

CHAPTER 22

Concepts and Categorization

ROBERT L. GOLDSTONE AND ALAN KERSTEN

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

Exemplars	606
Category Boundaries	608
Theories	610
Summary to Representation Approaches	611
CONNECTING CONCEPTS	611
Connecting Concepts to Perception	611
Connecting Concepts to Language	613
THE FUTURE OF CONCEPTS AND CATEGORIZATION	615
REFERENCES	616

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. This act of interpretation, 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 something 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 subject. 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. Nine 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

The authors are grateful to Alice Healy, Robert Proctor, Brian Rogosky, and Irving Weiner for helpful comments on earlier drafts of this chapter. This research was funded by NIH Grant MH56871, NSF Grant 0125287, and a Gill fellowship to the first author.

TABLE 22.1 A Common Category Structure, Originally Used by Medin and Schaffer (1978)

Category	Stimulus	Dimension			
		D1	D2	D3	D4
Category A	A1	1	1	1	0
	A2	1	0	1	0
	A3	1	0	1	1
	A4	1	1	0	1
	A5	0	1	1	1
Category B	B1	1	1	0	0
	B2	0	1	1	0
	B3	0	0	0	1
	B4	0	0	0	0

arbitrary. For example, for a color dimension, red may be assigned a value of 0 and blue a value 1. The exact category structure of Table 22.1 has been used in at least 30 studies (reviewed by J. D. Smith & Minda, 2000), instantiated by stimuli as diverse as geometric forms (Nosofsky, Kruschke, & 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 structures 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 generic 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 generic 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 deals with what is being represented and how 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, then one will tend to view concept learning as the enterprise of inductively creating mental structures that predict these categories. One 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 external supplied examples of the concept to which one has been exposed. If one assumes the primacy of internal 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 sense or meaning 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 equivalence classes, as Schusterman, Reichmuth, and Kastak (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 deemphasized. 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 (Keil, 1989;

Rips, 1989). People may never be able to transcend superficial appearances when categorizing objects (Goldstone, 1994a), nor is it clear that they would want to (Jones & Smith, 1993). Still, one of the most powerful aspects of concepts is their ability to make superficially different things alike (Sloman, 1996). If one has the concept Things to remove from a burning house, even children and jewelry become similar (Barsalou, 1983). The spoken phonemes /d/ /o/ /g/, the French word *chien*, the written word *dog*, and a picture of a dog can all trigger one's concept of **dog** (Snodgrass, 1984), and although they may trigger slightly different representations, much of the core information will be the same. Concepts are particularly useful when we need to make connections between things that have different apparent forms.

WHAT DO CONCEPTS DO FOR US?

Fundamentally, concepts function as filters. We do not have direct access to our external world. We have access to our world only as filtered through our concepts. Concepts are useful when they provide informative or diagnostic ways of structuring this world. An excellent way of understanding the mental world of an individual, group, scientific community, or culture is to find out how they organize their world into concepts (Lakoff, 1987; Medin & Atran, 1999; Wolff, Medin, & Pankratz, 1999).

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, & Parkes, 1980), but a considerable body of evidence suggests that concepts typically are acquired elements that are themselves decomposable into semantic elements (McNamara & Miller, 1989).

Once a concept has been formed, it can enter into compositions with other concepts. Several researchers have studied how novel combinations of concepts are produced and comprehended. For example, how does one interpret the term buffalo paper when one first hears it? Is it paper in the shape of buffalo, paper used to wrap buffaloes 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 connects the two concepts. In Murphy's (1988) concept specialization model, one interprets noun-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 **eats** slot in the **snake** concept. Wisniewski (1997, 1998; Wisniewski & Love, 1998) has argued that properties from one concept are often transferred to another concept, and that this is more likely to occur if the concepts are similar, with parts that can be easily aligned. By this account, a robin snake may be interpreted as a snake with a red belly, once the attribute **red breast** from the robin is transferred to the snake.

