George Lakoff

Women, Fire, and Dangerous Things

What Categories Reveal about the Mind

The University of Chicago Press, Chicago 60637
The University of Chicago Press, Ltd., London

Printed in the United States of America
95 94 93 92 91 90 89 88 87 54321

The University of Chicago Press
Chicago and London

Preface

Cognitive science is a new field that brings together what is known about the mind from many academic disciplines: psychology, linguistics, anthropology, philosophy, and computer science. It seeks detailed answers to such questions as: What is reason? How do we make sense of our experience? What is a conceptual system and how is it organized? Do all people use the same conceptual system? If so, what is that system? If not, exactly what is there that is common to the way all human beings think? The questions aren't new, but some recent answers are.

This book is about the traditional answers to these questions and about recent research that suggests new answers. On the traditional view, reason is abstract and disembodied. On the new view, reason has a bodily basis. The traditional view sees reason as literal, as primarily about propositions that can be objectively either true or false. The new view takes imaginative aspects of reason — metaphor, metonymy, and mental imagery — as central to reason, rather than as a peripheral and inconsequential adjunct to the literal.

The traditional account claims that the capacity for meaningful thought and for reason is abstract and not necessarily embodied in any organism. Thus, meaningful concepts and rationality are *transcendental*, in the sense that they transcend, or go beyond, the physical limitations of any organism. Meaningful concepts and abstract reason may happen to be embodied in human beings, or in machines, or in other organisms — but they exist abstractly, independent of any particular embodiment. In the new view, meaning is a matter of what is meaningful to thinking, functioning beings. The nature of the thinking organism and the way it functions in its environment are of central concern to the study of reason.

Both views take categorization as the main way that we make sense of experience. Categories on the traditional view are characterized solely by the properties shared by their members. That is, they are characterized

(*a*) independently of the bodily nature of the beings doing the categorizing and (*b*) literally, with no imaginative mechanisms (metaphor, metonymy, and imagery) entering into the nature of categories. In the new view, our bodily experience and the way we use imaginative mechanisms are central to how we construct categories to make sense of experience.

Cognitive science is now in transition. The traditional view is hanging on, although the new view is beginning to take hold. Categorization is a central issue. The traditional view is tied to the classical theory that categories are defined in terms of common properties of their members. But a wealth of new data on categorization appears to contradict the traditional view of categories. In its place there is a new view of categories, what Eleanor Rosch has termed *the theory of prototypes and basic-level categories.* We will be surveying that data and its implications.

The traditional view is a philosophical one. It has come out of two thousand years of philosophizing about the nature of reason. It is still widely believed despite overwhelming empirical evidence against it. There are two reasons. The first is simply that it is traditional. The accumulated weight of two thousand years of philosophy does not go away overnight. We have all been educated to think in those terms. The second reason is that there has been, until recently, nothing approaching a well-worked-out alternative that preserves what was correct in the traditional view while modifying it to account for newly discovered data. This book will also be concerned with describing such an alternative.

We will be calling the traditional view *objectivism* for the following reason: Modern attempts to make it work assume that rational thought consists of the manipulation of abstract symbols and that these symbols get their meaning via a correspondence with the world, *objectively construed,* that is, independent of the understanding of any organism. A collection of symbols placed in correspondence with an objectively structured world is viewed as a *representation* of reality. On the objectivist view, *all* rational thought involves the manipulation of abstract symbols which are given meaning only via conventional correspondences with things in the external world.

Among the more specific objectivist views are the following:

- Thought is the mechanical manipulation of abstract symbols.
- The mind is an abstract machine, manipulating symbols essentially in the way a computer does, that is, by algorithmic computation.
- Symbols (e.g., words and mental representations) get their meaning via correspondences to things in the external world. All meaning is of this character.

- Symbols that correspond to the external world are *internal representations of external reality.*
- Abstract symbols may stand in correspondence to things in the world independent of the peculiar properties of any organisms.
- Since the human mind makes use of internal representations of external reality, the mind is *a mirror of nature,* and correct reason mirrors the logic of the external world.
- It is thus incidental to the nature of meaningful concepts and reason that human beings have the bodies they have and function in their environment in the way they do. Human bodies may play a role in *choosing* which concepts and which modes of transcendental reason human beings actually employ, but they play no essential role in *characterizing* what constitutes a concept and what constitutes reason.
- Thought is *abstract* and *disembodied,* since it is independent of any limitations of the human body, the human perceptual system, and the human nervous system.
- Machines that do no more than mechanically manipulate symbols that correspond to things in the world are capable of meaningful thought and reason.
- Thought is *atomistic,* in that it can be completely broken down into simple "building blocks"—the symbols used in thought—which are combined into complexes and manipulated by rule.
- Thought is *logical* in the narrow technical sense used by philosophical logicians; that is, it can be modeled accurately by systems of the sort used in mathematical logic. These are abstract symbol systems defined by general principles of symbol manipulation and mechanisms for interpreting such symbols in terms of "models of the world."

Though such views are by no means shared by all cognitive scientists, they are nevertheless widespread, and in fact so common that many of them are often assumed to be true without question or comment. Many, perhaps even most, contemporary discussions of the mind as a computing machine take such views for granted.

The idea of a *category* is central to such views. The reason is that most symbols (i.e., words and mental representations) do not designate particular things or individuals in the world (e.g., Rickey Henderson or the Golden Gate Bridge). Most of our words and concepts designate categories. Some of these are categories of things or beings in the physical world—chairs and zebras, for example. Others are categories of activities and abstract things—singing and songs, voting and governments, etc. To a very large extent, the objectivist view of language and thought rests on

the nature of categories. On the objectivist view, things are in the same category if and only if they have certain properties in common. Those properties are necessary and sufficient conditions for defining the category.

On the objectivist view of meaning, the symbols used in thought get their meaning via their correspondence with things—particular things or categories of things—in the world. Since categories, rather than individuals, matter most in thought and reason, a category must be the sort of thing that 'can fit the objectivist view of mind in general. All conceptual categories must be symbols (or symbolic structures) that can designate categories in the real world, or in some possible world. And the world must come divided up into categories of the right kind so that symbols and symbolic structures can refer to them. "Categories of the right kind" are classical categories, categories defined by the properties common to all their members.

In recent years, conceptual categories have been studied intensively and in great detail in a number of the cognitive sciences—especially anthropology, linguistics, and psychology. The evidence that has accumulated is in conflict with the objectivist view of mind. Conceptual categories are, on the whole, very different from what the objectivist view requires of them. That evidence suggests a very different view, not only of categories, but of human reason in general:

- Thought is *embodied,* that is, the structures used to put together our conceptual systems grow out of bodily experience and make sense in terms of it; moreover, the core of our conceptual systems is directly grounded in perception, body movement, and experience of a physical and social character.
- Thought is *imaginative,* in that those concepts which are not directly grounded in experience employ metaphor, metonymy, and mental imagery—all of which go beyond the literal mirroring, or *representation,* of external reality. It is this imaginative capacity that allows for "abstract','thought and takes the mind beyond what we can see and feel. The imaginative capacity is also embodied—indirectly—since the metaphors, metonymies, and images are based on experience, often bodily experience. Thought is also imaginative in a less obvious way: every time we categorize something in a way that does not mirror nature, we are using general human imaginative capacities.
- Thought has *gestalt properties* and is thus not atomistic; concepts have an overall structure that goes beyond merely putting together conceptual "building blocks'' by general rules.
- Thought has an *ecological structure.* The efficiency of cognitive pro-

cessing, as in learning and memory, depends on the overall structure of the conceptual system and on what the concepts mean. Thought is thus more than just the mechanical manipulation of abstract symbols.

- Conceptual structure can be described using ***cognitive models*** that have the above properties.
- The theory of cognitive models incorporates what was right about the traditional view of categorization, meaning, and reason, while accounting for the empirical data on categorization and fitting the new view overall.

I will refer to the new view as ***experiential realism*** or alternatively as ***experientialism***. The term ***experiential realism*** emphasizes what experientialism shares with objectivism: ***(a)*** a commitment to the existence of the real world, (*b*) a recognition that reality places constraints on concepts, (*c*) a conception of truth that goes beyond mere internal coherence, and (*d*) a commitment to the existence of stable knowledge of the world.

Both names reflect the idea that thought fundamentally grows out of embodiment. "Experience" here is taken in a broad rather than a narrow sense. It includes everything that goes to make up actual or potential experiences of either individual organisms or communities of organisms —not merely' perception, motor movement, etc., but ***especially*** the internal genetically acquired makeup of the organism and the nature of its interactions in both its physical and its social environments.

