CHAPTER 8

Categorization and Concepts

ROBERT L. GOLDSTONE, ALAN KERSTEN, AND PAULO F. CARVALHO

INTRODUCTION

Concepts are the building blocks of thought. They are critically involved when we reason, make inferences, and try to generalize our previous experiences to new situations. Behind every word in every language lies a concept, although there are concepts, like the small plastic tubes attached to the ends of shoelaces, that we are familiar with and can think about even if we do not know that they are called *aglets*. Concepts are indispensable to human cognition because they take the "blooming, buzzing confusion" (James, 1890, p. 488) of disorganized sensory experiences and establish order through mental categories. These mental categories allow us to make sense of the world and predict how worldly entities will behave. We see, hear, interpret, remember, understand, and talk about our world through our concepts, and so it is worthy of reflection time to establish where concepts come from, how they work, and how they can best be learned and deployed to suit our cognitive needs.

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 (Kurtz, 2015; Murphy, 2002). Identifying the person sitting across from you at the breakfast table involves categorizing something as 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.

AQ1

We are grateful to Brian Rogosky, Robert Nosofsky, John Kruschke, Linda Smith, and David Landy for helpful comments on earlier drafts of this chapter. This research was funded by National Science Foundation REESE grant DRL-0910218, and Department of Education IES grant R305A1100060.

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 hopefully 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 8.1 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 and objects from Category B usually have values of 0. The dimensions are typically unrelated to each other, and assigning values of 0 and 1 to a dimension is 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 8.1 has been used in at least 30 studies (reviewed by J. D. Smith & Minda, 2000) and instantiated by stimuli as diverse as geometric forms, yearbook photographs, cartoons of faces (Medin & Schaffer, 1978), and line drawings of rocket ships. These researchers are not particularly interested in the category structure of Table 8.1 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 of category learning, and (d) potentially generalizable to natural categories that people learn. Work on generic concepts is 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.

Table 8.1 A Common Category Structure

		Dimension			
Category	**Stimulus**	**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

WHAT ARE CONCEPTS?

Concepts, Categories, and Internal Representations

A good starting place is Edward 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. On the one hand, 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. One extreme version of this view is the exemplar model of concept learning (Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1984), in which one's internal representation of 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, on the other hand, one assumes the primacy of internal mental concepts, then one tends to view external categories as the end product of using these internal concepts to organize observed entities. Some practitioners of a "concepts first" approach 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). Recent research indicates that concepts' extensions (the class of items to which the concept applies) and intensions (the features that distinguish that class of items) do not always cohere with each other (Hampton & Passanisi, 2016). For example, dolphins and whales are often judged to have many of the features characteristic of an internal representation of fish (e.g., swims, lives in oceans, and has fins), but are still placed in the extensional set of mammals rather than fish. The implication is that a complete model of concepts may require at least partially separate representations for intensions and extensions, rather than a more parsimonious model in which a concept's intension determines whether particular objects belong, and how well, to the concept's extension.

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; Connell & Lynott, 2014; Thelen & Smith, 1994). Still, the extent to which perceptually dissimilar things can be treated equivalently given the appropriate conceptualization is impressive. To the biologist armed with a strong "mammal" concept, even whales and dogs may be treated as commensurate in many situations related to biochemistry, child rearing, and thermoregulation.

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 people 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). When people classify objects into familiar, labeled categories such as *chair*, then their memory for the individuating information about the objects is markedly worse (Lupyan, 2008a). People may never be able to transcend superficial appearances when categorizing objects (Goldstone, 1994b), nor is it clear that they would want to (S. S. 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 only have access to our world 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; Malt & Wolff, 2010; Medin & Atran, 1999; Ojalehto & Medin, 2015).

Components of Thought

Concepts are cognitive elements that combine together to generatively produce an infinite variety of thoughts. Just as an endless variety of architectural structures can be constructed out of a finite set of building blocks, so 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 out of other concepts. Some theorists have argued that concepts such as *bachelor*, *kill*, and *house* are primitive in this sense (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 *buffalo paper* when one first hears it? Is it paper in the shape of a buffalo, paper used to wrap buffaloes presented as gifts, an essay on the subject of buffaloes, coarse paper, or is it like flypaper 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 the fill the *eats* slot in the *snake* concept.

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. However, 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 *black hair*, but *gray cloud* is more similar to *black cloud* than *white cloud* (Medin & Shoben, 1988). *Wooden spoons* are judged to be fairly large (for spoons), even though this property is not generally possessed by wood objects or spoons (Medin & Shoben, 1988). Still, there have been 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, 1997). A reasonable reconciliation of these results is that when concepts combine together, 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 inductively predicted. 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, oak tree), man-made artifacts (hammer, airplane, chair), ad hoc categories (things to take out of a burning house, 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, 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-base 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) 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 do not conform to these clumps (e.g., ad hoc categories), but many of our most natural-seeming categories do. Neural network models have been proposed that take advantage of these clumps to learn hierarchies of categories (Rogers & Patterson, 2007).

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 somebody'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/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 an "as needed" 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 only created 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 concept *peaches, llamas, telephone answering machines, or Ringo Starr* is an unlikely concept because belonging in this concept predicts very little. Researchers have empirically found that the categories that we create when we strive to maximize inferences are different from those that we create when we strive to sort the objects of our world into clearly separate groups (Yamauchi & Markman, 1998). Several researchers have been formally developing the notion that the concepts we possess are those that maximize inductive potential (Anderson, 1991; Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Tenenbaum, 1999). An implication of this approach is that there are degrees of concepthood, with concepts falling on a continuum of inductive power (Wixted, personal communication, November 2016). Most psychologists studying concept learning do not believe that most of our everyday concepts are defined by rules or discrete boundaries (see Rules section below), but we may well be guilty of treating the concept *concept* as more rule based than it actually is. If concepts vary crucially in terms of their inductive power, they are very likely to be fuzzy and graded rather than discrete objects that people either do or do not possess.

Communication

Communication between people is enormously facilitated if the people can count upon a set of common concepts being shared. 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 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 efficiently communicate. They find that people's concepts become more consistent and systematic over time in order to unambiguously establish reference for another individual with whom they need to communicate (see also Garrod & Doherty, 1994).

Cognitive Economy

We can discriminate far more stimuli than 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" in a detailed, perceptually rich code, but that this code rapidly (within 2 seconds) gives way to a more abstract conceptual code that "A" and "a" share. Huttenlocher, Hedges, and Vevea (2000) developed 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, & Engebretson, 2000). Together with studies showing that, over time, people tend to preserve the gist of a category rather than the exact members that comprise it (e.g., Posner & Keele, 1970), these results suggest that by preserving category-level information rather than individual-level information, efficient long-term representations can be maintained. In fact, it has been argued that our perceptions of an object represent a nearly optimal combination of evidence based on the object's individuating information and the categories to which it belongs (N. H. Feldman, Griffiths, & Morgan, 2009). By using category-level information, one will occasionally overgeneralize and make errors. Rattlesnakes may be dangerous in general, but one may stumble upon a congenial one in the Arizona desert. One makes an error when one is unduly alarmed by its presence, but it is an error that stems from a healthful, life-sustaining generalization.

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 8.1 shows 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 just storing the positions of the four triangles in Figure 8.1.

In addition to conserving memory storage requirements, 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 about it. Conversely, if an object can be identified as belonging to a preestablished category, then typically less cognitive processing is 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

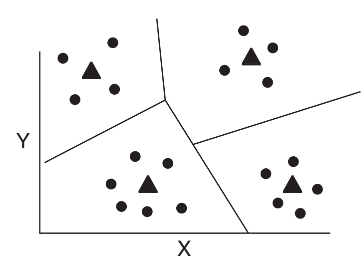

Figure 8.1 Alternative proposals have suggested that categories are represented by the individual exemplars in the categories (the circles), the prototypes of the categories (the triangles), or the category boundaries (the lines dividing the categories).

