CHAPTER 22

Concepts and Categorization

ROBERT L. GODSTONE AN ALAN KERSTEN

WHAT ARE CONCEPTS? 600
Concepts, Categories, and Internal Representations 600
Equivalence Classes 606
WHAT DO CONCEPTS DO FOR US? 601
Components of Thought 601
Inductive Predictions 601
Categorization 602
Cognitive Economy 602
HOW ARE CONCEPTS REPRESENTED? 603
Rules 604
Prototypes 605

Issues related to concepts and categorization are nearly ubiquitous in psychology because of people's natural tendency to perceive a thing as something. We have a powerful impulse to interpret our world, and act of interpretation is an act of "seeing something as X" rather than simply seeing it (Wittgenstein, 1953), is fundamentally an act of categorization.

The attraction of research on concepts is that an extremely wide variety of cognitive acts can be understood as categorizations. Identifying the person sitting across from you at the breakfast table involves categorizing someone as (for example) your spouse. Diagnosing the cause of someone's illness involves a disease categorization. Interpreting a painting as a Picasso, an artifact as Mayan, a geometry as non-Euclidean, a fugue as baroque, a conversationalist as charming, a wine as a Bordeaux, and a government as socialist are categorizations at various levels of abstraction. The typically unspoken assumption of research on concepts is that these cognitive acts have something in common. That is, there are principles that explain many or all acts of categorization. This assumption is controversial (see Medin, Lynch, & Solomon, 2000), but is perhaps justified by the potential payoff of discovering common principles governing concepts in their diverse manifestations.

The desirability of a general account of concept learning has led the field to focus its energy on what might be called generic concepts. Experiments typically involve artificial categories that are (it is hoped) unfamiliar to the subjects. Formal models of concept learning and use are constructed to be able to handle any kind of concept irrespective of its content. Although there are exceptions to this general trend (Malt, 1994; Ross & Murphy, 1999), much of the mainstream empirical and theoretical work on concept learning is concerned not with explaining how particular concepts are created, but with how concepts in general are represented and processed.

One manifestation of this approach is that the members of a concept are often given an abstract symbolic representation. For example, Table 22.1 shows a typical notation used to describe the stimuli seen by a subject in a psychological experiment or presented to a formal model of concept learning. Notice objects belong to two categories, and each object is defined by its value along four binary dimensions. In this notation, objects from Category A typically have values of 1 on each of the four dimensions, whereas objects from Category B have values of 0. The dimensions are typically unrelated to each other, and assigning values of 0 and 1 to a dimension is
TABLE 22.1 A Common Category Structure, Originally Used by Medin & Schaffer (1979)

<table>
<thead>
<tr>
<th>Category</th>
<th>Dimension D1</th>
<th>Dimension D2</th>
<th>Dimension D3</th>
<th>Dimension D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Category B</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

arbitrary. For example, for a color dimension, red may be assigned a value of 0 and blue a value of 1. The exact category structure of Table 22.1 has been used in at least 30 studies reviewed by J. D. Smith & Minda, 2003, instantiated by stimuli as diverse as geometric forms (Nosofsky, Kroisckke, & McKinley, 1992), cartoons of faces (Medin & Schaffer, 1978), yearbook photographs (Medin, Dewey, & Murphy, 1983), and line drawings of rocket ships (Nosofsky, Palmeri, & McKinley, 1994). These authors are not particularly interested in the category structure of Table 22.1 and are certainly not interested in the categorization of rocket ships per se. Instead, they choose their stimuli and stimuli so as to be (a) unfamiliar (so that learning is required), (b) well controlled (dimensions are approximately equally salient and independent), (c) diagnostic with respect to theories, and (d) potentially generalizable to natural categories that people learn. Work on genetic concepts is very valuable if it turns out that there are domain-general principles underlying human concepts that can be discovered. Still, there is no a priori reason to assume that all concepts will follow the same principles, or that we can generalize from genetic concepts to naturally occurring concepts.

WHAT ARE CONCEPTS?

Concepts, Categories, and Internal Representations

A good starting place is Edward E. Smith's (1989) characterization that a concept is "a mental representation of a class or individual and dots with what is being represented and bow that information is typically used during the categorization" (p. 502). It is common to distinguish between a concept and a category. A concept refers to a mentally possessed idea or notion, whereas a category refers to a set of entities that are grouped together. The concept dog is whatever psychological state signifies thoughts of dogs. The category dog consists of all the entities in the real world that are appropriately categorized as dogs. The question of whether concepts determine categories or vice versa is an important foundational controversy. If one assumes the primacy of external categories of entities, then one will tend to view concept learning as the enterprise of inductively creating mental structures that predict these categories. An extreme version of this view is the exemplar model of concept learning (Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1984; see also Capaldi's chapter in this volume), in which one's internal representation for a concept is nothing more than the set of all of the externally supplied examples of the concept to which one has been exposed. If one assumes the primacy of mental concepts, then one tends to view external categories as the end product of applying these internal concepts to observed entities. An extreme version of this approach is to argue that the external world does not inherently consist of rocks, dogs, and tables; these are mental concepts that organize an otherwise unstructured external world (Lakoff, 1987).

Equivalence Classes

Another important aspect of concepts is that they are equivalence classes. In the classical notion of an equivalence class, distinguishable stimuli come to be treated as the same thing once they have been placed in the same category (Sidman, 1994). This kind of equivalence is too strong when it comes to human concepts because even when we place two objects into the same category, we do not treat them as the same thing for all purposes. Some researchers have stressed the intrinsic variability of human concepts—variability that makes it unlikely that a concept has the same meaning or reasoning each time it is used (Barsalou, 1987; Thelen & Smith, 1994). Still, it is impressive the extent to which perceptually dissimilar things can be treated equivalently, given the appropriate conceptualization. To the biologist armed with a strong mammal concept, even whales and dogs may be treated as equivalent in many situations related to biochemistry, child rearing, and thermoregulation. Even sea lions may possess equivalent classes, as Schusterman, Reichmuth, and Katok (2000) have argued that these animals show free substitution between two entities once they have been associated together.

Equivalence classes are relatively impervious to superficial similarities. Once one has formed a concept that treats all skunks as equivalent for some purposes, irrelevant variations among skunks can be greatly de-emphasized. When subjects are told a story in which scientists discover that an animal that looks exactly like a raccoon actually contains the internal organs of a skunk and has skunk parents and skunk children, they often categorize the animal as a skunk (Kell, 1997).
Rips, 1989). People may never be able to transcend superficial
- nal appearances when categorizing objects (Goldstone, 1984),
or it is clear that they would want to (Jones & Smith, 1993).

Revolution, one of the most powerful aspects of con-
ccepts is their ability to make superficially different things
like (Scherer, 1996). If one has the concept Things to re-
mune from a burning house, even children and jewelry be-

Components of Thought

Concepts are cognitive elements that combine to generatively
produce an infinite variety of thoughts. Just as a finite set
of building blocks can be constructed into an endless variety
of architectural structures, so can concepts act as building
blocks for an endless variety of complex thoughts, claiming
that concepts are cognitive elements does not entail that they
are primitive elements in the sense of existing without being
learned and without being constructed from other concepts.
Some theorists have argued that concepts such as bachelor,
kill, and house are primitive in this sense (Fodor, 1975;
Fodor, Garrett, Walker, & Parke, 1980), but a considerable
body of evidence suggests that concepts typically are ac-
sic elements that are themselves decomposable into ele-
matic elements (McNaughton & Miller, 1985).

Once a concept has been formed, it can enter into compo-
sitions with other concepts. Several researchers have studied
how novel combinations of concepts are produced and com-
prehended. For example, how does one interpret the term
buffalo paper when one first hears it? Is it in the shape of
a buffalo, paper used to wrap buffalo robes presented as gifts, an
essay on the subject of buffalo, coarse paper, or something
like fly paper but used to catch bison? Interpretations of word
combinations are often created by finding a relation that con-
nects the two concepts. In Murphy's (1988) concept special-
ization model, one interprets nouns noun combinations by
finding a variable that the second noun has that can be filled
by the first noun. By this account, a robin snake might be
interpreted as a snake that eats robins once robin is used to
fill the slot eat in the snake concept. Willsenski (1987, 1988;
Willsenski & Lovel, 1988) has argued that properties from
one concept are often transformed to another concept, and
that this is more likely to occur if the concepts are simi-
lar, with parts that can be easily aligned. By this account, a
robin snake might be interpreted as a snake with a red belly,
once the attribute red bellied from the robin is transferred to
the snake.

