Categories and Concepts

Edward E. Smith and Douglas L. Medin

Chapter 1
pp. 1-6.

Chapter 2
pp. 7-21

Harvard University Press
Cambridge, Massachusetts
London, England
1981
Categories and Concepts

Edward E. Smith and Douglas L. Medin

Chapter 1
pp. 1-6.

Chapter 2
pp. 7-21

Harvard University Press
Cambridge, Massachusetts
London, England
1981
INTRODUCTION

Without concepts, mental life would be chaotic. If we perceived each entity as unique, we would be overwhelmed by the sheer diversity of what we experience and unable to remember more than a minute fraction of what we encounter. And if each individual entity needed a distinct name, our language would be staggeringly complex and communication virtually impossible. Fortunately, though, we do not perceive, remember, and talk about each object and event as unique, but rather as an instance of a class or concept that we already know something about. When entering a new room, we experience one particular object as a member of the class of chairs, another as an instance of desks, and so on. Concepts thus give our world stability. They capture the notion that many objects or events are alike in some important respects, and hence can be thought about and responded to in ways we have already mastered. Concepts also allow us to go beyond the information given: for once we have assigned an entity to a class on the basis of its perceptible attributes, we can then infer some of its nonperceptible attributes. Having used perceptible properties like color and shape to decide an object is an apple, we can infer the object has a core that is currently invisible but that will make its presence known as soon as we bite into it. In short, concepts are critical for perceiving, remembering, talking, and thinking about objects and events in the world.

The point of this discussion is that a great deal in psychology hinges on how people acquire and use concepts, which in turn depends on the structure of concepts. And lately many psychologists seem to be changing their views about such structure. Until recently the dominant position—which we will call the classical view—held that all instances of a concept shared common properties, and that these common properties were necessary and sufficient to define the
concept. This view, which dates back to Aristotle, has always had its critics, but in the past decade the criticisms have become more frequent and intense, and new views have emerged. Perhaps the most prominent of these assumes that instances of a concept vary in the degree to which they share certain properties, and consequently vary in the degree to which they represent the concept. This view has sometimes gone by the name of prototype, but since this label has been used to mean many different things, we prefer to call it the probabilistic view. Another emerging view, which offers an even more extreme departure from the classical one, holds that there is no single representation of an entire class or concept, but only specific representations of the class's exemplars. This we call the exemplar view.

The development of alternatives to the classical view has been accompanied by a bustle of research activity. Indeed, the activity has been so frenzied that we think it is time to take a hard look at what has been learned. This book attempts to provide that look. It offers a systematic analysis of the three views of concepts, the processing models they have generated, and the empirical findings that the views and models endeavor to capture. In particular, we will emphasize the major problems that are responsible for the recent shift away from the classical view, show how the newer views take account of these problems, and indicate some of the costs incurred by such an accounting. But before beginning such a survey, it will be useful to provide some introductory examples of the three views.

To illustrate the classical view, we can consider the geometric concept of squares. Suppose that people in general represented this concept in terms of four properties: (1) closed figure, (2) four sides, (3) sides equal (in length), and (4) angles equal. Since these four properties, or criteria, would be applied to any object whose squareness is at issue, we have a unitary description of the concept "square." Moreover, the four properties that make up this concept are precisely those that any square must have. Roughly, then, to have a classical-view concept is to have a unitary description of all class members, where this description specifies the properties that every member must have.

Generalizing this approach, let us try to come up with a plausible classical-view concept that people might have of a cup. Such a concept might consist of the following five properties: (1) concrete object, (2) concave, (3) can hold liquids, (4) has a handle, and (5) can be used to drink hot liquid out of. These properties offer a unitary description of cups, but are all the properties true of everything people would call a cup? Properties 1-3 seem to be, but 4 and 5 are debatable. The teacups commonly used in Chinese restaurants typically do not have handles, yet they are still cups; and one can certainly imagine a poorly manufactured cup which conducts heat and so is useless for drinking hot liquids, but which we would still call a cup. But if we give up the last two properties, we are left with properties 1-3, which are true of some non-cups—bowls, for example—and so do not uniquely describe cups. Considerations like these led Labov (1973) to argue that people's concept of a cup does not conform to the classical view because the properties that make up the concept are not common to all members. Then what kind of view would capture people's concept of cup? According to Labov and others, one needs a view that posits a unitary description of cups, but where the properties in this description are true of most though not all members. This is the probabilistic view. Given such a view, some instances of the class are going to have more of the critical properties than others, and those that do will seem more representative of the concept.

