - and spatial representation, 44–45
- Temporal lobe
 - memory function, 277
 - in naming task, 537
 - and object naming deficit, 358
 - and object properties, 557–558
- Temporal prepositions, 301–305. *See also* entries for each preposition
- Tenejapan language, 104n2, 110. *See also* Tzeltal language
- frameworks, 156
- gesture, 124–125
- transitive inference, 120
- Tense, 258, 264, 301
- Texture, integration of, 44
- Thalamus (rat), 557
- Theory of reasoning, 437
- Theta* activation, 557
- Thinking. *See* Cognition
- Thinking for speaking, 77–78, 104, 427n8
- Third-person, in narratives, 473–478
- Three-way system, 111
- Through*, 301, 323–324, 324
 - properties of, 266
- Tight-fit concept, 379–380
- Time, 52. *See also* Temporality
 - and axial systems, 23
 - conceptualization of, 48, 438
 - linguistic mapping, 268–269
 - neonate concept for, 374–375
 - spatial connection, 214
 - and spatial representation, 44–45
 - and temporal prepositions, 301–305
- To*, in semantic mapping, 305
- Token-type distinctions, and CS, 6
- Top*, 352
- Topological relators, 134
- Topological representation, 317–320
- Tour, 469. *See also* Descriptions; Descriptors; Gaze tour
- Transformation, 286, 314
- Transitive inference. *See* Inference
- Transitive verbs, 377
- Transitivity, 82–83
- Transmodality, 45. *See also* Modalities
- Transposed deictic center (Bühler's), 147
- Travel books, perspective in, 479
- Turkish language, 405
- Typology, 68–69, 143
- Tzeltal language, 62, 105n8, 110–113, 134, 176, 321, 325, 395
 - and absolute directions, 569–570
 - body-part system, 352
 - concept of *side*, 140
 - gestures in, 570
 - locative constructions in, 70
 - and mixed perspective system, 80
 - positional adjectives in, 73
 - relative coding, 115–116
 - spatial predicates in, 331–333
 - spatial preposition in, 324
 - unique characteristics, 350
 - and vision-bound perspective, 88
- Tzotzil Mayan language, 418, 428n18
- Under*, 134, 289–293, 308–311, 389, 390–391
 - in causal relation use, 311–313

to express social status, 306
in German, 32
spatial template for, 505–526
Underextensions, 377, 390–391. *See also* Extensions; Overextensions
Universal structure, 33–35, 73, 125
Until, 304, 305
Up, 111, 142, 287–289, 390
in perspective taking, 470

Vector grammar, 277–314, 280, 281–314
Verbal representation, 532–548
Verbs, 405
acquisition of, 375, 376–377
agentive, 225
in American Sign Language, 176
in Atsugewi, 324
and CVA recall, 543
positional, 62
Veridicality, discrepancy in, 211–213
Verification tasks, in spatial apprehension, 503
Verticality, 88, 544
Vertical prepositions, 282–294, 295–296. *See also* entries for each preposition
Vestibular cues, 92
Via, 301
Viewer-based frame. *See* Viewer-centered frame of reference
Viewer-centered frame of reference, 130–131, 132, 134, 254. *See also* Egocentric direction coding
Virtual motion, 215
Virtual structures, 265
Visual system, 256–257
ception theory, 261–262
fictivity pattern in, 213
imagery in, 130
and neuropsychological impairments, 533–540
processing, 557–558
recall in, 123–124
representation from, 2, 101, 532–548
semantics of, 532–533
and spatial development, 387
specialization relative to auditory stream, 173
theory of, 151

What system, 10, 268, 354, 355, 557, 561, 562, 563
Where system, 10, 268, 354, 355, 356, 399, 561, 562
Whorfian hypothesis, 114, 125, 377, 379, 571
definition, 404
of spatial representation, 102
Wik Mungan language, 147
Word count, 336
Word learning, 563

X-ray vision, in emanation theory, 233
in map recall, 479

Young learners. *See* Learners

Z-dimension, in vector notation, 282
Zinacantan language, 158n5