In addition to promoting creative thought, the combinato-
rial power of concepts is required for cognitive systematicity
(Fodor & Pylyshyn, 1988). The notion of systematicity is that
a system's ability to maintain complex thoughts is instrin-
sely connected to its ability to entertain the combinations of
those thoughts. In the field of conceptual combination, this
has appeared as the issue of whether the meaning of a combi-
nation of concepts can be deduced on the basis of the mean-
ings of its constituents. On the one hand, there are some
salient violations of this type of systematicity. When adja-
cutive and noun concepts are combined, there are sometimes
emergent interactions that cannot be predicted by the "main
effects" of the concepts themselves. For example, the concept
grey hair is more similar to white hair than to black hair,
but grey cloud is more similar to black cloud than to white
cloud (Medin & Shoben, 1988). Wooden spoons are judged
to be fairly large (for spoons), even though this property is
not generally possessed by wooden objects or spoons in gen-
eral (Medin & Shoben, 1988). On the other hand, there have
been notable successes in predicting how well an object fits a
cognitive description based on how well it fits the individ-
ual descriptions that comprise the conjunction (Hampton,
1987, 1997; Norms, De Rauck, Flamson, & Van-Michelen,
1999). A restorable recalculation of these results is that
when concepts are combined the concepts' meanings system-
atically determine the meaning of the conjunction, but emer-
gent interactions and real-world plausibility also shape the
conjunction's meaning.

Inductive Prediction

Concepts allow us to generalize our experiences with some
objects to other objects from the same category. Experience
with one shaving dog may lead us to suspect that an
unfamiliar dog may have the same proclivity. These inductive generalizations may be wrong and can lead to unfair stereotypes if inadequately supported by data, but if an organism is to survive in a world that has some systematics, it must “go beyond the information given” (Bronner, 1973) and generalize what it has learned. The concepts we use most often are useful because they allow many properties to be predicted inductively. To see why this is the case, we must digress slightly and consider different types of concepts. Category tax can be arranged roughly in order of their grounding by saliency: natural kinds (dog and oak tree), man-made artifacts (hammer, airplane, and chair), ad hoc categories (things to take out of a burning house, and things that could be stood on to reach a lightbulb), and abstract schemas or metaphors (e.g., events in which a kind action is repaid with cruelty, metaphorical prisons, and problems that are solved by breaking a large fight into parts that converge on a target). For the latter categories, members need not have very much in common at all. An un rewarded job and a relationship that cannot be ended may both be metaphorical prisons, but the situations may share little other than this.

Unlike ad hoc and metaphor-based categories, most natural kinds and many artifacts are characterized by members that share many features. In a series of studies, Rosch (Rosch, 1975; Rosch & Mervis, 1975; see also the chapter in this volume by Palmer and by Treiman, Clifton, Meyer, & Wurm) has shown that the members of natural kind and artifact “basic-level” categories such as chair, trout, bus, apple, saw, and guitar are characterized by high within-category overall similarity. Subjects listed features for basic-level categories, as well as for broader superordinate (e.g., furniture) and narrower subordinate (e.g., kitchen chair) categories. An index of within-category similarity was obtained by tallying the number of features listed by subjects that were common to items in the same category. Items within a basic-level category tend to have several features in common, far more than items within a superordinate category and almost as many as items that share a subordinate category. Rosch (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Yogues-Braem, 1976) argues that categories are defined by family resemblance; category members need not all share a definitional feature, but they tend to have several features in common. Furthermore, she argues that people’s basic-level categories preserve the intrinsic correlational structure of the world. All feature combinations are not equally likely. For example, in the animal kingdom, flying is correlated with laying eggs and possessing a beak. There are “clumps” of features that tend to occur together. Some categories (e.g., ad hoc categories) do not conform to these clumps, but many of our most natural-appearing categories do.

These natural categories also permit many inductive references. If we know something belongs to the category dog, then we know that it probably has four legs and two eyes, eats dog food, is someone’s pet, pants, barks, is bigger than a breadbox, and so on. Generally, natural-kind objects, particularly those at Rosch’s basic level, permit many inferences. Basic-level categories allow many inferences because their members share similarities across many dimensions or features. Ad hoc categories and highly metaphorical categories permit few inductive inferences, but in certain situations the inferences they allow are so important that the categories are created on a “by-need” basis. Our interesting possibility is that ad hoc concepts are created to fulfill an inductive need, and that standard taxonomic categories such as bird and hammer simply become automatically triggered because they have been used often, whereas ad hoc categories are created only when specifically needed (Baranski, 1982, 1991). In any case, evaluating the inductive potential of a concept goes a long way toward understanding why we have the concepts that we do. The single concept peaches, llamas, telephone answering machines, or Ringo Starr is an unlikely concept because belonging in this concept predicts very little. Several researchers have been developing the notion that the concepts we possess are those that maximize inductive potential (Anderson, 1991; Heer 2000; Oaksford & Chater, 1998).

Communication
Communication between people is enormously facilitated if the people can count upon sharing a set of common concepts. By uttering a simple sentence such as “Ed is a football player” one can transmit a wealth of information to a college student, dealing with the probabilities of Ed’s being strong, having visible tendencies, being a college physics or physical education major, and having a history of steroid use. Markman and Makin (1996) argue that a major factor in shaping our concepts is the need to communicate efficiently. The more a person’s concepts become more consistent and systematic over time in order to establish reference unambiguously to another individual with whom they need to communicate (see also Garrod & Doherty, 1994).

Cognitive Economy
We can discriminate far more stimuli than those for which we have concepts. For example, estimates suggest that we perceptually discriminate at least 10,000 colors from one another, but we have far fewer color concepts that can be achieved.
encoding concepts rather than entire raw (unprocessed) inputs. A classic study by Posner and Keele (1967) found that subjects code letters such as A by a raw, physical code, but that this code rapidly (within 2 s) gives way to a more abstract conceptual code that A and a share. Huttenlocher, Hedges, and Veeve (2000) develop a formal model in which judgments about a stimulus are based on both its category membership and its individuating information. As predicted by the model, when subjects are asked to reproduce a stimulus, their reproductions reflect a compromise between the stimulus itself and the category to which it belongs. When a delay is introduced between seeing the stimulus and reproducing it, the contribution of category-level information relative to individual-level information increases (Crawford, Huttenlocher, & Englebretson, 2000). Together with studies showing that, over time, people tend to preserve the gist of a category rather than the exact elements that constitute it (e.g., Posner & Keele, 1970), these results suggest that through the preservation of category-level information rather than individual-level information, efficient long-term representations can be maintained.

From an information-theory perspective, storing a category in memory rather than a complete description of an individual is efficient because fewer bits of information are required to specify the category. For example, Figure 22.1 depicts a set of objects (shown by circles) described along two dimensions. Rather than preserving the complete description of each of the 19 objects, one can create a reasonably faithful representation of the distribution of objects by storing only the positions of the four triangles in Figure 22.1. This kind of information reduction is particularly significant because computational algorithms exist that can automatically form these categories when supplied with the objects (Kohonen, 1995). For example, the competitive learning algorithm (Rumelhart & Zipser, 1985) begins with random positions for the triangles, and when an object is presented, the triangle that is closest to the object moves its position even closer to the object. The other triangles move less quickly, or do not move at all, leaving them free to specialize for other classes of objects. In addition to showing a way in which efficient category representations can be created, this algorithm has been put forth as a model of how a person creates categories even when there is no teacher, parent, or label that tells the person what, or how many, categories there are.

