George Lakoff

Women, Fire, and Dangerous Things

What Categories Reveal about the Mind

The University of Chicago Press
Chicago and London

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 functioning. 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: (1) gestalt perception, (2) image formation, (3) motor movement, (4) knowledge organization, (5) ease of cognitive processing (learning, recognition, memory, etc.), and (6) 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.

Wittgenstein

Family Resemblances

The first major crack in the classical theory is generally acknowledged to have been noticed by Wittgenstein (1953, 1:66–71). The classical category has *clear boundaries,* which are defined by *common properties.* Wittgenstein pointed out that a category like *game* does not fit the classical mold, since there are no common properties shared by all games. Some games involve mere amusement, like ring-around-the-rosy. Here there is no competition—no winning or losing—though in other games there is. Some games involve luck, like board games where a throw of the dice determines each move. Others, like chess, involve skill. Still others, like gin rummy, involve both.

Though there is no single collection of properties that all games share, the category of games is united by what Wittgenstein calls *family resemblances.* Members of a family resemble one another in various ways: they may share the same build or the same facial features, the same hair color, eye color, or temperament, and the like. But there need be no single collection of properties shared by everyone in a family. Games, in this respect, are like families. Chess and go both involve competition, skill, and the use of long-term strategies. Chess and poker both involve competition. Poker and old maid are both card games. In short, games, like family members, are similar to one another in a wide variety of ways. That, and not a single, well-defined collection of common properties, is what makes *game* a category.

Extendable Boundaries

Wittgenstein also observed that there was no fixed boundary to the category *game.* The category could be extended and new kinds of games introduced, provided that they resembled previous games in appropriate ways. The introduction of video games in the 1970s was a recent case in history where the boundaries of the *game* category were extended on a large scale. One can always impose an artificial boundary for some purpose; what is important for his point is that extensions are possible, as well as artificial limitations. Wittgenstein cites the example of the category *number.* Historically, numbers were first taken to be integers and were then extended successively to rational numbers, real numbers, complex numbers, transfinite numbers, and all sorts of other kinds of numbers invented by mathematicians. One can for some purpose limit the category *number* to integers only, or rational numbers only, or real numbers only. But the category *number* is not bounded in any natural way, and it can be limited or extended depending on one's purposes.

In mathematics, intuitive human concepts like *number* must receive precise definitions. Wittgenstein's point is that different mathematicians give different precise definitions, depending on their goals. One can define *number* to include or exclude transfinite numbers, infinitesimals, inaccessible ordinals, and the like. The same is true of the concept of a *polyhedron.* Lakatos (1976) describes a long history of disputes within mathematics about the properties of polyhedra, beginning with Euler's conjecture that the number of vertices minus the number of edges plus the number of faces equals two. Mathematicians over the years have come up with counterexamples to Euler's conjecture, only to have other mathematicians claim that they had used the "wrong" definition of *polyhedron.* Mathematicians have defined and redefined *polyhedron* repeatedly to fit their goals. The point again is that there is no single well-defined intuitive category *polyhedron* that includes tetrahedra and cubes and some fixed range of other constructs. The category *polyhedron* can be given precise boundaries in many ways, but the intuitive concept is not limited in any of those ways; rather, it is open to both limitations and extensions.

Central and Noncentral Members

According to the classical theory, categories are uniform in the following respect: they are defined by a collection of properties that the category members share. Thus, no members should be more central than other members. Yet Wittgenstein's example of *number* suggests that integers are central, that they have a status as numbers that, say, complex numbers or transfinite numbers do not have. Every precise definition of *number* must include the integers; not every definition must include transfinite numbers. If anything is a number, the integers are numbers; that is not true of transfinite numbers. Similarly, any definition of polyhedra had better include tetrahedra and cubes. The more exotic polyhedra can be included or excluded, depending on your purposes. Wittgenstein suggests that the same is true of games. "Someone says to me: 'Show the children a game.' I teach them gaming with dice, and the other says 'I didn't mean that sort of game'" (1:70). Dice is just not a very good example of a game. The fact that there can be good and bad examples of a category does not follow from the classical theory. Somehow the goodness-of-example structure needs to be accounted for.

Austin

Wittgenstein assumed that there is a single category named by the word *game,* and he proposed that that category and other categories are struc-

tured by family resemblances and good and bad examples. Philosopher J. L. Austin extended this sort of analysis to the study of words themselves. In his celebrated paper, "The Meaning of a Word," written in 1940 and published in 1961, Austin asked, "Why do we call different [kinds of] things by the same name?" The traditional answer is that the kinds of things named are similar, where "similar" means "partially identical." This answer relies on the classical theory of categories. If there are common properties, those properties form a classical category, and the name applies to this category. Austin argued that this account is not accurate. He cited several classes of cases. As we will see in the remainder of this book, Austin's analysis prefigured much of contemporary cognitive semantics—especially the application of prototype theory to the study of word meaning.

If we translate Austin's remarks into contemporary terms, we can see the relationship between Austin's observation and Wittgenstein's: the senses of a word can be seen as forming a category, with each sense being a member of that category. Since the senses often do not have properties in common, there is no classical category of senses that the word could be naming. However, the senses can be viewed as forming a category of the kind Wittgenstein described. There are central senses and noncentral senses. The senses may not be similar (in the sense of sharing properties), but instead are related to one another in other specifiable ways. It is such relationships among the senses that enable those senses to be viewed as constituting a single category: the relationships provide an explanation of why a single word is used to express those particular senses. This idea is far from new. Part of the job of traditional historical semanticists, as well as lexicographers, has been to speculate on such relationships. Recent research has taken up this question again in a systematic way. The most detailed contemporary study along these lines has been done by Brugman (1981), and it will be discussed below in case study 2.

Let us now turn to Austin's examples:

> The adjective 'healthy': when I talk of a healthy body, and again of a healthy complexion, of healthy exercise: the word is *not* just being used *equivocally* . . . there is what we may call a *primary nuclear sense* of 'healthy': the sense in which 'healthy' is used of a healthy body: I call this *nuclear* because it is 'contained as a part' in the other two senses which may be set out as 'productive of healthy bodies' and 'resulting from a healthy body'. . . . Now are we content to say that the exercise, the complexion, and the body are all called 'healthy' because they are similar? Such a remark cannot fail to be misleading. Why make it? (P. 71)

Austin's *primary nuclear sense* corresponds to what contemporary linguists call *central* or *prototypical* senses. The contained-as-a-part relationship is an instance of what we will refer to below as metonymy—where the part stands for the whole. Thus, given the relationships "productive of" and "resulting from," Austin's examples can be viewed in the following way:

Exercise of type *B* is productive of bodies of type *A*.
Complexion of type *C* results from bodies of type *A*.
The word *healthy* names *A*.
With respect to naming, *A* stands for *B*. (Metonymy)
With respect to naming, *A* stands for *C*. (Metonymy)

Thus, the word "healthy" has senses *A*, *B*, and *C*. *A*, *B*, and *C* form a category whose members are related in the above way. *A* is the central member of this category of senses (Austin's *primary nuclear sense*). *B* and *C* are extended senses, where metonymy is the principle of extension.

I am interpreting Austin as making an implicit psychological claim about categorization. In the very act of pointing out and analyzing the *differences* among the senses, Austin is presupposing that these senses form a natural collection for speakers—so natural that the senses have to be differentiated by an analyst. No such analysis would be needed for true homonyms, say, *bank* (where you put your money) and *bank* (of a river), which are not part of a natural collection (or category) of senses. In pointing out the existence of a small number of mechanisms by which senses are related to one another, Austin is implicitly suggesting that those mechanisms are psychologically real (rather than being just the arbitrary machinations of a clever analyst). He is, after all, trying to explain why people naturally use the same words for different senses. His implicit claim is that these mechanisms are *principles* which provide a "good reason" for grouping the senses together by the use of the same word. What I have referred to as "metonymy" is just one such mechanism.

From metonymy, Austin turns to what Johnson and I (Lakoff and Johnson 1980) refer to as metaphor, but which Austin, following Aristotle, terms "analogy."

> When *A*:*B*::*X*:*Y* then *A* and *X* are often called by the same name, e.g., the foot of a mountain and the foot of a list. Here there is a good reason for calling the things both "feet" but are we to say they are "similar"? Not in any ordinary sense. We may say that the relations in which they stand to *B* and *Y* are similar relations. Well and good: but *A* and *X* are not the relations in which they stand. (Pp. 71–72)

Austin isn't explicit here, but what seems to be going on is that both mountains and lists are being structured in terms of a metaphorical projection of the human body onto them. Expanding somewhat on Austin's analysis and translating it into contemporary terminology, we have:

A is the bottom-most part of the body.
X is the bottom-most part of the mountain.
X' is the bottom-most part of a list.
Body is projected onto mountain, with A projected onto X. (Metaphor)
Body is projected onto list, with A projected onto X'. (Metaphor)
The word "foot" names A.
A, X, and X' form a category, with A as central member. X and X' are noncentral members related to A by metaphor.

Austin also notes examples of what we will refer to below as *chaining* within a category.

> Another case is where I call B by the same name as A, because it resembles A, C by the same name because it resembles B, D . . . and so on. But ultimately A and, say D do not resemble each other in any recognizable sense at all. This is a very common case: and the dangers are obvious when we search for something 'identical' in all of them! (P. 72)

Here A is the *primary nuclear sense*, and B, C, and D are extended senses forming a chain. A, B, C, and D are all members of the same category of senses, with A as the central member.

> Take a word like 'fascist': this originally connotes a great many characteristics at once: say, x, y, and z. Now we will use 'fascist' subsequently of things which possess only *one* of these striking characteristics. So that things called 'fascist' in these senses, which we may call 'incomplete' senses, need not be similar at all to each other. (P. 72)

This example is very much like one Fillmore (1982*a*) has recently given in support of the use of prototype theory in lexical semantics. Fillmore takes the verb *climb*, as in

– John climbed the ladder.

Here, "climbing" includes both motion upward and the use of the hands to grasp onto the thing climbed. However, climbing can involve just motion upwards and no use of the hands, as in

– The airplane climbed to 20,000 feet.

Or the motion upward may be eliminated if there is grasping of the appropriate sort, as in

– He climbed out onto the ledge.