To illustrate a third possible concept, let us consider the class of all psychiatric patients who have suicidal tendencies. One could try to construct a classical-view concept that clinicians might have for this class, but years of futile effort suggests that this cannot be done. What about a probabilistic concept—can it capture a clinician's concept of the suicide-prone? No doubt a probabilistic concept would work better, but even it might fail to capture the knowledge that a clinician uses in deciding that a particular patient is suicidal. For a probabilistic concept provides a single description of all people with suicidal tendencies, yet clinicians may sometimes decide that a patient is suicidal by comparing him to other specific patients known to be suicidal (this suggestion is taken from Tversky and Kahneman, 1973). That is, the class of people with suicidal tendencies may be represented not by a single description, but rather by separate descriptions for various patients known to be members of the class. This corresponds to the exemplar view of concepts.

The examples given above suggest that the various views of concepts can be partly understood in terms of two fundamental questions: (1) Is there a single or unitary description for all members of the class? and (2) Are the properties specified in a unitary description true of all members of the class? The classical view says yes to both questions: the probabilistic view says yes to the first but no to the second; and the exemplar view says no to the first question, thereby making the second one irrelevant. These distinctions, which we will refine and augment later in the book, are summarized in Figure 1.

Though our principal concern in this book is with the three
views, there are important preliminary matters to be considered first. Some of these have to do with what we consider a concept to be in the first place; others concern the nature of the properties that are used to describe concepts. These preliminary issues will be discussed in Chapter 2. In Chapter 3 we discuss the classical view. After describing it in detail, we review numerous arguments and experimental findings that have been offered as evidence against it. We will emphasize that the classical view is a theory about representations, and that to evaluate the view against experimental findings one must add processing assumptions to it, thereby converting a theory about concepts into a model about categorization. In Chapters 4, 5, and 6 we move on to the probabilistic view. These three chapters correspond to the method of describing concepts in the probabilistic view — by qualitative features, quantitative dimensions, or holistic patterns — for the three modes of description have somewhat different consequences. Again we will be concerned not only with the view itself, but with processing models that have been generated from it and with their ability to account for empirical findings. Chapter 7 deals with the exemplar view, again with an eye toward the specific models that instantiate the view and the findings they purport to explain. Finally, Chapter 8 summarizes our main conclusions and raises some unexplored issues.

What exactly are the empirical phenomena with which we intend to measure the views and models? Our major source of phenomena deals with categorization. Specifically, we will be concerned primarily with data on how adults use natural concepts with one-word names to classify things. Such "things" might be pictures of objects, say of a dog or a cup, and subjects might be asked if a pictured object belongs in a particular category. Or the things to be categorized might be denoted by words, as when subjects are asked, "Is a raisin a fruit?" and timed as they make their decisions.

(Our reasons for so emphasizing the categorization aspect of concepts will be discussed further at the beginning of the next chapter.)

In almost all cases we will be concerned with object concepts — animals, plants, human artifacts, and so on. Our main reason for choosing this domain is that it has been the most extensively studied in the last decade of experimental research. There is also an ancillary benefit of working with this domain — namely, it is a particularly interesting test case for the three views. That is, had we chosen geometric concepts, like "square," as a target domain, we might have prejudiced things in favor of the classical view; for we know that mathematicians have constructed classical-view descriptions of these concepts and have at least partially succeeded in inculcating these descriptions into many people's conceptual lives. Similarly, had we chosen as our domain abstract concepts, such as "love" or "brilliance," we might have prejudiced the case against the classical view: for no mathematician or metaphysician has come even close to constructing a classical-view description of such concepts. Thus natural objects and human artifacts offer an in-between case — between concepts that any schoolboy can define and concepts that no scholar can grapple with.

There is another rationale for concentrating on object concepts. There are two questions we can ask about any potential psychological concept: (1) Can it be given a classical-view definition in any language? and (2) Can it be given a classical-view definition in the language of the mind—that is, do people have a mental representation that corresponds to a classical-view definition of the concept? While question 2 is clearly a psychological one, question 1 is more the province of philosophy and logic. However, question 1 has important implications for a psychological analysis of concepts, for a no answer to it virtually necessitates a no answer to question 2. (We hedge with "virtually" just in case all things in the world turn out not to have classical-view definitions but our minds impose classical-view structure on them.) It is for this reason that we ignore abstract concepts like love and brilliance: since there is good reason to doubt they can be classically defined in any language, to study whether they can be classically defined psychologically may be beside the point. In contrast, the object concepts that we will focus on seem more likely to be definable in some language, that is, more likely to provide a yes answer to question 1, thereby making question 2 a worthwhile topic of study.