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, and if it cannot, then reconceptualization is triggered (Grossberg, 1982). When a category instance is consistent with a simple category description, then people are 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. In general, concept learning proceeds far more quickly than would be predicted by a naïve associative learning process. Our concepts accelerate the acquisition of object information at the same time that our knowledge of objects accelerates concept formation (Griffiths & Tenenbaum, 2009; Kemp & Tenenbaum, 2009).

HOW ARE CONCEPTS REPRESENTED?

Much 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. 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 a fox, one needs to know more than how one's *cat* and *fox* concepts are represented. One needs to know how the information in these representations is integrated together 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 (Fific, Little, & Nosofsky, 2010)? 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 frame theories in terms of dynamic processes alone (Thelen & Smith, 1994; van Gelder, 1998). However, others 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's entry: "A carnivorous mammal (*Felis catus*) long domesticated and kept by man as a pet or for catching rats and mice." Still, this account claims that a concept is represented by some rule that allows one to determine whether or not an entity belongs within the category.

The most influential rule-based approach to concepts may be Bruner et al.'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 or not. The researchers documented different strategies for selecting cards, and a considerable body of subsequent work showed large differences in how easily acquired are different categorization rules (e.g., Bourne, 1970). 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*.

The assumptions of these rule-based models have been vigorously challenged for several decades now. Douglas Medin and Edward 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*, *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 fixed by adding in an *adult* precondition, but an indefinite number of other preconditions are 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 not clear. People disagree on whether or not a starfish is a fish, a camel is a vehicle, a hammer is a weapon, and 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 people give surprisingly inconsistent category-membership judgments when asked the same questions at different times. There is either variability in how to apply a categorization rule to an object, or 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, people do not treat the objects as equally good members of the 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 *robin* and *eagle* for the category *bird*, or *chair* and *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 (Nosofsky & Palmeri, 1998; 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. A strong version of a rule-based model predicts that people create categories that have the minimal possible description length (J. Feldman, 2006).

One approach to making the rule-governed approach to concepts more psychologically plausible is to discard the assumption that rule-governed implies deterministic. The past few years have seen a new crop of rule-based models that are intrinsically probabilistic (Piantadosi & Jacobs, 2016; Piantadosi, Tenenbaum, & Goodman, 2016; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). These models work by viewing categorization as the result of integrating many discrete rules, each of which may be imperfect predictors on its own. A remaining challenge for these models is that there often is a psychological connection between rule use and determinism. Rules, particularly ones that involve relations between elements, are often very difficult to learn when they are probabilistic rather than applied without exception (Jung & Hummel, 2015). In addition, even formal concepts such as *triangle* have graded and flexible structures—structures that are tied less to strict definitions when the concepts are activated by their labels (Lupyan, 2017).

In defending a role for rule-based reasoning in human cognition, E. E. Smith, Langston, and Nisbett (1992) proposed eight criteria for determining whether or not people use abstract rules in reasoning. These criteria include "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 categorizations involve anatomically separate brain regions (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; 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 three-inch, round object, and then asked whether the object was more similar to a quarter or a pizza, and whether the object was 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 inch 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 three-inch 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' 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). 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 the 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 above generates category prototypes by finding the most common attribute values shared among category members. An alternative conception views prototypes 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 inches, 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 in categorizing 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.

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 only of the 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 (L. R. Brooks, 1978; Estes, 1994; Hintzman, 1986; Kruschke, 1992; Lamberts, 2000; Logan, 1988; Medin & Schaffer, 1978; Nosofsky, 1984, 1986).

While 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 of exploiting the power of storing individual exposures to stimuli in a relatively raw, unabstracted form. The 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 pattern. Consistent with these models, psychological evidence suggests that people show good transfer to new stimuli in perceptual tasks just to the extent that the new stimuli superficially resemble previously learned stimuli (Palmeri, 1997).

The frequent inability of human generalization to transcend superficial similarities might be considered as evidence of either human stupidity or laziness. To the contrary, if a strong theory about what stimulus features promote valid inductions is lacking, the strategy of least commitment is to preserve the entire stimulus in its full richness of detail (L. R. 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 becomes available at a later point that specifies what 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 why 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 (L. R. Brooks, Norman, & Allen, 1991). Even when people 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 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 among 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 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 by computing the similarity between objects by multiplying rather than adding the similarities in each of their features. In Hintzman's (1986) MINERVA 2 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 (1968) 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 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 register every event as distinct, people seem to naturally consolidate events together 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 together, resulting in an exemplar representation that is a blend of two instances (Love, Medin, & Gureckis, 2004).

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 8.1 in terms of the triangles. The exemplar model would represent 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. One recurring empirical result that provides some prima facie evidence for representing categories in terms of boundaries is that oftentimes the most effectively categorized object is not the prototype of a category, but rather is a *caricature* of the category (Davis & Love, 2010; Goldstone, 1996; Goldstone, Steyvers, & Rogosky, 2003; Heit & Nicholson, 2010). A caricature is an item that is systematically distorted away from the prototype for the category in the direction opposite to the boundary that divides the category from another category.

An interesting phenomenon to consider with respect to whether centers or peripheries of concepts are representationally privileged is *categorical perception*. Due 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). 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 the 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, 1994a). 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 be because people represent items in terms of the prototype to which they are closest. Items that fall on different sides of the 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, category boundaries are 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 8.2, have been based on different ways of partitioning categories (Ashby & Maddox, 1998). With independent decision boundaries, categories boundaries must be perpendicular to a dimensional axis, forming rules such as *Category A items are larger than 3 centimeters, irrespective of their color*. This kind of boundary is appropriate when the dimensions that make up a stimulus are hard 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

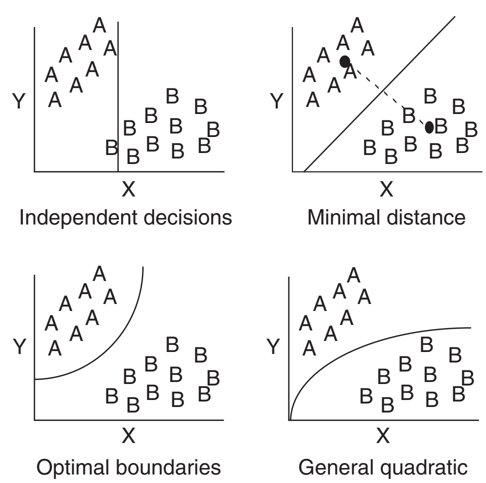

Figure 8.2 The notion that categories are represented by their boundaries can be constrained in several ways. Boundaries can be constrained to be perpendicular to a dimensional axis, to be equally close to prototypes for neighboring categories, to produce optimal categorization performance, or may be loosely constrained to be a quadratic function.

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 optimal boundary will be a straight line. If the categories differ in their 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 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 or not, but this is done by comparing the evidence in favor of the object being a cat to its being something else.

Theories

The representation approaches thus far considered 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 T-tests are only appropriate for certain domains of inquiry (e.g., testing political differences, but not disease differences), many researchers have argued that the use of purely data-driven, inductive methods for concept learning are strongly limited and modulated 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 Komatsu, 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 the category *drunk*'s features. 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 (1985) 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 spleens, are more likely to infer that the worm has a spleen than that the toy monkey does. Thus, the categorization of objects into *spleen* and *no spleen* 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 (Gelman & Markman, 1986).

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 internal organs 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 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 waste) 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, while the similarity judgments seems 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 people to form a coherent category if background 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, 1998).