The above argument suggests that concepts can be used to conserve memory. An equally important economicizing advantage of concepts is to reduce the need for learning (Bruner, Goodnow, & Austin, 1956). An unfamiliar object that has not been placed in a category attracts attention because the observer must figure out how to think of it. Conversely, if an object can be identified as belonging to a preestablished category, then less cognitive processing is typically necessary. One can simply treat the object as another instance of something that is known, updating one’s knowledge slightly, if at all. The difference between events that require altering one’s concepts and those that do not was described by Piaget (1952) in terms of accommodation (adjusting concepts on the basis of a new event) and assimilation (applying already known concepts to an event). This distinction has also been incorporated into computational models of concept learning that determine whether an input can be assimilated into a previously learned concept. If it cannot, then reconceptualization is triggered (Grossberg, 1982). When a category instance is consistent with a simple category description, then an individual is less likely to store a detailed description of it than if it is an exceptional item (Palmeri & Nosofsky, 1995), consistent with the notion that people simply use an existing category description when it suffices.

**HOW ARE CONCEPTS REPRESENTED?**

Much of the research on concepts and categorization revolves around the issue of how concepts are mentally represented. As with all discussion of representations, the standard caveat must be issued—mental representations cannot be determined or used without processes that operate on these representations (Anderson, 1978). Rather than discussing the representation of a concept such as cat, we should discuss a representation-process pair that allows for the use of this concept. Empirical results interpreted as favoring a particular representation format should almost always be interpreted as supporting a particular representation given particular processes that use the representation. As a simple example, when trying to decide whether a shadowy figure briefly glimpsed was a cat or fox, one needs to know more than how one’s cat and fox concepts are represented. One needs to

---

**Figure 22.1** Alternative proposals have suggested that categories are represented by the individual exemplars in the categories (the circles), the prototypes of the categories (the triangles), or the category boundaries (the lines dividing the categories).
There is considerable intuitive appeal to the notion that concepts are represented by something like dictionary entries. By a rule-based account of concept representation, to possess this concept cat is to know the dictionary entry for it. A person’s cat concept may differ from Webster’s Dictionary entry: “a carnivorous mammal (Felis catus) long-domesticated and kept as a pet and for catching rats and mice.” Still, this account claims a concept is represented by some rule that allows one to determine whether an entity belongs within the category (see also the chapter by Leighton & Sternberg in this volume).

The most influential rule-based approach to concepts may be Bruner, Goodnow, and Austin’s (1956) hypothesis-testing approach. Their theorizing was, in part, a reaction against behaviorist approaches (Hull, 1920) in which concept learning involves the relatively passive acquisition of an association between a stimulus (an object) to be categorized and a response (such as a verbal response, key press, or labeling). Instead, Bruner et al. argued that concept learning typically involves active hypothesis formation and testing. In a typical experiment, their subjects were shown flash cards that had different shapes, colors, quantities, and borders. The subjects’ task was to discover the rule for categorizing the flash cards by selecting cards to be tested and by receiving feedback from the experimenter indicating whether the selected card fit the categorizing rule. The researchers documented different strategies for selecting cards, and a considerable body of subsequent work (e.g., Bourne, 1970) showed large differences in how easily acquired are different categorization rules. For example, a conjunctive rule such as white and square is more easily learned than a conditional rule such as if white, then square, which is in turn more easily learned than a biconditional rule such as white if and only if square. A parallel development to these laboratory studies of rules and their categorial consequences was Katz and Todorov’s (1963) semantic memory theory of compositional semantics within linguistics. In this theory, a word’s semantic representation consists of a list of atomic semantic markers such as + Male, + Adul, + Physical, and + Married for the word bachelor. These markers serve as the components of a rule that specifies what a word is appropriately used. Each of the semantic markers for a word is assumed to be necessary for something to belong to the word category, and the markers are assumed to be jointly sufficient to strike the categorization.

The assumptions of these rule-based models have been vigorously challenged for several decades now (see the chapter by Tversky et al. in this volume). Douglas Medin and Edward E. Smith (Medin & Smith, 1984; E. E. Smith & Medin, 1981) dubbed this rule-based approach “the classical view” and characterized it as holding that all instances of a concept share common properties that are necessary and sufficient for defining the concept. At least three criticisms have been levied against this classical view. First, it has proved to be very difficult to specify the defining rules for most concepts. Wittgenstein (1953) raised this point with his famous example of the concept game. He argued that some of the candidate definitions of this concept, such as activity engaged in for fun, activity with certain rules, or competitive activity with winners and losers, is adequate to identify Frisbee, professional baseball, and roulette as games, while simultaneously excluding wars, debates, television viewing, and leisure walking from the game category. Even a seemingly well-defined concept such as bearheller seems to involve more than its simple definition of unmarried male. The counterexample of a 5-year-old child (who does not really seem to be a bachelor) may be solved by adding an adult precondition to the unmarried male condition, but an indefinite number of other preconditions is required to exclude a man in a long-term but unmarried relationship, the Pope, and a 80-year-old widower with four children (Lakoff, 1987). Wittgenstein argued that instead of expecting knowing a concept with knowing a definition, it is better to think of the members of a category as being related by family resemblance, not just any particular feature in common, but will have several features that are characteristic or typical of the set. Secondly, the category membership for some objects is unclear. People may disagree on whether a starfish is a fish, a camel is a vehicle, a hammer is a weapon, or a stroke is a
Performance on rule-governed items is as accurate with abstract as with concrete material; "performance on rule-governed items is as accurate with unfamiliar as with familiar material;" and "performance on a rule-governed item of problem determines as a function of the number of rules that are required for solving the problem." Based on the full set of criteria, they argue that rule-based reasoning does occur, and that it may be a mode of reasoning distinct from association-based or similarity-based reasoning. Similarly, Parker (1973) argued for distinct rule-based and association-based models for determining linguistic categories. Neurophysiological support for this distinction comes from studies showing that rule-based and similarity-based categorization involve anatomically separate brain regions (Ashby, Alfonso-Reese, Turken, & Waldeon, 1998; Ashby & Waldron, 2000; E. Smith, Patlánov, & Jonides, 1998).

In developing a similar distinction between similarity-based and rule-based categorization, Sloan (1990) introduced the notion that the two systems can simultaneously generate different solutions to a reasoning problem. For example, Rips (1990; see also Rips & Collins, 1993) asked subjects to judge a 3 in. (7.62 cm) round object, and then asked whether the object is more similar to a quarter or a pizza, and whether the object is more likely to be a pizza or a quarter. There is a tendency for the object to be judged as more similar to a quarter, but as more likely to be a pizza. The rule that quarters must not be greater than 1 in. plays a larger role in the categorization decision than in the similarity judgment, causing the two judgments to dissociate. By Sloan's analysis, the tension we feel about the categorization of the 3-in. object stems from the two different systems' indicating incomparable categorizations. Sloan argues that the rule-based system can surmise the similarity-based system but cannot completely suppress it. When Rips' experiment is repeated with a richer description of the object to be categorized, categorization again tracks similarity, and people tend to choose the quarter for both the categorization and similarity choices (E. E. Smith & Sloan, 1994).

Prototypes
Just as the active hypothesis-testing approach of the classical view was a reaction against the passive stimulus-response association approach, so the prototype model was developed as a reaction against what was seen as the overly analytic, rule-based classical view. Central to Eleanor Rosch's development of prototype theory is the notion that concepts are organized around family resemblances rather than features that are individually necessary and jointly sufficient for categorization (Morris & Rosch, 1981; Rosch, 1975; Rosch &
Mervis, 1975, see also the chapters in this volume by Capaldi, by Palmer, and by Treiman et al. The prototypes, for a category consists of the most common attribute values as associated with the members of the category, and can be empirically derived by the previously described method of asking subjects to generate a list of attributes for several members of a category. Once prototypes for a set of concepts have been determined, categorizations can be predicted by determining how similar an object is to each of the prototypes. The likelihood of placing an object into a category increases as it becomes more similar to the category’s prototype and less similar to other category prototypes (Rosch & Merivis, 1975).

This prototype model can neutrally deal with the three problems that confronted the classical view. ‘tis no problem if defining rules for a category are difficult or impossible to devise. If concepts are organized around prototypes, then only characteristic (not necessary or sufficient) features are expected. Unclear category boundaries are expected if objects are presented that are approximately equally similar to prototypes from more than one concept. Objects that clearly belong to a category may still vary in their typicality because they may be more similar to the category’s prototype than to any other category’s prototype, but they still may differ in how similar they are to the prototype. Prototype models do not require “fuzzy” boundaries around concepts (Hampton, 1993), but prototype similarities are based on computational across many attributes and are consequently graded, and lead naturally to categories with graded membership.