In addition to work on natural concepts, we will also consider some studies of artificial concepts. By an artificial concept we mean an equivalence class constructed for a particular experiment, say a class of schematic faces that tend to have similar properties. There...
Categories and Concepts

There are two reasons why such artificial classes are particularly useful for testing proposals about concepts. First, if natural concepts are hypothesized to have a particular structure, one can build this structure into an artificial class and see if people use this class in the same way they use a natural one; if they do, we have added support for the hypothesized structure of natural concepts. Second, the use of artificial concepts allows the experimenter to have precise control over the experiences or instances that the learner is exposed to in acquiring the concept. There are also drawbacks to using artificial concepts, which we will discuss later in the book. But first we need some discussion of what a concept is all about.

Preliminary Issues

Functions of a Concept

Though one's specific notion of a concept depends on which of the views one endorses, there are some aspects or functions of a concept that seem generally agreed upon. Most would agree that people use concepts both to provide a taxonomy of things in the world and to express relations between classes in that taxonomy (Woods, 1981). The taxonomic function can itself be split into the following two generally agreed upon component functions:

1. **Categorization.** This function involves determining that a specific instance is a member of a concept (for example, this particular creature is a guppy) or that one particular concept is a subset of another (for example, guppies are fish).
2. **Conceptual combination.** This function is responsible for enlarging the taxonomy by combining existent concepts into novel ones (for example, the concepts pet and fish can be combined into the conjunction pet-fish).

Similarly, the notion of using concepts to express relations can be subdivided into the following two component functions:

1. **Constructing propositional representations.** This function is at the heart of language understanding. The concepts denoted by the words in a sentence are mapped into a representation of the proposition expressed by that sentence (for example, the sentence "fish are friendly" is mapped into a proposition in which the concept of "friendly"—or a token of it—is predicated of the concept "fish").
2. Interrogating propositional representations. Here a relation between concepts is typically used as a basis for drawing certain inferences from representations (for example, given that one knows that guppies are fish and that fish have scales, one can infer that guppies have scales).

Though this fourfold breakdown has its rough edges, it gives some idea of the various research areas that bear on the nature of concepts. It also clearly suggests that the last three functions of concepts involve combinatorial procedures that play no role in the categorization function.

In this book we will focus almost exclusively on what we have called the categorization function. The main reason for so limiting our inquiry is that most of the critical work in psychology that bears on a shift from the classical to the probabilistic and exemplar views is categorization research. True, one can find analyses of constructing and interrogating representations that explicitly opt for the probabilistic or exemplar views over the classical one (for example, Rips, 1975; Clark and Clark, 1977; Collins, 1978), but the rationale usually given for this move is based on categorization studies. This is not to say that things have to be this way. Indeed, it is quite possible that research on conceptual combinations and on the construction and interrogation of propositional representations will ultimately be essential for constraining a view of concepts: we will return to this possibility at the end of the book. But until then, we are mainly concerned with what research on categorization can tell us about the nature of concepts.

Categorization and Inference

To say that concepts have a categorization function is to acknowledge that concepts are essentially pattern-recognition devices, which means that concepts are used to classify novel entities and to draw inferences about such entities. To have a concept of X is to know something about the properties of entities that belong to the class of X, and such properties can be used to categorize novel objects. Conversely, if you know nothing about a novel object but are told it is an instance of X, you can infer that the object has all or many of X's properties; that is, you can "run the categorization device in reverse."

Consider the concept of a hat. Let us assume that you have two properties for this concept: (1) it has an aperture that is the size of a human head, and (2) it was manufactured with the intent of a human wearing it. If we give you a novel object and ask, "Is it a hat?" presumably you would try to determine if it has properties 1 and 2. While you can check property 1 by means of perceptual tests, you need more than that to establish whether property 2 applies. Hence, classification may require recourse to nonperceptual information, and when we say that a concept is a pattern recognition device we do not necessarily mean that it uses only perceptual information. Now suppose that instead of asking you to classify the novel object, we hide it from view, tell you that it is a hat, and ask you to tell us something about it. Presumably you would say something about an aperture that was head-sized, and something like "you're supposed to wear it." These statements would be inferences, activated by gaining access to your concept via the word that denotes it, and something like this may happen whenever you hear the word.

Our example neatly separates classification from inference, but this is an oversimplification: most classification situations involve some inferences as well. If you have a prior reason to believe an object is a hat (it is on a hat rack, for example), you might infer that it has a head-sized opening and then perform only minimal perceptual checks to confirm your inference. More generally, where context suggests that an unexamined object belongs to a particular concept, inferences may be drawn about that object's properties, and such inferences will reduce the effort that need be put into classification. A similar phenomenon can occur even without context. Having determined that an object has some perceptual properties of a hat, you might tentatively assume that the object is a hat and then infer the less perceptual properties. This is like the apple example we used earlier—having determined that an object is small, round, and red, you assume that it is an apple and infer that it has a core.