Forming categories on the basis of data-driven, statistical evidence, and forming them based upon 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 in the field. 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. Theories in the form of prior knowledge about a domain are recruited in order to account for empirically observed categorizations, and one mechanism for this is the process of subjects trying to form explanations for the observations (Williams & Lombrozo, 2013; Williams, Lombrozo, & Rehder, 2013). Heit (1994, 1997) describes a similarity-based, exemplar model of categorization that incorporates background knowledge by storing category members as they are observed (as with all exemplar models), but also storing never-seen instances that are consistent with the background knowledge. Choi, McDaniel, and Busemeyer (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. Rehder and Murphy (2003) propose a bidirectional neural network model in which observations affect, and are affected by, background knowledge. Hierarchical Bayesian models allow theories, incorporated as prior probabilities on specific structural forms, to guide the construction of knowledge, oftentimes forming knowledge far more rapidly than predicted if each observation needed to be separately learned (Kemp & Tenenbaum, 2008, 2009; Lucas & Griffiths, 2010). All of these computational 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 starts with a theory, and tends to find evidence consistent with the theory—a result that is meta-analytically consistent with a theory-based approach! Although this state of affairs is typical throughout psychology, it is particularly rife in concept-learning research because researchers have a significant amount of flexibility in choosing what concepts they will use experimentally. Evidence for rule-based categories tends to be found with categories that are created from simple rules (Bruner et al., 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, Palmeri, & McKinley, 1994). Evidence for theories 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 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 (Weiskopf, 2009) 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 strategies are both present but separated—rule-based and similarity-based processes (Erickson & Kruschke, 1998; Pinker, 1991; Sloman, 1996). There is even recent evidence for reliable individual differences in terms of these strategies, with different groups of people naturally inclined toward either rule-based or exemplar learning processes (McDaniel, Cahill, Robbins, & Wiener, 2014). Other researchers have argued for separate processes for storing exemplars and extracting prototypes (Knowlton & Squire, 1993; J. D. Smith & Minda, 2000, 2002). Some researchers have argued for a computational rapprochement between exemplar and prototype models in which prototypes are formed around statistically supported clusters of exemplars (Love et al., 2004). Even if one holds out 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 (Lenat & Feigenbaum, 1991), it is important to remember that concepts are connected to both perception and language. Concepts' connections to perception serve to ground them (Goldstone & Rogosky, 2002; Harnad, 1990), and their connections to language allow them to transcend direct experience and to be easily transmitted.

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 starts to categorize that object. Categorization processes can then act upon the features that result from this analysis, largely unconcerned with the specific perceptual information that led to the identification of those features. This may be a useful simplifying assumption, allowing a researcher to test theories of how features are combined to form concepts. There is mounting evidence, however, that conceptual processes may act directly on modality-specific perceptual information, eliminating the need to first transduce that information into amodal feature lists before categorization can begin.

Evidence for a role of perceptual information in conceptual processes comes from research relating to Barsalou's (1999, 2008) theory of perceptual symbol systems. According to this theory, sensorimotor 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 symbolic theories of cognition.

In support of this theory, neuroimaging research has revealed that conceptual processing leads to activation in sensorimotor regions of the brain, even when that activated information is not strictly necessary to perform the conceptual task. For example, Simmons, Martin, and Barsalou (2005) revealed in a functional magnetic resonance imaging (fMRI) study that pictures of appetizing foods led to activation in areas associated with the perception of taste, even though the task simply involved visual matching and thus taste was irrelevant. This suggests that food concepts may be represented by activation in a variety of areas representing the different sensory modalities associated with those concepts. Furthermore, Simmons et al. (2007) revealed that brain areas responsive to color information in a visual task were also activated by judgments of color properties associated with concepts, even though the concepts and properties were presented only in verbal form. Representing the various properties associated with a concept apparently recruits the same brain mechanisms that are involved in directly perceiving those properties, consistent with the theory of perceptual symbol systems.

Although findings such as these demonstrate that perceptual representations are activated when a concept is instantiated, they leave open the possibility that the concepts themselves are represented in amodal format, and that the activation of perceptual representations reflects visual imagery processes subsequent to the identification of a concept. This account remains possible because the coarse temporal resolution of fMRI makes unclear the exact sequence of neural events leading to successful task performance. To address this concern, Kiefer, Sim, Herrnberger, Grothe, and Hoenig (2008) examined the activation of perceptual representations using not only fMRI but also event-related potentials (ERPs), a brain-imaging technique with much better temporal resolution. Neuroimaging with fMRI revealed that auditory areas were activated in response to visual presentation of words referring to objects with strongly associated acoustic features (e.g., telephone). Moreover, results with ERP suggested that these auditory areas were activated within 150 ms of stimulus presentation, quite similar to durations previously established for accessing meanings for visually presented words (e.g., Pulvermüller, Shtyrov, & Ilmoniemi, 2005). This combination of results suggests that activation of perceptual representations occurs early in the process of concept access, rather than reflecting later visual imagery processes.

Although neuroimaging studies with fMRI and ERP reveal that activation of perceptual areas is associated with conceptual processing, these are correlational methods, and thus it still remains possible that activation of perceptual areas does not play a direct causal role in concept access and use. For example, on the basis of these results alone, it remains possible that presentation of a word leads one to access an amodal symbolic representation of the associated concept, but that perceptual areas are also activated via a separate pathway. These perceptual representations would thus have no causal role in accessing or representing the concept. To establish causality, it is necessary to manipulate activation of perceptual areas and demonstrate an effect on concept access. An example of such a manipulation was carried out by Vermeulen, Corneille, and Niedenthal (2008). They presented participants with a memory load presented either in the visual or auditory modality. While maintaining these items, participants were given a property verification task in which either a visual or an auditory property of a concept had to be verified. Participants were slower to verify visual properties while simultaneously maintaining a visual memory load, whereas they were slower to verify auditory properties while simultaneously maintaining an auditory memory load. These results suggest that concept access involves activating various modalities of sensorimotor information associated with the concept, and thus that simultaneous processing of other unrelated perceptual information within the same modality can interfere with concept access (see also Witt, Kemmerer, Linkenauger, & Culham, 2010).

The results described so far in this section suggest that concepts are represented not in terms of amodal symbols transduced from perceptual experience, but rather in terms of perceptual information itself. The specific patterns of perceptual information that one has experienced when learning about a concept may thus influence the representation of that concept (e.g., see Kiefer, Sim, Liebich, Hauk, & Tanaka, 2007), possibly even for putatively abstract concepts that one may not expect to be associated with perceptual information, such as *truth* and *freedom* (Barsalou & Wiemer-Hastings, 2005). Moreover, if there is overlap in perceptual and conceptual representations, then not only may perceptual information affect concept access and use, but one's concepts may also influence one's perceptual sensitivities, with concept activation leading to top-down influences on perceptual discrimination abilities (Brooks & Freeman, Chapter 13 in Volume 4 of this *Handbook*; Lupyan, 2008b; but see Firestone & Scholl, 2016 for skepticism regarding top-down effects). The relationship between perceptual and conceptual processes may thus be bidirectional (Goldstone & Barsalou, 1998), with the identification of perceptual features influencing the categorization of an object, and the categorization of an object influencing the perception of features (Bassok, 1996).

Classic evidence for an influence of concepts on perception 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, & Studdert-Kennedy, 1967). For example, listeners can hear the difference in voice onset time between *bill* and *pill*, even when this difference is no greater than the difference between two /b/ sounds that cannot be distinguished. One may argue that categorical perception simply provides further evidence of an influence of perception on concepts. 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 chinchillas are sensitive to many of the same sound contrasts as are humans, even though chinchillas obviously have no language (Kuhl & Miller, 1975). 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 only hear 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 /l/ and /r/ phonemes, whereas children growing up in the United States retain this ability (Miyawaki, 1975). The categories of language thus influence 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 discovered in vision (Goldstone & Hendrickson, 2010). For example, Goldstone (1994a) trained participants to make a category discrimination either in terms of the size or brightness of an object. He then presented those participants with a same/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 be 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 (see also Ozgen & Davies, 2002). Other research has shown that prolonged experience with domains such as dogs (Tanaka & Taylor, 1991), cars and birds (Gauthier, Skudlarski, Gore, & Anderson, 2000), faces (Levin & Beale, 2000; O'Toole, Peterson, & Deffenbacher, 1995) or even novel "Greeble" stimuli (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999) leads to a perceptual system that is tuned to these domains. Goldstone et al. (2000; Goldstone, Landy, & Son, 2010) and Lupyan (2015) review other evidence for conceptual influences on visual perception. Concept learning appears to be effective both in combining stimulus properties together to create perceptual chunks that are diagnostic for categorization (Goldstone, 2000), and in splitting apart and isolating perceptual dimensions if they are differentially diagnostic for categorization (Goldstone & Steyvers, 2001; M. Jones & Goldstone, 2013). In fact, these two processes can be unified by the notion of creating perceptual units in a size that is useful for relevant categorizations (Goldstone, 2003).