A considerable body of data has been amassed that suggests that prototypes have cognitively important functions. The similarity of an item to its prototype (in terms of featural overlap) predicts the results from several converging tasks. Somewhat obviously, it is correlated with the average rating the item receives when subjects are asked to rate how good an example the item is of its category (Rosch, 1975). It is correlated with subjects’ speed in verifying statements of the form “An [item] is a [category name]” (E. E. Smith, Shoben, & Rips, 1974). It is correlated with subjects’ frequency and speed of listing the item when asked to supply members of a category (Mervis & Rossch, 1981). It is correlated with the probability of inductively extending a property from the item to other members of the category (Rips, 1975). Taken in total, these results indicate that different members of the same category differ in how typical they are of the category, and that these differences have a strong cognitive impact. Many natural categories seem to be organized not around definitional boundaries, but by graded typicality to the category’s prototype.

The prototype model described previously generates category prototypes by finding the most common attribute values shared among category members. An alternative conception views a prototype as the central tendency of continuously varying attributes. If the four observed members of a lizard category had tail lengths of 3, 3, 3, and 7 in., the former prototype model would store a value of 3 (the modal value) as the prototype’s tail length, whereas the central tendency model would store a value of 4 (the average value). The central tendency approach has proved useful in modeling categories composed of artificial stimuli that vary on continuous dimensions. For example, Pomer and Keeler’s (1969) classic dot-pattern stimulus consisted of one dots positioned randomly or in familiar configurations on a 30 x 30 invisible grid. Each prototype was a particular configuration of dots, but during categorization training, subjects never saw the prototypes themselves. Instead, they saw distortions of the prototypes obtained by shifting each dot randomly by a small amount. Categorization training involved subjects’ seeing dot patterns, guessing their category assignment, and receiving feedback indicating whether their guesses were correct or not. During a transfer stage, Poster and Keeler found that subjects were better able to categorize the never-before-seen category prototypes than they were to categorize new distortions of those prototypes. In addition, subjects’ accuracy in categorizing distortions of category prototypes was strongly correlated with the proximity of those distortions to the never-before-seen prototypes. The authors interpreted these results as suggesting that prototypes are extracted from distortions and used as a basis for determining categorizations (see also Horna, Sorling, & Trepsi, 1981).

Exemplars

Exemplar models deny that prototypes are explicitly extracted from individual cases, stored in memory, and used to categorize new objects. Instead, in exemplar models, a conceptual representation consists of only those actual, individual cases that one has observed. The prototype representation for the category bird consists of the most typical bird, or an assemblage of the most common attribute values across all birds, or the central tendency of all attribute values for observed birds. By contrast, an exemplar model represents the category bird by representing all of the instances (exemplars) that belong to this category (Brooks, 1978; Fiske, 1986; 1994; Huntzmann, 1986; Kruschke, 1982; Lathberts, 1988; 2000; Logan, 1988; Medin & Schaffer, 1978; Nosofsky, 1984, 1986; see also the chapter by Capaldi in this volume).

Although the prime motivation for these models has been to provide good fit to results from human experiments, computer scientists have pursued similar models with the aim to exploit the power of storing individual experiences to stimuli in
a relatively raw, unabstracted form. Exemplar, instance-based (Aha, 1992), view-based (Tirum & Gauthier, 1998), case-based (Shank, 1982), nearest neighbor (Ripley, 1996), outperform the (Clark & Breant, 1990), and vector-quantization (Kohonen, 1995) models all share the fundamental insight that novel patterns can be identified, recognized, or categorized by giving the novel patterns the same response that was learned for similar, previously presented patterns. By existing as representations of repeated processes, not only is it possible to respond to repetitions of these patterns; it is also possible to give responses to novel patterns that are likely to be correct by sampling responses to old patterns, weighted by their similarity to the novel patterns. Consistent with these models, psychological evidence suggests that people show good transfer to new stimuli in perceptual tasks only to the extent that the new stimuli superficially resemble previously learned stimuli (Kolers & Roddiger, 1984; Palmeri, 1997).

The frequent instability of human generalization is transcended superficial similarities might be considered evidence for either human stupidity or laziness, 'to the contrary, if a strong theory about which stimulus features promote valid inductions is lacking, the strategy of least commitment is to preserve the entire stimulus in its full richness of detail (Brooks, 1978). That is, by storing entire instances and basing generalizations on all of the features of these instances, one can be confident that one's generalizations are not systematically biased. It has been shown that in many situations, categorizing new instances by their similarity to old instances maximizes the likelihood of categorizing the new instances correctly (Ashby & Maddox, 1993; McKinley & Nosofsky, 1985; Ripley, 1966). Furthermore, if information later becomes available that specifies which properties are useful for generalizing appropriately, then preserving entire instances will allow these properties to be recovered. Such properties might be lost and unrecoverable if people were less "lazy" in their generalization from instances.

Given these considerations, it is understandable that people often use all of the attributes of an object even when a task demands the use of specific attributes. Doctors' diagnoses of skin disorders are facilitated when they are similar to previously presented cases, even when the similarity is based on attributes that are known to be irrelevant for the diagnosis (Brooks, Norman, & Allen, 1991). Even when subjects know a simple, clear-cut rule for a perceptual classification, performance is better on frequently presented items than rare items (Allen & Brooks, 1991). Consistent with exemplar models, responses to stimuli are frequently based on their overall similarity to previously exposed stimuli.

The exemplar approach assumes that a category is represented by the category exemplars that have been encountered, and that category-determination decisions are based on the similarity of the object to be categorized to all of the exemplars of each relevant category. As such, an item becomes more similar to the exemplars of Category A (or less similar to the exemplars of other categories) then the probability that it will be placed in Category A increases. Category-determination judgments may shift if an item is approximately equally close to two sets of exemplars, because probabilistic decision rules are typically used. Items may be categorized partially to a category as long as they vary in their similarity to the aggregate set of exemplars.

The exemplar approach to categorization raises a number of questions. First, once one has decided that concepts are to be represented in terms of sets of exemplars, the obvious question remains: How are the exemplars to be represented? Some exemplar models use a feature or attribute-value representation for each of the exemplars (Hinton, 1986; Medin & Schaffer, 1978). Another popular approach is to represent examples in a multidimensional psychological space. These points are obtained by measuring the subjective similarity of every object in a set to every other object. Once an N x N matrix of similarities between N objects has been determined by similarity judgments, perceptual confusion, spontaneous sorings, or other methods, a statistical technique called multidimensional scaling (MDS) finds coordinates for the objects in a D-dimensional space that allow the N x N matrix of similarities to be reconstructed with as little error as possible (Nosofsky, 1992). Given that D is typically smaller than N, a reduced representation is created in which each object is represented by terms of its values on D dimensions. Distances between objects in these quantitatively derived spaces can be used in the input to exemplar models to determine in-class vs. exemplar similarities. These MDS representations are useful for generating quantitative exemplar models that can be fit to human categorizations and similarity judgments, but these still beg the question of how a stand-alone computer program or a person would generate these MDS representations. Presumably, there is some human process that compares object representations and can derive object-to-object similarities from them, but this process is not currently modeled by exemplar models (for steps in this direction, see Edelman, 1999).

A second question for exemplar models is, if exemplar models do not explicitly extract prototypes, how can they account for results that concepts are organized around prototypes? A useful place to begin is by considering Posner and Keele's (1968) result that the new-beer-seen prototype is categorized better than new distortions based on the prototype. Exemplar models have been able to model this result because a categorization of an object is based on its summed
similarity to all previously stored exemplars (Medin & Schaffer, 1978; Nosofsky, 1986). The prototype of a category will, on average, be more similar to the training distortions of many exemplars in an exemplar model can create an emergent, epiphenomenal advantage for the prototype.