While all three views of concepts—classical, probabilistic, and exemplar—would acknowledge that concepts have the twin functions of categorization and inference, the views differ with respect to the confidence one can have in the accomplishments of these functions. According to the classical view, aside from perceptual limitations one can have complete confidence in any categorization that considers all relevant properties of the concept, for such properties offer sufficient conditions for class membership. The case is different with the probabilistic and exemplar views. In these views, one can never be certain about a categorization, for categorization is always probabilistic when there are no sufficient conditions. A similar story holds for inferences. In the classical view, the inferences are deductions because the inferred properties are necessary ones (if we tell you "it's a square," you can deduce that it has four sides); in the other two views, many of the inferred properties will be probabilistic inferences because they do not conform to nec-
essay properties. These distinctions will become clearer when we consider the views in detail.

Stability of Concepts

A good deal of work presupposes that concepts are relatively stable mental representations. They are stable in two ways: within an individual and across individuals. A concept is stable within an individual to the extent that once a person has a concept, then, except for early developmental changes or physiological ones, that person will always have the same set of properties in mind. A concept is stable across individuals to the extent that when any two people have the same concept, they have the identical sets of properties in mind. If concepts are indeed stable in these two respects, it seems reasonable to think of a concept as a very bounded unit of knowledge, something of which we can say, “He has it completely, she has it partially, and this child doesn’t have it at all.” This kind of thinking seems to lie behind much of the literature on conceptual and semantic development, where researchers talk of a child’s acquiring and mastering concepts.

It turns out that this way of thinking is view-dependent. Believing in the stability of concepts is more consistent with the classical view than with the other two approaches. To illustrate, if one believes that concepts are represented by exemplars rather than by necessary and sufficient conditions, then frequent experience with a new exemplar, say a derby hat, can alter one’s concept of hats, which constitutes a breakdown in within-individual stability. Similarly, if one person has intensive experience with a particular exemplar while another person does not, they may end up with very different concepts, a breakdown in across-individual stability.

Because of this view-dependence, we caution the reader not to prejudge issues of concept stability. If in the next chapter we sound as if we are treating concepts as stable, bounded entities, it is because we are describing the classical view, which thinks of concepts in this way. If in Chapter 7 we sometimes sound as if we think concepts are capable of constantly changing with individual experience, again it is the view under consideration—this time the exemplar view—that is dictating our description.

Kinds of Properties for Concepts

Component versus Holistic Properties

In talking about the knowledge contained in a concept, we have thus far used the term property with no attempt to specify exactly what we mean. In our attempt to unpack some of the meaning of this term, we will rely heavily on distinctions previously drawn by Garner (1978) and Tversky (1977).

Following Garner (1978), we first distinguish between component and holistic properties of an object concept. A component property is, roughly, one that helps to describe an object but does not usually constitute a complete description of the object. Some examples should help to get this across. For one’s concept of a car, the component properties might include the following: having wheels, having a motor, the average shape of a car, and the fact that its major function is transportation. Note that while some of these components refer to parts of the object (like wheels), others depict global aspects of the entire object (shape), or the purpose or function of the object (transportation). In contrast, a holistic property offers a complete description of the object. For example, your concept of a car might be represented by some sort of template of an ideal car.

Intuitive though it may be, there is a problem with this distinction—namely, how do we distinguish global component properties from holistic ones? What, for example, is the difference between a car’s shape when it is treated as a global component and when it is given by a holistic template? We think there are two answers. First, usually when we talk about components there will be more than one of them, and we assume that each component is processed as a unit. This contrasts with a holistic property, like that given in a template, where one property suffices to represent the object class and that property is the only unit that gets processed. Second, when a global property like shape is treated as a component, we intend it as an abstraction from the object, not as a point-for-point isomorphism with the object. In contrast, when shape is given by a template, we often intend it more as a point-for-point isomorphism. Thus a component version of a property is usually more abstract than the corresponding holistic version.1

Rough as it is, we think the distinction between component and holistic properties is a useful one, and we employ it in what follows.

Dimensions and Features

If we decide to represent object concepts in terms of components, we have a choice of how to characterize these components—either by quantitative components, called dimensions, or by qualitative components, called features. To illustrate this distinction, let us consider one’s possible concepts of weapons. You could represent all weapons in terms of a few dimensions, like degree of potential damage, with, say, penknife near one end and atomic bomb near
the other. Alternatively, you could represent each weapon by a set of features: for knife, such features might include (1) sharp, (2) has a handle, and (3) metallic. The key difference is that dimensions naturally capture quantitative variations, while features indicate qualitative ones. Thus if two concepts differ with respect to a particular dimension, one concept must have more of that dimension (a higher value) than the other; for example, one weapon is more damaging than the other. But if two concepts differ with respect to a feature, then one concept has "it" while the other does not. Or as Garner (1978) puts it, a feature is a component that either exists or does not exist.