The evidence reviewed in this section suggests that there is a strong interrelationship between concepts and perception, with perceptual 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 attributes of a stimulus as a given and try to account for how these attributes are used to categorize that stimulus.

One area of research that provides an exception to this rule is research on object recognition. As pointed out by Schyns (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 proposed to account for object recognition have characteristics in common with theories of categorization. For example, structural description theories of object recognition (e.g., Biederman, 1987; Hummel & Biederman, 1992) are similar to prototype theories of categorization in that a newly encountered exemplar is compared to a summary representation of a category in order to determine whether or not the exemplar is a member of that category. In contrast, multiple views theories of object recognition (e.g., Edelman, 1998; Riesenhuber & Poggio, 1999; Tarr & Bülthoff, 1995) are similar to exemplar-based theories 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 determined either by the exemplar in memory that most closely matches it or by computing 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 potentially 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. Furthermore, theories of object recognition could be adapted to provide a better account of the role of perceptual information in concept formation and use (Palmeri, Wong, & Gauthier, 2004). The rapid recent developments in object recognition research, including the development of detailed computational, neurally based models (e.g., Jiang et al., 2006; Yamins & DiCarlo, 2016), suggest that a careful consideration 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 influence the types of word meanings that one learns, whereas the language that one speaks may influence the types of concepts 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 correlational structure in the environment of the young child, as suggested by Rosch et al. (1976). For example, a child may form a concept of dog around the correlated properties of four legs, tail, wagging, slobbering, and so forth. The subsequent learning of a word meaning 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 & Boroditsky, 2001) has proposed that nouns can be mapped straightforwardly onto existing object concepts, and thus nouns are learned relatively early by children. The relation of verbs to prelinguistic event categories, on the other hand, may be less straightforward. The nature of prelinguistic event categories is not very well understood, but the available evidence suggests that they are structured quite differently from verb meanings. For example, research by Kersten and Billman (1997) demonstrated that when adults learned event categories in the absence of category labels, they formed those categories around a rich set of correlated properties, including the characteristics of the objects in the event, the motions of those objects, and the outcome of the event. Research by Casasola (2005, 2008) has similarly demonstrated that 10- to 14-month-old infants form unlabeled event categories around correlations among different aspects of an event, in this case involving particular objects participating in particular spatial relationships (e.g., containment, support) with one another. 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 one or a small number of properties of an event (Huttenlocher & Lui, 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 the collision.

It may thus be difficult to directly map verbs onto existing event categories. Instead, language-learning experience may be 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 (Bornstein et al., 2004; Gentner, 1982; Gentner & Boroditsky, 2001; Golinkoff & Hirsh-Pasek, 2008; 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. As one example, words in a language provide convenient "handles" for referring to patterns of correlated features that would otherwise be overwhelmingly complex (Lupyan, 2012). As a second 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 in a supervised classification task, a common finding is one of competition among correlated cues for predictive strength (Gluck & Bower, 1988; Shanks, 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; Shanks, 1991).

When categories are learned in the absence of category labels in unsupervised or observational categorization tasks, on the other hand, there is facilitation rather than competition among correlated predictors of category membership (Billman, 1989; Billman & Knutson, 1996, Kersten & Billman, 1997). The learning of unlabeled categories can be 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 Knutson (1996) used this unsupervised categorization method to examine the learning of unlabeled categories of novel animals. They found that participants were more likely to learn the predictiveness of an attribute when other correlated predictors were also present.

Related findings come from Chin-Parker and Ross (2002, 2004). They compared category learning in the context of a classification task, in which the goal of the participant was to predict the category label associated with an exemplar, to an inference-learning task, in which the goal of the participant was to predict a missing feature value. When the members of a category shared multiple feature values, one of which was diagnostic of category membership and others of which were nondiagnostic (i.e., they were also shared with members of the contrasting category), participants who were given a classification task homed in on the feature that was diagnostic of category membership, failing to learn the other feature values that were representative of the category but were nondiagnostic (see also Yamauchi, Love, & Markman, 2002). In contrast, participants who were given an inference-learning task were more likely to discover all of the feature values that were associated with a given category, even those that were nondiagnostic.

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 learns. 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 make no distinction between the past and present because the Hopi language provides no mechanism for talking about this distinction. Many of Whorf's linguistic analyses have since been debunked (see Pinker, 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) classic finding that speakers of Dani, a language with only two color words, learned new color concepts in a similar fashion to English speakers suggested that concepts were determined by perception rather than by language. More recently, however, Roberson and colleagues (Roberson, Davidoff, Davies, & Shapiro, 2005; Roberson, Davies, & Davidoff, 2000) attempted to replicate Heider's findings with other groups of people with limited color vocabularies, namely speakers of Berinmo in New Guinea and speakers of Himba in Namibia. In contrast to Heider's findings, Roberson et al. (2000) found that the Berinmo speakers did no better at learning a new color category for a color that was easy to name in English than for a hard to name color. Moreover, speakers of Berinmo and Himba did no better at learning a category discrimination between green and blue (a distinction not made in either language) than they did at learning a discrimination between two shades of green. This result contrasted with the results of English-speaking participants who did better at the green/blue discrimination. It also contrasted with superior performance in Berinmo and Himba speakers on discriminations that were present in their respective languages. 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 than was originally believed.

Regardless of one's interpretation of the Heider (1972) and Roberson et al. (2000, 2005) results, there are straightforward reasons to expect at least some influence of language on one's concepts. Research dating back to Homa and Cultice (1984) has demonstrated that people are better at learning concepts when category labels are provided as feedback. Verbally labeling a visual target exaggerates the degree to which conceptual categories penetrate visual processing (Lupyan, 2008b). 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 unlabeled. 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 (1994a), 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. Kersten, Goldstone, and Schaffert (1998) provided evidence for this phenomenon and referred to it as attentional persistence. This attentional persistence, in turn, would 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 sensitivities.

This proposal is consistent with L. B. Smith and Samuelson's (2006) account of the apparent shape bias in children's word learning. They proposed that children learning languages such as English discover 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. Consistent with this proposal, children learning English come to attend to shape more strongly and in a wider variety of circumstances than do speakers of Japanese, a language in which shape is marked less prominently and other cues, such as material and animacy, are more prominent (Imai & Gentner, 1997; Yoshida & Smith, 2003).

Although languages differ to some extent in the ways they refer to object categories, languages differ perhaps even more dramatically in their treatment of less concrete domains, such as time (Boroditsky, 2001), number (Frank, Everett, Fedorenko, & Gibson, E., 2008), space (Levinson, Kita, Haun, & Rasch, 2002), motion (Gentner & Boroditsky, 2001; Kersten, 1998a, 1998b, 2003), and blame (Fausey & Boroditsky, 2010, 2011). For example, when describing motion events, languages differ in the particular aspects of motion that are most prominently labeled by verbs. 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 adverbial, 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.