Such contemporary semantic analyses using prototype theory are very much in the spirit of Austin.

Fillmore's frame semantics is also prefigured by Austin.

> Take the sense in which I talk of a cricket bat and a cricket ball and a cricket umpire. The reason that all are called by the same name is perhaps that each has its part —its *own special* part—to play in the activity called cricketing: it is no good to say that *cricket* means simply 'used in cricket': for we cannot explain what we mean by 'cricket' *except* by explaining the special parts played in cricketing by the bat, ball, etc. (P. 73)

Austin here is discussing a holistic structure—a gestalt—governing our understanding of activities like cricket. Such activities are structured by what we call a cognitive model, an overall structure which is more than merely a composite of its parts. A modifier like *cricket* in *cricket bat, cricket ball, cricket umpire,* and so on does not pick out any common property or similarity shared by bats, balls, and umpires. It refers to the structured activity as a whole. And the nouns that *cricket* can modify form a category, but not a category based on shared properties. Rather it is a category based on the structure of the activity of cricket and on those things that are part of the activity. The entities characterized by the cognitive model of cricket are those that are in the category. What defines the category is our structured understanding of the activity.

Cognitive psychologists have recently begun to study categories based on such holistically structured activities. Barsalou (1983, 1984) has studied such categories as *things to take on a camping trip, foods not to eat on a diet, clothes to wear in the snow*, and the like. Such categories, among their other properties, do not show family resemblances among their members.

Like Wittgenstein, Austin was dedicated to showing the inadequacies of traditional philosophical views of language and mind—views that are still widely held. His contribution to prototype theory was to notice for words the kinds of things that Wittgenstein noticed for conceptual categories. Language is, after all, an aspect of cognition. Following Austin's lead, we will try to show how prototype theory generalizes to the linguistic as well as the nonlinguistic aspects of mind.

Zadeh

Some categories do not have gradations of membership, while others do. The category *U.S. Senator* is well defined. One either is or is not a senator. On the other hand, categories like *rich people* or *tall men* are graded, simply because there are gradations of richness and tallness. Lotfi Zadeh (1965) devised a form of set theory to model graded categories. He called it *fuzzy set theory*. In a classical set, everything is either in the set (has membership value 1) or is outside the set (has membership value 0). In a

fuzzy set, as Zadeh defined it, additional values are allowed between 0 and 1. This corresponds to Zadeh's intuition that some men are neither clearly tall nor clearly short, but rather in the middle—tall to some degree.

In the original version of fuzzy set theory, operations on fuzzy sets are simple generalizations of operations on ordinary sets:

> Suppose element x has membership value v in fuzzy set A and membership value w in fuzzy set B.
> Intersection: The value of x in $A \cap B$ is the minimum of v and w.
> Union: The value of x in $A \cup B$ is the maximum of v and w.
> Complement: The value of x in the complement of A is $1 - v$.

It is a natural and ingenious extension of the classical theory of sets.

Since Zadeh's original paper, other definitions for union and intersection have been suggested. For an example, see Goguen 1969. The best discussion of attempts to apply fuzzy logic to natural language is in McCawley 1981.

Lounsbury

Cognitive anthropology has had an important effect on the development of prototype theory, beginning with Floyd Lounsbury's (1964) studies of American Indian kinship systems. Take the example of Fox, in which the word *nehcihsähA* is used not only to refer to one's maternal uncle—that is, one's mother's mother's son—but also to one's mother's mother's son's son, one's mother's mother's father's son's son, one's mother's brother's son, one's mother's brother's son's son, and a host of other relatives. The same sort of treatment also occurs for other kinship categories. There are categories of "fathers," "mothers," sons," and "daughters" with just as diverse a membership.

The Fox can, of course, distinguish uncles from great-uncles from nephews. But they are all part of the same kinship category, and thus are named the same. Lounsbury discovered that such categories were structured in terms of a "focal member" and a small set of general rules extending each category to nonfocal members. The same rules apply across all the categories. The rules applying in Fox are what Lounsbury called the "Omaha type":

> Skewing rule: Anyone's father's sister, as a linking relative, is equivalent to that person's sister.
>
> Merging rule: Any person's sibling of the same sex, as a linking relative, is equivalent to that person himself.

> Half-sibling rule: Any child of one of one's parents is one's sibling.

The condition "as a linking relative" is to prevent the rule from applying directly; instead, there must be an intermediate relative between ego (the reference point) and the person being described. For example, the skewing rule does not say that a person's paternal aunt is equivalent to his sister. But it does say, for example, that his father's paternal aunt is equivalent to his father's sister. In this case, the intermediate relative is the father.

These rules have corollaries. For example,

> Skewing corollary: The brother's child of any female linking relative is equivalent to the sibling of that female linking relative. (For example, a mother's brother's daughter is equivalent to a mother's sister.)

Lounsbury illustrates how such rules would work for the Fox maternal uncle category. We will use the following abbreviations: M: mother, F: father, B: brother, S: sister, d: daughter, s: son. Let us consider the following examples of the *nehcihsähA* (mother's brother) category, and the equivalence rules that make them part of this category. Lounsbury's point in these examples is to take a very distant relative and show precisely how the same general rules place that relative in the MB (mother's brother) category. Incidentally, all the intermediate relatives in the following cases are also in the MB category—e.g., MMSs, that is, mother's mother's sister's son, etc. Let "→" stand for "is equivalent to."

1. Mother's mother's father's sister's son: MMFSs

MMFSs	→	MMSs	[by the skewing rule]
MMSs	→	MMs	[by the merging rule]
MMs	→	MB	[by the half-sibling rule]

2. Mother's mother's sister's son's son: MMSss

MMSss	→	MMss	[by the merging rule]
MMss	→	MBs	[by the half-sibling rule]
MBs	→	MB	[by the skewing corollary]

3. Mother's brother's son's son's son: MBsss

MBsss	→	MBss	[by the skewing corollary]
MBss	→	MBs	[by the skewing corollary]
MBs	→	MB	[by the skewing corollary]

Similarly, the other "uncles" in Fox are equivalent to MB.

Not all conceptual systems for categorizing kinsmen have the same skewing rules. Lounsbury also cites the Crow version of the skewing rule:

> Skewing rule: Any woman's brother, as a linking relative, is equivalent to that woman's son, as a linking relative.

Skewing corollary: The sister of any male linking relative is equivalent to the mother of that male linking relative.

These rules are responsible for some remarkable categorizations. One's paternal aunt's son is classified as one's "father." But one's paternal aunt's daughter is classified as one's "grandmother"! Here are the derivations:

Father's sister's son: FSs

FSs	→	FMs	[by skewing corrollary]
FMs	→	FB	[by half-sibling rule]
FB	→	F	[by merging rule]

Father's sister's daughter: FSd

FSd	→	FMd	[by skewing corollary]
FMd	→	FS	[by half-sibling rule]
FS	→	FM	[by skewing corollary]

Moreover, Lounsbury observed that these categories were not mere matters of naming. Such things as inheritance and social responsibilities follow category lines.

Categories of this sort—with a central member plus general rules—are by no means the norm in language, as we shall see. Yet they do occur. We will refer to such a category as a *generative* category and to its central member as a *generator*. A generative category is characterized by at least one generator plus something else: it is the "something else" that takes the generator as input and yields the entire category as output. It may be either a general principle like similarity or general rules that apply elsewhere in the system or specific rules that apply only in that category. In Lounsbury's cases, the "something else" is a set of rules that apply throughout the kinship system. The generator plus the rules generate the category.

In such a category, the generator has a special status. It is the best example of the category, the model on which the category as a whole is built. It is a special case of a prototype.

Berlin and Kay

The next major contribution of cognitive anthropology to prototype theory was the color research of Brent Berlin and Paul Kay. In their classic, *Basic Color Terms* (Berlin and Kay 1969), they took on the traditional view that different languages could carve up the color spectrum in arbitrary ways. The first regularity they found was in what they called *basic color terms*. For a color term to be basic,

- It must consist of only one morpheme, like *green*, rather than more than one, as in *dark green* or *grass-colored*.
- The color referred to by the term must not be contained within another color. *Scarlet* is, for example, contained within *red*.
- It must not be restricted to a small number of objects. *Blond*, for example, is restricted to hair, wood, and perhaps a few other things.
- It must be common and generally known, like *yellow* as opposed to *saffron*.

Once one distinguishes basic from nonbasic color terms, generalizations appear.

- Basic color terms name basic color *categories*, whose central members are the same universally. For example, there is always a psychologically real category RED, with focal red as the best, or "purest," example.
- The color categories that basic color *terms* can attach to are the equivalents of the English color categories named by the terms *black, white, red, yellow, green, blue, brown, purple, pink, orange* and *gray*.
- Although people can *conceptually* differentiate all these color categories, it is not the case that all languages make all of those differentiations. Many languages have fewer basic categories. Those categories include *unions* of the basic categories; for example, BLUE + GREEN, RED + ORANGE + YELLOW, etc. When there are fewer than eleven basic color terms in a language, one basic term, or more, names such a union.
- Languages form a hierarchy based on the number of basic color terms they have and the color categories those terms refer to.

Some languages, like English, use all eleven, while others use as few as two. When a language has only two basic color terms, they are *black* and *white*—which might more appropriately be called *cool* (covering black, blue, green, and gray) and *warm* (covering white, yellow, orange, and red). When a language has three basic color terms, they are *black, white*, and *red*. When a language has four basic color terms, the fourth is one of the following: *yellow, blue*, or *green*. The possibilities for four-color-term languages are thus: *black, white, red, yellow; black, white, red, blue*; and *black, white, red, green*. And so on, down the following hierarchy:

black, white
red
yellow, blue, green
brown
purple, pink, orange, gray

What made it possible for Berlin and Kay to find these regularities was their discovery of *focal colors*. If one simply asks speakers around the world to pick out the portions of the spectrum that their basic color terms refer to, there seem to be no significant regularities. The boundaries between the color ranges differ from language to language. The regularities appear only when one asks for the *best example* of a basic color term given a standardized chart of 320 small color chips. Virtually the same best examples are chosen for the basic color terms by speakers in language after language. For example, in languages that have a basic term for colors in the blue range, the best example is the same focal blue for all speakers no matter what language they speak. Suppose a language has a basic color term that covers the range of both *blue* and *green*; let us call that color *grue*. The best example of grue, they claim, will not be turquoise, which is in the middle of the blue-to-green spectrum. Instead the best example of grue will be either focal blue or focal green. The focal colors therefore allow for comparison of terms across languages.

