BRIEF REPORT

# Is there an exemplar theory of concepts?

Gregory L. Murphy [1]

**Abstract** It is common to describe two main theories of concepts: prototype theories, which rely on some form of summary description of a category, and exemplar theories, which claim that concepts are represented as remembered category instances. This article reviews a number of important phenomena in the psychology of concepts, arguing that they have no proposed exemplar explanation. In some of these cases, it is difficult to see how an exemplar theory would be adequate. The article concludes that exemplars are certainly important in some categorization judgments and in category-learning experiments, but that there is no exemplar theory of human concepts in a broad sense.

**Keywords** Concepts · Categories · Semantic memory · Induction · Categorization

The two main theories of concepts are prototype and exemplar models—all the textbooks say so, including my own chapter in an introductory textbook (Murphy, 2013) and standard cognitive psychology texts such as Goldstein (2014). Prototype models are summary descriptions of the category as a whole, based on properties that are often found in the category. They include feature-based models (e.g., Rogers et al., 2004; Rosch & Mervis, 1975; Smith, Rips, & Shoben, 1974) and network models, in which category nodes are linked to the category's features as well as to other categories (e.g., Collins & Quillian, 1969). Categories formed by rules are also summary descriptions, because the rule describes the category as a whole.

✉ Gregory L. Murphy
   gregory.murphy@nyu.edu

[1]   Department of Psychology, New York University, 6 Washington
      Place, 8th floor, New York, New York 10003, USA

Exemplar models assume that this summary description is not needed, instead representing categories by memories of instances (exemplars) that are labeled with their category name (Medin & Schaffer, 1978; Smith & Medin, 1981). Research has suggested that "instances" are not category members but are encounters with category members. That is, if you encounter a mouse three times (identifying it as a mouse), that is effectively three exemplars, in terms of its influence on later classification (Barsalou, Huttenlocher, & Lamberts, 1998; Nosofsky, 1988).[1] In discussing this distinction, I will focus on the issue of summary representations versus memory for instances and will refer to such theories as *prototype* and *exemplar models*, ignoring other important differences between the theories.

Prototype theory arose through the work of Eleanor Rosch on natural categories (e.g., Rosch, 1975; Rosch & Mervis, 1975). As a result, it has focused on familiar natural categories and their structure. Exemplar theory arose from the learning tradition; Douglas Medin was a researcher on discrimination learning in animals when he applied the theory to human learning. Theoretically, exemplar models seem to make the most sense within category-learning experiments, in which a small number of exemplars are presented over and over again in a classification task. Under those circumstances, it seems reasonable that individual exemplars would be remembered, along with their categorizations. Furthermore, evidence for such processes also seems strongest when the categories are smaller and poorly structured, when prototypes cannot suffice

---

[1] It is important to understand that *exemplars* refers to actual instances that are members of a category (or encounters with such instances). Sometimes the term is loosely used to refer to subcategories—for example, "swallow is an exemplar of the category of birds." However, swallows are not an exemplar but are a category with many thousands of individual members.

for learning the categories (Blair & Homa, 2001; Smith & Minda, 1998). Indeed, many of the earliest demonstrations of the better performance of exemplar theory were in studies in which the categories were extremely difficult to learn (e.g., Medin & Schaffer, 1978; Medin & Schwanenflugel, 1981; see Murphy, 2005, for a discussion).

Another source of exemplar theory came about through a very different rationale and experimental paradigm. The late Lee Brooks (1987) argued that particularly salient and familiar exemplars could influence categorization above and beyond any category description. For example, if you see a dog that looks quite similar to your own dog, you would likely be much faster to identify it as a dog and, Brooks suggested, attribute characteristics to it that your own dog has. However, Brooks did not say that you were likely to remember every dog you encountered or that every dog in memory would influence your classification. Indeed, his experiments with novel categories generally involved intense study of a few exemplars, which then were shown to influence the categorization of novel items even when this influence was harmful (Allen & Brooks, 1991; Thibaut & Gelaes, 2006). In those experiments, people learned both a category rule (not quite a prototype, but an experimental analogue) and specific exemplars. Memory of similar exemplars could overcome use of the rule. Although Brooks did not specify a detailed model, it seems likely that he had in mind that people used both specific, very familiar exemplars and the rule, to varying degrees. Furthermore, he often emphasized that the similarity between known exemplars and new items should be great in order to observe exemplar effects.