Initial tests of English and Spanish speakers' sensitivities to manner and path of motion (e.g., Gennari, Sloman, Malt, & Fitch, 2002; Papafragou, Massey, & Gleitman, 2002) only revealed differences between the two groups when they were asked to describe events in language. These results thus provide evidence only of an influence of one's prior language-learning history on one's subsequent language use, rather than an influence of language on one's nonlinguistic concept use. More recently, however, Kersten et al. (2010) revealed effects of one's language background on one's performance in a supervised classification task in which either manner of motion or path served as the diagnostic attribute. In particular, monolingual English speakers, monolingual Spanish speakers, and Spanish/English bilinguals performed quite similarly when the path of an alien creature was diagnostic of category membership. Differences emerged when a novel manner of motion of a creature (i.e., the way it moved its legs in relation to its body) was diagnostic, however, with monolingual English speakers performing better than monolingual Spanish speakers. Moreover, Spanish/English bilinguals performed differently depending upon the linguistic context in which they were tested, performing like monolingual English speakers when tested in an English language context but performing like monolingual Spanish speakers when tested in a Spanish language context (see also Lai, Rodriguez, & Narasimhan, 2014). Importantly, the same pattern of results was obtained regardless of whether the concepts to be learned were given novel linguistic labels or were simply numbered, suggesting an influence of native language on nonlinguistic concept formation.

Thus, although the notion that language influences concept use remains controversial, there is a growing body of evidence that speakers of different languages perform differently in a variety of different categorization tasks. Proponents of the universalist viewpoint (e.g., Li, Dunham, & Carey, 2009; Pinker, 1994) have argued that such findings simply represent attempts by research participants to comply with experimental demands, falling back on overt or covert language use to help them solve the problem of "What does the experimenter want me to do here?" According to these accounts, speakers of different languages all think essentially the same way when they leave the laboratory. Unfortunately, we do not have very good methods for measuring how people think outside of the laboratory, so it is difficult to test these accounts. Rather than arguing about whether a given effect of language observed in the laboratory is sufficiently large and sufficiently general to count as a Whorfian effect, perhaps a more constructive approach may be to document the various conditions under which language does and does not influence concept use. This strategy may lead to a better understanding of the bidirectional relationship between concepts and language, and the three-way relationship among concepts, language, and perception (Winawer et al., 2007).

HOW TO IMPROVE CATEGORY LEARNING?

A diverse set of research has shown that there are many factors that influence the effectiveness of category learning. A central question in this research is how to best configure a category-learning context to promote the acquisition of the relevant concept and the ability to apply that knowledge to new exemplars or situations where that knowledge is useful. As an example, students and teachers in every school, university, and training center struggle to make learning efficient and generalizable. Proposals for how to optimize category learning range from selecting the appropriate type of examples to be studied (Gibson, Rogers, & Zhu, 2013), choosing the right type of task, and shifting how a student should approach the task.

An important way to promote learning and transfer of concepts is by establishing comparisons between similar items and making use of analogical reasoning while learning. When learners study two or more instances of the same concept side by side, transfer to more remote instances (e.g., Gick & Holyoak, 1983; Meagher, Carvalho, Goldstone, & Nosofsky, 2017; Omata & Holyoak, 2012; Quilici & Mayer, 2002) or acquisition of a new category (e.g., Gentner & Namy, 1999) is more likely than when only one instance is studied at a time. Moreover, if the instances being studied (even if individually) include a high level of variation along the irrelevant dimensions, learners are generally more likely to transfer their learning to novel situations (e.g., Ben-Zeev & Star, 2001; Chang, Koedinger, & Lovett, 2003; Braithwaite & Goldstone, 2015; H. S. Lee, Betts, & Anderson, 2015). One explanation for this benefit is that studying a superficially diverse set of items allows the learner to notice and extrapolate the common properties, central to the concept being learned (Belenky & Schalk, 2014; Day & Goldstone, 2012; Gentner, 1983; Gick & Holyoak, 1983). However, item variability has also been shown to have no effect on learning effectiveness (e.g., Reed, 1989). Braithwaite and Goldstone (2015) have presented empirical evidence and computational modeling showing that learners with a lower initial knowledge level benefit from less variation among study items, while learners with high levels of prior knowledge benefit from working through examples with more variability (see also Novick, 1988).

Similarly, Elio and Anderson (1984) proposed that learning should start with low-variability items (e.g., items that do not differ much from one another or from the central tendency of the category), and items with greater variability should be introduced later (for similar evidence with young learners see Sandhofer & Doumas, 2008). However, not all learners benefit from this approach. The authors also show that if the learners' approach to the task is to consciously generate hypotheses for category membership, the pattern of results is reversed (Elio & Anderson, 1984). One possibility is that how the learner approaches the task changes what type of information gets encoded and, consequently, what information is most relevant (Elio & Anderson, 1984). It has also been suggested that for optimal transfer of category learning, the study situation should emphasize items that promote a coherent generalization based on the properties that frequently occur (Elio & Anderson, 1981). Moreover, in situations where one needs to learn several items that promote different types of generalizations, the best learning is achieved by studying items that promote similar generalizations close in time (Elio & Anderson, 1981; Mathy & Feldman, 2009).

From the brief survey just provided a question surfaces: Should learning start with difficult items and progress toward easier ones or the other way around? In general, research seems to indicate that learning benefits from study with examples organized in increasing order of complexity or difficulty, that is, from the easiest and simplest to the hardest and more complex (Ahissar & Hochstein, 1997; Baddeley, 1992; Hull, 1920; Terrace, 1964; Wilson, Baddeley, Evans, & Shiel, 1994). However, this might be the case only for categories requiring integration across different dimensions (Spiering & Ashby, 2008), but the reverse may hold for categories organized around a clear rule. Consistent with this latter provision, E. S. Lee, MacGregor, Bavelas, and Mirlin (1988) showed that learners who start by studying examples that other learners classified incorrectly made fewer errors during classification tests than learners who studied the examples in the opposite order.

When it is not possible to change type of examples, their difficulty, or the task, category learning can nonetheless be improved by a careful sequential organization of the examples (Carvalho & Goldstone, 2015). It has been shown before that alternating presentation of items from different hard-to-discriminate categories improves category learning and transfer (Birnbaum, Kornell, Bjork, & Bjork, 2013; Carvalho & Goldstone, 2014; Rohrer, Dedrick, & Stershic, 2015). On the other hand, when the goal is to acquire an independent characterization of each concept, presenting several examples of the same category in close succession improves transfer to a greater degree than frequent alternation (Carpenter & Mueller, 2013; Carvalho & Goldstone, 2014, 2015; Zulkiply & Burt, 2013). Moreover, because greater delays between presentations make it harder to retrieve previous encounters, spaced presentation of items of a single category can promote learning and constitute a desirable difficulty (Bjork, 1994). Consistent with this idea, Birnbaum et al. (2013) showed that while introducing a temporal delay between successive presentations of different categories hindered learning and transfer, increasing the temporal delay between successive presentations of items of the same category improved it (see also Kang & Pashler, 2012). Similarly, research with young children has shown that category generalization benefits from play periods between successive presentations of items of the same category (Vlach, Sandhofer, & Kornell, 2008), and from study with increasingly greater temporal delays between successive repetitions of the same category (Vlach, Sandhofer, & Bjork, 2014).

Finally, an increasingly important issue is whether category learning can be improved by just leaving it in the hands of those who might care the most about it—the learners themselves. Is self-regulated learning better than being given the best study format and organization as determined by an informed teacher? On the one hand, it has been shown that learners are often unaware of how to improve concept learning and fall victim to a series of biases (Bjork, Dunlosky, & Kornell, 2013). On the other hand, self-regulated study is often relatively effective. For example, learners deciding how to sample information outperform those who receive a random sampling of examples or who are yoked to the selections of another learner in a categorization task (e.g., Markant & Gureckis, 2014) or a recognition memory task (Markant, DuBrow, Davachi, & Gureckis, 2014). The mechanism behind this advantage is still an open question. It might be the result of efficient sampling of information given the data available, the process of sampling itself, the greater effort afforded by actively learning, or decisional processes (Gureckis & Markant, 2012).

Understanding how concept learning can be optimized will contribute to designing better instructional techniques and systems (Koedinger, Booth, & Klahr, 2013). Furthermore, it will also contribute to our foundational understanding of the mechanisms underlying concept learning (Carvalho & Goldstone, 2015). Methods for constructing optimal sets of training items given a known learning algorithm and well-defined educational goal have proven informative for understanding both human and machine learning, their commonalities, and their differences (Zhu, 2015).