In addition to promoting creative thought, the combinatorial power of concepts is required for cognitive systematicity (Fodor & Pylyshyn, 1988). The notion of systematicity is that a system's ability to entertain complex thoughts is intrinsically connected to its ability to entertain the components of those thoughts. In the field of conceptual combination, this has appeared as the issue of whether the meaning of a combination of concepts can be deduced on the basis of the meanings of its constituents. On the one hand, there are some salient violations of this type of systematicity. When adjective 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 **gray hair** is more similar to **white hair** than to **black hair**, but **gray 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 general (Medin & Shoben, 1988). On the other hand, there have been notable successes in predicting how well an object fits a conjunctive description based on how well it fits the individual descriptions that comprise the conjunction (Hampton, 1987, 1997; Storms, De Boeck, Hampton, & Van Mechelen, 1999). A reasonable reconciliation of these results is that when concepts are combined the concepts' meanings systematically determine the meaning of the conjunction, but emergent interactions and real-world plausibility also shape the conjunction's meaning.

Inductive Predictions

Concepts allow us to generalize our experiences with some objects to other objects from the same category. Experience with one slobbering dog may lead one 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 systematicity, it must "go beyond the information given" (Bruner, 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. Categories can be arranged roughly in order of their grounding by similarity: 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 force into parts that converge on a target**). For the latter categories, members need not have very much in common at all. An unrewarding 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 chapters 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 categorization. Rosch (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-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-seeming categories do.

These natural categories also permit many inductive inferences. 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 inductions because their members share similarities across many dimensions or features. Ad hoc categories and highly metaphorical categories permit fewer inductive inferences, but in certain situations the inferences they allow are so important that the categories are created on a "by-need" basis. One interesting possibility is that all 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 (Barsalou, 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 formally developing the notion that the concepts we possess are those that maximize inductive potential (Anderson, 1991; Heit, 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 colleague, dealing with the probabilities of Ed's being strong, having violent tendencies, being a college physics or physical education major, and having a history of steroid use. Markman and Makin (1998) have argued that a major force in shaping our concepts is the need to communicate efficiently. They find that a person's concepts become more consistent and systematic over time in order to establish reference unambiguously for 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 can perceptually discriminate at least 10,000 colors from each other, but we have far fewer color concepts than this. Dramatic savings in storage requirements can be achieved by

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 Vevea (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, & Engbretson, 2000). Together with studies showing that, over time, people tend to preserve the gist of a category rather than the exact members 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 economizing 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

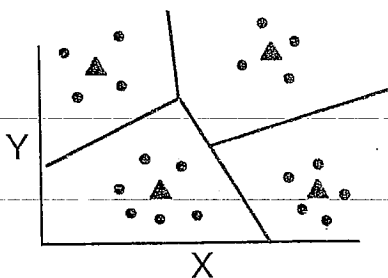


Figure 22.1 Alternative proposals have suggested that categories are represented by the individual exemplars in the categories (the circles), the prototypes of the categories (the triangles), or the category boundaries (the lines dividing the categories).

know how the information in these representations is integrated to make the final categorization. Does one wait for the amount of confirmatory evidence for one of the animals to rise above a certain threshold (Busmeyer & Townsend, 1993)? Does one compare the evidence for the two animals and choose the more likely (Luce, 1959)? Is the information in the candidate animal concepts accessed simultaneously or successively? Probabilistically or deterministically? These are all questions about the processes that use conceptual representations. One reaction to the insufficiency of representations alone to account for concept use has been to dispense with all reference to independent representations, and instead to frame theories in terms of dynamic processes alone (Thelen & Smith, 1994; van Gelder, 1998). However, some researchers feel that this is a case of throwing out the baby with the bath water, and insist that representations must still be posited to account for enduring, organized, and rule-governed thought (Markman & Dietrich, 2000).

Rules

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 the 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 that 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 involved 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 artificial categories was Katz and Fodor's (1963) semantic marker 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, +Adult, +Physical, and -Married for the word bachelor. These markers serve as the components of a rule that specifies when 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 make the categorization.

The assumptions of these rule-based models have been vigorously challenged for several decades now (see the chapter by Treiman 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 conditions for defining the concept. At least three criticisms have been levied against this classical view.

First, it has proven 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 none 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 **bachelor** 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 equating knowing a concept with knowing a definition, it is better to think of the members of a category as being related by family resemblance. A set of objects related by family resemblance need not have any particular feature in common, but will have several features that are characteristic or typical of the set.