The differences between Brooks's approach and exemplar models arising from classification-learning experiments led me (Murphy, 2002, chap. 4) to distinguish the two, calling the Medin-derived models *exemplar models* and the Brooks-like demonstrations *exemplar effects*. I suggested that Brooks's exemplar effects were essentially cases of implicit memory, in which prior experience influenced later judgments. Like implicit memory effects, they are found even after a gap of some weeks, suggesting that explicit recall of exemplars might not be occurring (Brooks, Norman, & Allen, 1991). Furthermore, such effects are quite similar to reminding effects found in problem solving (e.g., Ross, 1984). Given all this, the evidence for exemplar *effects* seems quite strong and very consistent with other research in cognitive psychology.

The evidence for exemplar *models*, however, did not seem as strong to me when I reviewed the field over a decade ago (Murphy, 2002, chap. 13). Although it seems very likely that such models are good descriptions of many (but not all) category-learning experiments, there were many questions about how the models could be applied to real-world category learning and concept use. Furthermore, the proponents of this theory did not seem to be attempting to broaden its application

to other topics in the psychology of concepts outside of category-learning experiments. This absence, which continues to this day, makes me propose that in fact *there is no exemplar theory of concepts*, when we understand *concepts* to be the mental representations of things in the world that are involved in planning, reasoning, communication, learning, and all the things that concepts are supposed to be involved in. There is an exemplar theory of category-learning experiments of the sort that psychologists often do, but such experiments are a tool to explore theories—they are not the actual domain of behavior we are trying to explain when we develop theories of concepts.

The remainder of this article outlines this argument. I discuss some basic phenomena in the psychology of concepts and ask whether each has an exemplar explanation and whether that explanation is plausible. In many of these cases, the problem is not that the theory is not correct, but that no account has been put forward at all. In each case, some kind of prototype model has been proposed (or can be immediately devised) to address that phenomenon. Unfortunately, space does not allow me to develop the prototype explanations here.

## Phenomena of concepts

### Hierarchical structure

Since at least Linnaeus, it has been known that many categories are structured into hierarchies, in which more-specific categories are nested underneath general ones, often forming a tree structure. Even outside the Linnaean realm of biology we have nested categories, such as side-door refrigerator, refrigerator, kitchen appliance, appliance, machine, and artifact. Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) documented this in a set of familiar concepts and also introduced the notion that a mid-level category was generally *basic*—that is, the easiest and preferred way to categorize objects. When people walk into a kitchen, they see a refrigerator rather than a machine or a side-door refrigerator, unless a specific task induces a different categorization. Furthermore, children learn words at this level sooner than at more-specific and more-general levels (Anglin, 1977). Within prototype theory, there is a generally accepted explanation of this advantage—namely that basic-level categories are informative but also distinct from other categories (Murphy & Brownell, 1985; Rosch et al., 1976).

In prototype theory, hierarchical structure can be represented in terms of links between concepts, or else by inference from typical features (see Murphy, Hampton, & Milovanovic, 2012, for a recent discussion). In the context of exemplar theory, however, it is not clear how to represent hierarchical structure itself. One of the properties of such hierarchies is that people are able to affirm statements such as "All refrigerators are appliances" or "A squirrel is an animal," and many

experiments have tested sentence verification of this sort since Collins and Quillian (1969; Murphy et al., 2012). If knowledge of squirrels is in terms of remembered exemplars, how is the hierarchy to be represented? The simplest way would be to encode each exemplar with all its categories. Of course, this would be rather inefficient, since every squirrel I see would have to be remembered as a squirrel, rodent, mammal, vertebrate, and animal. And I see a *lot* of squirrels. Also, it is puzzling that each exemplar should be remembered this way, because when I see squirrels, it very seldom crosses my mind that they are mammals or animals, much less vertebrates. Since I am not in a learning experiment in which I have to respond with what categories each squirrel is in, it seems unlikely that I would encode these categories with my memories of these squirrel encounters.