CONCLUSION

The field of concept learning and representation is noteworthy for its large number of directions and perspectives. While 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 last section suggests, we believe that much of the progress of research on concepts will be to connect concepts to other concepts (Goldstone, 1996; Landauer & Dumais, 1997), to the perceptual world, and to language. One of the risks of viewing concepts as represented by rules, prototypes, sets of exemplars, or category boundaries is that one can easily imagine that one 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 where each concept's meaning depends on other concepts, as well as 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 reveal an important dual nature of concepts. Concepts are used both to recognize objects and to ground word meanings. Working out the details of this dual nature will go a long way toward understanding how human thinking can be both concrete and symbolic.

A second direction is the development of more sophisticated formal models of concept learning. One important recent trend in mathematical models has been the extension of rational models of categorization (Anderson, 1991) to Bayesian models that assume that categories are constructed to maximize the likelihood of making legitimate inferences (Goodman et al., 2008; Griffiths & Tenenbaum, 2009; Kemp & Tenenbaum, 2009). In contrast to this approach, other researchers are continuing to pursue neural network models that offer process-based accounts of concept learning on short and long timescales (M. Jones, Love, & Maddox, 2006; Rogers & McClelland, 2008), and still others chastise Bayesian accounts for inadequately describing how humans learn categories in an incremental and memory-limited fashion (M. Jones & Love, 2011). Progress in neural networks, mathematical models, statistical models, and rational analyses can be gauged by several measures: goodness of fit to human data, breadth 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 process real-world stimuli without relying on human judgments or hand coding (Goldstone & Landy, 2010).

A third direction for research is to tackle more real-world concepts rather than laboratory-created categories, which are often motivated by considerations of controlled construction, ease of analysis, and fit to model assumptions. Some researchers have, instead, tried to tackle particular concepts in their subtlety and complexity, such as the concepts of food (Ross & Murphy, 1999), water (Malt, 1994), and political party (Heit & Nicholson, 2010). Others have made the more general point that how a concept is learned and represented will depend on how it is used to achieve a benefit while interacting with the world (Markman & Ross, 2003; Ross, Wang, Kramer, Simons, & Crowell, 2007). Still others have worked to develop computational techniques that can account for concept formation when provided with large-scale, real-world data sets, such as library catalogs or corpuses of one million words taken from encyclopedias (Glushko, Maglio, Matlock, & Barsalou, 2008; Griffiths, Steyvers, & Tenenbaum, 2007; Landauer & Dumais, 1997). All of these efforts share a goal of applying our theoretical knowledge of concepts to understand how specific conceptual domains of interest are learned and organized, and in the process of so doing, challenging and extending our theoretical knowledge.

A final important direction will be to apply psychological research on concepts. 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

Acker, B. E., Pastore, R. E., & Hall, M. D. (1995). Within-category discrimination of musical chords: Perceptual magnet or anchor? *Perception and Psychophysics*, *57*, 863–874.

Aha, D. W. (1992). Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, *36*, 267–287.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*, 401–406.

Ahn, W-K, Brewer, W. F., & Mooney, R. J. (1992). Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 391–412.

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*, 3–19.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Anderson, J. R., Kline, P. J., & Beasley, C. M., Jr. (1979). A general learning theory and its application to schema abstraction. *Psychology of Learning and Motivation*, *13*, 277–318.

Ashby, F. G. (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *10*, 442–481.

Ashby, F. G., & Gott, R. (1988). Decision rules in perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.

Ashby, F. G., & Maddox, W. T. (1993). Relations among prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *38*, 423–466.

Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making: Handbook of perception and cognition* (pp. 251–301). San Diego, CA: Academic Press.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.

Baddeley, A.D. (1992). Implicit memory and errorless learning: A link between cognitive theory and neuropsychological rehabilitation? In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (2nd ed., pp. 309–314). New York, NY: Guilford Press.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition*, *10*, 82–93.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition*, *11*, 211–227.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development* (pp. 101–140). New York, NY: Cambridge University Press.

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27). New York, NY: Academic Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology*, *36*, 203–272.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge, United Kingdom: Cambridge University Press.

Bassok, M. (1996). Using content to interpret structure: Effects on analogical transfer. *Current Directions in Psychological Science*, *5*, 54–58.

Belenky, D. M., & Schalk, L. (2014). The effects of idealized and grounded materials on learning, transfer, and interest: An organizing framework for categorizing external knowledge representations. *Educational Psychology Review*, *26*, 27–50.

Ben-Zeev, T., & Star, J. R. (2001). Spurious correlations in mathematical thinking. *Cognition and Instruction*, *19*(3), 253–275.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Billman, D. (1989). Systems of correlations in rule and category learning: Use of structured input in learning syntactic categories. *Language and Cognitive Processes*, *4*, 127–155.

Billman, D., & Knutson, J. F. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 458–475.

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*(3), 392–402.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*(November 2012), 417–444.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S., Pascual, L., & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75*, 1115–1139.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*, 1–22.

Bourne, L. E., Jr. (1970). Knowing and using concepts. *Psychological Review*, *77*, 546–556.

Bourne, L. E., Jr. (1982). Typicality effect in logically defined categories. *Memory and Cognition*, *10*, 3–9.

Braithwaite, D. W., & Goldstone, R. L. (2015). Effects of variation and prior knowledge on abstract concept learning. *Cognition and Instruction*, *33*, 226–256.

Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General*, *120*, 278–287.

Bruner, J. S. (1973). *Beyond the information given: Studies in the psychology of knowing*. New York, NY: Norton.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York, NY: Wiley.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, *3*, 81–117.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, *41*(5), 671–682.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481–495.

Carvalho, P.F., & Goldstone, R.L. (2015). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6*(505), 1–12.

Casasola, M. (2005). When less is more: How infants learn to form an abstract categorical representation of support. *Child Development*, *76*, 1818–1829.

Casasola, M. (2008). The development of infants' spatial categories. *Current Directions in Psychological Science*, *17*, 21–25.

Chang, N. M., Koedinger, K. R., & Lovett, M. C. (2003). Learning spurious correlations instead of deeper relations. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Cognitive Science Society*. Boston, MA: Cognitive Science Society.

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition*, *30*, 353–362.

Chin-Parker, S., & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: A comparison of inference learning and classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 216–226.

Choi, S., McDaniel, M. A., & Busemeyer, J. R. (1993). Incorporating prior biases in network models of conceptual rule learning. *Memory & Cognition*, *21*, 413–423.

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*, 390–406.

Crawford, L. J., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, *11*, 284–288.

Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science*, *21*, 234–242.

Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist*, *47*(3), 153–176.

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, *21*, 449–498.

Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA: Bradford Books.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 397–417.

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, *12*(1), 20–30.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*, 107–140.

Estes, W. K. (1994). *Classification and cognition*. New York, NY: Oxford University Press.

Fausey, C. M., & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, *17*, 644–650.

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, *18*(1), 150–157.

Feldman, J. (2006) An algebra of human concept learning. *Journal of Mathematical Psychology*, *50*, 339–368.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*, 752–782.

Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review*, *117*, 309–348.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, 1–77.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fodor, J., Garrett, M., Walker, E., & Parkes, C. M. (1980). Against definitions. *Cognition*, *8*, 263–367.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*, 819–824.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*, 181–215.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*, 191–197.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*, 568–573.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*, 49–79.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory*, *4*, 161–178.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language* *development: Vol. 2. Language, thought, and culture* (pp. 301–334). Hillsdale, NJ: Erlbaum.

Gentner, D. (1983). Structure-Mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge, United Kingdom: Cambridge University Press.

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, *14*(4), 487–513.

Gibson, B. R., Rogers, T. T., & Zhu, X. (2013). Human semi-supervised learning. *Topics in Cognitive Science*, *5*(1), 132–172.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38.

Gluck, M.A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 227–247.

Gluck, M. A., & Bower, G. H. (1990). Component and pattern information in adaptive networks. *Journal of Experimental Psychology: General*, *119*, 105–109.