Each way of describing components appears to have a natural limitation: dimensions cannot handle qualitative variations while features will not work with quantitative variations. But these appearances are somewhat misleading. For featural descriptions can clearly be extended to cover quantitative variations, while a limiting case of what we mean by a dimension is extendable to qualitative variations. These ideas require some amplification.

Consider first the extendability of featural descriptions. As Atkinson and Estes (1963) pointed out some time ago, the values for any dimension may be expressed by a set of nested features. It is simplest to illustrate with the dimension of line length. Suppose that the shortest length you can detect is one millimeter; we can represent the length of this standard line by a single feature, call it F₁. Suppose the next shortest line you can detect as different from our standard is two millimeters long: that is, a just-noticeable-difference, or JND, is one millimeter. We represent that JND by another feature, F₂, and represent the second line by the feature set F₁ and F₂. Hence the feature representing the length of the first line is nested within the feature set representing the length of the second line. By continuing in this way — representing each successive JND by a new feature, and nesting the feature sets of shorter lines in those of longer ones — we can represent any value of the dimension of length by a set of nested features. The same procedure will work for any quantitative dimension. And the resulting feature representation captures the important intuition that the closer two objects are on a supposed dimension, the more features they share and the fewer features there are to distinguish between them. So feature descriptions can clearly cover quantitative variations, and can do so in a manner that captures intuitions about quantitative similarity in terms of common and distinctive features (Tversky, 1977). (We note, though, that some implications of this extension of feature descriptions may prove incorrect: for example, since the descriptions of entities at the upper end of a continuum need to contain more features than those at the lower end, the former entities would be expected to be more complex.)

The case for the extendability of dimensions is more precarious. Note first that while dimensions are typically quantitative, they need not be continuous: discrete dimensions are still dimensions (Garner, 1978). To return to our weapon example, your dimension of potential damage may only include five discriminable values, but it is still quantitative as long as you can reliably judge one value to be greater than another. This raises the possibility of a dimension with only two values, where one value indicates some positive amount and the other, or zero value, indicates the absence of the dimension. Such a limiting case of a dimension captures something of what we mean by a feature, with the positive dimension value mapping onto the presence of a feature and the zero value mapping onto the absence of the feature. There are two problems with this extendability argument. First, a zero value on a dimension may not capture what people mean by the absence of a feature: thus it is one thing to say that the feature of "red" does not apply to the concept of truth, and quite another to say that truth has a zero value of redness. Second, our limiting case of a dimension may no longer constitute a "dimension" in any rigorous sense. Following Beals, Krantz, and Tversky (1968), it seems reasonable to insist that for something to be called a dimension it must have the property of betweenness — roughly that there exist some triple of values along the dimension such that one value appears (psychologically) to be between the others. Given this constraint, binary-valued contrasts are not dimensions, and true featural variations cannot be captured by dimensional descriptions. The upshot is that only a weak notion of dimensions can be fully extended to cover featural variations, and in this respect dimensional descriptions may be less powerful and more constrained than their featural counterparts.

Let us now return to differences between dimensions and features as they are typically used. In addition to the quantitative versus qualitative distinction, there is another difference in how dimensions and features are employed to represent objects. In describing a set of related objects in terms of dimensions, a person (1) typically uses a small number of dimensions and (2) assumes that every object has some value on each one. For example, a dimensional description of weapons might involve only three to five dimensions and assume that each weapon can be placed on each dimension. Feature descriptions, in contrast, are likely (1) to use many features and (2) to include features for some objects that are not applicable to others. One weapon, say a knife, might have dozens of features corresponding to the size, shape, and structure of its blade and handle, while another weapon, say a grenade, might be described in terms of fewer features that need not correspond to any of those used to describe the previous instance (what feature of a grenade
can possibly correspond to the pointedness of a knife's blade). The feature approach again seems less constrained than the dimensional one.\textsuperscript{3}

Regardless of whether dimensions or features are used to describe concepts, the goal is the same: to make apparent the relations between concepts. There are numerous ways this can be done. With the dimensional approach, each concept may be represented by a list of the values it takes on a set of dimensions, and the relation between any pair of concepts is determined by the proximity of their values along all dimensions; alternatively, the concepts may be represented as points in a multidimensional space and the relation between concepts given by the distance between the corresponding points. With the featural approach, concepts are often represented by feature lists, and the relation between concepts is given in terms of common and distinctive features. Figure 2 illustrates the latter approach. It contains (partial) featural descriptions of the concepts of robin, chicken, collie, and daisy. The features are of various types—animate, feathered, furry, and so on—but they are all potentially useful in describing relations between concepts. For example, Figure 2 makes it apparent why robin and chicken are more similar than robin and collie: robin and chicken have more common features and fewer distinctive ones than robin and collie, and the similarity of any two items should increase with the number of shared features and decrease with the number of distinctive ones (Tversky, 1977). An alternative to simple feature lists is to depict not only the features of a concept but relations between them as well—for example, to indicate explicitly that the features "animate" and "feathered" are correlated. When this is done many of the features are interconnected, and a network (rather than a list) of features results.