Therefore, it seems more likely that in my memory of squirrel exemplars, many are encoded as "squirrel," a few as "mammal," a few more as "animal," and probably none as "vertebrate." How from this set of exemplars do we derive a hierarchy? We cannot follow the standard practice of exemplar models in experiments, of checking each exemplar to see which categories it is in: All the "squirrels" are not also encoded as "mammal," and so forth. So, on the basis of exemplars, we would have to answer *false* to the question of whether all squirrels are mammals. Of course, some squirrels are in fact encoded both as mammals and squirrels, so perhaps we could accept this partial overlap as evidence for a hierarchical relation. The problem is that such partial overlap also occurs for categories that do not have a hierarchical relationship, like pets or New York City residents, and people do not claim that all squirrels are pets or New York residents. So, it is very difficult to see how the conceptual hierarchy would be represented in terms of exemplars and how people draw inferences such as that squirrels must breathe because they are animals. This seems like a huge issue for the exemplar theory.

The only study I know taking the exemplar approach that has tested hierarchically organized categories is Palmeri (1999). However, he avoided the problem I just noted by making two separate models: One learned the categories at one level and another at the higher level, analogous to one model learning to classify birds versus mammals, and the other learning to classify dogs, horses, moose, and so on. So, in fact, neither model represented a hierarchy or the relation between categories at the different levels. Palmeri's goal was to compare different learning models' performance on the two levels rather than to construct a model of category hierarchies; my point is that his solution suggests the difficulties of trying to build an exemplar model of hierarchical concepts.

Another challenge for exemplar theories is explaining the basic-level advantage. People are faster at identifying something as a squirrel than as a mammal or as a chair than as a dining-room chair (Murphy & Brownell, 1985; Rosch et al., 1976; Smith, Balzano, & Walker, 1978). As was mentioned earlier,

this is typically explained by the fact that chair is an informative category that will closely match its members, whereas furniture is a vague category that does not match its members as closely. (I will not address subordinate classification here.)

When classification is by virtue of exemplars, it is not clear that the explanation of the slowness of superordinate classification applies. Categorization is not performed by using the vague representation of mammals but instead by using exemplars encoded as mammals. Assuming that some squirrels are encoded as mammals (or else one would not have represented the hierarchy), a new squirrel is very likely to match those squirrels just as much as it matches squirrels encoded as squirrels. Remembered exemplars are concrete, and when they represent abstract categories, they do so concretely. Thus, the slowness of superordinate classification that prototype theory ascribes to the abstract properties and vagueness of general categories simply does not apply for exemplar theory: There are no abstract properties, only remembered squirrels. The new squirrel matches my remembered squirrel exemplars encoded as mammals just as much as those encoded as squirrels, so the basic-level advantage should not occur.

There are ways to try to address this problem. For example, one could point out that squirrels are not very similar to *other* mammal exemplars (e.g., seals, humans, or cows), which slows down their classification as mammals. However, this is inconsistent with the theme of exemplar models that close similarity is used in classification, dating from Medin and Schaffer's (1978) multiplicative similarity rule, and continued in Shepard's (1987) widely adopted exponential similarity function. That is, moderate similarity to multiple exemplars generally counts for less in classification than very close similarity to a single exemplar. Exemplar reminding effects also seem to work only with high levels of similarity (Thibaut & Gelaes, 2006). As Smith and Minda (1998) pointed out, exemplar models of experimental results can have such a fast similarity drop-off that only one learned exemplar is effectively influencing the categorization of a given item. Thus, it seems awkward that very close similarity of exemplars is required in order to explain categorization in learning experiments, but the dissimilarity of squirrels to cows is necessary to explain why squirrels are not easily classified as mammals. Cows should not affect how a squirrel is classified into any category, according to exemplar theory.

My argument, then, is in the form of a puzzle: It is simply unclear to me how exemplar models propose to represent hierarchical structure and the basic-level advantage in learning and category use. I have not proven that it cannot do so; it is possible that an exemplar theory can be developed to account for these phenomena. But what puzzles me is, where is that theory? The hierarchical nature of categories has been known roughly forever, and the basic level since 1976. If there is an exemplar theory of concepts, what is its explanation of these very familiar phenomena that are in all the textbooks?