Experientialism is thus defined in contrast with objectivism, which holds that the characteristics of the organism have nothing essential to do with concepts or with the nature of reason. On the objectivist view, human reason is just a limited form of transcendental reason. The only roles accorded to the body are ***(a)*** to provide access to abstract concepts, (*b*) to provide "wetware," that is, a biological means of mimicking patterns of transcendental reason, and (*c*) to place limitations on possible concepts and forms of reason. On the experientialist view, reason is made possible by the body—that includes abstract and creative reason, as well as reasoning about concrete things. Human reason is not an instantiation of transcendental reason; it grows out of the nature of the organism and all that contributes to its individual and collective experience: its genetic inheritance, the nature of the environment it lives in, the way it functions in that environment, the nature of its social functioning, and the like.

The issue is this:

> Do meaningful thought and reason concern merely the manipulation of abstract symbols and their correspondence to an objective reality, independent of any embodiment (except, perhaps, for limitations imposed by the organism)?

> Or do meaningful thought and reason essentially concern the nature of the organism doing the thinking—including the nature of its body, its interactions in its environment, its social character, and so on?

Though these are highly abstract questions, there does exist a body of evidence that suggests that the answer to the first question is no and the answer to the second is yes. That is a significant part of what this book is about.

Why does all this matter? It matters for our understanding of who we are as human beings and for all that follows from that understanding. The capacity to reason is usually taken as defining what human beings are and as distinguishing us from other things that are alive. If we understand reason as being disembodied, then our bodies are only incidental to what we are. If we understand reason as mechanical—the sort of thing a computer can do—then we will devalue human intelligence as computers get more efficient. If we understand rationality as the capacity to mirror the world external to human beings, then we will devalue those aspects of the mind that can do infinitely more than that. If we understand reason as merely literal, we will devalue art.

How we understand the mind matters in all these ways and more. It matters for what we value in ourselves and others—for education, for research, for the way we set up human institutions, and most important for what counts as a humane way to live and act. If we understand reason as embodied, then we will want to understand the relationship between the mind and the body and to find out how to cultivate the embodied aspects of reason. If we fully appreciate the role of the imaginative aspects of reason, we will give them full value, investigate them more thoroughly, and provide better education in using them. Our ideas about what people can learn and should be learning, as well as what they should be doing with what they learn, depend on our concept of learning itself. It is important that we have discovered that learning for the most part is neither rote learning nor the learning of mechanical procedures. It is important that we have discovered that rational thought goes well beyond the literal and the mechanical. It is important because our ideas about how human minds should be employed depend on our ideas of what a human mind is.

It also matters in a narrower but no less important way. Our understanding of what reason is guides our current research on the nature of reason. At present, that research is expanding faster than at any time in history. The research choices made now by the community of cognitive scientists will shape our view of mind for a long time to come. We are at present at an important turning point in the history of the study of the mind. It is vital that the mistaken views about the mind that have been with us for two thousand years be corrected.

This book attempts to bring together some of the evidence for the view that reason is embodied and imaginative—in particular, the evidence that comes from the study of the way people categorize. Conceptual systems are organized in terms of categories, and most if not all of our thought involves those categories. The objectivist view rests on a theory of categories that goes back to the ancient Greeks and that even today is taken for granted as being not merely true, but obviously and unquestionably true. Yet contemporary studies of the way human beings actually categorize things suggest that categorization is a rather different and more complex matter.

What is most interesting to me about these studies is that they seem to provide evidence for the experientialist view of human reason and against the objectivist view. Taken one by one, such studies are things only scholars could care about, but taken as a whole, they have something magnificent about them: evidence that the mind is more than a mere mirror of nature or a processor of symbols, that it is not incidental to the mind that we have bodies, and that the capacity for understanding and meaningful thought goes beyond what any machine can do.

CHAPTER 1

The Importance of Categorization

Many readers, I suspect, will take the title of this book as suggesting that women, fire, and dangerous things have something in common—say, that women are fiery and dangerous. Most feminists I've mentioned it to have loved the title for that reason, though some have hated it for the same reason. But the chain of inference—from conjunction to categorization to commonality—is the norm. The inference is based on the common idea of what it means to be in the same category: things are categorized together on the basis of what they have in common. The idea that categories are defined by common properties is not only our everyday folk theory of what a category is, it is also the principal technical theory—one that has been with us for more than two thousand years.

The classical view that categories are based on shared properties is not entirely wrong. We often do categorize things on that basis. But that is only a small part of the story. In recent years it has become clear that categorization is far more complex than that. A new theory of categorization, called *prototype* theory, has emerged. It shows that human categorization is based on principles that extend far beyond those envisioned in the classical theory. One of our goals is to survey the complexities of the way people really categorize. For example, the title of this book was inspired by the Australian aboriginal language Dyirbal, which has a category, *balan*, that actually includes women, fire, and dangerous things. It also includes birds that are not dangerous, as well as exceptional animals, such as the platypus, bandicoot, and echidna. This is not simply a matter of categorization by common properties, as we shall see when we discuss Dyirbal classification in detail.

Categorization is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action, and speech, Every time we see something as a ***kind*** of thing, for example, a tree, we are categorizing. Whenever we reason about ***kinds*** of things—chairs, nations, illnesses, emotions, any kind of thing at all—we

are employing categories. Whenever we intentionally perform any ***kind*** of action, say something as mundane as writing with a pencil, hammering with a hammer, or ironing clothes, we are using categories. The particular action we perform on that occasion is a ***kind*** of motor activity (e.g., writing, hammering, ironing), that is, it is in a particular category of motor actions. They are never done in exactly the same way, yet despite the differences in particular movements, they are all movements of a kind, and we know how to make movements of that kind. And any time we either produce or understand any utterance of any reasonable length, we are employing dozens if not hundreds of categories: categories of speech sounds, of words, of phrases and clauses, as well as conceptual categories. Without the ability to categorize, we could not function at all, either in the physical world or in our social and intellectual lives. An understanding of how we categorize is central to any understanding of how we think and how we function, and therefore central to an understanding of what makes us human.

Most categorization is automatic and unconscious, and if we become aware of it at all, it is only in problematic cases. In moving about the world, we automatically categorize people, animals, and physical objects, both natural and man-made. This sometimes leads to the impression that we just categorize things as they are, that things come in natural kinds, and that our categories of mind naturally fit the kinds of things there are in the world. But a large proportion of our categories are not categories of ***things;*** they are categories of abstract entities. We categorize events, actions, emotions, spatial relationships, social relationships, and abstract entities of an enormous range: governments, illnesses, and entities in both scientific and folk theories, like electrons and colds. Any adequate account of human thought must provide an accurate theory for *all* our categories, both concrete and abstract.

From the time of Aristotle to the later work of Wittgenstein, categories were thought be well understood and unproblematic. They were assumed to be abstract containers, with things either inside or outside the category. Things were assumed to be in the same category if and only if they had certain properties in common. And the properties they had in common were taken as defining the category.

This classical theory was not the result of empirical study. It was not even a subject of major debate. It was a philosophical position arrived at on the basis of a priori speculation. Over the centuries it simply became part of the background assumptions taken for granted in most scholarly disciplines. In fact, until very recently, the classical theory of categories was not even thought of as a ***theory.*** It was taught in most disciplines not as an empirical hypothesis but as an unquestionable, definitional truth.

In a remarkably short time, all that has changed. Categorization has moved from the background to center stage because of empirical studies in a wide range of disciplines. Within cognitive psychology, categorization has become a major field of study, thanks primarily to the pioneering work of Eleanor Rosch, who made categorization an issue. She focused on two implications of the classical theory:

> First, if categories are defined only by properties that all members share, then no members should be better examples of the category than any other members.
>
> Second, if categories are defined only by properties inherent in the members, then categories should be independent of the peculiarities of any beings doing the categorizing; that is, they should not involve such matters as human neurophysiology, human body movement, and specific human capacities to perceive, to form mental images, to learn and remember, to organize the things learned, and to communicate efficiently.

Rosch observed that studies by herself and others demonstrated that categories, in general, have best examples (called "prototypes") and that all of the specifically human capacities just mentioned do play a role in categorization.

In retrospect, such results should not have been all that surprising. Yet the specific details sent shock waves throughout the cognitive sciences, and many of the reverberations are still to be felt. Prototype theory, as it is evolving, is changing our idea of the most fundamental of human capacities—the capacity to categorize—and with it, our idea of what the human mind and human reason are like. Reason, in the West, has long been assumed to be disembodied and abstract—distinct on the one hand from perception and the body and culture, and on the other hand from the mechanisms of imagination, for example, metaphor and mental imagery.

In this century, reason has been understood by many philosophers, psychologists, and others as roughly fitting the model of formal deductive logic:

> Reason is the mechanical manipulation of abstract symbols which are meaningless in themselves, but can be given meaning by virtue of their capacity to refer to things either in the actual world or in possible states of the world.

Since the digital computer works by symbol manipulation and since its symbols can be interpreted in terms of a data base, which is often viewed as a partial model of reality, the computer has been taken by many as essentially possessing the capacity to reason. This is the basis of the contem-

porary mind-as-computer metaphor, which has spread from computer science and cognitive psychology to the culture at large.