Though all the alternatives discussed above can be used to represent concepts and their interrelations, the feature list approach has

\begin{center}
\begin{tabular}{ l l l l l }
Robin & Chicken & Collie & Daisy \\
\hline
$F_1$ animate & $F_1$ animate & $F_1$ animate & $F_1$ inanimate \\
$F_2$ feathered & $F_2$ feathered & $F_2$ furry & $F_2$ stem \\
$F_3$ flies & $F_3$ pecks$^1$ & $F_3$ brown-grey & $F_3$ white \\
\hline
\end{tabular}
\end{center}

Figure 2 Some featural descriptions

been the most extensively used, particularly for classical-view representations of concepts. For this reason, the rest of this chapter focuses on features per se. Later, when we take up the probabilistic view, we will return to other kinds of descriptions.

Contents of Features

Constraints on Features

If we are going to describe concepts in terms of features, we need some guidelines as to what can count as a feature. If any property, no matter how arbitrary or complex, could serve as a feature, we would soon be faced with all kinds of difficulties. For example, one could have a "concept" that consisted of the pseudofeatures "saw all of Tuesday Weld's movies in one week" and "prefers Bloody Marys that contain two parts Worcestershire sauce for one part Tabasco." Such a "concept" seems totally unnatural and points to the need to constrain features.

One constraint follows directly from the main purpose of features. If features make apparent relations between concepts, then a property is a useful feature to the extent that it reveals many relations between concepts. The feature of being male, for example, is useful because it brings out the relation between various classes of animate beings, like that between son and father, or between boy and colt. In contrast, neither of the pseudofeatures mentioned above would reveal many relations between concepts.

The above constraint can be strengthened. A set of features should not only make relations apparent but ideally should exhaust all potential relations between the concepts of interest. To illustrate, suppose someone postulated that young-male served as a single feature in describing concepts of animate beings. Although this would make apparent the relation between boy and colt (both contain the feature of young-male), it would not exhaust all potential relations of interest; for example, it would not show the relation between boy and girl. To capture the latter relation we obviously need to decompose young-male into the component features of (1) young (in the sense of a quality, not a quantity) and (2) male, where feature 1 can now be used to show the relation between boy and girl. To ask for features that exhaust all possible relations is to ask for features that are not themselves decomposable, that is, for primitives or all-or-none features. This is a difficult constraint to meet (we never really invoke it), but it is surely powerful since it unequivocally excludes a pseudofeature like "saw all of Tuesday Weld's movies in one week."\textsuperscript{4}

As a second constraint we seek features with some generality, in
Categories and Concepts

the sense that a feature should apply to many concepts within a domain rather than to a few. (Of course, we would not want the feature to apply to every concept in the domain, for then it would have no discriminative value.) A kind of corollary of this generality constraint is that the number of features needed to describe a conceptual domain should be small relative to the number of concepts in that domain; that is, all other things being equal, the greater the applicability of a set of features, the fewer the features needed to characterize a domain. Although this constraint of generality rests on rough intuitions, it again seems sufficient to rule out properties like "saw all of Tuesday Weld's movies in one week." (Note that even though we seek features that are as general as possible, they still may be more numerous than the dimensions needed to describe the same conceptual domain.)

The two constraints discussed above deal with structural aspects of concepts. Consider now a processing constraint: the features posited should serve as the inputs for categorization processes. Thus, for the feature of male to be a useful one, it must not only reveal relations between concepts in an economical fashion (our two structural constraints), but it must also be used by people in reaching decisions about categorization. Though this constraint may carry little force with those interested in a purely structural approach to concepts (many linguists and philosophers), it is essential for psychologists interested in the processing of concepts. Indeed, it must take precedence over other constraints, as we will readily accept a nonprimitive and nongeneral property as a feature if there is convincing evidence that it is used in categorization.

