

George Lakoff

Women,

Fire, and

Dangerous

Things

*What Categories Reveal
about the Mind*

Notice

This Material May Be Returned by Call or Mail
to (703) 371-1111, Ext. 1111

RESERVE DESK

DEC 1 0 2001

DEC 1 0 2001

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 to 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.