Second, 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

disease. By itself, this is not too problematic for a rule-based approach. People may use rules to categorize objects, but different people may have different rules. However, it turns out that people not only disagree with each other about whether a bat is mammal—they also disagree with themselves! McCloskey and Glucksberg (1978) showed that subjects give surprisingly inconsistent category membership judgments when asked the same questions at different times. Either there is variability in how to apply a categorization rule to an object, people spontaneously change their categorization rules, or (as many researchers believe) people simply do not represent objects in terms of clear-cut rules.

Third, even when a person shows consistency in placing objects in a category, he or she might not treat all the objects as equally good members of that category. By a rule-based account, one might argue that all objects that match a category rule would be considered equally good members of the category (but see Bourne, 1982). However, when subjects are asked to rate the typicality of animals such as a robin and an eagle for the category bird, or a chair and a hammock for the category furniture, they reliably give different typicality ratings for different objects. Rosch and Mervis (1975) were able to predict typicality ratings with respectable accuracy by asking subjects to list properties of category members, and measuring how many properties possessed by a category member were shared by other category members. The magnitude of this so-called “family resemblance measure” is positively correlated with typicality ratings.

Despite these strong challenges to the classical view, the rule-based approach is by no means moribund. In fact, in part due to the perceived lack of constraints in neural network models that learn concepts by gradually building up associations, the rule-based approach experienced a rekindling of interest in the 1990s after its low point in the 1970s and 1980s (Marcus, 1998). Nosofsky and Palmeri (1998; Nosofsky et al., 1994; Palmeri & Nosofsky, 1995) have proposed a quantitative model of human concept learning that learns to classify objects by forming simple logical rules and remembering occasional exceptions to those rules. This work is reminiscent of earlier computational models of human learning that created rules such as **if white and square, then Category 1** from experience with specific examples (Anderson, Kline, & Beasley, 1979; Medin, Wattenmaker, & Michalski, 1987). The models have a bias to create simple rules, and are able to predict entire distributions of subjects' categorization responses rather than simply average responses.

In defending a role for rule-based reasoning in human cognition, E. E. Smith, Langston, and Nisbett (1992) proposed eight criteria for determining whether people use abstract rules in reasoning. These criteria include the following:

“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 or problem deteriorates 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, Pinker (1991) argued for distinct rule-based and association-based modes 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, & Waldron, 1998; Ashby & Waldron, 2000; E. E. Smith, Patalano, & Jonides, 1998).

In developing a similar distinction between similarity-based and rule-based categorization, Sloman (1996) introduced the notion that the two systems can simultaneously generate different solutions to a reasoning problem. For example, Rips (1989; see also Rips & Collins, 1993) asked subjects to imagine 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 Sloman's analysis, the tension we feel about the categorization of the 3-in. object stems from the two different systems' indicating incompatible categorizations. Sloman argues that the rule-based system can suppress the similarity-based system but cannot completely suspend it. When Rips's 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 & Sloman, 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 (Mervis & Rosch, 1981; Rosch, 1975; Rosch &

Mervis, 1975; see also the chapters in this volume by Capaldi, by Palmer, and by Treiman et al.). The prototype for a category consists of the most common attribute values 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 & Mervis, 1975).

This prototype model can naturally deal with the three problems that confronted the classical view. It is 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 commonalities 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 category 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 & Rosch, 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 definitive 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 proven useful in modeling categories composed of artificial stimuli that vary on continuous dimensions. For example, Posner and Keele's (1968) classic dot-pattern stimuli consisted of nine dots positioned randomly or in familiar configurations on a 30 × 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, Posner and Keele 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 Homa, Sterling, & Trepel, 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; Estes, 1986, 1994; Hintzman, 1986; Kruschke, 1992; Lamberts, 1998, 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 fits to results from human experiments, computer scientists have pursued similar models with the aim to exploit the power of storing individual exposures to stimuli in

a relatively raw, unabstracted form. Exemplar, instance-based (Aha, 1992), view-based (Tarr & Gauthier, 1998), case-based (Schank, 1982), nearest neighbor (Ripley, 1996), configural cue (Gluck & Bower, 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 creating representations for presented patterns, 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 & Roediger, 1984; Palmeri, 1997).