## Knowledge effects

Concepts are part of our knowledge of the world. They are the way that we gain access to information that we can then apply to the objects and events around us. Prototype theory directly represents that knowledge, by its very nature as a description of classes of objects, as well as by the relations linking different conceptual entities. Exemplars represent such knowledge more implicitly, by embodying the experiences of individual instances, from which we can infer generalizations.

Concepts not only contain knowledge of the world, they are also formed and processed in the light of our other knowledge. For example, we learn concepts that are consistent with our knowledge faster than we do arbitrary concepts (Murphy & Allopenna, 1994). When constructing concepts without feedback, we discover the category structure if the concept's properties go together in a way that fits our expectations (Spalding & Murphy, 1996). We learn many more of a concept's properties when they are related to a consistent theme than if they are not (Hoffman, Harris, & Murphy, 2008).

One of the most exciting topics in higher-level cognition has been the discovery of people's knowledge of causal structures and the reasoning that they engage in with such structures (e.g., Sloman, 2005). Researchers have documented that the properties of concepts are causally related and then showed that such relations influenced classification, both with artificially constructed (e.g., Rehder & Hastie, 2001) and natural (Ahn, 1998; Kim & Ahn, 2002) concepts. Investigation into the internal causal structure of concepts continues today (see the next section).

Finally, within both developmental and adult studies of concepts, people's beliefs about domains have been shown to influence their judgments about categories. The most prominent such proposal is Medin and Ortony's (1989) *psychological essentialism*, which claims that people think of some categories as having an invisible essence that determines category membership. This remarkably productive idea has led to advances as far-flung as early conceptual development (Gelman, 2003) and the psychology of prejudice and social groups (Hirschfeld, 1996; Rhodes, Leslie, & Tworek, 2012).

The knowledge referred to by the writers in these fields is almost always generic knowledge—that is, knowledge of an entire category or set of objects possessing a property. For example, I believe that wings enable flying, and animals that fly can live in trees. Such knowledge could influence my classification of something as a bird (Rehder, 2009). However, this is not knowledge about a particular bird or about the object I am evaluating—it is knowledge about the billions of animals with wings. Similarly, people's belief that something hatched from an egg that was laid by a robin must have the robin's underlying essence is not a belief about a specific exemplar but about robins (or birds or species) in general.

Prototypes represent such general knowledge very easily, because they are descriptions of whole classes of entities, and therefore can contain generalizations about those entities. If you learn a new fact about birds, you have a summary representation of that category to link it to. Other knowledge can be inferred by hierarchical inference (e.g., robins have the properties of birds) or via causal inferences involving generic properties (e.g., wings usually enable flying). Although much work is still to be done to specify the structure of all that knowledge and how such inferences work, representing concepts as general descriptions causes no problems in constructing such explanations.

Exemplar theory does not store generalizations, by definition, but instead yields them through some kind of processing. For example, it does not say that feathers are an important property of birds, but that fact is implicitly represented by most bird exemplars having feathers represented with them, and few other animal concepts having feathers. As a result, a new exemplar having feathers is likely to be classified as a bird, because it will be much more similar to things in memory labeled as *bird* than to other things.

There are two main problems with this approach to knowledge representation, however. The first is how to represent knowledge that is not tied to an experience with exemplars. I firmly believe that mammals have four-chambered hearts, but this knowledge is not tied to any exemplar of a whale, bat, cow, or human I have ever seen. I might also believe that mammals evolved from dinosaur ancestors, but that is obviously also something not associated with individual exemplars.

Wings enable flying in my view of the world, but that is not only because of the pairing of wings and flying things in many exemplars. First of all, I may well not notice either the wings or flying of some things that have both properties. When I see pictures of birds in books, they are often perched; when I swat at flying insects, I often do not perceive their wings. Furthermore, I have some (very sketchy) beliefs about how wings enable flying. I do not think that if I attached wings to a rock or a pig that it would fly. I have some idea of why ostriches and penguins do not fly, in spite of having wings. In order to make a prediction about a new animal with wings, I must reason through my beliefs about wings, how they support a body in the air, the relative sizes of the wings and body, and so on. How is such knowledge represented in exemplar representations?