Glushko, R. J., Maglio P., Matlock, T., & Barsalou, L. (2008). Categorization in the wild. *Trends in Cognitive Sciences*, *12*, 129–135.

Goldstone, R. L. (1994a). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*, 178–200.

Goldstone, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*, 125–157.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory and Cognition*, *24*, 608–628.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 86–112.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. In R. Kimchi, M. Behrmann, and C. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives* (pp. 233–278). Mahwah, NJ: Erlbaum.

Goldstone, R. L., & Barsalou, L. (1998). Reuniting perception and conception. *Cognition*, *65*, 231–262.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Interdisciplinary Reviews: Cognitive Science*, *1*, 65–78.

Goldstone, R. L., & Landy, D. H. (2010). Domain-creating constraints. *Cognitive Science*, *34*, 1357–1377.

Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science*, *2*, 265–284.

Goldstone, R. L., & Rogosky, B. J. (2002). Using relations within conceptual systems to translate across conceptual systems. *Cognition*, *84*, 295–320.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*, 116–139.

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, *31*, 169–180.

Goldstone, R. L., Steyvers, M., Spencer-Smith, J., & Kersten, A. (2000). Interactions between perceptual and conceptual learning. In E. Diettrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 191–228). Mahwah, NJ: Erlbaum.

Golinkoff, R., & Hirsh-Pasek, K. (2008). How toddlers learn verbs. *Trends in Cognitive Science*, *12*, 397–403.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*, 108–154

Gopnik, A., & Choi, S. (1995). Names, relational words, and cognitive development in English and Korean speakers: Nouns are not always learned before verbs. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs*. Hillsdale, NJ: Erlbaum.

Gureckis, T. M., & Markant, D. B. (2012). Self-Directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation, *Psychological Review*, *114*, 211–244.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*, 661–716.

Grossberg, S. (1982). Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review*, *89*, 529–572.

Hampton, J. A. (1993). Prototype models of concept representation. In I. V. Mecheln, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67–95). London, United Kingdom: Academic Press.

Hampton, J. A. (1997). Conceptual combination: Conjunction and negation of natural concepts. *Memory & Cognition*, *25*, 888–909.

Hampton, J.A., & Passanisi, A. (2016). When intensions do not map onto extensions: Individual differences in conceptualization. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(4), 505–523

Harnad, S. (1987). *Categorical perception*. Cambridge, United Kingdom: Cambridge University Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Harnad, S., Hanson, S. J., & Lubin, J. (1995). Learned categorical perception in neural nets: Implications for symbol grounding. In V. Honavar & L. Uhr (Eds.), *Symbolic processors and connectionist network models in artificial intelligence and cognitive modelling: Steps toward principled integration* (pp. 191–206). Boston, MA: Academic Press.

Heider, E.R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*, 10–20.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1264–1282.

Heit, E. (1997). Knowledge and concept learning. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 7–41). Hove, United Kingdom: Psychology Press.

Heit, E., & Nicholson, S. (2010). The opposite of Republican: Polarization and political categorization. *Cognitive Science*, *34*, 1503–1516.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–429.

Homa, D., & Cultice, J. C. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 234–257.

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs*, *28*, Whole No. 123.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480–517.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*, 220–241.

Huttenlocher, J., & Lui, F. (1979). The semantic organization of some simple nouns and verbs. *Journal of Verbal Learning and Verbal Behavior*, *18*, 141–162.

Imai, M. & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, *62*, 169–200.

James, W. (1890). *The principles of psychology, in two volumes*. New York, NY: Henry Holt.

Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. *Neuron*, *50*, 159–172.

Jones, M., & Goldstone, R. L. (2013). The structure of integral dimensions: Contrasting topological and Cartesian representations. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 111–132.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*, 169–188.

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 316–332.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development*, *8*, 113–139.

Jung, W. & Hummel, J. E. (2015). Making probabilistic relational categories learnable. *Cognitive Science*, *39*, 1259–1291.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: Bradford Books.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences, USA*, *105*, 10687–10692.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*, 20–58.

Kersten, A. W. (1998a). A division of labor between nouns and verbs in the representation of motion. *Journal of Experimental Psychology: General*, *127*, 34–54.

Kersten, A. W. (1998b). An examination of the distinction between nouns and verbs: Associations with two different kinds of motion. *Memory & Cognition*, *26*, 1214–1232.

Kersten, A. W. (2003). Verbs and nouns convey different types of motion in event descriptions. *Linguistics*, *41*, 917–945.

Kersten, A. W., & Billman, D. O. (1997). Event category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 638–658.

Kersten, A. W., Goldstone, R. L., & Schaffert, A. (1998). Two competing attentional mechanisms in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1437–1458.

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General*, *139*, 638–653.

Kiefer, M., Sim, E., Herrnberger, B., Grothe, J. & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, *28*, 12224–12230.

Kiefer, M., Sim, E., Liebich, S., Hauk, O., & Tanaka, J. (2007). Experience-dependent plasticity of conceptual representations in human sensory-motor areas. *Journal of Cognitive Neuroscience*, *19*, 525–542.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(6161), 935–937.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, *112*, 500–526.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.

Kohonen, T. (1995). *Self-organizing maps*. Berlin, Germany: Springer.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voice-voiceless distinction in alveolar plosive consonants. *Science*, *190*, 69–72.

Kurtz, K. J. (2015). Human category learning: Toward a broader explanatory account. *Psychology of Learning and Motivation*, *63*, 77–114.

Lai, V. T., Rodriguez, G. G., & Narasimhan, B. (2014). Thinking-for-speaking in early and late bilinguals. *Bilingualism: Language and Cognition*, *17*, 139–152.

Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*, 227–260.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Lassaline, M. E., & Logan, G. D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 561–581.

Lee, E. S., MacGregor, J. N., Bavelas, A., & Mirlin, L. (1988). The effects of error transformations on classification performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 66–74.

Lee, H. S., Betts, S., & Anderson, J. R. (2015). Not taking the easy road: When similarity hurts learning. *Memory & Cognition*, *43*(6), 939–952.

Lenat, D. B., & Feigenbaum, E. A. (1991). On the thresholds of knowledge. *Artificial Intelligence*, *47*, 185–250.

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception and Psychophysics*, *62*, 386–401.

Levinson, S. C., Kita, S., Haun, D. B. M., & Rasch, B. H. (2002). Returning the tables: Language effects spatial reasoning. *Cognition*, *84*, 155–188.

Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology*, *58*, 487–524.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.

Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science*, *34*, 113–147

Luce, R. D. (1959). *Individual choice behavior*. New York, NY: Wiley.

Lupyan, G. (2008a). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*(2), 348–369.

Lupyan, G. (2008b). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, *108*, 566–577

Lupyan, G., (2012). What do words do? Toward a theory of language-augmented thought. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, pp. 255–297). Cambridge, MA: Academic Press.

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*, *6*, 547–569.

Lupyan, G. (2017). The paradox of the universal triangle: Concepts, language, and prototypes. *Quarterly Journal of Experimental Psychology*, *70*, 389–412.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*, 49–70.

Meagher, B. J., Carvalho, P. F., Goldstone, R. L., & Nosofsky, R. M. (2017). Organized simultaneous displays facilitate learning of complex natural science categories. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-017-1251-6

Malt, B. C. (1994). Water is not $H_2$0. *Cognitive Psychology*, *27*, 41–70.

Malt, B. C. & Wolff, P. (Eds.). (2010). *Words and the mind: How words capture human experience*. New York, NY: Oxford University Press.

Marcus, G. F. (1998). Rethinking eliminativist connectionism. *Cognitive Psychology*, *37*, 243–282.

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94–122.

Markant, D., DuBrow, S., Davachi, L., & Gureckis, T. M. (2014). Deconstructing the effect of self-directed study on episodic memory. *Memory & Cognition*, *42*(8), 1211–1224.

Markman, A. B., & Dietrich, E. (2000). In defense of representation. *Cognitive Psychology*, *40*, 138–171.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, *127*, 331–354.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592–613.