Since we reason not just about individual things or people but about categories of things and people, categorization is crucial to every view of reason. Every view of reason must have an associated account of categorization. The view of reason as the ***disembodied*** manipulation of abstract symbols comes with an implicit theory of categorization. It is a version of the classical theory in which categories are represented by sets, which are in turn defined by the properties shared by their members.

There is a good reason why the view of reason as disembodied symbol-manipulation makes use of the classical theory of categories. If symbols in general can get their meaning only through their capacity to correspond to things, then ***category*** symbols can get their meaning only through a capacity to correspond to ***categories*** in the world (the real world or some possible world). Since the symbol-to-object correspondence that defines meaning in general must be independent of the peculiarities of the human mind and body, it follows that the symbol-to-category correspondence that defines meaning for category symbols must also be independent of the peculiarities of the human mind and body. To accomplish this, categories must be seen as existing in the world independent of people and defined only by the characteristics of their members and not in terms of any characteristics of the human. The classical theory is just what is needed, since it defines categories only in terms of shared properties of the ***members*** and not in terms of the peculiarities of human understanding.

To question the classical view of categories in a fundamental way is thus to question the view of reason as disembodied symbol-manipulation and correspondingly to question the most popular version of the mind-as-computer metaphor. Contemporary prototype theory does just that—through detailed empirical research in anthropology, linguistics, and psychology.

The approach to prototype theory that we will be presenting here suggests that human categorization is essentially a matter of both human experience and imagination—of perception, motor activity, and culture on the one hand, and of metaphor, metonymy, and mental imagery on the other. As a consequence, human reason crucially depends on the same factors, and therefore cannot be characterized merely in terms of the manipulation of abstract symbols. Of course, certain aspects of human reason can be isolated artificially and modeled by abstract symbol-manipulation, just as some part of human categorization does fit the classical theory. But we are interested not merely in some artificially isolatable subpart of the human capacity to categorize and reason, but in the

full range of that capacity. As we shall see, those aspects of categorization that do fit the classical theory are special cases of a general theory of cognitive models, one that permits us to characterize the experiential and imaginative aspects of reason as well.

To change the very concept of a category is to change not only our concept of the mind, but also our understanding of the world. Categories are categories of things. Since we understand the world not only in terms of individual things but also in terms of *categories* of things, we tend to attribute a real existence to those categories. We have categories for biological species, physical substances, artifacts, colors, kinsmen, and emotions and even categories of sentences, words, and meanings. We have categories for everything we can think about. To change the concept of *category* itself is to change our understanding of the world. At stake is our understanding of everything from what a biological species is (see chap. 12) to what a word is (see case study 2).

The evidence we will be considering suggests a shift from classical categories to prototype-based categories defined by cognitive models. It is a change that implies other changes: changes in the concepts of truth, knowledge, meaning, rationality—even grammar. A number of familiar ideas will fall by the wayside. Here are some that will have to be left behind:

- Meaning is based on truth and reference; it concerns the relationship between symbols and things in the world.
- Biological species are natural kinds, defined by common essential properties.
- The mind is separate from, and independent of, the body.
- Emotion has no conceptual content.
- Grammar is a matter of pure form..
- Reason is transcendental, in that it transcends—goes beyond—the way human beings, or any other kinds of beings, happen to think. It concerns the inferential relationships among all possible concepts in this universe or any other. Mathematics is a form of transcendental reason.
- There is a correct, God's eye view of the world—a single correct way of understanding what is and is not true.
- All people think using the same conceptual system.

These ideas have been part of the superstructure of Western intellectual life for two thousand years. They are tied, in one way or another, to the classical concept of a category. When that concept is left behind, the others will be too. They need to be replaced by ideas that are not only more accurate, but more humane.

Many of the ideas we will be arguing against, on empirical grounds, have been taken as part of what *defines* science. One consequence of this study will be that certain common views of science will seem too narrow. Consider, for example, scientific rigor. There is a narrow view of science that considers as rigorous only hypotheses framed in first-order predicate calculus with a standard model-theoretic interpretation, or some equivalent system, say a computer program using primitives that are taken as corresponding to an external reality. Let us call this the predicate calculus (or "PC") view of scientific theorizing. The PC view characterizes explanations only in terms of deductions from hypotheses, or correspondingly, in terms of computations. Such a methodology not only claims to be rigorous in itself, it also claims that no other approach can be sufficiently precise to be called scientific. The PC view is prevalent in certain communities of linguists and cognitive psychologists and enters into many investigations in the cognitive sciences.

Such a view of science has long been discredited among philosophers of science (for example, see Hanson 1961, Hesse 1963, Kuhn 1970, 1977, and Feyerabend 1975). As we will see (chaps. 11–20), the PC view is especially inappropriate in the cognitive sciences since it *assumes* an a priori view of categorization, namely, the classical theory that categories are sets defined by common properties of objects. Such an assumption makes it impossible to ask, as an empirical question, whether the classical view of categorization is correct. The classical view is assumed to be correct, because it is built into classical logic, and hence into the PC view. Thus, we sometimes find circular arguments about the nature of categorization that are of the following form:

Premise (often hidden): The PC view of scientific rigor is correct.
. . .
. . .
. . .
Conclusion: Categories are classical.

The conclusion is, of course, presupposed by the premise. To avoid vacuity, the empirical study of categorization cannot take the PC view of scientific rigor for granted.

A central goal of cognitive science is to discover what reason is like and, correspondingly, what categories are like. It is therefore especially important for the study of cognitive science not to assume the PC view, which presupposes an a priori answer to such empirical questions. This, of course, does not mean that one cannot be rigorous or precise. It only means that rigor and precision must be characterized in another way—a

way that does not stifle the empirical study of the mind. We will suggest such a way in chapter 17.

The PC view of rigor leads to rigor mortis in the study of categorization. It leads to a view of the sort proposed by Osherson and Smith (1981) and Armstrong, Gleitman, and Gleitman (1983) and discussed in chapter 9 below, namely, that the classical view of categorization is correct and the enormous number of phenomena that do not accord with it are either due to an "identification" mechanism that has nothing to do with reason or are minor "recalcitrant" phenomena. As we go through this book, we will see that there seem to be more so-called recalcitrant phenomena than there are phenomena that work by the classical view.

This book surveys a wide variety of rigorous empirical studies of the nature of human categorization. In concluding that categorization is not classical, the book implicitly suggests that the PC view of scientific rigor is itself not scientifically valid. The result is not chaos, but an expanded perspective on human reason, one which by no means requires imprecision or vagueness in scientific inquiry. The studies cited, for example, those by Berlin, Kay, Ekman, Rosch, Tversky, Dixon, and many others, more than meet the prevailing standards of scientific rigor and accuracy, while challenging the conception of categories presupposed by the PC view of rigor. In addition, the case studies presented below in Book II are intended as examples of empirical research that meet or exceed the prevailing standards. In correcting the classical view of categorization, such studies serve to raise the general standards of scientific accuracy in the cognitive sciences.

The view of categorization that I will be presenting has not arisen all at once. It has developed through a number of intermediate stages that lead up to the cognitive model approach. An account of those intermediate steps begins with the later philosophy of Ludwig Wittgenstein and goes up through the psychological research of Eleanor Rosch and her associates.

CHAPTER 2

From Wittgenstein to Rosch

The short history I am about to give is not intended to be exhaustive. Its purpose, instead, is to give some sense of the development of the major themes I will be discussing. Here are some of those themes.

Family resemblances: The idea that members of a category may be related to one another without all members having any properties in common that define the category.

Centrality: The idea that some members of a category may be "better examples" of that category than others.

Polysemy as categorization: The idea that related meanings of words form categories and that the meanings bear family resemblances to one another.

Generativity as a prototype phenomenon: This idea concerns categories that are defined by a generator (a particular member or subcategory) plus rules (or a general principle such as similarity). In such cases, the generator has the status of a central, or "prototypical," category member.

Membership gradience: The idea that at least some categories have degrees of membership and no clear boundaries.

Centrality gradience: The idea that members (or subcategories) which are clearly within the category boundaries may still be more or less central.

Conceptual embodiment: The idea that the properties of certain categories are a consequence of the nature of human biological capacities and of the experience of functioning in a physical and social environment. It is contrasted with the idea that concepts exist independent of the bodily nature of any thinking beings and independent of their experience.

Functional embodiment: The idea that certain concepts are not merely *understood intellectually;* rather, they are *used* automatically, unconsciously, and without noticeable effort as part of normal func-

tioning. Concepts used in this way have a different, and more important, psychological status than those that are only thought about consciously.

Basic-level categorization: The idea that categories are not merely organized in a hierarchy from the most general to the most specific, but are also organized so that the categories that are cognitively basic are "in the middle" of a general-to-specific hierarchy. Generalization proceeds "upward" from the basic level and specialization proceeds "downward."

Basic-level primacy: The idea that basic-level categories are functionally and epistemologically primary with respect to the following factors: gestalt perception, image formation,' motor movement, knowledge organization, ease of cognitive processing (learning, recognition, memory, etc.), and ease of linguistic expression.