This illustrates the second issue with exemplar theories, that it is not clear what reasoning processes would allow people to make specific predictions or generalizations about concepts using exemplar knowledge. Most exemplar models focus on classification and do not provide standard mechanisms for feature inference. Furthermore, given that much of the knowledge that we apply is based on feature relations (such as the causal research cited above), it is not clear how exemplar

models would represent and use that knowledge. Is "wings enable flying" simply a statistical relationship between exemplars with wings and those that fly? That is the most obvious exemplar representation, but it does not seem likely to be adequate. Feathers and beaks are very highly correlated but have no direct causal connection. How would biases such as essentialism apply to exemplar representations to yield the results in the literature? One might argue that the use of knowledge and causal reasoning are simply separate processes from the learning of categories, but research suggests that categorization and generalization are closely tied to the causal relations of categories' features (Kim & Ahn, 2002; Rehder, 2009). Initial category learning is also strongly influenced by prior knowledge (Hoffman et al., 2008; Murphy & Allopenna, 1994).

I cannot argue that exemplar representations are in some way incompatible with all of these knowledge effects, but I simply do not have a clue how exemplar theories would attempt to account for them. And, as in the first section, I am not aware of a detailed attempt to account for them within exemplar theory. Instead, people who study these issues almost never talk about exemplars, and researchers who work on exemplar theory almost never talk about knowledge or how their concept representations could explain knowledge effects. So, I cannot say that knowledge effects are inconsistent with exemplar theories, but I can ask again, "Where is the exemplar theory of all these phenomena?"

## Induction

It is now often repeated (at least, by me) that the purpose of having concepts is to make inductions. We do not need concepts in order to recall facts about individual entities we have encountered; we need them in order to understand and make predictions about new objects and situations. This function was important enough to appear in the first paragraph of Smith and Medin's (1981) classic book. Since then, research on *category-based induction* (CBI) has exploded (Feeney & Heit, 2007).

Many of the problems of exemplar theory for explaining induction are inherent in the problems discussed above. For example, I have not experienced seeds in many apples I have encountered (I avoid them), so why am I so certain that a new apple I have bought will have seeds inside it? Much research in CBI involves the transfer of properties from one or more *premise* categories to a *conclusion* category (Osherson, Smith, Wilkie, López, & Shafir, 1990). For example, if grizzly bears and brown bears love onions, what is the probability that all bears love onions? It is hard to know how to represent such a question in terms of exemplars, because the premises are usually universally quantified, but one's exemplar representations of grizzly bears does not include the information that each one loves onions, nor does it seem likely that this information could be easily attached to each exemplar. Because the

properties used are usually hypothetical or even imaginary, people should not (and presumably do not) store the properties with actual exemplars. (I have discussed the bears loving onions example for many years, but I do not actually believe that grizzly bears love onions.)

Research on CBI has shown that two variables are important in such problems. First, the similarity between the premise and conclusion categories influences the degree to which the properties from one will be attributed to the other (Rips, 1975). It is not very difficult to think of an exemplar process that would explain this. In hearing the question, exemplars representing the premise categories could be activated, and their similarity to exemplars of the conclusion category could be computed. Of course, this could run into some of the problems mentioned above, such as the very few exemplars I have in memory for grizzly bears or larks. Also, calculating the similarity of two categories is not trivial if there are many exemplars in each. So, I think it would be possible to propose an exemplar account of the typicality effect, though I do not know whether it would be psychologically plausible.

The second variable is the *coverage* of the premise categories to their superordinate category (I am simplifying a bit here). That is, people draw stronger inferences when the premise categories are more representative of their category as a whole. People will generalize a property from a robin to a penguin more than they will from an owl, because robins are more similar to birds in general (Dunsmoor & Murphy, 2014; Rips, 1975). For multiple premises, the better the categories cover the entire superordinate category, the stronger the inference will be (Osherson et al., 1990). As a result, inference is stronger when the premise categories are diverse, because they would as a whole be more similar to their superordinate category.

I am not sure how an exemplar model would explain this effect. Indeed, the effect seems incompatible with the general goal of exemplar models to explain category phenomena by the similarity of test items to learned exemplars. Why should a robin's representativeness of the bird category, which was not mentioned in the question, affect its induction to penguins? Why is the similarity of a robin to a penguin not sufficient?