Mathy, F., & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review*, *16*(6), 1050–1057.

McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy-sets? *Memory & Cognition*, *6*, 462–472.

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General*, *143*, 668–693.

McFadden, D., & Callaway, N. L. (1999). Better discrimination of small changes in commonly encountered than in less commonly encountered auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 543–560.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 128–148.

McNamara, T. P, & Miller, D. L. (1989). Attributes of theories of meaning. *Psychological Bulletin*, *106*, 355–376.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, *44*, 1469–1481.

Medin, D. L., & Atran, S. (1999). *Folkbiology*. Cambridge, MA: MIT Press.

Medin, D. L., Lynch. E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual Review of Psychology*, *51*, 121–147.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*, 158–190.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, *35*, 113–138.

Medin, D. L., Wattenmaker, W. D., & Michalski, R. S. (1987). Constraints and preferences in inductive learning: An experimental study of human and machine performance. *Cognitive Science*, *11*, 299–339.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89–115.

Miyawaki, K. (1975). An effect of linguistic experience. The discrimination of (r) and (l) by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331–340.

Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science*, *12*, 529–562.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 904–919.

Murphy, G. L., & Kaplan, A. S. (2000). Feature distribution and background knowledge in category learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *53A*, 962–982.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *10*, 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, *43*, 25–53.

Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin and Review*, *5*, 345–369.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 510–520.

Ojalehto, B. L., & Medin, D. (2015). Perspectives on culture and concepts. *Annual Review of Psychology*, *66*, 249–275.

Omata, T., & Holyoak, K. J. (2012). The role of comparison processes in the induction of schemas for design styles. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 2144–2149). Austin, TX: Cognitive Science Society.

O'Toole, A. J., Peterson, J., & Deffenbacher, K. A. (1995). An "other-race effect" for categorizing faces by sex. *Perception*, *25*, 669–676.

Ozgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, *131*, 477–493.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 324–354.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 548–568.

Palmeri, T. J., Wong, A. C. N., & Gauthier, I. (2004). Computational approaches to the development of visual expertise. *Trends in Cognitive Sciences*, *8*, 378–386.

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, *84*, 189–219.

Pastore, R. E. (1987). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical perception* (pp. 29–52). Cambridge, United Kingdom: Cambridge University Press.

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 416–432.

Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). New York, NY: International Universities Press.

Piantadosi, S. T., & Jacobs, R. (2016). Four problems solved by the probabilistic language of thought. *Current Directions in Psychological Science*, *25*, 54–59.

Piantadosi, S. T., Tenenbaum, J. & Goodman, N. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review*, *123*(4), 392–424.

Pinker, S. (1991). Rules of language. *Science*, *253*, 530–535.

Pinker, S. (1994). *The language instinct*. New York, NY: William Morrow.

Posner, M. I., & Keele, S. W. (1967). Decay of visual information from a single letter. *Science*, *158*, 137–139.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304–308.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, *17*, 884–892.

Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, *16*(3), 325–342.

Reed, S. K. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology*, *81*(4), 532–540.

Rehder, B., & Murphy, G. L. (2003). A knowledge-resonance (KRES) model of knowledge-based category learning. *Psychonomic Bulletin & Review*, *10*, 759–784.

Repp, B. H., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In S. R. Harnad (Ed.), *Categorical Perception* (pp. 89–112). New York, NY: Cambridge University Press.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.

Ripley B. D. (1996). *Pattern recognition and neural networks*. New York, NY: Cambridge University Press.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*, 665–681.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadu & A. Ortony (Eds.), *Similarity, analogy, and thought* (pp. 21–59). Cambridge, United Kingdom: Cambridge University Press.

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General*, *122*, 468–486.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*, 378–411.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*, 369–398.

Rogers, T. T., & McClelland, J. L. (2008). A précis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences*, *31*, 689–714.

Rogers, T. T., & Patterson, K. (2007) Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*, 451–469.

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, *107*(3), 900–908.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–232.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, *38*, 495–553.

Ross, B. H., Wang, R. F., Kramer, A. F., Simons, D. J., & Crowell, J. A. (2007). Action information from classification learning. *Psychonomic Bulletin & Review*, *14*, 500–504.

Sandhofer, C. M., & Doumas, L. A. A. (2008). Order of presentation effects in learning color categories. *Journal of Cognition and Development*, *9*(2), 194–221.

Schank, R. C. (1982). *Dynamic memory: A theory of reminding and learning in computers and people*. Cambridge, United Kingdom: Cambridge University Press.

Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition*, *67*, 147–179.

Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 433–443.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*, 278–304.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA: Authors Cooperative.

Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex*, *15*, 1602–1608.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*, 2802–2810.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Sloman, S. A., Love, B. C., & Ahn, W.-K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*, 189–228.

Smith, E. E. (1989). Concepts and induction. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 501–526). Cambridge, MA: MIT Press.

Smith, E. E., Langston, C., & Nisbett, R. (1992). The case for rules in reasoning. *Cognitive Science*, *16*, 1–40.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smith. E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*, 167–196.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*, 214–241.

Smith, E. E., & Sloman, S. A. (1994). Similarity-versus rule-based categorization. *Memory and Cognition*, *22*, 377–386.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 3–27.

Smith J. D., Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 800–811.

Smith, L. B., & Samuelson, L. (2006). An attentional learning account of the shape bias: Reply to Cimpian and Markman (2005) and Booth, Waxman, and Huang (2005). *Developmental Psychology*, *42*, 1339–1343.

Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior*, *23*, 3–22.

Spalding, T. L., & Murphy, G. L. (1999). What is learned in knowledge-related categories? Evidence from typicality and feature frequency judgments. *Memory & Cognition*, *27*, 856–867.

Spiering, B. J., & Ashby, F. G. (2008). Initial training with difficult items facilitates information integration, but not rule-based category learning. *Psychological Science*, *19*(11), 1169.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description: Vol. 3. Grammatical categories and the lexicon*. New York, NY: Cambridge University Press.

Tanaka, J., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*, 457–482.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, *32*, 492–504.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple-views? *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1494–1505.

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition*, *67*, 73–110.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems 11* (pp. 59–65). Cambridge, MA: MIT Press.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Terrace, H.S. (1964). Wavelength generalization after discrimination learning with and without errors. *Science*, *144*, 78– 80.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*, 615–665.

Vermeulen, N., Corneille, O., & Niedenthal, P. M. (2008). Sensory load incurs conceptual processing costs. *Cognition*, *109*, 287–294.

Vlach, H. A, Sandhofer, C. M., & Bjork, R. A. (2014). Equal spacing and expanding schedules in children's categorization and generalization. *Journal of Experimental Child Psychology*, 1–9.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167.

Weiskopf, D. (2009). The plurality of concepts. *Synthese*, *169*, 145–173.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Whorf, B. L. (1956). The relation of habitual thought and behavior to language. In J. B. Carroll (Ed.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (pp. 35–270). Cambridge, MA: MIT Press.

Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, *66*, 55–84.

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*, *142*, 1006–1014.

Wilson, B. A., Baddeley, A., Evans, J., & Shiel, A. (1994). Errorless learning in the rehabilitation of memory impaired people. *Neuropsychological Rehabilitation*, *4*, 307–326.

Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, USA*, *104*, 7780–7785.

Witt, J. K., Kemmerer, D., Linkenauger, S. A., & Culham, J. (2010). A functional role for motor simulation in identifying tools. *Psychological Science*, *21*, 1215–1219.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). New York, NY: Macmillan.

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 585–593.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language*, *39*, 124–148.

Yamins D., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*, 356–365.

Yoshida, H., & Smith, L. B. (2003). Correlation, concepts, and cross-linguistic differences. *Developmental Science*, *16*, 90–95.

Zhu, X. (2015). Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In *Twenty-Ninth AAAI Conference on Artificial Intelligence (Senior Member Track, AAAI)*. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9487/9685

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*, 16–27.