Reference-point, or "metonymic," reasoning: The idea that a part of a category (that is, a member or subcategory) can stand for the whole category in certain reasoning processes.

What unites these themes is the idea of a cognitive model:

- Cognitive models are directly *embodied* with respect to their content, or else they are systematically linked to directly embodied models. Cognitive models structure thought and are used in forming categories and in reasoning. Concepts characterized by cognitive models are understood via the embodiment of the models.
- Most cognitive models are embodied with respect to use. Those that are not are only used consciously and with noticeable effort.
- The nature of conceptual embodiment leads to *basic-level categorization* and *basic-level primacy*.
- Cognitive models are used in *reference-point, or "metonymic," reasoning.*
- *Membership gradience* arises when the cognitive model characterizing a concept contains a scale.
- *Centrality gradience* arises through the interaction of cognitive models.
- *Family resemblances* involve resemblances among models.
- *Polysemy* arises from the fact that there are systematic relationships between different cognitive models and between elements of the same model. The same word is often used for elements that stand in such cognitive relations to one another.

Thus it is the concept of a cognitive model, which we will discuss in the remainder of the book, that ties together the themes of this section.

The scholars we will be discussing in this section are those I take to be most representative of the development of these themes:

- Ludwig Wittgenstein is associated with the ideas of family resemblance, centrality, and gradience.
- J. L. Austin's views on the relationships among meanings of words are both a crystalization of earlier ideas in lexicography and historical semantics and a precursor of the-contemporary view of polysemy as involving family resemblances among meanings.
- Lotfi Zadeh began the technical study of categories with fuzzy boundaries by conceiving of a theory of fuzzy sets as a generalization of standard set theory.
- Floyd Lounsbury's generative analysis of kinship categories is an important link between the idea that a category can be generated by a generator plus rules and the idea that a category has central members (and subcategories).
- Brent Berlin and Paul Kay are perhaps best known for their research on color categories, which empirically established the ideas of centrality and gradience.
- Paul Kay and Chad McDaniel put together color research from anthropology and neurophysiology and established the importance of the embodiment of concepts and the role that embodiment plays in determining centrality.
- Roger Brown began the study of what later became known as "basic-level categories." He observed that there is a "first level" at which children learn object categories and name objects, which is neither the most general nor most specific level. This level is characterized by distinctive actions, as well as by shorter and more frequently used names. He saw this level of categorization as "natural," whereas he viewed higher-level and lower-level categorization as "achievements of the imagination."
- Brent Berlin and his associates, in research on plant and animal naming, empirically established for these domains many of the fundamental ideas associated with basic-level categorization and basic-level primacy. They thereby demonstrated that embodiment determines some of the most significant properties of human categories.
- Paul Ekman and his Co-workers have shown that there are universal basic human emotions that have physical correlates in facial expressions and the autonomic nervous system. He thereby confirmed such ideas as basic-level concepts, basic-level primacy, and centrality while demonstrating that emotional concepts are embodied.

– Eleanor Rosch saw the generalizations behind such studies of particular cases and proposed that thought in general is organized in terms of prototypes and basic-level structures. It was Rosch who saw categorization itself as one of the most important issues in cognition. Together with Carolyn Mervis and other Co-workers, Rosch established research paradigms in cognitive psychology for demonstrating centrality, family resemblance, basic-level categorization, basic-level primacy, and reference-point reasoning, as well as certain kinds of embodiment. Rosch is perhaps best known for developing experimental paradigms for determining subjects' ratings of how good an example of a category a member is judged to be. Rosch ultimately realized that these ratings do not in themselves constitute models for representing category structure. They are effects that are inconsistent with the classical theory and that place significant constraints on what an adequate account of categorization must be.

These scholars all played a significant role in the history of the paradigm we will be presenting. The theory of cognitive models, which we will discuss later, attempts to bring their contributions into a coherent paradigm.

There are some notable omissions from our short survey. Since graded categories will be of only passing interest to us, I will not be mentioning much of the excellent work in that area. Graded categories are real. To my knowledge, the most detailed empirical study of graded categories is Kempton's thoroughly documented book on cognitive prototypes with graded extensions (Kempton 1981). It is based on field research in Mexico on the categorization of pottery. I refer the interested reader to that superb work, as well as to Labov's classic 1973 paper. I will also have relatively little to say about fuzzy set theory, since it is also tangential to our concerns here. Readers interested in the extensive literature that has developed on the theory of fuzzy sets and systems should consult (Dubois and Prade 1980). There is also a tradition of research in cognitive psychology that will not be surveyed here. Despite Rosch's ultimate refusal to interpret her goodness-of-example ratings as constituting a representation of category structure, other psychologists have taken that path and have given what I call an **EFFECTS = STRUCTURE INTERPRETATION** to Rosch's results. Smith and Medin (1980) have done an excellent survey of research in cognitive psychology that is based on this interpretation. In chapter 9 below, I will argue that the **EFFECTS = STRUCTURE INTERPRETATION** is in general inadequate.

Let us now turn to our survey.

Summary

The basic results of prototype theory leading up to the cognitive models approach can be summarized as follows:

- Some categories, like *tall man* or *red*, are graded; that is, they have inherent degrees of membership, fuzzy boundaries, and central members whose degree of membership (on a scale from zero to one) is one.
- Other categories, like *bird*, have clear boundaries; but within those boundaries there are graded prototype effects—some category members are better examples of the category than others.
- Categories are not organized just in terms of simple taxonomic hierarchies. Instead, categories "in the middle" of a hierarchy are the most *basic*, relative to a variety of psychological criteria: gestalt perception, the ability to form a mental image, motor interactions, and ease of learning, remembering, and use. Most knowledge is organized at this level.
- The basic level depends upon perceived part-whole structure and corresponding knowledge about how the parts function relative to the whole.
- Categories are organized into systems with contrasting elements.
- Human categories are not objectively "in the world," external to human beings. At least some categories are *embodied*. Color categories, for example, are determined jointly by the external physical world, human biology, the human mind, plus cultural considerations. Basic-level structure depends on human perception, imaging capacity, motor capabilities, etc.
- The properties relevant to the description of categories are *interactional properties*, properties characterizable only in terms of the interaction of human beings as part of their environment. Prototypical members of categories are sometimes describable in terms of *clusters* of such interactional properties. These clusters act as gestalts: the cluster as a whole is psychologically simpler than its parts.
- Prototype effects, that is, asymmetries among category members such as goodness-of-example judgments, are superficial phenomena which may have many sources.

The cognitive models approach to categorization is an attempt to make sense of all these observations. It is motivated by

- a need to understand what kinds of prototype effects there are and what their sources are

- a need to account for categorization not merely for physical objects but in abstract conceptual domains—emotions, spatial relations, social relationships, language, etc.
- a need for empirical study of the nature of cognitive models
- a need for appropriate theoretical and philosophical underpinnings for prototype theory.

These needs will be addressed below. But before we begin, it is important to see that prototype effects occur not only in nonlinguistic conceptual structure, but in linguistic structure as well. The reason is that linguistic structure makes use of general cognitive apparatus, such as category structure. Linguistic categories are kinds of cognitive categories.

CHAPTER 3

Prototype Effects in Language

One of the principal claims of this book is that language makes use of our general cognitive apparatus. If this claim is correct, two things follow:

- Linguistic categories should be of the same type as other categories in our conceptual system. In particular, they should show prototype and basic-level effects.
- Evidence about the nature of linguistic categories should contribute to a general understanding of cognitive categories in general. Because language has such a rich category structure and because linguistic evidence is so abundant, the study of linguistic categorization should be one of the prime sources of evidence for the nature of category structure in general.

Thus, we need to ask the general question: What evidence is there that language shows prototype and basic-level effects?

The issue is a profound one, because it is by no means obvious that the language makes use of our general cognitive apparatus. In fact, the most widely accepted views of language within both linguistics and the philosophy of language make the opposite assumption: that language is a separate "modular" system *independent* of the rest of cognition. The independence of grammar from the rest of cognition is perhaps the most fundamental assumption on which Noam Chomsky's theory of language rests. As we shall see in chapter 14, the very idea that language is a "formal system" (in the technical mathematical sense used by Chomsky and many other linguistic theorists) requires the assumption that language is independent of the rest of cognition. That formal-system view also embodies the implicit assumption that categories are classical (and hence can be characterized by distinctive features). Such views are also the norm in the philosophy of language, especially in the work of Richard Montague, Donald Davidson, David Lewis, Saul Kripke, and many others.

Thus, the question of what linguistic categories are like is important in two ways.

First, it affects our understanding of what language is. Does language make use of general cognitive mechanisms? Or is it something separate and independent, using only mechanisms of its own? How this question is answered will determine the course of the future study of language. Entirely different questions will be asked and theories proposed depending on the answer.

Second, the answer will affect the study of cognition, since it will determine whether linguistic evidence is admissible in the study of the mind in general.

It is for these reasons that it is important to look closely at studies that have revealed the existence of prototype effects in language.