Assuming that coverage is important, how would an exemplar model calculate it? Osherson et al. (1990) calculated coverage by using similarity judgments between pairs of categories used in their stimuli, but they did this at the level of subcategories (e.g., comparing robin to swallow, lark, etc.). A true exemplar model would have to compare the exemplars in these categories—that is, all of the robins in memory to all of the swallows, all of the larks, all of the chickens, and so on. For large categories and multiple premises, that seems computationally unrealistic.

Another problem is related to the previous section's discussion—namely, that the content of the features and categories is important. In an important demonstration, Heit and Rubinstein (1994) found that the nature of the property changed the

inductive strength. For example, people generalized a biological property from whales to rabbits more than they did a behavioral property. This is easily explained by people's beliefs about the biological relationship between whales and rabbits (both being mammals), but it cannot be explained by any account that relies on exemplar similarity, which would be equal in the two inductions. Similarly, Gelman and Markman (1986) found that children sensibly generalize some properties on the basis of category membership (e.g., behavioral tendencies), and others on the basis of appearance (e.g., weight). Rehder (2009) discussed evidence that induction is based on causal relations between features (when they are known). In addition to impressive evidence for such effects, he found that the similarity between exemplars in induction problems has little effect when causal links can explain the induction. This is contrary to the usual mechanism of exemplar models, which is based on similarity between items.[2] In all these cases, some kind of generic information is accessed during induction, and (as was discussed in the previous section) such knowledge is a summary representation.

Perhaps as a result of these problems, no exemplar model of CBI seems to exist. In a review of the developmental literature on induction, Hayes (2007) outlined the main theories of CBI. This is useful here as a checklist to make sure that I have not missed an exemplar account. Although some of the models are quite task-specific, to the degree that they make claims about category representation, they refer to features associated with a category, to knowledge-related features, and to other summary representations. I do not see any exemplar theories in his list. Recent Bayesian models of category knowledge and inference (e.g., Kemp & Tenenbaum, 2009) have also assumed that features are associated with categories and do not seem to use exemplars as a way of representing category knowledge and implementing induction.

Unlike with some of the other examples I have considered here, I think it would be possible to develop an exemplar account that could yield some of the observed phenomena of CBI (though it would have to address some of the earlier representational issues). My guess is that this account would not be as convincing as the current proposals, but the more concrete problem is that no such account seems to have been put forward. That is, where is the exemplar theory of category-based induction? When knowledge effects, which are rife in this domain, are introduced, I do not see how exemplar theories could account for the results, but right now there is no extant theory of those effects for us to evaluate.

_____

[2] To be clear, some prototype models also do not account for content effects, such as Osherson et al.'s (1990) seminal model. To explain such effects, any theory would have to refer to more general knowledge beyond the categories involved. My point is that there is no exemplar account of how such knowledge would be represented or accessed.

### The rest of the concepts literature

Much of the research on concepts has been rather different from the standard category-learning study that exemplar theory has focused on. Perhaps the most productive example is research on conceptual development. Throughout the 1980s and 1990s, research on adult concepts focused on the prototype-versus-exemplar debate. However, it is remarkable how little that debate impinged on research in conceptual development during the same time period. There, the focus was on topics such as how children induce categories on the basis of very little evidence, the use of categories in induction, and word learning. None of this work was phrased in terms of exemplar knowledge. It might be possible to recast some of it in exemplar terms, though some of it is explicitly about biases and knowledge that apply to categories as a whole (e.g., essentialism).

Conceptual development is a very large and complex area, and I cannot make a broad statement about whether its results would or would not be explicable in terms of exemplar representations. However, it simply does not seem that exemplar theory has made significant inroads into this large field, which presumably is about the same content matter as adult research on concepts.

Concepts are also intimately linked to the semantic representations of words (Murphy, 2002, chap. 11). However, within the psycholinguistic literature, there has also been little influence of exemplar theory. Has an exemplar account proposed how we interpret words in context or how we choose which word to utter? The accounts that I am familiar with are all description-based—that is, words are associated with semantic properties that describe the things referred to by the word. Theories of aphasia and loss of semantic knowledge in dementia also assume category-level descriptions (Rogers et al., 2004).