The frequent inability of human generalization to transcend 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, 1995; Ripley, 1996). 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 generalizations 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 categorization decisions are based on the similarity of the object to be categorized to all of the exemplars of each relevant category. As such, as 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. Categorization judgments may shift if an item is approximately equally close to two sets of exemplars, because probabilistic decision rules are typically used. Items will vary in their typicality 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 featural or attribute-value representation for each of the exemplars (Hintzman, 1986; Medin & Schaffer, 1978). Another popular approach is to represent exemplars as points 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 $N \times N$ matrix of similarities between N objects has been determined by similarity ratings, perceptual confusions, spontaneous sortings, or other methods, a statistical technique called multidimensional scaling (MDS) finds coordinates for the objects in a D -dimensional space that allow the $N \times N$ 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 in terms of its values on D dimensions. Distances between objects in these quantitatively derived spaces can be used as the input to exemplar models to determine item-to-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 computes 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 never-before-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 than are new distortions because the prototype was used to generate all of the training distortions. Without our positing the explicit extraction of the prototype, the cumulative effect 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 summed 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 Hintzman's (1986) Minerva 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 does 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 (Shin & Nosofsky, 1992).

A third question for exemplar models is, In what way are concept representations economical if every experienced exemplar is stored? It is certainly implausible with large real-world categories to suppose that every instance ever experienced is stored in a separate trace. However, more realistic exemplar models may either store only part of the information associated with an exemplar (Lassaline & Logan, 1993), or only some of the exemplars (Aha, 1992; Palmeri & Nosofsky, 1995). One particularly interesting way of conserving space that has received empirical support (Barsalou, Huttenlocher, & Lamberts, 1998) is to combine separate events that all constitute a single individual into a single representation. Rather than passively registering every event as distinct, people seem naturally to consolidate events that refer to the same individual. If an observer fails to register the difference between a new exemplar and a previously encountered exemplar (e.g., two similar-looking chihuahuas), 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 boundary around a category. The prototype model would represent the four categories of Figure 22.1 in terms of the triangles. The exemplar model represents the categories by the circles. The category boundary model would represent the categories by the four 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 /be/ to /de/. 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, Etcoff, & 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 (Pastore, 1987).

Computational models have been developed that operate on both principles. Following the prototype approach, Harnad, Hanson, and Lubin (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, Spencer-Smith, and Kersten (2000) describe a neural network that learns to strongly represent critical boundaries between categories by shifting perceptual detectors to these regions. Empirically, the results are mixed. Consistent with prototypes' being represented, some researchers have found particularly good discriminability close to a familiar prototype (Acker, Pastore, & Hall, 1995; McFadden & 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 arbitrarily fixed, a category boundary is most likely to 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, 1998). 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 dimensions that make up a stimulus are difficult to integrate (Ashby & Gott, 1988). With minimal distance boundaries, a Category A response is given if and only if an object 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 bisects and is orthogonal to this 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

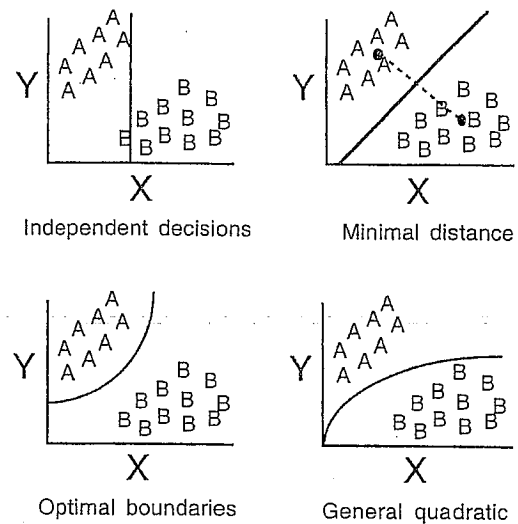


Figure 22.2 The notion that categories are represented by their boundaries can be constrained in several ways. Boundaries can be constrained to be perpendicular to a dimensional axis, to be equally close to prototypes for neighboring categories, to produce optimal categorization performance, or (loosely constrained) to be a quadratic function.