There are actually two bodies of relevant studies. One is a body of research based on Phases I and II of Rosch's research on prototype theory. It is concerned with demonstrating the existence of prototype effects in language. The second body of research focuses on the cognitive model interpretation of prototype effects that we will be discussing below. The present chapter is a survey of the first body of results, which show little more than the existence of prototype effects in language. Chapters 4 through 8 and the three case studies at the end of the book will survey the second body of results, which focus more on the nature of the effects.

Prototype Effects in Linguistic Categories

The study of prototype effects has a long tradition in linguistics. The kinds of effects that have been studied the most are asymmetries within categories and gradations away from a best example.

Markedness

The study of certain types of asymmetries within categories is known within linguistics as the study of *markedness*. The term *markedness* arises from the fact that some morphological categories have a "mark" and others are "unmarked." Take the category of number in English. Plural number has a "mark," the morpheme *-s*, as in *boys*, while singular number lacks any overt "mark," as in *boy*. The singular is thus the unmarked member of the morphological category *number* in English. Thus, singular and plural—the two members of the *number* category—show an asymmetry; they are not treated the same in English, since singular has no overt mark. The intuition that goes along with this is that singular is, somehow, cognitively simpler than plural and that its cognitive simplicity is reflected

in its shorter form. The idea here is that simplicity in cognition is reflected in simplicity of form. Zero-marking for a morpheme is one kind of simplicity.

In phonology, markedness is often understood in terms of some notion of relative ease of articulation. For example, the consonants *p*, *t*, and *k* are voiceless, that is, they do not involve the vibration of the vocal chords, while the minimally contrasting voiced consonants *b*, *d*, and *g* do involve vocal cord vibration. Thus, one can understand voicing as a "mark" added to voiceless consonants to yield voiced consonants, except between vowels where the vocal cords are vibrating to produce the vowels. In that situation, the voiced consonants are unmarked and the voiceless consonants are marked. Thus, there is an asymmetry in terms of relative ease of articulation. Voiced and voiceless consonants also show an asymmetry in the way they pattern in the sound systems of languages. For example, many languages do not have both voiced and voiceless consonants. If voicing and voicelessness were symmetric, one might expect an equal number of languages to have only voiceless or only voiced consonants. But in such a situation, the norm is for such a language to have voiceless consonants. Similarly, within a language, there are environments where it is impossible to have both voiced and voiceless consonants. For example, in English, after initial *s-*, there is no contrast between voiced and voiceless consonants. Only voiceless consonants may occur. English has words like *spot*, but no contrasting words like *sbot.* Similarly, at the end of words in German, there is no contrast between voiced and voiceless stop consonants. Only the voiceless consonants can occur. Thus, for example, /d/ is pronounced as [t]. In general, where the contrast is neutralized (that is, only one member of the pair can occur), the one which occurs is "unmarked" in that environment.

Neutralization of contrasts can also occur in semantics. Consider contrasts like *tall-short, happy-sad*, etc. These pairs are not completely symmetric. For example, if one asks *How tall is Harry?* one is not suggesting that Harry is tall, but if one asks *How short is Harry?* one is suggesting that Harry is short. Only one member of the pair *tall-short* can be used with a neutral meaning, namely, *tall.* Since it occurs in cases where the contrast is neutralized, *tall* is referred to as the "unmarked" member of the *tall-short* contrast set. Correspondingly, it is assumed that tallness is cognitively more basic than shortness and the word marking the cognitively basic dimension occurs in neutral contexts.

In general, markedness is a term used by linguists to describe a kind of prototype effect—an asymmetry in a category, where one member or subcategory is taken to be somehow more basic than the other (or

others). Correspondingly, the unmarked member is the default value—the member of the category that occurs when only one member of the category can occur and all other things are equal.

Other Prototype Effects

Prototype effects have shown up in all areas of language—phonology, morphology, syntax, and semantics. In all cases, they are inconsistent with the classical theory of categories and are in conflict with current orthodoxies in the field which assume the correctness of the classical theory. Here is a sampling of studies which have shown prototype effects.

Phonology

There is no more fundamental distinction in linguistics than the distinction between a *phone* and a *phoneme.* A phone is a unit of speech sound, while a phoneme is a cognitive element understood as occurring "at a higher level" and usually represented by a phone. For example, English has a phoneme /k/ (sometimes spelled with the letter *c* in English orthography) which occurs in the words *cool, keel, key, school,* and *flak.* If attention is payed to details of pronunciation, it turns out that /k/ is pronounced differently in these words: aspirated velar [k^h] in *cool,* aspirated palatal [k'^{h}] in *keel,* unaspirated velar [k] in *school,* and unaspirated palatal [k'] in *ski.* English speakers perceive these, despite their differences in pronunciation, as being instances of the same phoneme /k/. However, there are other languages in which [k^h] and [k] are instances of different phonemes, and others still in which [k'] and [k] are instances of different phonemes.

Jeri Jaeger (1980) has replicated Rosch's experiments in the domain of phonology. She suggests, on the basis of experimental evidence, that phonemes are prototype-based categories of phones. Thus, the phoneme /k/ in English is the category consisting of the phones [k], [k^h], [k'], and [k'^{h}] with [k] as the prototypical member. Phonemic categories in general are understood in terms of their prototypical members. The nonprototypical phones are related to the prototype by phonological rules. Jaeger's results, if correct, indicate that phonological categorization, like other cognitive categorization, shows prototype effects. Her results contradict most contemporary phonological theories, which take the classical theory of categorization for granted. They point in the direction of a unification of phonology and other aspects of cognition.

Jaeger's other experimental results show:

- In English, the [k] after word-initial [s] is part of the /k/ phoneme and not either the /g/ phoneme or some velar archiphoneme.

- In English, the affricates [tš] and [dž] are unitary phonemes from a cognitive point of view.
- English speakers consider the following vowel pairs to belong together in a psychologically unified set: [ey-ae], [i-ɛ], [ow-a], [u-ʌ]. The source of the speaker's knowledge about this set of alternations is the orthographic system of English.
- Phonetic features in general have psychological reality, but not all the features proposed in various theories do. [Continuant], [sonorant], and [voice] are confirmed as real by the experiments, but [anterior] is brought into question.
- Phonetic features are not binary, but consist of a dimension along which segments can have varying values.
- A psychologically real theory must allow for the possibility of more than one correct feature assignment for a segment.

The application of Rosch's experimental techniques to phonology is a real innovation that requires a thorough reevaluation of phonological theory.

Morphology

Bybee and Moder (1983) have shown that English strong verbs like *string/strung* form a morphological category that displays prototype effects. They argue that verbs that form their past tense with ʌ (spelled *u* in English orthography) form a prototype-based category. The verbs include: *spin, win, cling, fling, sling, sting, string, swing, wring, hang, stick, strike, slink, stick, sneak, dig*, and some others that have recently developed similar past tense forms in certain dialects, e.g., *bring, shake.* On the basis of experimental results, they argue that the category has a prototype with the following properties:

It begins with *s* followed by one or two consonants: sC(C)-.
It ends with the velar nasal: /ŋ/.
It has a lax high front vowel: *I.*

Although the verbs in the category cannot be defined by common features, they all bear family resemblances to this prototype. *String, sling, swing*, and *sting* fit it exactly. The following have what Bybee and Moder analyze as "one" difference from the prototype: *cling, fling*, and *bring* have two initial consonants, but no *s*; *spin* and *stick* have the right initial consonant cluster and vowel, but differ from the final consonant by one phonological property each—*spin* has a dental instead of a velar nasal and *stick* has a velar stop instead of a velar nasal. *Win* has two minimal differences: no initial *s* and a final dental nasal instead of a velar. *Strike* also has two differences: a nonnasal final consonant and a different vowel.

This category can be categorized by a central member plus something else. In this case the "something else" is a characterization of "minimal" phonological differences: the lack of an initial *s*, the lack of nasalization, a different vowel, the difference between a velar and a dental consonant, etc. Bybee and Moder have investigated this case only and do not claim that these "minimal" differences will always count as minimal, either in English or in all languages. Without a theory of what counts as a minimal difference for morphological categorization, Bybee and Moder simply have a list of relevant differences that hold in this case. It would be interesting to see if a more general theory could be developed.

Syntax

In a number of studies ranging widely over English syntax, John Robert Ross (1972, 1973*a*, *b*, 1974, 1981) has shown that just about every syntactic category in the language shows prototype effects. These include categories like noun, verb, adjective, clause, preposition, noun phrase, verb phrase, etc. Ross has also demonstrated that syntactic constructions in English show prototype effects, for example, passive, relative WH-preposing, question WH-preposing, topicalization, conjunction, etc.