It is instructive to illustrate these three constraints in relation to recent work on the perception of letters. This work essentially treats each letter of the alphabet as a separate concept and asks what features should be posited to define the concept. Introspection as well as early physiological evidence (for example, Hubel and Wiesel, 1962) suggested that the features may be line segments, angles, and curves. But exactly which lines, angles, and curves are posited as features depends on the three constraints. Just about everyone would agree that the features must account for why, under limited viewing conditions, a particular letter (for example, P) is confused with some letters (B, R) but not with others (X, V; see Gibson, 1969; Rumelhart and Siple, 1974; Townsend and Ashby, 1976). Since the probability of confusing a particular pair of letters is assumed to be a measure of the similarity of that pair, accounting for patterns of confusions amounts to making similarity relations apparent, which is our first constraint. Furthermore, all serious proposals in this domain posit features that apply to many letters—a vertical line is a feature of B, D, E, F, and so on—and this is in keeping with our second constraint. Finally, researchers have tried to show that these features are the inputs to recognition processes by demonstrating, for example, that an estimate of the exact features extracted from a letter can be used to predict the accuracy, confidence, and speed with which a letter can be categorized (see Rumelhart and Siple, 1974; Townsend and Ashby, 1976). Failure to meet this third constraint would usually make researchers discard their featural definitions.

This example can also be used to address a fundamental issue about features, one that arises whenever any component properties are posited. The issue, in a nutshell, is that there are no a priori means for distinguishing between a concept and a feature, for any feature can itself be treated as a concept. In the letter perception work, for example, the concept for the letter E might include the feature of a vertical line, but a vertical line is also a possible concept, so what reasons are there for calling it a "feature" and reserving the term "concept" for E? Or to put it another way, do not all the problems we face in explaining how E is represented, accessed, and processed also apply to how a vertical line is represented, accessed, and processed? If so, explaining E's in terms of vertical lines is simply begging the question, and the business of seeking constraints on features is beside the point.

There is a way out of this conceptual morass, we think, that is nicely demonstrated by the research on letter perception. In essence the solution is this: true, there may be no conceptual reasons for distinguishing between concepts and features, but there may well be empirical reasons for making this distinction. Thus, when it comes to explaining results in letter perception, there are many empirical reasons for assuming that a vertical line is a feature while the letter E is not. To restate a previously mentioned finding, treating a vertical line as a feature helps to explain why some letter pairs are more likely than others to be confused under limited viewing conditions; for example, E and F are more likely to be confused than E and V because the former pair shares a vertical line. No such account of confusions is forthcoming if one treats E and F as features here. This is just one example among many of how treating components such as vertical lines as features has led to an account of empirical findings about letter perception that might have remained a mystery had researchers not reduced the letters to their component properties.

These empirical successes in distinguishing visual features from visual concepts provides a rationale for thinking that the same kind of distinction can be supported with semantic concepts and their
features. Indeed, we think there is already some evidence that describing concepts in terms of their component features will lead to accounts of empirical phenomena that would not be forthcoming otherwise.

One important caveat must be added to the foregoing discussion. If empirical results determine whether we should assign an entity the role of feature or that of concept, then a major change in the set of results to be accounted for can lead to a change in role assignment. To illustrate, if we change the empirical results of interest from findings on letter perception to findings on word perception, we may want to change the status of $E$ from concept to feature. We may, for example, need to assume that $E$ is a feature of word concepts in order to explain why certain words, such as bed and ted, are likely to be confused under limited viewing conditions. When such a change in role assignments occurs, it may indicate we are dealing with a system that has several levels, so that what is a concept at level $n$ is best thought of as a feature at level $n + 1$.

Perceptibility of Features

Perceptual versus Abstract Features

Even with the constraints mentioned above, the features of object concepts can vary a great deal. One particular source of variation, the perceptibility of features, will be of great concern to us. At one extreme are true perceptual features — those corresponding to outputs from the perceptual system; examples might include the presence of a line or the presence of some degree of curvature. At the other extreme are abstract features that have minimal connection to perceptual experience: as an example, a feature of the concept “clothing” might be “manufactured with the intent of human usage.” And then there are myriad in-between cases, including the familiar features used with animal concepts like “animate” and “male,” and functional features like “used for transportation.”

This perceptibility distinction raises a number of issues. The first is whether the features of object concepts should be restricted to more perceptual features, that is, whether perceptibility should be another constraint on features. This constraint has frequently been tried, particularly in conjunction with the classical view, for there is a long tradition in psychology to define stimulus objects in physical terms. This move always seems to end in failure, however. The most recent effort, and it is a valiant one, can be found in the work of Miller and Johnson-Laird (1976). After systematically reviewing what is known about the possible outputs of the human perceptual system, they attempt to use only these outputs in their feature de-

scription of certain object concepts (such as “table”). They convincingly argue that this enterprise is doomed.