### Conclusion

My conclusion to this very brief review is that, surprisingly, there is not really an exemplar theory of concepts. There is an exemplar theory of category learning, within the usual paradigm of repeated exemplars. That theory seems to have strong support in some experimental circumstances. Furthermore, we also have good evidence that specific salient exemplars can influence one's classification. However, no well-developed exemplar theory of concepts as a whole seems to have addressed the broader phenomena of natural concepts.

There are two obvious explanations for this gap: One is that it would be difficult to devise such a theory, and the other is that the proponents of exemplar models have simply not turned their attention to issues beyond category learning—and researchers on those other topics have for some reason

neglected exemplar models. The first seems more plausible to me. Theories of general knowledge and hierarchical representation are not trivial issues that a theory of concepts might neglect. If it were easy to construct exemplar accounts that straightforwardly address those phenomena, I think they would have been constructed. Concept and word learning in children are central topics in cognitive science; it is hard to believe that exemplar theories of these things are simply waiting to be brought into existence but that no one has thought to do so for the past 25 years.

My inference is only an inference, and it might be wrong. But if it is wrong, it seems reasonable to require exemplar theorists to expand their theories to give serious accounts of the well-known phenomena in the field. If strong claims about how categories are represented are to be taken seriously, then the full breadth of findings in the field needs to be addressed. So, if an exemplar theory of concepts is to be made, it is time for one to be put forward.

## Recommendations

Textbooks and review articles have almost universally discussed two main approaches to concepts: exemplar and prototype theories. In reality, each approach is a large class of possible models, with considerable diversity. My argument that there is no exemplar theory should not be taken to imply that any single prototype account explains everything. Rather, my claim is that most of the accounts put forward to explain the conceptual behaviors reviewed here are, at least implicitly, prototype accounts; they rely on category-level descriptions and (often structured) associations between properties and categories.

Concepts are important because they represent vast amounts of complex knowledge and allow us to understand and draw inferences about new objects and classes of things. They are also fascinating because of how quickly children acquire a vast fund of concepts, on the basis of less evidence than one might expect. The standard category-learning task is not in and of itself a central topic in the psychology of concepts: It is a tool to understand how concepts are acquired and represented. And although it is a very important tool, I think we have to recognize that a theory that explains this task but that has not been extended to the central cases of real-world concept use is not a true theory of concepts.

So, if I were writing an introductory textbook on concepts today, how would I frame this debate? Following Smith and Medin's (1981) example, I would introduce Rosch's prototype notion as the alternative to the classical view of concepts. Then I would point out that in addition to such general representations, people are influenced by specific exemplars, describing some of the classic Brooks research. I would describe the category-learning task and point out that in category-learning experiments, people can learn individual exemplars and classify new items in terms of similarity to them. However, I do not think I would describe this as "Some believe concepts are prototypes and some believe that they are exemplars," because that debate could not be carried over into the rest of the chapter, that talked about induction, conceptual development, word meaning, and so on. It makes little sense to claim that there are two theories if only one of them has anything to say about two-thirds of the material in the chapter.

Rather than phrasing the issue as a debate between two incompatible theories, it might be more productive to develop psychologically plausible accounts of how the two kinds of information work together to produce our rich store of conceptual knowledge, allowing each kind of knowledge to explain the tasks that are most suited for it. Early connectionist models (e.g., Knapp & Anderson, 1984) and more recent clustering models (Love, Medin, & Gureckis, 2004) have emphasized that exemplar and summary information can blur into one another. A distinct exemplar can be remembered as a separate item. As more exemplars similar to it are encoded, however, the details of that exemplar may become lost (through interference) or less relevant, as they are combined with new information. Such models can represent both "exemplar-like" and "prototype-like" effects, while not restricting themselves to one kind of representation. Perhaps such approaches can serve as the basis for a more encompassing theory of concepts.

## References

Ahn, W.-K. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition, 69,* 135–178.

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3–19. doi:10.1037/0096-3445.120.1.3

Anglin, J. M. (1977). *Word, object and conceptual development.* New York, NY: W. W. Norton.