Let us consider one of Ross's examples: nouns. Ross's basic insight is that normal nouns undergo a large range of grammatical processes in English, while less nouny nouns do not undergo the full range of processes that apply to nouns in general. Moreover, even nouns that, in most constructions, are excellent examples of nouns may be less good examples in special constructions. Consider the nouns *toe, breath, way*, and *time*, as they occur in the expressions:

to stub one's toe
to hold one's breath
to lose one's way
to take one's time

These all look superficially as if they have the same structure. But, as Ross demonstrates, within these expressions *toe* is nounier than *breath*, which is nounier than *way*, which is nounier than *time*. Ross (1981) gives three syntactic environments that demonstrate the hierarchy. Starred sentences indicate ill-formedness.

I. Modification by a passive participle
A stubbed toe can be very painful.
**Held breath* is usually fetid when released.
**A lost way* has been the cause of many a missed appointment.
**Taken time* might tend to irritate your boss.

II. Gapping
I stubbed my toe, and she hers.
I held my breath, and she hers.
*I lost my way, and she hers.
*I took my time, and she hers.

III. Pluralization
Betty and Sue stubbed their toes.
*Betty and Sue stubbed their toe.

Betty and Sue held their breaths.
Betty and Sue held their breath.

*Betty and Sue lost their ways.
Betty and Sue lost their way.

*Betty and Sue took their times.
Betty and Sue took their time.

Ross's tests do not differentiate *way* and *time*. Here is a further test environment that confirms Ross's judgment:

IV. Pronominalization
I stubbed my toe, but didn't hurt *it*.
Sam held his breath for a few seconds and then released *it*.
Harry lost his way, but found *it* again.
*Harry took his time, but wasted *it*.

In each of these cases, the nounier nouns follow the general rule (that is, they behave the way one would expect nouns to behave), while the less nouny nouns do not follow the rule. As the sentences indicate, there is a hierarchy of nouniness among the examples given. Rules differ as to how nouny a noun they require. As Ross has repeatedly demonstrated, examples like these are rampant in English syntax.

More recently, Hopper and Thompson (1984) have proposed that the prototypical members of the syntactic categories *noun* and *verb* can be defined in terms of semantic and discourse functions. They provide an account with examples from a wide range of languages that indicate that nouns and verbs have prototypical functions in discourses.

Subject, Agent, and Topic

Bates and MacWhinney (1982) proposed on the basis of language acquisition data that prototype theory can be used to characterize the grammatical relation SUBJECT in the following way:

– A prototypical SUBJECT is both AGENT and TOPIC.

Van Oosten (1984) has found a wide range of evidence in English substantiating this hypothesis and expanding it to include the following:

- AGENT and TOPIC are both natural categories centering around prototypes.
- Membership in the category SUBJECT cannot be completely predicted from the properties of agents and topics.

As usual in prototype-based categories, things that are very close to prototypical members will most likely be in the category and be relatively good examples. And as expected, the boundary areas will differ from language to language. Category membership will be motivated by (though not predicted from) family resemblances to prototypical members.

- Noun phrases that are neither prototypical agents nor prototypical topics can be subjects—and relatively good examples of subjects—providing that they have important agent and topic properties.
- This permits what we might call a "prototype-based universal." SUBJECT IS A CATEGORY WHOSE CENTRAL MEMBERS ARE BOTH PROTOTYPICAL AGENTS AND PROTOTYPICAL TOPICS.

This characterization of subject is semantically based, but not in the usual sense; that is, it does not attempt to predict all subjects from semantic and pragmatic properties. But it does define the prototype of the category in semantic and pragmatic terms. Noncentral cases will differ according to language-particular conventions. The subject category is thus what we will refer to in chapter 6 as a *radial category*. In this case, the center, or prototype, of the category is predictable. And while the noncentral members are not predictable from the central member, they are "motivated" by it, in the sense that they bear family resemblances to it. *Motivation* in this sense will be discussed in great detail below.

Perhaps the most striking confirmation of the Bates-MacWhinney hypothesis comes from Van Oosten's study of the uses of the passive in English. Van Oosten picked out passive sentences as they occurred in transcribed conversation and compiled a list of all the uses. The list seemed random. She then compared her list of uses of the passive with her list of the properties of prototypical agents and topics. What she noticed was a remarkable correlation. According to the Bates-MacWhinney hypothesis, the subjects of simple active sentences should be capable of displaying all the properties of agents and topics. We can view this as a conjunction of the following form, where each P_i is either an agent property or a topic property:

$$P_1 \ \& \ P_2 \ \& \ \ldots \ \& \ P_n.$$

Passive sentences are used for various reasons—whenever no single noun phrase has all the agent and topic properties. Thus, passives (on the Bates-MacWhinney hypothesis) should occur when the subject of the passive sentence *fails* to have one of the prototypical agent or topic properties. Thus, the uses of the passive should be a disjunction of the form:

not P_1 or not P_2 or . . . or not P_n.

This was in fact just the list of uses of the passive that Van Oosten had compiled in her empirical study!

For example, among the agent properties are volition (call it P_1) and primary responsibility for the action (call it P_2). Correspondingly, passives can be used to indicate that an action was accidental (not P_1) or to avoid placing responsibility on the person performing the action (not P_2). Similarly, one of the topic properties of a prototypical simple active sentence is that the actor is already under discussion in the discourse (call this P_3). Correspondingly, a passive may be used to introduce (not P_3) the actor into the discourse, by placing the actor in the *by*-phrase. In this way, prototype theory enables Van Oosten to explain why the passive is used as it is. Van Oosten's analysis also provides evidence that supports the conception of subject as a category whose prototypical subcategory is predictable from semantic and pragmatic considerations.

Basic Clause Types

Just about all of the considerable number of contemporary theories of grammar recognize an asymmetry among types of clauses in a given language. In certain clauses, there is a "natural" or "direct" relationship between the meaning of the clause and the grammar of the clause. In English, for example, simple active declarative sentences—*Sam ate a peach, Max is in the kitchen, Harry drives a sports car, That fact is odd*, etc.—are usually taken as examples of that natural (or direct) relationship. Other kinds of clause types are usually considered as deviations from the basic clause type. Here is a handful of standard examples of such "deviations":

Passive: The peach was eaten by Sam.
Existential *There*-sentences: There is a man in the kitchen.
Patient subject sentences: This car drives easily.
Extrapositions: It is odd that Maxine eats pears.
WH-questions: What did Sam eat?

Different theories of grammar treat such basic clause types by different theoretical means. Harris (1957) hypothesized "kernel sentences." Chomsky (1965) hypothesized "deep structures." And virtually every theory of grammar since then has made some such distinction. What is of

interest in this context is the asymmetry. The basic clauses show a privileged relationship between meaning and grammar; the nonbasic clause types do not show that relationship. Within the category of clause types in a language, the subcategory of basic clause types has a privileged status. This asymmetry between basic clause types and other clause types is a kind of prototype effect. Within the theory of grammatical constructions, described in case study 3 below, such prototype effects in grammar are characterized in the same way as other prototype effects, using the general theory of cognitive models, which is set out in the remainder of this book.

Summary

Linguistic categories, like conceptual categories, show prototype effects. Such effects occur at every level of language, from phonology to morphology to syntax to the lexicon. I take the existence of such effects as prima facie evidence that linguistic categories have the same character as other conceptual categories. At this point I will adopt it as a working hypothesis that language does make use of general cognitive mechanisms—at least categorization mechanisms. Under this working hypothesis, we will use linguistic evidence to study the cognitive apparatus used in categorization. On the basis of all of the available evidence, I will argue in chapters 9–17 that our working hypothesis is indeed correct and that as a result our understanding of both language and cognition in general must be changed considerably.

CHAPTER 4

Idealized Cognitive Models

Sources of Prototype Effects

The main thesis of this book is that we organize our knowledge by means of structures called *idealized cognitive models*, or ICMs, and that category structures and prototype effects are by-products of that organization. The ideas about cognitive models that we will be making use of have developed within cognitive linguistics and come from four sources: Fillmore's frame semantics (Fillmore 1982*b*), Lakoff and Johnson's theory of metaphor and metonymy (Lakoff and Johnson 1980), Langacker's cognitive grammar (Langacker 1986), and Fauconnier's theory of mental spaces (Fauconnier 1985). Fillmore's frame semantics is similar in many ways to schema theory (Rumelhart 1975), scripts (Schank and Abelson 1977), and frames with defaults (Minsky 1975). Each ICM is a complex structured whole, a gestalt, which uses four kinds of structuring principles:

- propositional structure, as in Fillmore's frames
- image-schematic structure, as in Langacker's cognitive grammar
- metaphoric mappings, as described by Lakoff and Johnson
- metonymic mappings, as described by Lakoff and Johnson

Each ICM, as used, structures a mental space, as described by Fauconnier.

Probably the best way to provide an idea of what ICMs are and how they work in categorization is to go through examples. Let us begin with Fillmore's concept of a *frame*. Take the English word *Tuesday*. *Tuesday* can be defined only relative to an idealized model that includes the natural cycle defined by the movement of the sun, the standard means of characterizing the end of one day and the beginning of the next, and a larger seven-day calendric cycle—the week. In the idealized model, the week is a whole with seven parts organized in a linear sequence; each part is called a *day*, and the third is *Tuesday*. Similarly, the concept *weekend* re-

quires a notion of a *work week* of five days followed by a break of two days, superimposed on the seven-day calendar.