Given the exemplar model’s account of prototype categorization, one might ask whether predictions from exemplar and prototype models differ. In fact, they typically do, in large part because categorizations in exemplar models are not simply based on quantified similarity to category exemplars, but to similarities weighted by the proximity of an exemplar to the item to be categorized. In particular, exemplar models have mechanisms to bias categorization decisions so that they are more influenced by exemplars that are similar to items to be categorized. In Medin and Schaffer’s (1978) context model, this is achieved through computing the similarity between objects by multiplying rather than adding their similarities on each of their features. In Hesseman’s (1986) mineva model, this is achieved by raising object-to-object similarities to a power of 3 before summing them together. In Nosofsky’s Generalized Context Model (1986), this is achieved by basing object-to-object similarities on an exponential function of the objects’ distance in an MDS space. With these quantitative biases for close exemplars, the exemplar model may do a better job of predicting categorization accuracy for Posner and Keele’s experiment than the prototype model because it can also predict that familiar distortions will be categorized more accurately than novel distortions that are equally far removed from the prototype (Ship & Nosofsky, 1992).

A third question for exemplar models is, In what way are concept representations economical if every experienced exemplar is stored? Is it certainly implausible with large real-world categories to suppose that every instance ever experienced is stored in a separate place. However, more realistic exemplar models may either store only part of the information associated with an exemplar (Lassalme & Logan, 1993), or only some of the exemplars (Aha, 1992; Palmen & Nosofsky, 1995). One particularly interesting way of conserving space that has received empirical support (Barsalou, Huttonlocher, & Lambeert, 1998) is to combine separate events that all constitute a single individuum into a single representation. Rather than passively registering every event as distinct, people seem naturally to consolidate events that refer to the same individuum. If an observer fails to register the difference between a new exemplar and a previously encountered exemplar (e.g., two similar-looking chihuhuaas), then he or she may combine the two, resulting in an exemplar representation that is a blend of two instances.

Category Boundaries

Another notion is that a concept representation describes the category boundary around a category. The prototype model would represent the four categories of Figure 2.1 in terms of the tringles. The exemplar model represents the categories by the circles. The category boundary model would represent the categories by the 4 × 4 dividing lines between the categories. This view has been most closely associated with the work of Ashby and his colleagues (Ashby, 1992; Ashby et al., 1998; Ashby & Gott, 1988, Ashby & Maddox, 1993; Ashby & Townsend, 1986; Maddox & Ashby, 1993). It is particularly interesting to contrast the prototype and category boundary approaches, because their representational assumptions are almost perfectly complementary. The prototype model represents a category in terms of its most typical member—the object in the center of the distribution of items included in the category. The category boundary model represents categories by their periphery, not their center.

An interesting phenomenon to consider with respect to whether centers or peripheries of concepts are representationally privileged is categorical perception. According to this phenomenon, people are better able to distinguish between physically different stimuli when the stimuli come from different categories than when they come from the same category (see Harnad, 1987, for several reviews of research; see also the chapters in this volume by Fowler and by Treiman et al.). The effect has been best documented for speech phoneme categories. For example, Liberman, Harris, Hoffman, and Griffith (1957) generated a continuum of equally spaced consonant-vowel syllables going from /ba/ to /dæ/. Observers listened to three sounds—A followed by B followed by X—and indicated whether X was identical to A or B. Subjects performed the task more accurately when syllables A and B belonged to different phonemic categories than when they were variants of the same phoneme, even when physical differences were equated.

Categorical perception effects have been observed for visual categories (Calder, Young, Perrett, Eccoff, & Rowland, 1996) and for arbitrarily created laboratory categories (Goldstone, 1994b). Categorical perception could emerge from either prototype or boundary representations. An item to be categorized might be compared to the prototypes of two candidate categories. Increased sensitivity at the category boundary would exist because people represent items in terms of the prototypes to which they are closest. Items that fall on different sides of a boundary would have very different
representations because they would be closest to different prototypes (Liberman et al., 1957). Alternatively, the boundary itself might be represented as a reference point, and as pairs of items move closer to the boundary, it becomes easier to discriminate between them because of their proximity to this reference point (Paivio, 1987).

Computational models have been developed that operate on both principles. Following the prototype approach, Holyoak and Thagard (1995) describe a neural network in which the representation of an item is "pulled toward the prototype of the category to which it belongs. Following the boundaries approach, Goldstone, Steyvers, Spranger-Smith, and Keren (2000) describe a neural network that learns to strongly represent critical boundaries between categories by shifting perceptual vectors to these regions. Ergonomically, the results are mixed. Consistent with prototypes being represented, some researchers have found particularly good discriminating ability close to a familiar prototype (Ackert, Panton, & Hall, 1995; McKadden & Callaway, 1999). Consistent with boundaries being represented, other researchers have found that the sensitivity peaks associated with categorical perception heavily depend on the saliency of perceptual cues at the boundary (Kuhl & Miller, 1975). Rather than being abstractive fixed, a category boundary may likely occur at a location where a distinctive perceptual cue, such as the difference between an aspirated and unaspirated speech sound, is present. A possible reconciliation is that information about either the center or periphery of a category can be represented, and that boundary information is more likely to be represented when two highly similar categories must be frequently discriminated and there is a salient reference point for the boundary.

Different versions of the category boundary approach, illustrated in Figure 22.2, have been based on different ways of partitioning categories (Ashby & Maddox, 1996). With independent decision boundaries, category boundaries must be perpendicular to a dimensional axis, forming rules such as Category A items are larger than 3 cm, irrespective of their color. This kind of boundary is appropriate when the dimension makes up a stimulus are difficult to integrate (Ashby & Gott, 1988). With minimal distance boundaries, a category A prototype is closer to the Category A prototype than the Category B prototype. The decision boundary is formed by finding the line that connects the two categories' prototypes, and creating a boundary that blue's and is orthogonal to the line. The optimal boundary is the boundary that maximizes the likelihood of correctly categorizing an object. If the two categories have the same patterns of variability on their dimensions, and people use information about variance to form their boundaries, then the optimal boundary will be a straight line. If the categories differ in variability, then the optimal boundary will be described by a quadratic equation (Ashby & Maddox, 1995, 1996). A general quadratic boundary is any boundary that can be described by a quadratic equation.

One difficulty with representing a concept by a boundary is that the location of the boundary between two categories depends on several contextual factors. For example, Rey and Liberman (1987) argue that categories of speech sounds are influenced by order effects, adaptation, and the surrounding speech context. The same sound that is halfway between [p] and [b] will be categorized as [p] if preceded by several [b] repetitions of a prototypical [b] sound, but categorized as [b] if preceded by several [p] sounds. For a category boundary representation to accommodate this, two category boundaries would need to be hypothesized—relatively permanent category boundary between [p] and [b], and a second boundary that shifts depending upon the immediate context. The relatively permanent boundary is needed because the contextualized boundary must be based on some earlier information. In many cases, it is more parsimonious to hypothetize representations for the category members themselves, and to view category boundaries as side effects of the competition between neighboring categories. Context effects are then explained simply by changes to the strengths associated with different categories. By this account, there may be no fixed boundary around one's cat concept that causally affects categorizations. When asked about a particular object we can decide whether it is a cat, but this is done by comparing
the evidence in favor of the object's being a cat to its being something else.

Theories
The representation approaches considered thus far all work irrespectively of the actual meaning of the concepts. This is both an advantage and a liability. It is an advantage because it allows the approaches to be universally applicable to any kind of material. They share with inductive statistical techniques the property that they can operate on any data set once the data set is formally described in terms of numbers, features, or coordinates. However, the generality of these approaches is also a liability if the meaning or semantic content of a concept influences how it is represented. While few would argue that statistical tests are appropriate only for certain domains of inquiry (e.g., testing political differences, not disease differences), many researchers have argued that the use of purely data-driven, inductive methods for concept learning are strongly limited and mediated by the background knowledge one has about a concept (Carey, 1985; Gelman & Markman, 1986; Keil, 1989; Medin, 1989; Murphy & Medin, 1985).

People's categorizations seem to depend on the theories they have about the world (for reviews, see Komau, 1992; Medin, 1989). Theories involve organized systems of knowledge. In making an argument for the use of theories in categorization, Murphy and Medin (1985) provide the example of a man jumping into a swimming pool fully clothed. This man may be categorized as drunk because we have a theory of behavior and inebriation that explains the man's action. Murphy and Medin argue that the categorization of the man's behavior does not depend on matching the man's features to those of the category drunk. It is highly unlikely that the category drunk would have such a specific feature as jumps into pools fully clothed. It is not the similarity between the instance and the category that determines the instance's classification; it is the fact that our category provides a theory that explains the behavior.