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, 1993, 1998). 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, Repp 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 [pa] and [ba] will be categorized as /pa/ if preceded by several repetitions of a prototypical [ba] sound, but categorized as /ba/ if preceded by several [pa] sounds. For a category boundary representation to accommodate this, two category boundaries would need to be hypothesized—a relatively permanent category boundary between /ba/ and /pa/, 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 hypothesize 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 reified 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

- Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 31–60). London: Routledge.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608–631.
- Over, D. E., & Evans, J. St. B. T. (1999). The meaning of mental logic. *Cahiers de Psychologie*, 18, 99–104.
- Park, N. S., & Robertson, D. (1997). A localist network architecture for logical inference. In R. Sun & F. Alexandre (Eds.), *Connectionist-symbolic integration: From unified to hybrid approaches* (pp. 245–263). Mahwah, NJ: Erlbaum.
- Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, 102, 533–566.
- Pollard, P. (1990). Natural selection for the selection task: Limits to social-exchange theory. *Cognition*, 36, 195–204.
- Pollard, P., & Evans, J. St. B. T. (1987). Content and context effects in reasoning. *American Journal of Psychology*, 100, 41–60.
- Priest, A. G., & Lindsay, R. O. (1992). New light on novice-expert differences in physics problem solving. *British Journal of Psychology*, 83, 389–405.
- Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- Rips, L. J. (1995). Deduction and cognition. In E. E. Smith & D. N. Osherson (Eds.), *Thinking: Vol. 3. An invitation to cognitive science* (2nd ed., pp. 297–343). Cambridge, MA: MIT Press.
- Roberts, M. J. (1993). Human reasoning: Deduction rules or mental models, or both? *Quarterly Journal of Experimental Psychology*, 46A, 569–589.
- Roberts, M. J. (2000). Individual differences in reasoning strategies: A problem to solve or an opportunity to seize? In W. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 23–48). Mahwah, NJ: Erlbaum.
- Rumain, B., Connell, J., Braine, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: *If* is not the biconditional. *Developmental Psychology*, 19, 471–481.
- Schaeken, W., De Vooght, G., Vandierendonck, A., & d'Ydewalle, G. (Eds.). (2000). *Deductive reasoning and strategies*. Mahwah, NJ: Erlbaum.
- Shepard, R. N. (1990). *Mind sights*. New York: W. H. Freeman.
- Simon, H. A. (1991). Cognitive architectures and rational analysis: Comment. In K. VanLehn (Ed.), *Architectures for intelligence* (pp. 25–39). Hillsdale, NJ: Erlbaum.
- Simon, H. A. (1999a). Problem solving. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 674–676). Cambridge, MA: MIT Press.
- Simon, H. A. (1999b). Production systems. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 676–677). Cambridge, MA: MIT Press.
- Smith, S. M., & Blankenship, S. E. (1989). Incubation effects. *Bulletin of the Psychonomic Society*, 27, 311–314.
- Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, 57, 31–95.
- Stenning, K., & Oaksford, M. (1993). Rational reasoning and human implementations of logic. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 136–176). Florence, KY: Taylor & Francis/Routledge.
- Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, 19, 97–140.
- Stenning, K., & Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology*, 34, 109–159.
- Sternberg, R. J. (1980a). Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General*, 109, 119–159.
- Sternberg, R. J. (1980b). The development of linear syllogistic reasoning. *Journal of Experimental Child Psychology*, 29, 340–356.
- Sternberg, R. J. (1981). Reasoning with determinate and indeterminate linear syllogisms. *British Journal of Psychology*, 72, 407–420.
- Sternberg, R. J. (1983). Components of human intelligence. *Cognition*, 15, 1–48.
- Sternberg, R. J. (1999). *Cognitive psychology* (2nd ed.). Ft. Worth, TX: Harcourt Brace.
- Sternberg, R. J., & Ben Zeev, T. (2001). *Complex cognition: The psychology of human thought*. New York: Oxford University Press.
- Sternberg, R. J., & Davidson, J. E. (Eds.). (1995). *The nature of insight*. Cambridge, MA: MIT Press.
- Torrens, D., Thompson, V. A., & Cramer, K. M. (1999). Individual differences and the belief bias effect: Mental models, logical necessity, and abstract reasoning. *Thinking and Reasoning*, 5, 1–28.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Tversky, A., & Kahneman, D. (1986). Judgment under uncertainty: Heuristics and biases. In H. R. Arkes & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 38–55). Cambridge, UK: Cambridge University Press.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth, Middlesex, England: Penguin.
- Wason, P. C. (1983). Realism and rationality in the selection task. In J. St. B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches* (pp. 44–75). London: Routledge & Kegan Paul.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.
- Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition*, 26, 716–730.