The existence of focal colors shows that color categories are not uniform. Some members of the category RED are better examples of the category than others. Focal red is the best example. Color categories thus have central members. There is no general principle, however, for predicting the boundaries from the central members. They seem to vary, somewhat arbitrarily, from language to language.

Kay and McDaniel

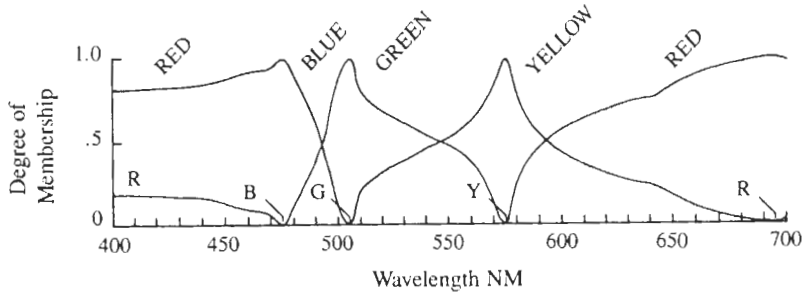
The Berlin-Kay color research raised questions that were left unanswered. What determines the collection of universal focal colors? Why should the basic color terms pick out just those colors? Kay and McDaniel (1978) provided an answer to these questions that depended jointly on research on the neurophysiology of color vision by DeValois and his associates and on a slightly revised version of Zadeh's fuzzy set theory.

DeValois and his associates (DeValois, Abramov, and Jacobs 1966; DeValois and Jacobs 1968) had investigated the neurophysiology of color vision in the macaque, a monkey with a visual system similar to man's. Their research concentrated on the neural pathways between the eye and the brain. They found six types of cell. Four of these, called *opponent response cells*, determine hue, while the other two determine brightness. The opponent response cells are grouped into two pairs, one having to do with the perception of blue and yellow, the other having to do with the perception of red and green. Each opponent response cell has a spontaneous rate of firing—a base response rate that it maintains without any external stimulation. There are two types of blue-yellow cells. The $+B-Y$ cells fire above their base rate in response to a blue stimulus, and below their base rate in response to a yellow stimulus. The $+Y-B$ cells do the reverse: they fire above their base rate in response to yellow and below their base rate in response to blue. Similarly, there are two types of red-green cells: $+G-R$ cells fire above their base rate in response to green and below in response to red, while $+R-G$ cells fire above in response to red and below in response to green. The two types of blue-yellow cells jointly determine a blue-yellow response, while the two kinds of red-green cells jointly determine a red-green response.

Focal blue is perceived when the blue-yellow cells show a blue response and when the red-green cells are firing at the neutral base rate. Purple is a combination of blue and red; it is perceived when the blue-yellow cells show a blue response and the red-green cells show a red response. Turquoise is perceived when the blue-yellow cells show a blue response and the red-green cells show a green response. Pure primary colors—blue, yellow, red, and green—are perceived when either the blue-yellow or red-green cells are firing at their neutral base rates. Nonprimary colors correspond to cases where no opponent cells are firing at neutral base rates.

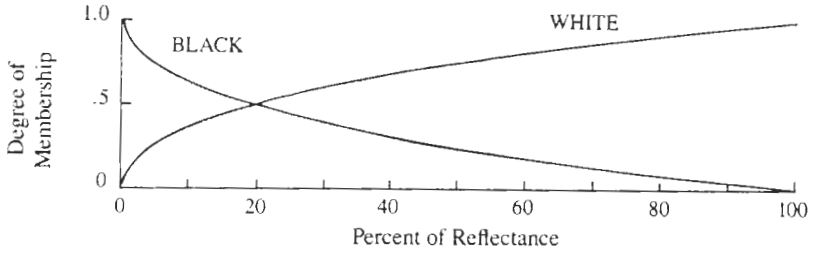
The remaining two kinds of cells are light-sensitive and darkness-sensitive. Pure black, white, and gray are perceived when the blue-yellow and red-green cells are all firing at their neutral base rates and making no color contribution. Pure black occurs when the darkness-sensitive cells are firing at their maximum rate and the light-sensitive cells are firing at their minimum rates. Pure white is the reverse.

Given these results from neurophysiological studies, Kay and McDaniel apply a version of fuzzy set theory to make sense of the Kay-Berlin results. For example, they define degree of membership in the category *blue* as the proportion of blue response on the part of the blue-yellow cells. Pure blue (degree of membership = 1) occurs when the red-green response is neutral. Blues in the direction of purple or green or white have an intermediate degree of membership in the blue category. Corresponding definitions are given for other primary colors. The accompanying dia-



grams give curves that correlate degree of membership in color categories with wavelengths in nanometers for hues and percentage of reflectance for black and white.

The neurophysiological account only characterizes the primary colors: black, white, red, yellow, blue, and green. What allows us to "see" other colors as being members of color categories? What about orange, brown,



purple, etc.? Some cognitive mechanism in addition to the neurophysiology is needed to account for those. Kay and McDaniel suggested that such a mechanism would make use of something akin to fuzzy set theory.

The postulation of a cognitive mechanism that has some of the effects of fuzzy set theory enables Kay and McDaniel to do two things that the neurophysiological account alone could not do. First, it enables them to characterize focal nonprimary colors (orange, purple, pink, brown, gray, etc.) in the following intuitive way:

ORANGE = RED and YELLOW
PURPLE = BLUE and RED
PINK = RED and WHITE
BROWN = BLACK and YELLOW
GRAY = BLACK and WHITE

Thus, ORANGE is characterized in terms of the fuzzy set intersection of the RED and YELLOW curves. (Actually, for technical reasons the definition is twice the fuzzy-set intersection value. See Kay and McDaniel 1978, pp. 634–35, for details.) Correspondingly, PURPLE is defined in terms of the fuzzy set intersection of BLUE and RED, and GRAY in terms of the fuzzy set intersection for BLACK and WHITE. PINK and BROWN require somewhat different functions based on fuzzy set intersections.

The second advantage of fuzzy set theory is that it permits an intuitive account of basic color categories that include more than one focal color. Dani, for example, has only two basic color terms: *mili* contains black and all the cool colors, the greens and blues; *mola* contains white and all the warm colors, the reds, oranges, yellows, pinks, and red-purples. Some

languages have basic color categories containing both blues and greens, while others have basic color categories containing both reds and yellows. Such cases can be accounted for intuitively by using fuzzy set union.

DARK-COOL = BLACK or GREEN or BLUE
LIGHT-WARM = WHITE or RED or YELLOW
COOL = GREEN or BLUE
WARM = RED or YELLOW

Thus, Kay and McDaniel make the claim that basic color categories are a product of both neurophysiology and cognitively real operations that can be partially modelled by fuzzy set intersection and union.

At present, this is the only plausible account we have of why the facts of basic color categories should be as they are. The Kay-McDaniel theory has important consequences for human categorization in general. It claims that colors are not objectively "out there in the world" independent of any beings. Color concepts are *embodied* in that focal colors are partly determined by human biology. Color categorization makes use of human biology, but color categories are more than merely a consequence of the nature of the world plus human biology. Color categories result from the world plus human biology plus a cognitive mechanism that has some of the characteristics of fuzzy set theory plus a culture-specific choice of which basic color categories there are.

The Kay-McDaniel theory seems to work well for characterizing the focal colors corresponding to basic color categories. But it does not work as well at the boundaries between colors. According to the Kay-McDaniel account, the boundaries, as well as the focal colors, should be uniform across languages. But this is simply not the case. The most detailed work on the detailed mapping of color categories, especially in nonfocal areas, has been done by MacLaury (in preparation). Among the test cases for the Kay-McDaniel theory are cases where a language does not have a separate color category for nonprimary focal colors, like purple and orange, colors that, in the Kay-McDaniel account, are "computed" on the basis of fuzzy set theory plus the response curves for the primary colors. The Kay-McDaniel theory predicts that colors like purple and orange should be treated uniformly across languages and that they should always be on the boundaries between basic color categories in languages that do not have separate categories for them.

But MacLaury has found cases where purple is entirely within the cool color range (a single color with focal points at blue and green) and other cases where purple is on the boundary between cool and red. He has also found cases where brown is subsumed by yellow and other cases where it is subsumed by black. That is, what we call "brown" falls within the range

of a category with a center at pure yellow in some languages, and it falls within the range of a category with a center at pure black in other languages.

In Kay-McDaniel terms, this means that the fuzzy-set-theoretical functions that compute conjunctions and disjunctions for color categories are not exactly the same for all people; rather they vary in their boundary conditions from culture to culture. They are thus at least partly conventional, and not completely a matter of universal neurophysiology and cognition. What this requires is a revision of the Kay-McDaniel theory to permit conceptual systems for color to vary at the boundaries, by having the exact nature of the disjunction function be somewhat different in different systems. Such differences may not only be at the boundaries but at the focal peaks. Kay and McDaniel's theory implied that each binary disjunctive color category (e.g., COOL = BLUE or GREEN) should have two focal peaks (e.g., both focal blue and focal green). MacLaury has found cases where there is a cool category covering blue and green, but where there is a skewing effect such that the center of the category is at pure green alone or pure blue alone. Thus, in Kay-McDaniel terms, conceptual systems seem to have disjunction functions that take the blue and green response curves as input and yield an output curve with only one focal center. This would require a cognitive mechanism with more than just something akin to the operation of union in fuzzy set theory.

Color categories, thus, are generative categories in the same sense in which kinship categories characterized by Lounsbury are. They have generators plus something else. The generators are the neurophysiologically determined distribution functions, which have peaks where the primary colors are pure: black, white, red, yellow, blue, and green. These generators are universal; they are part of human neurophysiology. The "something else" needed to generate a system of basic color categories consists of a complex cognitive mechamism incorporating some of the characteristics of fuzzy set theory union and intersection. This cognitive mechanism has a small number of parameters that may take on different values in different cultures.