Other researchers have empirically supported the dissociation between theory-derived categorization and similarity. In one experiment, Carey (1953) observes that children choose a toy monkey over a worm as being more similar to a human, but that when they are told that humans have spines, they are more likely to infer that the worm has a spine than that the toy monkey does. Thus, the categorization of objects into spines and no-spines groups does not appear to depend on the same knowledge that guides similarity judgments. Carey argues that even young children have a theory of living things. Part of this theory is the notion that living things have self-propelled motion and rich internal, organizations. Children as young as 3 years of age make inferences about an animal's properties on the basis of its category label even when the label opposes superficial visual similarity (Carruthers & Markman, 1986, see also the chapter by Tichy in this volume).

Using different empirical techniques, Keil (1989) has come to a similar conclusion. In one experiment, children are told a story in which scientists discover that an animal that looks exactly like a raccoon actually contains the insides of a skunk and has skunk parents and skunk children. With increasing age, children increasingly claim that the animal is a skunk. That is, there is a developmental trend for children to categorize on the basis of theories of heredity and biology rather than on visual appearance. In a similar experiment, Rips (1989) shows an explicit dissociation between categorization judgments and similarity judgments in adults. An animal that is transformed (by toxic wasps) from a bird into something that looks like an insect is judged by subjects to be more similar to an insect, but is also judged to be a bird still. Again, the category judgment seems to depend on biological, genetic, and historical knowledge, whereas the similarity judgments seem to depend more on gross visual appearance.

Researchers have explored the importance of background knowledge in shaping our concepts by manipulating this knowledge experimentally. Concepts are more easily learned when a learner has appropriate background knowledge, indicating that more than "brute" statistical regularities underlie our concepts (Pazzani, 1991). Similarly, when the features of a category can be connected through prior knowledge, category learning is facilitated (Murphy & Allopenna, 1994; Spalding & Murphy, 1999). Even a single instance of a category can allow one to form a coherent category if the knowledge constrains the interpretation of this instance (Ahn, Brewer, & Mooney, 1992). Concepts are disproportionately represented in terms of concept features that are tightly connected to other features (Sloman, Love, Ahn, & Alplopenna, 1998).

Forming categories on the basis of data-driven, statistical evidence and forming them based on knowledge-rich theories of the world seem like strategies fundamentally at odds with each other. Indeed, this is probably the most basic difference between theories of concepts. However, these approaches need not be mutually exclusive. Even the most outspoken proponents of theory-based concepts do not claim that similarity-based or statistical approaches are not also needed (Murphy & Medin, 1985). Moreover, some researchers have suggested integrating the two approaches. Hatt (1994, 1997) describes a similarity-based, example
model of categorization that incorporates background knowledge by assigning category members as they are observed (as with all exemplar models), but also storing never-seen instances that are consistent with the background knowledge. Chiol, McDaniel, and Bussemeyer (1993) described a neural network model of concept learning that does not begin with random or neutral connections between features and concepts (as is typical), but begins with theory-consistent connections that are relatively strong. Both approaches allow domain-general category learners to also have biases toward learning categories consistent with background knowledge.

Summary to Representation Approaches

One cynical conclusion to reach from the preceding alternative approaches is that a researcher begins with a theory, then tends to find evidence consistent with the theory is real that is meta-analytically consistent with a theory-based approach!). Although this state of affairs is typical throughout the field of psychology, it is particularly rife in concept-learning research because researchers have a significant amount of feasibility in choosing what concepts they will experimentally test. Evidence for rule-based categories tends to be found with categories that are created from simple rules (Bruner, Goodman, & Austin, 1956). Evidence for prototypes tends to be found for categories made up of members that are distortions around single prototypes (Posner & Keele, 1968). Evidence for exemplar models is particular strong when categories include exceptional instances that must be individually memorized (Nosofsky & Palmeri, 1998; Nosofsky et al., 1994). Evidence for rules is found when categories are created that subjects already know something about (Murphy & Kaplan, 2000). The researcher's choice of representation seems to determine the experiment that is conducted, rather than the experiment's influencing the choice of representation.

There may be a grain of truth to this cynical conclusion, but our conclusions are instead that people use multiple representational strategies, and can flexibly deploy these strategies based upon the categories to be learned. From this perspective, representational strategies should be evaluated according to their trade-offs and for their fit to the real-world categories and empirical results. For example, exemplar representations are costly in terms of storage demands, but are sensitive to interactions between features and adaptable to new categorization demands. There is a growing consensus that at least two kinds of representational strategy are both present but separated—rule-based and similarity-based processes (Ericsson & Kneschke, 1998, Pinker, 1991; Sloman, 1996). Other researchers have argued for separate processes for storing exemplars and extracting prototypes (Kowalton & Squire, 1993; J. D. Smith & Minda, 2000). Even if one holds our hope for a unified model of concept learning, it is important to recognize these different representational strategies as special cases that must be achievable by the unified model given the appropriate inputs.

CONNECTING CONCEPTS

Although knowledge representation approaches have often treated conceptual systems as independent networks that gain their meaning by their internal connections (Leont & Feigenbaum, 1991), it is important to remember that concepts are connected to both perception and language. Concepts' connections to perception serve to ground them (Hurtado, 1990), and their connections to language allow them to transcend direct experience and to be transmitted easily.

Connecting Concepts to Perception

Concept formation is often studied as though it were a modular process (in the sense of Fodor, 1983). For example, participants in category-learning experiments are often presented with verbal feature lists representing the objects to be categorized. The use of this method suggests an implicit assumption that the perceptual analysis of an object into features is complete before one begins to categorize that object. This may be a useful simplifying assumption, allowing a researcher to test theories of how features are transmitted to the concept. There is nothing evidence, however, that the relationship between the formation of concepts and the identification of features is bidirectional (Goldstone & Barsalou, 1998). In particular, not only does the identification of features influence the categorization of an object, but also the categorization of an object influences the interpretation of features (Bassok, 1996).

In this section of the chapter, we will review the evidence for a bidirectional relationship between concept formation and perception. Evidence for an influence of perception on concept formation comes from the classic study of Heider (1972). She presented a paired-associate learning task involving colors and words to the Duni, a population in New Guinea that has only two color terms. Participants were given a different verbal label for each of 16 color chips. They were then presented with each of the chips and asked for the appropriate label. The correct label was given in feedback when participants made incorrect responses, allowing participants to learn the new color terms over the course of training.
The key manipulation in this experiment was that 8 of the color chips represented English focal colors, whereas 8 represented colors that were not prototypical examples of one of the basic English color categories. Both English speakers and Danish were found to be more accurate at providing the correct label for the focal color chips than for the nonfocal color chips, where focal colors are those that have a consistent and strong label in English. Heider's (1972) explanation for this finding was that the English division of the color spectrum into color categories is not arbitrary, but rather reflects the sensitivities of the human perceptual system. Because the Danish share these same perceptual sensitivities with English speakers, they were better at distinguishing focal colors than nonfocal colors, allowing them to learn color categories for focal colors more easily.

Further research provides evidence for a role of perceptual information not only in the formation but also in the use of concepts. This evidence comes from research relating to Barsalou's (1999) theory of perceptual symbol-systems. According to this theory, sensory orator areas of the brain that are activated during the initial perception of an event are reactivated at a later time by association areas, serving as a representation of one's prior perceptual experience. Rather than preserving a verbatim record of what was experienced, however, association areas only reactivate certain aspects of one's perceptual experience, namely those that received attention. Because these reactivated aspects of experience may be common to a number of different events, they may be thought of as symbols, representing an entire class of events. Because they are formed around perceptual experience, however, they are perceptual symbols, unlike the amodal symbols typically employed in symbols-theories of cognition.

Barsalou's (1999) theory suggests a powerful influence of perception on the formation and use of concepts. Evidence consistent with this proposal comes from property verification tasks. Solomon and Barsalou (1999) presented participants with a number of concept words, each followed by a property word, and asked participants whether each property was a part of the corresponding concept. Half of the participants were instructed to use visual imagery to perform the task, whereas half were given no specific instructions. Despite this difference in instructions, participants in both conditions were found to perform in a qualitatively similar manner. In particular, reaction times of participants in both conditions were predicted most strongly by the perceptual characteristics of properties. For example, participants were quicker to verify small properties of objects than to verify large properties. Findings such as this suggest that detailed perceptual information is represented in concepts and that this information is used when reasoning about those concepts.