Our model of a week is idealized. Seven-day weeks do not exist objectively in nature. They are created by human beings. In fact, not all cultures have the same kinds of weeks. Consider, for example, the Balinese calendric system:

> The two calendars which the Balinese employ are a lunar-solar one and one built around the interaction of independent cycles of day-names, which I shall call "permutational." The permutational calendar is by far the most important. It consists of ten different cycles of day-names, following one another in a fixed order, after which the first day-name appears and the cycle starts over. Similarly, there are nine, eight, seven, six, five, four, three, two, and even—the ultimate of a "contemporized" view of time—one day-name cycles. The names in each cycle are also different, and the cycles run concurrently. That is to say, any given day has, at least in theory, ten different names simultaneously applied to it, one from each of the ten cycles. Of the ten cycles, only those containing five, six, and seven day-names are of major cultural significance. . . . The outcome of all this wheels-within-wheels computation is a view of time as consisting of ordered sets of thirty, thirty-five, forty-two and two hundred and ten quantum units ("days"). . . . To identify a day in the forty-two-day set—and thus assess its practical and/or religious significance—one needs to determine its place, that is, its name in the six-name cycle (say *Ariang*) and in the seven-day cycle (say *Boda*): the day is Boda-Ariang, and one shapes one's actions accordingly. To identify a day in the thirty-five day set, one needs its place and name in the five-name cycle (for example, *Klion*) and in the seven-: for example, *Boda-Klion*. . . . For the two-hundred-and-ten-day set, unique determination demands names from all three weeks: for example, *Boda-Ariang-Klion*, which, it so happens, is the day on which the most important Balinese holiday, Galungan, is celebrated. (Geertz 1973, pp. 392–93)

Thus, a characterization of *Galungan* in Balinese requires a complex ICM which superimposes three week-structures—one five-day, one six-day, and one seven-day. In the cultures of the world, such idealized cognitive models can be quite complex.

The Simplest Prototype Effects

In general, any element of a cognitive model can correspond to a conceptual category. To be more specific, suppose schema theory in the sense of Rumelhart (1975) were taken as characterizing propositional models. Each schema is a network of nodes and links. Every node in a schema would then correspond to a conceptual category. The properties of the category would depend on many factors: the role of that node in the given

schema, its relationship to other nodes in the schema, the relationship of that schema to other schemas, and the overall interaction of that schema with other aspects of the conceptual system. As we will see, there is more to ICMs than can be represented in schema theory. But at least those complexities do arise. What is particularly interesting is that even if one set up schema theory as one's theory of ICMs, and even if the categories defined in those schemas were classical categories, there would still be prototype effects—effects that would arise from the interaction of the given schema with other schemas in the system.

A clear example of this has been given by Fillmore (1982*a*). The example is a classic: the category defined by the English word *bachelor.*

> The noun *bachelor* can be defined as an unmarried adult man, but the noun clearly exists as a motivated device for categorizing people only in the context of a human society in which certain expectations about marriage and marriageable age obtain. Male participants in long-term unmarried couplings would not ordinarily be described as bachelors; a boy abandoned in the jungle and grown to maturity away from contact with human society would not be called a bachelor; John Paul II is not properly thought of as a bachelor.

In other words, *bachelor* is defined with respect to an ICM in which there is a human society with (typically monogamous) marriage, and a typical marriageable age. The idealized model says nothing about the existence of priests, "long-term unmarried couplings," homosexuality, Moslems who are permitted four wives and only have three, etc. With respect to this idealized cognitive model, a *bachelor* is simply an unmarried adult man.

This idealized model, however, does not fit the world very precisely. It is oversimplified in its background assumptions. There are some segments of society where the idealized model fits reasonably well, and when an unmarried adult man might well be called a bachelor. But the ICM does not fit the case of the pope or people abandoned in the jungle, like Tarzan. In such cases, unmarried adult males are certainly not representative members of the category of bachelors.

The theory of ICMs would account for such prototype effects of the category *bachelor* in the following way: An idealized cognitive model may fit one's understanding of the world either perfectly, very well, pretty well, somewhat well, pretty badly, badly, or not at all. If the ICM in which *bachelor* is defined fits a situation perfectly and the person referred to by the term is unequivocally an unmarried adult male, then he qualifies as a member of the category *bachelor.* The person referred to deviates from prototypical bachelorhood if either the ICM fails to fit the world perfectly or the person referred to deviates from being an unmarried adult male.

Under this account *bachelor* is not a graded category. It is an all-or-none concept relative to the appropriate ICM. The ICM characterizes representative bachelors. One kind of gradience arises from the degree to which the ungraded ICM fits our knowledge (or assumptions) about the world.

This account is irreducibly cognitive. It depends on being able to take two cognitive models—one for *bachelor* and one characterizing one's knowledge about an individual, say the pope—and compare them, noting the ways in which they overlap and the ways in which they differ. One needs the concept of "fitting" one's ICMs to one's understanding of a given situation and keeping track of the respects in which the fit is imperfect.

This kind of explanation cannot be given in a noncognitive theory—one in which a concept either fits the world as it is or not. The background conditions of the *bachelor* ICM rarely make a perfect seamless fit with the world as we know it. Still we can apply the concept with some degree of accuracy to situations where the background conditions don't quite mesh with our knowledge. And the worse the fit between the background conditions of the ICM and our knowledge, the less appropriate it is for us to apply the concept. The result is a gradience—a simple kind of prototype effect.

Lie

A case similar to Fillmore's *bachelor* example, but considerably more complex, has been discussed by Sweetser (1984). It is the category defined by the English word *lie*. Sweetser's analysis is based on experimental results by Coleman and Kay (1981) on the use of the verb *lie*. Coleman and Kay found that their informants did not appear to have necessary and sufficient conditions characterizing the meaning of *lie*. Instead they found a cluster of three conditions, no one of which was necessary and all of which varied in relative importance:

> A consistent pattern was found: falsity of belief is the most important element of the prototype of *lie*, intended deception the next most important element, and factual falsity is the least important. Informants fairly easily and reliably assign the word *lie* to reported speech acts in a more-or-less, rather than all-or-none, fashion, . . . [and] . . . informants agree fairly generally on the relative weights of the elements in the semantic prototype of *lie*.

Thus, there is agreement that if you steal something and then claim you didn't, that's a good example of a lie. A less representative example of a lie is when you tell the hostess "That was a great party!" when you were bored stiff. Or if you say something true but irrelevant, like "I'm going to

the candy store, Ma" when you're really going to the pool hall, but will be stopping by the candy store on the way.

An important anomaly did, however, turn up in the Coleman-Kay study. When informants were asked to define a *lie*, they consistently said it was a false statement, even though actual falsity turned out consistently to be the least important element by far in the cluster of conditions. Sweetser has observed that the theory of ICMs provides an elegant way out of this anomaly. She points out that, in most everyday language use, we take for granted an idealized cognitive model of social and linguistic interaction. Here is my revised and somewhat oversimplified version of the ICM Sweetser proposes:

THE MAXIM OF HELPFULNESS

People intend to help one another.

This is a version of Grice's cooperative principle.

THE ICM OF ORDINARY COMMUNICATION

(*a*) If people say something, they're intending to help if and only if they believe it.

(*b*) People intend to deceive if and only if they don't intend to help.

THE ICM OF JUSTIFIED BELIEF

(*c*) People have adequate reasons for their beliefs.

(*d*) What people have adequate reason to believe is true.

These two ICMs and the maxim of helpfulness govern a great deal of what we consider ordinary conversation, that is, conversation not constrained by special circumstances. For example, if I told you I just saw a mutual friend, under ordinary circumstances you'd probably assume I was being helpful, that I wasn't trying to deceive you, that I believed I had seen the friend, and that I did in fact see the friend. That is, unless you have reason to believe that the maxim of helpfulness is not applying or that one of these idealized models is not applicable, you would simply take them for granted.

These ICMs provide an explanation of why speakers will define a lie as a false statement, when falsity is by far the least important of the three factors discovered by the Kay-Coleman study. These two ICMs each have an internal logic and when they are taken together, they yield some interesting inferences. For example, it follows from (*c*) and (*d*) that if a person believes something, he has adequate reasons for his beliefs, and if he has adequate reasons for believing the proposition, then it is true. Thus, in the idealized world of these ICMs if *X* believes a proposition *P*, then *P* is true. Conversely, if *P* is false, then *X* doesn't believe *P*. Thus, falsity entails lack of belief.