It is important to bear in mind that it is not just the names for colors that vary. The color names do not just attach to the neurophysiologically determined distribution functions directly. Cognitive mechanisms of the sort described above must be postulated in addition. There are general characteristics of the cognitive mechanisms, for example, the use of something like fuzzy set theory union and intersection. But, as MacLaury shows, color cognition is by no means all the same across cultures. Nor is it by any means arbitrarily different across cultures. The possible color ranges depend upon limited parameters within the cognitive mechanism.

Brown and Berlin: Glimpses of the Basic Level

The study of basic-level categories is usually traced to Roger Brown's classic paper, "How Shall a Thing Be Called?" (1958), and his textbook, *Social Psychology* (1965, pp. 317–21).

Brown observed that objects have many names: "The dime in my pocket is not only a *dime*. It is also *money*, a *metal object*, a *thing*, and, moving to subordinates, it is a 1952 dime, in fact, a *particular 1952 dime* with a unique pattern of scratches, discolorations, and smooth places. The dog on the lawn is not only a *dog* but is also a *boxer*, a *quadruped*, an *animate being*" (Brown 1958, p. 14). Brown also observed that of all the possible names for something in a category hierarchy, a particular name, at a particular level of categorization, "has a superior status." "While a dime *can* be called a *coin* or *money* or *a 1952 dime*, we somehow feel that *dime* is its real name. The other categorizations seem like achievements of the imagination" (Brown 1965, p. 320). Such "real names," Brown observed, seem to be shorter and to be used more frequently. They also seem to correlate with nonlinguistic actions.

> When Lewis' son first looked upon the yellow jonquils in a bowl and heard them named flowers he was also enjoined to smell them and we may guess that his mother leaned over and did just that. When a ball is named *ball* it is also likely to be bounced. When a cat is named *kitty* it is also likely to be petted. Smelling and bouncing and petting are actions distinctively linked to certain categories. We can be sure they are distinctive because they are able to function as symbols of these categories. In a game of charades one might symbolize *cat* by stroking the air at a suitable height in a certain fashion, or symbolize *flower* by inclining forward and sniffing.
>
> . . . . . .
>
> Flowers are marked by sniffing actions, but there are no actions that distinguish one species of flower from another. The first names given to things fall at the level of distinctive action but names go on to code the world at every level; non-linguistic actions do not.
>
> . . . . . .
>
> When something is categorized it is regarded as equivalent to certain other things. For what purposes equivalent? How are all dimes equivalent or all flowers or all cats? . . . Dimes are equivalent in that they can be exchanged for certain newspapers or cigars or ice cream cones or for any two nickels. In fact, they are equivalent for all purposes of economic exchange. Flowers are equivalent in that they are agreeable to smell and are pickable. Cats are equivalent in that they are to be petted, but gently, so that they do not claw. (Brown 1965, pp. 318–19)

The picture Brown gives is that categorization, for a child, begins "at the level of distinctive action," the level of flowers and cats and dimes, and then proceeds upward to superordinate categories (like *plant* and *animal*) and downward to subordinate categories (like *jonquil* and *Siamese*) by "achievements of the imagination." "For these latter categories there seem to be no characterizing actions" (Brown 1965, p. 321). This "first level" of categorization was seen by Brown as having the following converging properties:

- It is the level of distinctive actions.
- It is the level which is learned earliest and at which things are first named.
- It is the level at which names are shortest and used most frequently.
- It is a natural level of categorization, as opposed to a level created by "achievements of the imagination."

The next important impetus to the study of basic-level categories came from the work of Brent Berlin and his associates. Berlin's research can be viewed as a response to the classical philosophical view that THE CATEGORIES OF MIND FIT THE CATEGORIES OF THE WORLD, and to a linguistic version of this, THE DOCTRINE OF NATURAL KIND TERMS. That doctrine states that the world consists very largely of natural kinds of things and that natural languages contain names (called "natural kind terms") that fit those natural kinds. Typical examples of natural kinds are cows, dogs, tigers, gold, silver, water, etc.

Berlin takes these philosophical doctrines as empirically testable issues and asks: To what extent do the categories of mind (as expressed in language) fit the categories of the world? In particular, Berlin considers domains in which there *are* natural kinds of things: the domains of plants and animals. Moreover, botany and zoology can reasonably be taken to have determined to a high degree of scientific accuracy just what kinds of plants and animals there are. Since Berlin is an anthropologist who studies people who live close to nature and who know an awful lot about plants and animals, he is in an excellent position to test such philosophical doctrines empirically.

Berlin and his students and associates have studied folk classification of plants and animals in incredibly minute detail and compared those classifications with scientific classifications. Most of the research has been carried out with speakers of Tzeltal living in Tenejapa in the Chiapas region of Mexico. This enormous undertaking has been documented meticulously in *Principles of Tzeltal Plant Classification* (Berlin, Breedlove, and Raven 1974), *Tzeltal Folk Zoology* (Hunn 1977), and "Language Acquisition by Tenejapa Tzeltal Children" (Stross 1969). The results to date have been surprising and have formed the basis for the psychological research on basic-level categorization.

What Berlin and his co-workers discovered was that a single level of classification—the genus—was for Tzeltal speakers psychologically basic in a certain number of ways. Examples of plants and animals at the genus level are *oak, maple, rabbit, raccoon*, etc. The first way that the priority of the genus manifested itself was in a simple naming task. Berlin went out into the jungle with a native consultant, stopped on the path, and asked the consultant to name the plants he could see. The consultant could easily name forty or fifty, but he tended to name them at the level of the genus (oak, maple, etc.) instead of the level of the species (sugar maple, live oak), even though further study showed he could distinguish the species and had names for them. Nor did he name them at the level of the life form (tree), nor at an intermediate level (needle-bearing tree). The level of the genus is, incidentally, "in the middle" of the folk classification hierarchy, the levels being:

UNIQUE BEGINNER (plant, animal)
LIFE FORM (tree, bush, bird, fish)
INTERMEDIATE (leaf-bearing tree, needle-bearing tree)
GENUS (oak, maple)
SPECIES (sugar maple, white oak)
VARIETY (cutleaf staghorn sumac)

Further study revealed that this was no accident and that the level of the genus (what Berlin called the "folk-generic level") seems to be a psychologically basic level in the following respects:

- People name things more readily at that level.
- Languages have simpler names for things at that level.
- Categories at that level have greater cultural significance.
- Things are remembered more readily at that level.
- At that level, things are perceived holistically, as a single gestalt, while for identification at a lower level, specific details (called *distinctive features*) have to be picked out to distinguish, for example, among the kinds of oak.

In addition, Stross (1969), in a study of Tzeltal language acquisition, discovered that "the bulk of the child's first-learned plant names are generic names and that from this starting point he continues to differentiate nomenclaturally, while cognitively he continues to differentiate and generalize plants simultaneously." In other words, the basic-level (or ge-

neric) categories, which are in the middle of the taxonomic hierarchy, are learned first; then children work up the hierarchy generalizing, and down the hierarchy specializing. Thus, we can add the finding:

– Children learn the names for things at that level earlier.

But perhaps the most remarkable finding of all was this:

– Folk categories correspond to scientific categories extremely accurately at this level, but not very accurately at other levels.

This says something very remarkable about THE DOCTRINE OF NATURAL KIND TERMS: For the Tzeltal, this doctrine works very well at the level of the genus, but not very well at other levels of classification, e.g., the intermediate, the species, and the variety levels.

But now if one considers philosophical discussions of natural kinds, it turns out that this is not such a surprising result after all. In the literature on natural kinds, one finds that the usual examples of natural kinds are animals like *dog, cow, tiger*, and substances like *gold* and *water*. As it happens, they are all basic-level categories! In short, the examples on which the doctrine of natural kinds was based were all basic level, which is the level of the genus among plants and animals. At least for the Tzeltal, the doctrine works well for the kinds of examples that philosophers had in mind when they espoused the doctrine. For other kinds of examples, it does not work very well.

But if THE DOCTRINE OF NATURAL KIND TERMS fits well for the Tzeltal at even one level of categorization, it still seems to be quite a remarkable result. It suggests that there is one psychologically relevant level at which THE CATEGORIES OF THE MIND FIT THE CATEGORIES OF THE WORLD. However, Berlin's research into the history of biological classification shows this result to be much less remarkable. Scientific classification in biology grew out of folk classification. And when Linnaeus classified the living things of the world, he specifically made use of psychological criteria in establishing the level of the genus. This comes across particularly clearly in A. J. Cain's 1958 essay "Logic and Memory in Linnaeus's System of Taxonomy" (1958). The heart of the Linnaean system was the genus, not the species. It is the genus that gives the *general* characteristics and the species that is defined in terms of differentiating characteristics. But what is a *general* characteristic? As Cain observes, "The *Essential Character* of a genus is that which gives some characteristic peculiar to it, if there is one such, which will instantly serve to distinguish it from all others in the natural order" (p. 148). This is a psychologically defined notion of an "essential character"; which characteristics can be instantly distinguished

depends on the perceptual systems of the beings doing the distinguishing. As Linnaeus's son writes,

> My Father's secret art of determining (delimiting) genera in such a way the Species should not become genera? This was no other than his practice in knowing a plant from its external appearance (externa facie). Therefore he often deviated from his own principles in such a way that variation as to the number of parts . . . did not disturb him, if only the character of the genus . . . could be preserved. Foreigners don't do so, but as soon as a plant has a different splitting (cleavage) of the corolla and calyx, or if the number of stamens and pistils . . . varies, they make a new genus. . . . If possible he [Linnaeus] tried to build the character genericus on the cleavage of the fruit so that all species that constitute a genus should have the same shape of their fruit. (Cain, p. 159)

Why did Linnaeus use the shape of the fruit as a basis for defining the genus? As Cain observes, "The characters chosen from the fructification were clearly marked, readily appreciated, easily described in words, and usually determinable on herbarium specimens" (p. 152). In other words, the shape of the fruit was easy to perceive and describe. Genera, as Linnaeus conceived of them, were "practical units of classification" upon which all biologists should be able to agree; it was important that they should "not become confused and indistinct in the mind" (Cain, p. 156). Most of Linnaeus's rules of nomenclature "follow directly from [the] requirement that the botanist must know and remember all genera" (Cain, p. 162)—again a psychological requirement. "Linnaeus states explicitly and repeatedly that the botanist . . . [and] the zoologist too must know all genera and commit their names to memory" (Cain, p. 156). Linnaeus also assumed, of course, that this practical system would also be "natural," in short, a convergence between nature and psychology could be taken for granted at this level.