There is also evidence for an influence of concepts on perception. Classic evidence for such an influence comes from research on the previously described phenomenon of categorical perception. Listeners are much better at perceiving contrasts that are representative of different phoneme categories (Liberman, Cooper, Shankweiler, & Suddendorf-Kennedy, 1967). For example, listeners can hear the difference in voice onsets time between the words bill and pill, even when this difference is no greater than the difference between two thin sounds that cannot be distinguished. One may simply argue that categorical perception provides further evidence of an influence of concepts on perception. In particular, the phonemes of language may have evolved to reflect the sensitivities of the human perceptual system. Evidence consistent with this viewpoint comes from the fact that chimpanzees are sensitive so many of the same sound contrasts as are humans, even though chimpanzees obviously have no language (Kuhl & Miller, 1975; see also the chapter by Treiman et al. in this volume). There is evidence, however, that the phonemes to which a listener is sensitive can be modified by experience. In particular, although newborn babies appear to be sensitive to all of the sound contrasts present in all of the world's languages, a 1-year-old can hear only those sound contrasts present in his or her linguistic environment (Werker & Tees, 1984). Thus, children growing up in Japan lose the ability to distinguish between the /r/ and /l/ phonemes, whereas children growing up in the United States retain this ability (Miyawaki, 1975). The category of language thus influences one's perceptual sensitivities, providing evidence for an influence of concepts on perception.

Although categorical perception was originally demonstrated in the context of auditory perception, similar phenomena have since been identified in other senses. For example, Goldstone (1994b) trained participants to make a category discrimination in terms of either the size or the brightness of an object. He then presented those participants with a single different task, in which two briefly presented objects were either the same or varied in terms of size or brightness. Participants who had earlier categorized objects on the basis of a particular dimension were found to perform better at telling objects apart in terms of that dimension than were control participants who had been given no prior categorization training. Moreover, this sensitization of categorically relevant dimensions was most evident at those values of the dimension that straddled the boundary between categories.

These findings thus provide evidence that the concepts that one has learned influence one's perceptual sensitivities, in the visual as well as in the auditory modality. Other research has shown that prolonged experience with a domain such as dogs (Tanaka & Taylor, 1991) or faces (Levin &
Beale, 2000; O’Toole, Peterson, & Deffenbacher, 1995) lends
development of a perceptual system that is tuned to these
domains. Goldstone et al. (2000) review other evidence for
conceptual influences on visual perception. Concept learning
appears to be effective both in combing stimuli properties
to create perceptual chunks that are diagnostic for categorization
(Goldstone, 2000), and in splitting apart just those perceptual
dimensions if they are differentially diagnostic for categorization
(Goldstone & Swevers, 2001).

The evidence reviewed here suggests that there is a strong
interaction between concepts and perception, with percep-
tual information influencing the concepts that one forms
and conceptual information influencing how one perceives
the world. Most theories of concept formation fail to account
for this interrelationship. They instead take the perceptual at-
tributes of a stimulus as a given and try to account for how
these attributes are used to categorize that stimulus.

Our area of research that provides an exception to this rule is
research on object recognition. As pointed out by Schvets
(1998), object recognition can be thought of as an example of
object categorization, with the goal of the process being to
identify what kind of object one is observing. Unlike theories of
categorization, theories of object recognition place strong
emphasis on the role of perceptual information in identifying
an object.

Interestingly, some of the theories that have been pro-
posed to account for object recognition have characteristics in
common with theories of categorization. For example, struc-
tural description theories of object recognition (e.g.,
Biederman, 1987; Hummel & Biederman, 1992; Mat &
Nishihara, 1978; see also Palmer’s chapter in this volume) are similar to prototype theories of categorization in
that a newly encountered exemplar is compared to a sum-
mary representation of a category in order to determine
whether the exemplar is a member of that category. In con-
trast, multiple-views theories of object recognition (e.g.,
Edelman, 1998; Tarr & Blathoff, 1995; see also Palmer’s
chapter in this volume) are similar to exemplar-based theo-
ries of categorization in that a newly encountered exemplar is
compared to a number of previously encountered exemplars
stored in memory. The categorization of an exemplar is de-
termined either by the exemplar in memory that most closely
matches it or by a comparison of the similarities of the new
exemplar to each of a number of stored exemplars.

The similarities in the models proposed to account for
categorization and object recognition suggest that there is
considerable opportunity for cross-talk between these two
domains. For example, theories of categorization could po-
tentially be adapted to provide a more complete account for
object recognition. In particular, they may be able to provide
an account of not only the recognition of established object
categories, but also the learning of new ones, a problem not
typically addressed by theories of object recognition. Fur-
thermore, theories of object recognition could be adapted to
provide a better account of the role of perceptual information
in concept formation and use. The rapid recent developments
in object recognition research, including the development of
detailed computational, ontologically based models (e.g., Perrett,
Orsh, & Ashbridge, 1998), suggest that a careful consider-
tion of the role of perceptual information in categorization
can be a profitable research strategy.

Connecting Concepts to Language

Concepts also take part in a bidirectional relationship with
language. In particular, one’s repertoire of concepts may in-
fluence the types of word meanings that one learns, whereas
the language that one speaks may influence the types of con-
cepts that one forms.

The first of these two proposals is the less controversial. It
is widely believed that children come into the process of
vocabulary learning with a large set of unlabeled concepts.
These early concepts may reflect the referential structure in
the environment of the young child, as suggested by Rosch
(1976). For example, a child may form a concept of dog
around the correlated properties of four legs, tail, wagging,
and so forth. The subsequent learning of a word’s mean-
ing should be relatively easy to the extent that one can
map that word onto one of these existing concepts.

Different kinds of words may vary in the extent to which
they map directly onto existing concepts, and thus some
types of words may be learned more easily than others. For
example, Gentner (1981, 1982; Gentner & Markman, 2001)
has proposed that nouns can be mapped straightforwardly onto
existing object concepts, and that verbs are thus learned
relatively early by children. The relation of verbs to prelin-
guistic event categories, on the other hand, may be less
straightforward. The nature of children’s prelinguistic event
categories is not well understood, but the available evidence
suggests that they are structured quite differently from verb
meanings. In particular, research by Kersten and Billman
(1997) demonstrated that when adults learned event
categories in the absence of category labels, they formed
categories around a rich set of correlated properties, in-
cluding the characteristics of the objects in the event, the mo-
tions of those objects, and the outcome of the event. Research
by Cohen and Oakes (1993) has similarly demonstrated that
6-month-old infants learned unlabeled event categories in-
volving correlations among different aspects of an event, in
this case between the agent in an event and the outcome of a
causal interaction involving that agent. These unlabeled event categories learned by children and adults differ markedly from verb meanings. Verb meanings tend to have limited correlational structure, instead picking out only a small number of properties of an event (Huttonlocher & Levi, 1979; Talmy, 1985). For example, the verb collide involves two objects moving into contact with one another, irrespective of the objects involved or the outcome of this collision.

Verbs thus cannot be mapped directly onto existing event categories. Instead, language-learning expertise is necessary to determine which aspects of an event are relevant and which aspects are irrelevant to verb meanings. Perhaps as a result, children learning a variety of different languages have been found to learn verbs later than nouns (Au, Dapretto, & Song, 1994; Gentner, 1982; Gentner & Boroditsky, 2000; but see Gopnik & Choi, 1995, and Tardif, 1996, for possible exceptions). More generally, word meanings should be easy to learn to the extent that they can be mapped onto existing concepts.

There is greater controversy regarding the extent to which language may influence one's concepts. Some influences of language on concepts are fairly straightforward, however. For example, whether a concept is learned in the presence or absence of language (e.g., a category label) may influence the way in which that concept is learned. When categories are learned in the presence of a category label, a common finding is one of competition among correlated cues for predictive strength (Gluck & Bower, 1988; Shank, 1991). In particular, more salient cues may overshadow less salient cues, causing the concept learner to fail to notice the predictiveness of the less salient cue (Gluck & Bower, 1988; Kruschke, 1992; Shank, 1991).