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

Blair, M., & Homa, D. (2001). Expanding the search for a linear separability constraint on category learning. *Memory & Cognition, 29,* 1153–1164. doi:10.3758/BF03206385

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 141–174). Cambridge, UK: Cambridge University Press.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnosis task. *Journal of Experimental Psychology: General, 120,* 278–287. doi:10.1037/0096-3445.120.2.278

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 241–248.

Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus typicality determines how broadly fear is generalized. *Psychological Science, 25,* 1816–1821.

Feeney, A., & Heit, E. (2007). *Inductive reasoning: Experimental, developmental, and computational approaches*. Cambridge, UK: Cambridge University Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, UK: Oxford University Press.

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

Goldstein, E. B. (2014). *Cognitive psychology: Connecting mind, research, and everyday experience* (4th ed.). Stamford, CT: Cengage.

Hayes, B. K. (2007). The development of inductive reasoning. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches* (pp. 25–54). Cambridge, UK: Cambridge University Press.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 411–422. doi:10.1037/0278-7393.20.2.411

Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds*. Cambridge, MA: MIT Press.

Hoffman, A. B., Harris, H. D., & Murphy, G. L. (2008). Prior knowledge enhances the category dimensionality effect. *Memory & Cognition, 36,* 256–270. doi:10.3758/MC.36.2.256

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

Kim, N. S., & Ahn, W.-K. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131,* 451–476. doi:10.1037/0096-3445.131.4.451

Knapp, A. G., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 616–637. doi:10.1037/0278-7393.10.4.616

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111,* 309–332. doi:10.1037/0033-295X.111.2.309

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge, UK: Cambridge University Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238. doi:10.1037/0033-295X.85.3.207

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 355–368. doi:10.1037/0278-7393.7.5.355

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

Murphy, G. L. (2005). The study of concepts inside and outside the lab: Medin vs. Medin. In W.-K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the lab: Essays in honor of Douglas Medin* (pp. 179–195). Washington, DC: American Psychological Association.

Murphy, G. L. (2013). Categories and concepts. In E. Diener & R. Biswas-Diener (Eds.), *Psychology (Noba Textbook Series)* (pp. 813–842). Champaign, IL: DEF Publishers.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 904–919. doi:10.1037/0278-7393.20.4.904

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 70–84. doi:10.1037/0278-7393.11.1.70

Murphy, G. L., Hampton, J. A., & Milovanovic, G. S. (2012). Semantic memory redux: An experimental test of hierarchical category representation. *Journal of Memory and Language, 67,* 521–539. doi:10.1016/j.jml.2012.07.005

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65. doi:10.1037/0278-7393.14.1.54

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200. doi:10.1037/0033-295X.97.2.185

Palmeri, T. J. (1999). Learning categories at different hierarchical levels: A comparison of category learning models. *Psychonomic Bulletin & Review, 6,* 495–503.

Rehder, B. (2009). Causal-based property generalization. *Cognitive Science, 33,* 301–344.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130,* 323–360. doi:10.1037/0096-3445.130.2.323

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences, 109,* 13526–13531.

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

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111,* 205–235.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–233. doi:10.1037/0096-3445.104.3.192

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605. doi:10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439. doi:10.1016/0010-0285(76)90013-X

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16,* 371–416. doi:10.1016/0010-0285(84)90014-8

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323. doi:10.1126/science.3629243

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives*. New York, NY: Oxford University Press.

Smith, E. E., Balzano, G. J., & Walker, J. (1978). Nominal, perceptual, and semantic codes in picture categorization. In J. W. Cotton & R. L. Klatzky (Eds.), *Semantic factors in cognition* (pp. 137–168). Hillsdale, NJ: Erlbaum.

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

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1411–1436. doi:10.1037/0278-7393.24.6.1411

Smith, E. E., Rips, L. J., & Shoben, E. J. (1974). Semantic memory and psychological semantics. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 1–45). New York, NY: Academic Press.

Spalding, T. L., & Murphy, G. L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 525–538. doi:10.1037/0278-7393.22.2.525

Thibaut, J.-P., & Gelaes, S. (2006). Exemplar effects in the context of a categorization rule: Featural and holistic influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1403–1415. doi:10.1037/0278-7393.32.6.1403