Miller and Johnson-Laird first try to represent the concept of table by perceptual features like the following: (1) is connected and rigid, (2) has a flat and horizontal top, and (3) has vertical legs. They then spell out two serious drawbacks to this approach. First, some of these features are not true of all things we call tables (for example, many drafting tables do not have a horizontal top); this seems to be the case no matter how one alters the list of perceptual features. Second, if concept descriptions are confined to perceptual features, how can one ever explain that the same object may sometimes be treated as an instance of the concept “kitchen table” and other times as an instance of the concept “work table”? The solution to these problems, according to Miller and Johnson-Laird, is to include functional features in concept descriptions. For example, in the concept “table” if we replace the perceptual features “has a flat and horizontal top” with the functional feature “has a top capable of support,” we can see how drafting tables can be included in the general concept of tables. And once we accept functional features, there is no problem in explaining how the same object can sometimes be subcategorized a kitchen table and sometimes a work table: the more specific concepts of kitchen table and work table have different functional features, and sometimes use of the object will match one of these functions, while other times its use will match the other function. Given arguments like this, we take it as a starting point that the features of an object concept can contain abstract and functional features as well as perceptual ones.

But now we have a new problem. How can one decide that a visually presented object is an instance of a particular concept, since features of the object must be true perceptual ones whereas some features of the concept may be abstract? The only solution is that one must have some knowledge that is capable of mediating between the features at the two levels; that is, to determine whether an abstract feature is perceptually instantiated in an object, one must have recourse to ancillary knowledge about the relation between abstract and perceptual features.

Let us illustrate with the concept of boy. Following traditional linguistic wisdom, we assume that people may represent this concept by three abstract features: human, male, and young (relative to some criterion). How can one use these features to determine that some particular person (a clothed one) is in fact a boy? To establish that the person is male, one would presumably check things like body proportions, dress, and hairstyle, which are more perceptual features. (Even these features are more abstract than
outputs of the perceptual system, but we will ignore this complication in what follows.) To establish that the person is young, one would use perceptual features like height and weight. Hence, both male and young are instantiated by more perceptual features, and categorization therefore requires ancillary knowledge about which perceptual features instantiate which abstract ones.

The Core versus Identification Procedure of a Concept

In view of the foregoing discussion, let us distinguish between the core of the concept and the ancillary knowledge, which we will call the identification procedure. (A similar distinction is used by Miller and Johnson-Laird, 1976, and by Woods, 1981). The core of the concept “boy” would include the abstract features of human, male, and young, while the identification procedure would include perceptual features like height and weight as well as an indication that they instantiate the feature of young. The features of the core are primarily responsible for revealing certain relations between concepts, like that between boy and girl (both young and human) or that between boy and colt (both young and male), while the features of the identification procedure are used for categorizing real-world objects. To the extent that the features of a core are perceptual, there is no need for a separate identification procedure; for a geometric concept like “square,” the core contains perceptual features like “four-sided” which can be used for categorizing real-world objects, thereby making an identification procedure superfluous. Perceptual features in the core, then, shrink the requisite identification procedure. There is also another way in which the core takes psychological precedence over the identification procedure: abstract features in the core determine the nature of the perceptual features in the identification procedure. To return to the boy example, being male (a core feature) is partly a matter of hormones, and these hormones produce secondary sex characteristics like body proportions (a feature in the identification procedure). From this perspective, if one wants to understand the nature of concepts, it is best to go for the core.

The distinction discussed above is clearly related to Frege’s classic distinction (1892) between sense and reference: the sense of a concept is given by its relation to other concepts, while the reference of a concept is given by its relation to objects and events in the world. Thus Frege’s sense corresponds to our core, and Frege’s reference to our identification procedure. Just as we have assumed that the core dictates the contents of the identification procedure, so Frege postulated that the sense of a concept determines its reference. We prefer our terms to Frege’s because his distinction was intended as a philosophical rather than a psychological one, and hence it may be wrong to identify his notion of reference with the features people actually use in categorizing real-world objects.

Consider now how we can use our core versus identification distinction in evaluating experimental studies of concepts. As mentioned earlier, in some experiments subjects are asked to decide whether statements about subset relations (for example, “A raisin is a fruit”) are true or false. In this kind of semantic task, subjects can answer just by consulting the cores of the relevant concepts (do the core features of the concept fruit match those of the concept raisin?). They might also consider the identification procedures of the two concepts, but there is no necessity to do so. In other experiments, however, subjects are asked to decide whether a pictured object is an instance of a target concept (for example, “Is this specific object a fruit?”). In such a perceptual task, subjects must consult the identification procedure of the target concept unless the core consists entirely of perceptual features. Since the views of concepts we are interested in, particularly the classical view, are mainly concerned with concept cores, we will emphasize experimental studies that either use a semantic task or use a perceptual one where there is good reason to believe the core consists entirely of perceptual features. Later, at the end of our discussion of the classical view, we will have occasion to question the wisdom of our decision.