When categories are learned in the absence of a category label, on the other hand, there is facilitation rather than competition among correlated predictors of category membership (Bilzon, 1989; Billman & Knutsen, 1996; Cabrera & Billman, 1996; Kersten & Billman, 1997). The learning of unlabeled categories has been measured in terms of the learning of correlations among attributes of a stimulus. For example, one's knowledge of the correlation between a wagging tail and a slobbering mouth can be used as a measure of one's knowledge of the category dog. Billman and Knutsen (1996) used this method to examine the learning of unlabeled categories of novel animals. They found that participants were more likely to learn the predictive attributes of an attribute when other correlated predictors were also present.

The key difference between these two concept-learning situations may be that the learning of labeled categories, one piece of information, namely the category label, is singled out as being important to predict. Thus, when participants can adequately predict the category label on the basis of a single attribute, they need not look to additional attributes. On the other hand, when multiple pieces of information is singled out, as in the case of unlabeled categories, participants who have learned one predictive relation cannot be sure that they have learned all that they need to learn. As a result, they may continue looking for additional predictive relations. In doing so, they may preferentially attend to those attributes that have already been discovered to be useful, resulting in facilitated learning of further relations involving those attributes (Billman & Heit, 1988).

There is thus evidence that the presence of language influences the way in which a concept is learned. A more controversial suggestion is that the language that one speaks may influence the types of concepts that one is capable of learning. This suggestion, termed the linguistic relativity hypothesis, was first made by Whorf (1956) on the basis of apparent dramatic differences between English and Native American languages in their expressions of ideas such as time, motion, and color. For example, Whorf proposed that the Hopi have no concept of time because the Hopi language provides no mechanism for talking about time. Many of Whorf's linguistic analyses have since been debunked (see Pickel, 1994, for a review), but his theory remains a source of controversy.

Early experimental evidence suggested that concepts were relatively impervious to linguistic influences. In particular, Heider's (1972) finding that the Dani learned new color concepts in a similar fashion to English speakers, despite the fact that the Dani had only two color words, suggested that concepts were determined by perception rather than by language. More recently, however, Roberson, Davies, and Davidoff (2000) attempted to replicate Heider's findings with another group of people with limited color vocabulary, the Berinmo of New Guinea. In contrast to Heider's findings, Roberson et al. found that the Berinmo performed no better at learning a new color concept for a familiar color than for a nonfocal color. Moreover, the Berinmo performed no better at learning a category discrimination between green and blue (a distinction not made in their language) than they did at learning a discrimination between two shades of green. This result contrasted with the results of English-speaking participants, who performed better at the green-blue discrimination. It also contrasted with superior Berinmo performance on a discrimination that was present in their language. These results suggest that the English division of the color spectrum may be more a function of the English language and less a function of human color physiology that was originally believed.

Regardless of one's interpretation of the Heider (1972) and Roberson et al. (2000) results, there are straightforward reasons to expect at least some influence of language on one's
concepts. Homa and Colthe (1954) have demonstrated that people are better at learning concepts when category labels are provided as feedback. Thus, at the very least, one may expect that a concept will be more likely to be learned when it is labeled in a language than when it is not labeled. Although this may seem obvious, further predictions are possible when this finding is combined with the evidence for influences of concepts on perception reviewed earlier. In particular, on the basis of the results of Goldstone (1994b), one may predict that when a language makes reference to a particular dimension, thus causing people to form concepts around that dimension, people's perceptual sensitivities to that dimension will be increased. This, in turn, will make people who learn this language more likely to notice further contrasts along this dimension. Thus, language may influence people's concepts indirectly through one's perceptual abilities.

This proposal is consistent with L. B. Smith's (1999) account of the apparent shape bias in children's word learning. Smith proposed that children learn over the course of early language acquisition that the shapes of objects are important in distinguishing different nouns. As a result, they attend more strongly to shape in subsequent word learning, resulting in an acceleration in subsequent shape-word learning. Although this proposal is consistent with an influence of language on concepts, languages do not seem to differ very much in the extent to which they refer to the shapes of objects (Gentner, 1982; Gentner & Boroditsky, 2001), and thus one would not expect speakers of different languages to differ in the extent to which they are sensitive to shape.

Languages do differ in other respects, however, most notably in their use of verbs (Gentner & Boroditsky, 2001; Kersten, 1998). In English, the most frequently used class of verbs refers to the manner of motion of an object (e.g., running, skipping, sauntering), or the way in which an object moves around (Talmy, 1985). In other languages (e.g., Spanish), however, the most frequently used class of verbs refers to the path of an object (e.g., entering, exiting), or its direction with respect to some external reference point. In these languages, manner of motion is relegated to an adver-bial, if it is mentioned at all. If language influences one's perceptual sensitivities, it is possible that English speakers and Spanish speakers may differ in the extent to which they are sensitive to motion attributes such as the path and manner of motion of an object.

Suggestive evidence in this regard comes from a study by Naigles and Terrazas (1998). They found that English speakers were more likely to generalize a novel verb to an event involving the same manner of motion and a different path than an event involving the same path and a different manner of motion, whereas Spanish speakers showed the opposite tendency. One possible account of this result is that English speakers attended more strongly to manner of motion than did Spanish speakers, causing English speakers to be more likely to map the new verb onto manner of motion. If this were the case, it would have important implications for learning a second language. In particular, one may have difficulty attending to concepts in a second language that are not explicitly marked in one's native language.

Thus, although the evidence for influences of language on one's concepts is mixed, there are reasons to believe that some such influence may take place, if only at the level of attention to different attributes of a stimulus. Proponents of the universality viewpoint (e.g., Pinker, 1994) may argue that this level of influence is a fact of the strongest interpretation of Whorf's hypothesis that language determines the concepts that one is capable of learning. A more fruitful approach, however, may be to stop arguing about whether a given result supports Whorf's theory and start testing more specific theories regarding the relationship between language and concepts.

THE FUTURE OF CONCEPTS AND CATEGORIZATION

The field of concept learning and representation is noteworthy for its large number of directions and perspectives. Although the lack of closure may frustrate some outside observers, it is also a source of strength and resilience. With an eye toward the future, we describe some of the most important avenues for future progress in the field.

First, as the previous section suggests, we believe that much of our progress on research on concepts will be to connect concepts to other concepts (Goldzweig, 1996; Lambauer & Dainton, 1997), to the perceptual world, and to language. One of the key viewpoints as represented by rules, prototypes, sets of exemplars, or category boundaries is that one can easily imagine that such concept is independent of others. For example, one can list the exemplars that are included in the concept bird, or describe its central tendency, without making recourse to any other concepts. However, it is likely that all of our concepts are embedded in a network in which each concept's meaning depends on other concepts as well as on perceptual processes and linguistic labels. The proper level of analysis may not be individual concepts, as many researchers have assumed, but systems of concepts. The connections between concepts and perception on the one hand and between concepts and language on the other hand reveals an important dual nature of concepts. Concepts are used both to recognize objects and to ground word meanings.
A second direction is the development of more sophisticated formal models of concept learning. Progress in neural networks, mathematical models, statistical models, and rational analyses can be gauged by several measures; goodness-of-fit to human data, breath of empirical phenomena accommodated, model constraint and parsimony, and autonomy from human intervention. The current crop of models is fairly impressive in terms of fitting specific data sets, but there is much room for improvement in terms of their ability to accommodate rich sets of concepts and to process real-world stimuli without relying on human judgments or hand-coding.

A final important direction will be to apply psychological research on concepts (see also the chapter by Nickerson & Pew in this volume). Perhaps the most important and relevant application is in the area of educational reform. Psychologists have amassed a large amount of empirical research on various factors that impact the ease of learning and transferring conceptual knowledge. The literature contains excellent suggestions on how to manipulate category labels, presentation order, learning strategies, stimulus format, and category variability in order to optimize the efficiency and likelihood of concept attainment. Putting these suggestions to use in classrooms, computer-based tutorials, and multimedia instructional systems could have a substantial positive impact on pedagogy. This research can also be used to develop autonomous computer diagnosis systems, user models, information visualization systems, and databases that are organized in a manner consistent with human conceptual systems. Given the importance of concepts for intelligent thought, it is not unreasonable to suppose that concept learning research will be equally important for improving thought processes.

REFERENCES