In this idealized situation, falsity also entails an intent to deceive. As we have seen, falsity entails a lack of belief. By (*a*), someone who says something is intending to help if and only if he believes it. If he doesn't believe it, then he isn't intending to help. And by (*b*), someone who isn't intending to help in giving information is intending to deceive. Thus, in these ICMs, falsity entails both lack of belief and intent to deceive. Thus, from the definition of a lie as a false statement, the other properties of lying follow as consequences. Thus, the definition of *lie* does not need to list all these attributes. If *lie* is defined relative to these ICMs, then lack of belief and intent to deceive follow from falsity.

As Sweetser points out, the relative importance of these conditions is a consequence of their logical relations given these ICMs. Belief follows from a lack of intent to deceive and truth follows from belief. Truth is of the least concern since it is a consequence of the other conditions. Conversely, falsity is the most informative of the conditions in the idealized model, since falsity entails both intent to deceive and lack of belief. It is thus falsity that is the defining characteristic of a lie.

Sweetser's analysis provides both a simple, intuitive definition of *lie* and an explanation of all of the Coleman-Kay findings. The ICMs used are not made up just to account for *lie*. Rather they govern our everyday common sense reasoning. These results are possible because the ICMs have an internal logic. It is the *structure* of the ICMs that explains the Coleman-Kay findings.

Coleman and Kay discovered prototype effects for the category *lie*—situations where subjects gave uniform rankings of how good an example of a lie a given statement was. Sweetser's analysis explains these rankings on the basis of her ICM analysis, even though her ICM fits the classical theory! Nonprototypical cases are accounted for by imperfect fits of the lying ICM to knowledge about the situation at hand. For example, white lies and social lies occur in situations where condition (*b*) does not hold. A white lie is a case where deceit is not harmful, and a social lie is a case where deceit is helpful. In general, expressions such as *social lie, white lie, exaggeration, joke, kidding, oversimplification, tall tale, fiction, fib, mistake*, etc. can be accounted for in terms of systematic deviations from the above ICMs.

Although neither Sweetser nor anyone else has attempted to give a theory of complex concepts in terms of the theory of ICMs, it is worth considering what would be involved in doing so. As should be obvious, adjective-noun expressions like *social lie* do not work according to traditional theories. The category of social lies is not the intersection of the set of social things and the set of lies. The term *social* places one in a domain of experience characterized by an ICM that says that being polite is more

important than telling the truth. This conflicts with condition (*b*), that intent to deceive is not helpful, and it overrides this condition. Saying "That was a great party!" when you were bored stiff is a case where deception is helpful to all concerned. It is a prototypical social lie, though it is not a prototypical lie. The concept *social lie* is therefore represented by an ICM that overlaps in some respects with the lying ICM, but is different in an important way. The question that needs to be answered is whether the addition of the modifier *social* can account for this difference systematically. Any general account of complex concepts like *social lie* in terms of ICMs will have to indicate how the ICM evoked by *social* can cancel one condition of the ICM evoked by *lie*, while retaining the other conditions. An obvious suggestion would be that in conflicts between modifiers and heads, the modifiers win out. This would follow from the general cognitive principle that special cases take precedence over general cases.

CHAPTER 21

Overview

We have now completed everything but the case studies. Let us review the territory we have covered. We set out to argue for an experientialist view of reason and against the objectivist view. Here were the first things that had to be shown:

- Meaningful thought is not merely the manipulation of abstract symbols that are meaningless in themselves and get their meaning only by virtue of correspondences to things in the world.
- Reason is not abstract and disembodied, a matter of instantiating some transcendental rationality.
- The mind is thus not simply a "mirror of nature," and concepts are not merely "internal representations of external reality."

The argument is based on the nature of categorization. Most of our concepts concern categories, not individuals (e.g., *dog* as opposed to *Fido*). If the objectivist view were correct, the following would have to be true of categories:

- Conceptual categories would have to be symbolic structures that get their meaning only by virtue of corresponding to objectively existing categories in the world (the world as it actually is or some possible state of the world).
- Categories in the world would have to be characterized objectively, in terms of objective properties of their members and not in any way taking into account the nature of the beings doing the categorization.
- Conceptual categories could only be mental representations of categories in the world.
- Conceptual categories, being mental representations of categories in the world, would have to mirror the structure of categories in the world, excluding anything that was not a reflection of the properties of the category members. Otherwise, they would not be true internal

representations of external reality and could not represent true knowledge of the external world.
- Conceptual categories must thus have the same structure as categories in the world: the structure of classical categories.

The classical theory of categories is thus central to the objectivist view of mind. It views categories as being defined solely by the objectively given properties shared by the members of the category.

Our goal was to show that the classical theory was wrong (1) for conceptual categories, (2) for categories in the world, and (3) for the hypothesized relationship between conceptual categories and categories in the world. Our strategy was to demonstrate three things:

1. Conceptual categories are not merely characterized in terms of objective properties of category members. They differ in two respects:

- Human conceptual categories have properties that are, at least in part, determined by the bodily nature of the people doing the categorizing rather than solely by the properties of the category members.
- Human conceptual categories have properties that are a result of imaginative processes (metaphor, metonymy, mental imagery) that do not mirror nature.

2. The real world cannot be properly understood in terms of the classical theory of categories.

3. The relationship between conceptual categories and real-world categories cannot be as the objectivist view claims.

Part I of the book was dedicated to reviewing the research needed to demonstrate the first item in the list:

- Basic-level category structure reflects the *bodily nature* of the people doing the categorizing, since it depends on gestalt perception and motor movements. Color categories also depend on the nature of the human body, since they are characterized in part by human neurophysiology.
- Basic-level structure is partly characterized by human *imaginative processes*: the capacity to form mental images, to store knowledge at a particular level of categorization, and to communicate. Prototype structure also testifies to imaginative processes of many kinds: metonymy (the capacity to let one thing stand for another for some purpose), the ability to construct and use idealized models, and the ability to extend categories from central to noncentral members using imaginative capacities such as metaphor, metonymy, mythological associations, and image relationships.

Thus, we were able to show that *conceptual* categories do not fit the objectivist view of meaningful thought and reason.

Chapter 12 demonstrated the second item. By showing that biological species do not fit the classical account of categorization, we were able to show that species, which are taken to be categories in the world, are not classical categories.

Chapter 15 demonstrated the third item, namely, that the purported relationship between categories in the world and their "mental representations" could not hold. In other words, mental representations for categories cannot be given meaning via their relationship to categories in the world. This is a consequence of Putnam's theorem together with a fundamental constraint on the nature of meaning.

Having argued against the objectivist view of meaningful thought and reason, we put forth an alternative in chapter 17. On the experientialist account, meaningful thought and reason make use of symbolic structures *which are meaningful to begin with.* Those that are directly meaningful are of two sorts: basic-level concepts and kinesthetic image schemas. Basic-level concepts are directly meaningful because they reflect the structure of our perceptual-motor experience and our capacity to form rich mental images. Kinesthetic image schemas are directly meaningful because they preconceptually structure our experience of functioning in space. They also have an internal basic logic that we believe is sufficient to characterize human reason. With such a dual basis for directly meaningful symbolic structures, indirectly meaningful symbolic structures are built up by imaginative capacities (especially metaphor and metonymy). But despite the fact that we rely centrally on our bodily natures and our imaginative capacities, experientialism has maintained a form of basic realism, since our conceptual structures are strongly (though by no means totally) constrained by reality and by the way we function as an inherent part of reality.

Finally, we defended the experientialist view of reason against objections having to do with three issues—relativism, artificial intelligence, and mathematics:

- Relativism is commonly and falsely identified with total relativism. Experiential realism permits a form of relativism, though one that is not at all like total relativism. Chapter 18 surveyed the forms of relativism and showed that there is not only nothing wrong with the relativism that we propose, and that there is positive evidence for it.
- Artificial intelligence is often given an objectivist interpretation, especially by philosophers. If accepted, that interpretation would place the field at odds with the experientialist view of reason. We ob-

served in chapter 19 that such an interpretation of the endeavor of artificial intelligence is not only unnecessary but in fact goes against the practice of many researchers in the field. The study of artificial intelligence does not in any way conflict with an experientialist view of reason. It is only an interpretation of artificial intelligence in terms of objectivist philosophy that is in conflict with our views.

- The very existence of mathematical truth is sometimes cited in support of the existence of a single transcendental rationality that we can have access to. We argued in chapter 20 that if mathematics is assumed to be transcendentally true, it cannot be unique. For example, there are versions of algebra and topology that differ substantively from one another because they are based on different models of set theory. They are all transcendentally true, not absolutely, but relative to what is taken to be a "set." Thus, the mere existence of mathematical truths cannot provide evidence of a unique transcendental rationality. It is at least as plausible that mathematics arises out of *human* rational structures.

We have argued that the objectivist views on meaningful thought and ason are incorrect on both empirical and logical grounds. No doubt, denses of objectivism will be forthcoming. What is important is that obctivist views can no longer be taken for granted as being obviously true id beyond question. The questions have been asked and an alternative is been proposed. It is an alternative that opens up further inquiry into e nature of the human mind. The value of opening up such a path of inuiry can only be shown through detailed case studies of phenomena that veal something about the nature of human reason.