In short, the genus was established as that level of biological discontinuity at which human beings could most easily perceive, agree on, learn, remember, and name the discontinuities. The genus, as a scientific level of classification, was set up because it was the most psychologically basic level for the purposes of the study of taxonomic biology by human beings. It was assumed that this would also fit certain real discontinuities in nature. Berlin found that there is a close fit at this level between the categories of Linnaean biology and basic-level categories in folk biology. This fit follows in part from the criteria used to set up the level of the genus in Linnaean biology; those criteria correspond to the psychological criteria that characterize the basic level in folk biology.

At the level of the genus, the categories of mind of the biologists who set up the level of the genus correspond closely to the basic-level categories of mind of Tzeltal speakers. But this is not merely a fact about psychology. It is a fact about both psychology and biology. Here is the reason: Within scientific biology, the genus is one level above the species—the level defined by interbreeding possibilities: two populations that are members of the same species can breed and produce fertile offspring. Typically, members of two populations that can interbreed have pretty much the same overall shape. In the course of evolution, two populations of the same species may change sufficiently so that interbreeding is no longer possible. At the point at which they cease to be able to interbreed, they become different species. But at this point they may still have pretty much the same overall shape. They will no longer be members of the same species, but they will be members of the same genus—the category one level up. Thus, one level up from the species in scientific biology, it is common to find certain general shape similarities. It may not always happen, but it is common.

Now overall shape is a major determinant of the basic level in folk biology. The basic level is primarily characterized by gestalt perception (the perception of overall shape), by imaging capacity (which depends on overall shape), and by motor interaction (the possibilities for which are also determined by overall shape). It is anything but an accident that the level of the genus in scientific biology should correspond so well to the basic level in folk biology.

Moreover, given the experience of people like the Tzeltal, who are indigenous to a circumscribed geographical area, there is a good reason why divisions in nature at the level of the genus should be particularly striking. In the course of evolution, the species that survive in a particular geographical region are those that adapt most successfuly to the local environment. Thus, for each genus, it is common for there to be only one species representing the genus locally. This does not always happen, but it does happen frequently. Thus, there tend to be genus-sized gaps among the *species* that occur locally—and these are very striking and perceptible gaps. Thus, divisions at the basic level in folk biology correspond to very striking discontinuities in nature for people in a circumscribed geographical area.

In summary, ethnobiological research has established that there is, at least for biological categories, a basic level of categorization. Among the Tzeltal, who have an intimate familiarity with a large range of plants and animals, the categories of the mind fit discontinuities in the world very well at the level of the genus, though not very well at other levels. The reason for this is partly because the level of the genus, as a fundamental

level used in scientific biology, is a psychologically based level of categorization. But there are equally important biological reasons.

Basic-level categorization depends upon experiential aspects of human psychology: gestalt perception, mental imagery, motor activities, social function, and memory. (What I call "memory" here is the ability of a subject in a psychological test to recall on demand particular presented instances of the category.) To what extent is basic-level categorization universal? If we assume that human physiology and psychology are pretty much the same around the world, then any variation would most likely be due to culture and context. But how much variation would there be and what kind would it be?

Berlin has suggested (personal communication) that a distinction be made between a general human capacity for basic-level categorization (due to general physiological and psychological factors) and functional basic-level categorization, which adds in factors having to do with culture and specialized training. Berlin suggests that a given culture may underutilize certain human capacities used in basic-level categorization, for example, the capacity for gestalt perception. Thus, in urban cultures, people may treat the category *tree* as basic level. Such cases have been documented by Dougherty (1978). Moreover, there may be subpopulations of specialists in a culture who, through training, may achieve a more finely honed gestalt perception for a limited range of domains, e.g., breeds of horses, types of cars, etc. But this should be possible only in a limited number of domains, even for trained specialists. Berlin thus hypothesizes two kinds of nonuniversality: (*a*) one kind due to cultural underutilization of general human capacities, with the result that certain higher-level categories (e.g., tree) may be treated as basic, and (*b*) another kind due to special training, limited to subpopulations of experts who may treat a slightly more specific level as basic in some domains of expertise.

Berlin's hypothesis makes the following prediction: People from, say, an urban culture that treats trees as basic level should still have the general human capacity for gestalt perception and should thus be capable of learning to discriminate among trees readily at the level of the genus, but not so readily at the level of the species or variety. Berlin's hypothesis also predicts that there will be no whole cultures that will treat the level of the species or variety as basic, but that individuals may have a capacity for expertise in a limited range of domains and thus may be able to treat a small number of more specific categories as basic.

Berlin also predicts that there will be no culture where all the levels of categorization are different from ours or from the Tzeltal. In most domains, levels of categorization will be the same for all human beings, sim-

ply because human beings share the same general capacities for gestalt perception and for holistic motor movement. It is these capacities that have the major role in determining basic-level categorization.

Basicness in categorization has to do with matters of human psychology: ease of perception, memory, learning, naming, and use. Basicness of level has no objective status external to human beings. It is constant only to the extent that the relevant human capacities are utilized in the same way. Basicness varies when those capacities either are underutilized in a culture or are specially developed to a level of expertise.

As we shall see below, Berlin's results have a special philosophical importance. Berlin showed that human categorizations based on interactions with the environment are extremely accurate at the basic level. Basic-level interactions thus provide a crucial link between cognitive structure and real knowledge of the world. We will argue in chapter 17 that basic-level interactions can therefore form the basis of an epistemology for a philosophy of mind and language that is consistent with the results of prototype theory.

Ekman

In research spanning more than two decades, Paul Ekman and his associates have studied in detail the physiological correlates of emotions (Ekman 1971; Ekman, Friesen, and Ellsworth 1972). In a major crosscultural study of facial gestures expressing emotion, Ekman and his associates discovered that there were basic emotions that seem to correlate universally with facial gestures: happiness, sadness, anger, fear, surprise, and interest. Of all the subtle emotions that people feel and have words and concepts for around the world, only these have consistent correlates in facial expressions across cultures.

Although Ekman was by no means a prototype theorist, his research fits prototype research in the following way. The seven basic emotions appear to have prototype status. There are many shades and varieties of happiness, sadness, anger, etc. These form categories of emotions. Rage and annoyance, for example, are in the anger category. Basic happiness, anger, etc.—the emotions that correlate with the universal facial gestures—seem to function as central members of those categories. These emotions also appear to have basic-level status. They are readily recognizable by gestalt perception around the world. We have facial images and motor movements for them that represent the entire emotional category.

As we will see below in case study 1, emotional concepts are embodied, in that the physiology corresponding to each emotion has a great deal to do with how the emotion is conceptualized. We will see, for example, that anger is metaphorically understood in terms of heat and internal pressure. Ekman, Levenson, and Friesen (1983) have shown that there is autonomic nervous system (ANS) activity that corresponds to the basic emotions. The ANS activity that corresponds to anger is an increase in skin temperature and an increase in heart rate (experienced as internal pressure).

The experiments that demonstrated this involved two tasks. In the first, subjects were instructed to change their facial expressions, muscle by muscle, until their expressions matched the facial prototypes of emotions. In the second, subjects were asked to relive emotional experiences. Heart rate and left- and right-finger temperatures were recorded.

Two findings were consistent across tasks:

> (i) Heart rate increased more in anger (mean calculated across tasks ± standard error, +8.0 ± 1.8 beats per minute) and fear (+8.0 ± 1.6 beats per minute) than in happiness (+2.6 ± 1.0 beat per minute).
> (ii) Left- and right-finger temperatures increased more in anger (left, +0.10°C ±0.009°; right, +0.08° ± 0.008°C) than in happiness (left, −0.07°C ± 0.002°; right, −0.03° ± 0.002°). (Ekman, Levenson, and Friesen 1983, p. 1209)

Thus the metaphorical conceptualization of anger that we will explore in case study 1 is actually embodied in the autonomic nervous system, in that it is motivated by ANS activity that corresponds to the emotions as felt.

Rosch

The studies cited above are all special cases. It was Eleanor Rosch who first provided a general perspective on all these problems. She developed what has since come to be called "the theory of prototypes and basic-level categories," or "prototype theory." In doing so, she provided a full-scale challenge to the classical theory and did more than anyone else to establish categorization as a subfield of cognitive psychology. Before her work, the classical theory was taken for granted, not only in psychology, but in linguistics, anthropology, and philosophy, as well as other disciplines. In a series of electrifying papers, Rosch and her associates presented an overwhelming array of empirical studies that challenged the classical view.

The experimental contributions of Rosch and her associates are generally and justly recognized by cognitive psychologists as having revolutionized the study of categorization within experimental psychology. Rosch's

experimental results fall into two categories: prototype effects, which extend the Berlin-Kay color research, and basic-level effects, which generalize Brown's observations and Berlin's results.

Prototype Effects

If the classical theory were both correct and complete, no member of a category would have any special status. The reason is that, in the classical theory, the properties defining the category are shared by all members, and so all members have equal status as category members. Rosch's research on prototype effects has been aimed at showing asymmetries among category members and asymmetric structures within categories. Since the classical theory does not predict such asymmetries, either something more or something different must be going on.

Rosch's early studies were on color. She learned of the Berlin-Kay color research midway through her own research and found that their results meshed with her own work on Dani, a New Guinea language that has only two basic color categories: *mili* (dark-cool, including black, green, and blue) and *mola* (light-warm, including white, red, yellow). Berlin and Kay had shown that focal colors had a special status within color categoraies—that of the best example of the category. Rosch found that Dani speakers, when asked for the best examples of their two color categories, chose focal colors, for example, white, red, or yellow for *mola* with different speakers making different choices.

In a remarkable set of experiments, Rosch set out to show that primary color categories were psychologically real for speakers of Dani, even though they were not named. She set out to challenge one of Whorf's hypotheses, namely, that language determines one's conceptual system. If Whorf were right on this matter, the Dani's two *words* for colors would determine two and only two *conceptual categories* of colors. Rosch reasoned that if it was language alone that determined color categorization, then the Dani should have equal difficulty learning new words for colors, no matter whether the color ranges had a primary color at the center or a nonprimary color. She then went about studying how Dani speakers would learn new, made-up color terms. One group was taught arbitrary names for eight focal colors, and another group, arbitrary names for eight nonfocal colors (Rosch 1973). The names for focal colors were learned more easily. Dani speakers were also found (like English speakers) to be able to remember focal colors better than nonfocal colors (Heider 1972). In an experiment in which speakers judged color similarity, the Dani were shown to represent colors in memory the same way English speakers do (Heider and Olivier 1972). Rosch's color research also extended to children. When three-year-olds were presented with an array of color chips, and the experimenter turned her back and said "Show me a color," the children picked focal colors overwhelmingly over nonfocal colors (Heider 1971). And when four-year-olds were given a color chip and asked to pick from an assortment of chips the one that matched best, the children did best with focal colors.

Focal colors correspond to what Rosch in her later research called *cognitive reference points* and *prototypes*—subcategories or category members that have a special cognitive status—that of being a "best example." Rosch showed that a variety of experimental techniques involving learning, matching, memory, and judgments of similarity converged on cognitive reference points. And she extended the results from colors to other categories, primarily categories of physical objects. She developed other experimental paradigms for investigating categories of physical objects. In each case, asymmetries (called *prototype effects*) were found: subjects judged certain members of the categories as being more representative of the category than other members. For example, robins are judged to be more representative of the category BIRD than are chickens, penguins, and ostriches, and desk chairs are judged to be more representative of the category CHAIR than are rocking chairs, barber chairs, beanbag chairs, or electric chairs. The most representative members of a category are called "prototypical" members. Here are some of the experimental paradigms used in studying categories of physical objects. Subjects give consistent goodness-of-example ratings across these experimental paradigms.

Direct rating: Subjects are asked to rate, say on a scale from one to seven, how good an example of a category (e.g., BIRD) various members are (e.g., a robin, a chicken, etc.).

Reaction time: Subjects are asked to press a button to indicate true or false in response to a statement of the form "An [example] is a [category name]" (e.g., "A chicken is a bird"). Response times are shorter for representative examples.

Production of examples: When asked to list or draw examples of category members, subjects were more likely to list or draw more representative examples.

Asymmetry in similarity ratings: Less representative examples are often considered to be more similar to more representative examples than the converse. Not surprisingly, Americans consider the United States to be a highly representative example of a country. In experiments where subjects were asked to give similarity ratings for pairs of countries, the following asymmetry arose. Subjects considered Mexico to be more similar to the United States than the United States is to Mexico. See Rosch 1975*a* and Tversky and Gati 1978.

Asymmetry in generalization: New information about a representative category member is more likely to be generalized to nonrepresentative members than the reverse. For example, it was shown that subjects believed that a disease was more likely to spread from robins to ducks on an island, than from ducks to robins. (This result is from Rips 1975.)

Family resemblances: Wittgenstein had speculated that categories were structured by what he called "family resemblances." Rosch showed that what philosophers took as a matter for a priori speculation could be demonstrated empirically. Characterizing "family resemblances" as perceived similarities between representative and nonrepresentative members of categories, Rosch showed that there was a correlation between family resemblances and numerical ratings of best examples derived from the above experiments. (See Rosch and Mervis 1975 and Rosch, Simpson, and Miller 1976.)

Such studies have been replicated often by other experimenters. There is no doubt that prototype effects of this sort are real. However, there have been some misunderstandings and debates concerning the interpretation of these results. Some of the debates will be discussed in detail below. But before we go on, we ought to clear up some of the common misunderstandings.

Rosch's genius has two aspects: she both launched a general challenge to the classical theory and devised, with her co-workers, replicable experiments demonstrating prototype effects, as well as basic-level effects. These experiments demonstrate the inadequacy of the classical theory; the classical theory cannot account for such results. But prototype effects, in themselves, do not provide any specific alternative theory of mental representation. And, as a responsible experimenter, Rosch has consistently distinguished between what her experimental results show and any theories that might account for those results.

Rosch went through three phases in her thinking about categorization.

- Phase I (late 1960s to early 1970s): Because she was studying color, shape, and emotions, she assumed prototypes were primarily a matter of (*a*) perceptual salience, or which things are most readily noticed by people; (*b*) memorability, or which things are easiest for people to remember; and (*c*) stimulus generalization, or the ability of people to generalize from one thing to something else that is physically similar to it. As she says (Rosch, in press): "Suppose that there are perceptually salient colors which more readily attract attention and are more easily remembered than other colors. When category names are learned, they tend to become attached first to the salient stimuli; then, by means of the principle of stimulus generalization, they generalize to other, physically similar instances."
- Phase II (early to mid 1970s): Under the influence of information-processing psychology, Rosch considered the possibility that prototype effects, as operationalized by the experiments cited above, might provide a characterization of the internal structure of the category. Thus, for example, the goodness-of-example ratings might directly reflect the internal structure of the category in mental representation. Two natural questions arose:
  1. Do the EFFECTS, defined operationally, characterize the STRUCTURE of the category as it is represented in the mind?
  2. Do the PROTOTYPES constitute mental REPRESENTATIONS?

  Given the assumptions of information-processing psychology, the experimental data can be interpreted most straightforwardly by answering yes to both questions. Rosch (1975*b*) initially interpreted her data in just this way.
- Phase III (late 1970s): Rosch eventually gave up on these interpretations of her experimental results. Such interpretations were artifacts of an overly narrow view of information-processing psychology. She came to the conclusion that prototype effects, defined operationally by experiment, underdetermined mental representations. The effects constrained the possibilities for what representations might be, but there was no one-to-one correspondence between the effects and mental representations. The effects had "sources," but one could not determine the sources given the effects. As she says of the research in Phase II (Rosch, in press): "The type of conclusions generated by this approach were, however, very general; e.g., that the representation evoked by the category name was more like good examples than poor examples of the category; that it was in a form more general than either words or pictures, etc. On the whole other information-processing researchers have considered the concepts of prototypes and typicality functions underspecified and have provided a variety of precise models, mini-models, and distinctions to be tested."

It is often the case that positions taken early in one's career tend to be associated with a researcher long after he or she has given up those positions. Many of those who read Rosch's early works did not read her later works, where she gave up on her early interpretations of the experimental results. Consequently, it is not widely known that Rosch abandoned the ideas that prototype effects directly mirror category structure and that prototypes constitute representations of categories. Because of this,

Rosch has had to provide explicit admonitions against overly simplistic interpretations of prototype effects—interpretations of the sort that she herself made in Phase II of her research. For example, she states:

> The pervasiveness of prototypes in real-world categories and of prototypicality as a variable indicates that prototypes must have some place in psychological theories of representation, processing, and learning. However, prototypes themselves do not constitute any particular model of processes, representations, or learning. This point is so often misunderstood that it requires discussion:
>
> 1. To speak of a *prototype* at all is simply a convenient grammatical fiction; what is really referred to are judgments of degree of prototypicality. . . . For natural-language categories, to speak of a single entity that is the prototype is either a gross misunderstanding of the empirical data or a covert theory of mental representation.
> 2. Prototypes do not constitute any particular processing model for categories. . . . What facts about prototypicality do contribute to processing notions is a constraint—processing models should not be inconsistent with the known facts about prototypes. For example, a model should not be such as to predict equal verification times for good and bad examples of categories nor predict completely random search through a category.
> 3. Prototypes do not constitute a theory of representation for categories. . . . Prototypes can be represented either by propositional or image systems. . . . As with processing models, the facts about prototypes can only constrain, but do not determine, models of representation. A representation of categories in terms of conjoined necessary and sufficient attributes alone would probably be incapable of handling all of the presently known facts, but there are many representations other than necessary and sufficient attributes that are possible.
> 4. Although prototypes must be learned, they do not constitute any particular theory of category learning. (Rosch 1978, pp. 40–41)

Despite Rosch's admonitions to the contrary, and despite her minimal theorizing concerning the sources of prototype effects, her results on prototype effects are still sometimes interpreted as constituting a prima facie theory of representation of category structure, as she thought was possible during Phase II of her research.

For example, take her results showing prototype effects within the category *bird*. Her experimental ranking shows that subjects view robins and sparrows as the best examples of birds, with owls and eagles lower down in the rankings and ostriches, emus, and penguins among the worst examples. In the early to mid 1970s, during Phase II of Rosch's research, such empirical goodness-of-example ratings were commonly taken as

constituting a claim to the effect that membership in the category *bird* is graded and that owls and penguins are less members of the *bird* category than robins. (See Lakoff 1972 for a typical example.) It later became clear that that was a mistaken interpretation of the data. Rosch's ratings make no such claim; they are just ratings and do not make any claims at all. They are consistent with the interpretation that the category *bird* has strict boundaries and that robins, owls, and penguins are all 100 percent members of that category. However, that category must have additional internal structure of some sort that produces these goodness-of-example ratings. Moreover, that internal structure must be part of our concept of what a bird is, since it results in asymmetric inferences of the sort discussed above, described by Rips (1975).

This point is extremely important. Category structure plays a role in reasoning. In many cases, prototypes act as *cognitive reference points* of various sorts and form the basis for inferences (Rosch 1975*a*, 1981). The study of human inference is part of the study of human reasoning and conceptual structure; hence, those prototypes used in making inferences must be part of conceptual structure.

It is important to bear in mind that prototype effects are superficial. They may result from many factors. In the case of a graded category like *tall man*, which is fuzzy and does not have rigid boundaries, prototype effects may result from degree of category membership, while in the case of *bird*, which does have rigid boundaries, the prototype effects must result from some other aspect of internal category structure.

One of the goals of this book is to outline a general approach to the theory of categorization and to sketch the range of sources for superficial prototype effects. We will undertake this in chapters 4 through 6, where we discuss cognitive models. Our basic claim will be that prototype effects result from the nature of cognitive models, which can be viewed as "theories" of some subject matter.

One of the most interesting confirmations of this hypothesis has come through the work of Barsalou (1983, 1984). Barsalou has studied what he calls "ad hoc categories"—categories that are not conventional or fixed, but rather are made up on the fly for some immediate purpose. Such categories must be constructed on the basis of one's cognitive models of the subject matter under consideration. Examples of such categories are *things to take from one's home during a fire, what to get for a birthday present, what to do for entertainment on a weekend*, etc. Barsalou observes that such categories have prototype structure—structure that does not exist in advance, since the category is not conventional and does not exist in advance. Barsalou argues that in such cases, the nature of the

category is principally determined by goals and that such goal structure is a function of one's cognitive models. Such a view has also been advocated by Murphy and Medin (1984).

Basic-Level Effects

The classical theory of categories gives no special importance to categories in the middle of a taxonomic hierarchy. Yet, as Berlin (Berlin, Breedlove, Raven 1974) and Hunn (1977) have shown for Tzeltal plant and animal taxonomies, the level of the biological genus is psychologically basic. The genus stands in the middle of the hierarchy that goes from UNIQUE BEGINNER TO LIFE FORM to INTERMEDIATE TO GENUS TO SPECIES TO VARIETY. Their results show a discrepancy between the classical theory of categories and a cognitively adequate theory of categories.

Rosch and her associates have extended the study of basic-level effects from cognitive anthropology to the experimental paradigm of cognitive psychology. Like Berlin, they found that the psychologically most basic level was in the middle of the taxonomic hierarchies:

SUPERORDINATE	ANIMAL	FURNITURE
BASIC LEVEL	DOG	CHAIR
SUBORDINATE	RETRIEVER	ROCKER

Just as Hunn (1975) argued that the basic level for animal categories is the only level at which categorization is determined by overall gestalt perception (without distinctive feature analysis), so Rosch and others (1976) have found that the basic level is:

- The highest level at which category members have similarly perceived overall shapes.
- The highest level at which a single mental image can reflect the entire category.
- The highest level at which a person uses similar motor actions for interacting with category members.
- The level at which subjects are fastest at identifying category members.
- The level with the most commonly used labels for category members.
- The first level named and understood by children.
- The first level to enter the lexicon of a language.
- The level with the shortest primary lexemes.
- The level at which terms are used in neutral contexts. For example, *There's a dog on the porch* can be used in a neutral context, whereas special contexts are needed for *There's a mammal on the porch* or *There's a wire-haired terrier on the porch.* (See Cruse 1977.)
- The level at which most of our knowledge is organized.

Thus basic-level categories are basic in four respects:

Perception: Overall perceived shape; single mental image; fast identification.
Function: General motor program.
Communication: Shortest, most commonly used and contextually neutral words, first learned by children and first to enter the lexicon.
Knowledge Organization: Most attributes of category members are stored at this level.

The fact that knowledge is mainly organized at the basic level is determined in the following way: When subjects are asked to list attributes of categories, they list very few attributes of category members at the superordinate level (furniture, vehicle, mammal); they list most of what they know at the basic level (chair, car, dog); and at the subordinate level (rocking chair, sports car, retriever) there is virtually no increase in knowledge over the basic level.

Why should most information be organized at a single conceptual level and why should it be this level in particular? To me, the most convincing hypothesis to date comes from the research of Tversky and Hemenway (1984). Berlin (Berlin, Breedlove, Raven 1974) and Hunn (1977) had suggested that gestalt perception—perception of overall part-whole configuration—is the fundamental determinant of the basic level. The experimental evidence accumulated by Tversky and Hemenway supports the Berlin-Hunn hypothesis. Their basic observation is that the basic level is distinguished from other levels on the basis of the type of attributes people associate with a category at that level, in particular, attributes concerned with *parts*. Our knowledge at the basic level is mainly organized around part-whole divisions. The reason is that the way an object is divided into parts determines many things. First, parts are usually correlated with functions, and hence our knowledge about functions is usually associated with knowledge about parts. Second, parts determine shape, and hence the way that an object will be perceived and imaged. Third, we usually interact with things via their parts, and hence part-whole divisions play a major role in determining what motor programs we can use to interact with an object. Thus, a handle is not just long and thin, but it can be grasped by the human hand. As Tversky and Hemenway say, "We sit on the *seat* of a chair and lean against the *back*, we remove the *peel* of a banana and eat the *pulp*."

Tversky and Hemenway also suggest that we impose part-whole structure on events and that our knowledge of event categories is structured very much the way our knowledge of physical object categories is. Their suggestion is in the same spirit as Lakoff and Johnson (1980), where it is

suggested that event categories and other abstract categories are structured metaphorically on the basis of structures from the realm of physical experience.

Acquisition

One of the most striking results about basic-level categorization concerns the acquisition of concepts by children. If the classical theory of categorization were correct, then there should be no more to categorization than what one finds in the logic of classes: hierarchical categorization based on shared properties of the members of the categories. Before the work of Rosch and Mervis (Rosch et al. 1976), research on child development had not been informed by the idea of basic-level categorization. It had been concluded that, for example, three-year-old children had not mastered categorization, which was taken to be taxonomic categorization defined by the logic of classes. This conclusion was based on the performance of children in "sorting tasks," where subjects are asked to "put together the things that go together." Rosch and her associates observed that such studies tended to involve categorization at the *superordinate* level.

> The stimuli used in sorting tasks have tended to be of two types: If abstract (e.g., geometric forms varying in dimensions such as form, color, and size), they are typically presented in a set which has no structure (e.g., each attribute occurs with all combinations of all others); if representational (e.g., toy versions or pictures of real-world objects), the arrays are such that they can be grouped taxonomically only at the superordinate level. Thus, the representational stimuli used in sorting tasks are such that if the child were to sort the objects into those of like taxonomic category, he would have to put together such items as socks and shirt, dog and cow. Children do not seem to have been asked to sort together objects belonging to the same basic level category (e.g., several shoes or several dogs). We suspect this results from the fact that basic objects are so obviously the "same object" to adults that a task does not seem to be a problem of categorization to an adult experimenter unless objects are taken from different basic level categories. (Rosch et al. 1976, pp. 414–15)

Rosch and Mervis then compared sorting tasks for basic-level and superordinate categories. Basic-level sorting required being able to put together pictures of two different kinds of cows (compared to an airplane, say) or two different kinds of cars (compared to, say, a dog). Superordinate sorting required, for example, being able to put together a cow and a dog (compared to an airplane), or a motorcycle and an airplane (compared to a cow). At all age levels, from three years old up, subjects were

virtually perfect on basic-level sorting. But, as had been well-known, the three-year-olds had trouble with superordinate sorting. They were only 55 percent correct, while the four-year-olds were 96 percent correct.

It is not true that three-year-olds have not mastered categorization. They have mastered *basic-level* categorization perfectly. It is *superordinate* categorization that is not mastered till later. The ability to categorize at the basic level comes first; the general logic of classes is learned later. Learning to categorize is thus something rather different from learning to use the logic of classes. Therefore, categorization itself is not merely the use of classical taxonomies.

It is important to bear these results in mind throughout the remainder of the book. The reason is this: It is sometimes claimed that basic-level categorization is merely classical taxonomic classification with additional constraints on cognitive processing added (e.g., perceptual and motor constraints). The Rosch-Mervis acquisition results show that this is not the case. Basic-level categories develop prior to classical taxonomic categories. They therefore cannot be the result of classical taxonomic categories *plus* something of a sensory-motor nature. Basic-level categories have an integrity of their own. They are our earliest and most natural form of categorization. Classical taxonomic categories are later "achievements of the imagination," in Roger Brown's words.

As Rosch and her co-workers observe, basic-level distinctions are "the generally most useful distinctions to make in the world," since they are characterized by overall shape and motor interaction and are at the most general level at which one can form a mental image. Basic-level categorization is mastered by the age of three. But what about children at earlier ages? It is known, for example, that two-year-olds have different categories from adults. Lions and tigers as well as cats are commonly called "kitty" by two-year-olds. Round candles and banks are commonly called "ball." And some things that we call "chair" may not be chairs for two-year-olds, e.g., beanbag chairs. The categories of two-year-olds may be broader than adult categories, or narrower, or overlapping. Does this mean that two-year-olds have not mastered the ability to form basic-level categories?

Not at all. Mervis (1984) has shown that although two-year-olds may have different categories than adults have, those categories are determined by the same principles that determine adult basic-level categories. In short, two-year-olds have mastered basic-level categorization, but have come up with different categories than adults—for very good reasons.

The difference is determined by three factors:

1. The child may not know about culturally significant attributes. Thus, not knowing that a bank is used for storing money, the child may attend to its round shape and classify it as a *ball.*

2. The salience of particular attributes may be different for a child than for an adult. Thus, a child may know that a bank is for storing money, but may attend to its round shape more than to the slot and keyhole, and still call it a ball. Or the child may attend to both and classify it as both a *bank* and a *ball.*

3. The child may include false attributes in the decision process. Thus, if the child thinks a leopard says "meow," he or she may classify leopards as *kitties.*

The point is that the level of categorization is not independent of who is doing the categorizing and on what basis. Though the same principles may determine the basic level, the circumstances under which those principles are employed determine what system of categories results.

Clusters of Interactional Properties

What determines basic-level structure is a matter of correlations: the overall perceived part-whole structure of an object correlates with our motor interaction with that object and with the functions of the parts (and our knowledge of those functions). It is important to realize that these are not purely objective and "in the world"; rather they have to do with the world as we interact with it: as we perceive it, image it, affect it with our bodies, and gain knowledge about it.

This is, again, a matter which has often been misunderstood, and Rosch has written at length on the nature of the misunderstanding. "It should be emphasized that we are talking about a perceived world and not a metaphysical world without a knower" (Rosch 1978, p. 29). She continues:

> When research on basic objects and their prototypes was initially conceived (Rosch et al. 1976), I thought of such attributes as inherent in the real world. Thus, given an organism that had sensory equipment capable of perceiving attributes such as wings and feathers, it was a fact in the real world that wings and feathers co-occurred. The state of knowledge of a person might be ignorant of (or indifferent or inattentive to) the attributes or might know the attributes but be ignorant concerning their correlation. Conversely, a person might know of the attributes and their correlational structure but exaggerate that structure, turning partial into complete correlations (as when attributes true of only many members of a category are thought of as true of all members). However, the environment was thought to constrain categorizations in that human knowledge could not provide correlational structure where there was none at all. For purposes of the basic object

> experiments, perceived attributes were operationally defined as those attributes listed by our subjects. Shape was defined as measured by our computer programs. We thus seemed to have our system grounded comfortably in the real world.
>
> On contemplation of the nature of many of our attributes listed by our subjects, however, it appeared that three types of attributes presented a problem for such a realistic view: (1) some attributes, such as "seat" for the object "chair," appeared to have names that showed them not to be meaningful prior to the knowledge of the object as chair; (2) some attributes such as "large" for the object "piano" seemed to have meaning only in relation to categorization of the object in terms of a superordinate category—piano is large for furniture but small for other kinds of objects such as buildings; (3) some attributes such as "you eat on it" for the object "table" were functional attributes that seemed to require knowledge about humans, their activities, and the real world in order to be understood. That is, it appeared that the analysis of objects into attributes was a rather sophisticated activity that our subjects (and indeed a system of cultural knowledge) might be considered to be able to impose only *after* the development of a system of categories. (Rosch 1978, pp. 41–42)

Thus the relevant notion of a "property" is not something objectively in the world independent of any being; it is rather what we will refer to as an *interactional property*—the result of our interactions as part of our physical and cultural environments given our bodies and our cognitive apparatus. Such interactional properties form *clusters* in our experience, and prototype and basic-level structure can reflect such clusterings.

As Berlin has observed, interactional properties and the categories they determine seem objective in the case of properties of basic-level categories—categories like *chair, elephant,* and *water.* The reason is that, given our bodies, we perceive certain aspects of our external environment very accurately at the basic level, though not so accurately at other levels. As long as we are talking about properties of basic-level objects, interactional properties will seem objective.

Perhaps the best way of thinking about basic-level categories is that they are "human-sized." They depend not on objects themselves, independent of people, but on the way people interact with objects: the way they perceive them, image them, organize information about them, and behave toward them with their bodies. The relevant properties clustering together to define such categories are not inherent to the objects, but are interactional properties, having to do with the way people interact with objects.

Basic-level categories thus have different properties than superordinate categories. For example, superordinate categories seem not to be characterized by images or motor actions. For example, we have mental

images of chairs—abstract images that don't fit any particular chair—and we have general motor actions for sitting in chairs. But if we go from the basic-level category CHAIR to the superordinate category FURNITURE, a difference emerges. We have no abstract mental images of furniture that are not images of basic-level objects like chairs, tables, beds, etc. Try to imagine a piece of furniture that doesn't look like a chair, or table, or bed, etc., but is more abstract. People seem not to be able to do so. Moreover, we do not have motor actions for interacting with furniture in general that are not motor actions for interacting with some basic-level object—chairs, tables, beds, etc. But superordinate categories do have other human-based attributes—like purposes and functions.

In addition, the complements of basic-level categories are not basic level. They do not have the kinds of properties that basic-level categories have. For example, consider nonchairs, that is, those things that are not chairs. What do they look like? Do you have a mental image of a general or an abstract nonchair? People seem not to. How do you interact with a nonchair? Is there some general motor action one performs with nonchairs? Apparently not. What is a nonchair used for? Do nonchairs have general functions? Apparently not.

In the classical theory, the complement of a set that is defined by necessary and sufficient conditions is another set that is defined by necessary and sufficient conditions. But the complement of a basic-level category is not itself a basic-level category.

Cue Validity

One of the ideas that Rosch has regularly stressed is that categories occur in systems, and such systems include contrasting categories. Categorization depends to a large extent on the nature of the system in which a category is embedded. For example, within the superordinate category of things-to-sit-on, *chair* contrasts with *stool, sofa, bench*, etc. *Chair* would no doubt cover a very different range if one of the contrasting categories, say, *stool* or *sofa*, were not present.

Rosch has made use of contrasting categories in trying to give a theory of basic-level categorization. At the basic level, Rosch has claimed, categories are maximally distinct—that is, they maximize perceived similarity among category members and minimize perceived similarities across contrasting categories. Rosch and others (1976) attempted to capture this intuition by means of a quantitative measure of what they called *category cue validity*.

Cue validity is the conditional probability that an object is in a particular category given its possession of some feature (or "cue"). The best cues are those that work all of the time for categories at a given level. For example, if you see a living thing with gills, you can be certain it is a fish. *Gills* thus has a cue validity of 1.0 for the category *fish*, and a cue validity of 0 for other categories. Rosch and her associates suggested that one could extend this definition of cue validity to characterize basic-level categories. They defined *category cue validity* as the sum of all the individual cue validities of the features associated with a category.

The highest cue validities in a taxonomic hierarchy, they reasoned, should occur at the basic level. For example, subordinate categories like *kitchen chair* should have a low category cue validity because most of the attributes of kitchen chairs would be shared with other kinds of chairs and only a few attributes would differentiate kitchen chairs from other chairs. The individual attributes shared across categories would have low cue validities for the kitchen chair category; thus, seeing a chair with a back doesn't give you much reason for thinking it's a kitchen chair rather than some other kind of chair. Since most of the individual cue validities for attributes would be low, the sum should be low.

Correspondingly, they reasoned that category cue validity would be low for superordinate categories like *furniture*, since they would have few or no common attributes. Since basic-level categories have many properties in common among their members and few across categories, their category cue validities should be highest.

This idea was put forth during the earlier phase of Rosch's career when she still believed that the relevant attributes for characterizing basic-level categories were objectively existing attributes "in the world." Murphy (1982) has shown, however, that if category cue validity is defined for objectively existing attributes, then that measure cannot pick out basic-level categories. Murphy observes that individual cue validities for a superordinate category are always greater than or equal to those for a basic-level category; the same must be true for their sums. For example,

> (a) If people know that some trucks [basic-level] have air brakes, they know that having air brakes is a possible cue for being a vehicle [superordinate].
> (b) People know that some animals [superordinate] have beaks, but that fish [basic-level] do not (thereby giving *animal* a valid cue that the basic category does not have). (Murphy 1982, p. 176)

Murphy observes that his objection could be gotten around under the assumption that most attributes are not directly linked to superordinate categories in memory. This would be true, for example, given Tversky and Hemenway's characterization of the basic level as that level at which most knowledge is organized. But this would require a psychological definition of attribute (equivalent to our *interactional properties*), not a

notion of attributes as existing objectively in the external world. But such a notion would presuppose a prior characterization of basic-level category—that level at which most knowledge is organized. Category cue validity defined for such psychological (or interactional) attributes might *correlate* with basic-level categorization, but it would not *pick out* basic-level categories; they would already have to have been picked out in order to apply the definition of category cue validity so that there was such a correlation. Thus, it seems reasonable to conclude that basic-level categories are, in fact, most differentiated in people's minds; but they are most differentiated because of their other properties, especially because most knowledge is organized at that level.

Clustering and Causation

Two of the themes that emerge from the research just discussed are the clustering of properties and the nonobjective, or interactional, character of properties relevant to human categorization. One of the most interesting of human categories from a philosophical point of view is the category of causes. Causation is represented in the grammar of most languages—and usually not just one kind of causation, but a variety of kinds. I have suggested elsewhere (Lakoff 1977) that the category of kinds of causation shows prototype effects in the ways that they are represented in natural languages. These effects are relatively uniform across languages.

We can account for these effects if we assume that prototypical causation is understood in terms of a cluster of interactional properties. This hypothesis appears to account best for the relation between language and conceptual structure, as well as for the relationships among the varieties of causation. The cluster seems to define a prototypical causation, and nonprototypical varieties of causation seem to be best characterizable in terms of deviations from that cluster.

Prototypical causation appears to be direct manipulation, which is characterized most typically by the following cluster of interactional properties:

1. There is an agent that does something.
2. There is a patient that undergoes a change to a new state.
3. Properties 1 and 2 constitute a single event; they overlap in time and space; the agent comes in contact with the patient.
4. Part of what the agent does (either the motion or the exercise of will) precedes the change in the patient.
5. The agent is the energy source; the patient is the energy goal; there is a transfer of energy from agent to patient.
6. There is a single definite agent and a single definite patient.

7. The agent is human.
8. a. The agent wills his action.
   b. The agent is in control of his action.
   c. The agent bears primary responsibility for both his action and the change.
9. The agent uses his hands, body, or some instrument.
10. The agent is looking at the patient, the change in the patient is perceptible, and the agent perceives the change.

The most representative examples of humanly relevant causation have all ten of these properties. This is the case in the most typical kinds of examples in the linguistics literature: Max broke the window, Brutus killed Caesar, etc. Billiard-ball causation, of the kind most discussed in the natural sciences, has properties 1 through 6. Indirect causation is not prototypical, since it fails in number 3, and possibly other conditions. According to this account, indirect causes are less representative examples of causation than direct causes. Multiple causes are less representative than single causes. Involuntary causation is less representative than voluntary causation. Many languages of the world meet the following generalization: The more direct the causation, the closer the morphemes expressing the cause and the result. This accounts for the distinction between *kill* and *cause to die*. *Kill* expresses direct causation, with cause and result expressed in a single morpheme—the closest possible connection. When would anyone ever say "cause to die"? In general, when there is no direct causation, when there is causation at a distance or accidental causation. Hinton (1982) gives a similar case from Mixtec, an Otomanguean language of Mexico. Mixtec has three causative morphemes: the word *sá?à*, and the prefixes *sá-* and *s-*. The longest of these corresponds to the most indirect causation, and the shortest to the most direct causation. An explanation of this fact about the linguistic expression of kinds of causation is provided by Lakoff and Johnson (1980, chap. 20).

What is particularly interesting about this state of affairs is that the best example of the *conceptual category* of causation is typically marked by a grammatical construction or a morpheme and that the word *cause* is reserved for noncentral members of the conceptual category. There is a good reason for this. The concept of causation—prototypical causation—is one of the most fundamental of human concepts. It is a concept that people around the world use in thought. It is used spontaneously, automatically, effortlessly, and often. Such concepts are usually coded right into the grammar of languages—either via grammatical constructions or grammatical morphemes. For this reason, the prototypical concept of causation is built into the grammar of the language, and the word *cause* is relegated to characterizing noncentral causation.