Further Reading

Faugeras O (1993) *Three-Dimensional Computer Vision*. Cambridge, MA: MIT Press.

Fischler MA and Firschein O (1987) *Readings in Computer Vision: Issues, Problems, Principles and Paradigms*. Los Altos, CA: Morgan Kaufmann.

Marr D (1982) *Vision*. San Francisco, CA: WH Freeman.

Tsotsos JK (1990) Analyzing vision at the complexity level. *Behavioral and Brain Sciences* **13**: 423–445.

Tsotsos JK (1992) Image understanding. In: Shapiro S (ed.) *Encyclopedia of Artificial Intelligence*, 2nd edn, pp. 641–663. New York, NY: John Wiley.

Zucker SW (1992) Early vision. In: Shapiro S (ed.) *Encyclopedia of Artificial Intelligence*, 2nd edn, pp. 394–420. New York, NY: John Wiley.

Computers

See **Human–Computer Interaction**

Concept Learning

Introductory article

Bradley C Love, University of Texas, Austin, Texas, USA

CONTENTS

Concept learning is the process of acquiring knowledge structures that enable an agent to make predictive inferences.

INTRODUCTION

The human species evolves to meet challenges in the environment. Unfortunately, evolution is a slow 'learning' process. Evolution can only help us address aspects of our environment that are not very variable and that are stable over a long period of time. Of course, many aspects of our environment are constantly undergoing change. Accordingly, many concepts have to be learned *de novo* by each individual. For example, a radiologist is not born knowing how to interpret x-ray images. It is hard to imagine how that particular skill could evolve.

Concept learning is integral to the survival of any agent (e.g. a human, an animal, a robot, etc.) operating in a complex and changing environment. A concept is a mental representation that is often derived from experiences with specific instances. We often develop concepts of categories (i.e. collections of objects) in the world. Without acquired concepts, we would be unable to make sense of the world around us. Every new object encountered would appear completely novel and we would not know how to interact with it. For example, the first time a child encounters a hot stove he may get burned. When the child visits a friend's house and encounters another stove, it is unlikely the child will touch it, even though the new stove may differ in a number of ways from the original stove (e.g. size, color, design, etc.). If the child did not generalize from his experiences and form a concept of stoves, he would go through life with burned hands. (*See* **Categorization, Development of**; **Generalization**)

One basic question is how do we learn new concepts? Philosophers, psychologists, and computer scientists have all pondered this question. In the following sections, three basic views (i.e. models) of concept learning and concept representation (i.e. what is stored as a consequence of learning) will be examined. The first account posits that concepts consist of rules. A more recent account holds that concepts are represented as prototypes. A prototype can be thought of as the average example of a concept. A third account of concepts is the exemplar view. The exemplar view holds that concepts are nothing more than a collection of stored exemplars (i.e. examples of the concept). We will evaluate the relative merits of each of these accounts of human concept learning. All three accounts correctly characterize some aspects of human concept learning. After evaluating these three accounts, we will discuss more modern neural network models of concept learning. Neural network models embody some of the characteristics of rule, prototype, and exemplar approaches. (*See* **Concept Learning and Categorization: Models**; **Classifier Systems**; **Concepts, Philosophical Issues about**; **Conceptual Representations in Psychology**)

RULES

The classical view of concepts holds that categories are defined by logical rules. In Figure 1, any item that is a square is a member of category A. This simple rule determines category membership. According to the rule view, our concept of category A can be represented by this simple rule. Discovering this rule would involve a rational hypothesis-testing procedure. This procedure attempts to discover a rule that is satisfied by all of the positive examples of a concept, but none of the negative examples of the concept (i.e. items that are members of other categories). In trying to come up with such a rule for category A, one might first try the rule 'if dark, then in category A'. After rejecting this rule (because there are counterexamples), other rules would be tested (starting with simple rules and progressing towards more complex rules) until the correct rule is eventually discovered. For example, in learning about birds, one might first try the rule 'if it flies, then it is a bird.' This rule works pretty well, but not perfectly (penguins do not fly and bats do). Another simple rule like 'if it has feathers, then it is a bird' would not work either because a pillow filled with feathers is not a bird. Eventually, a more complex rule might be discovered like 'if it has feathers and lays eggs, then it is a bird'.

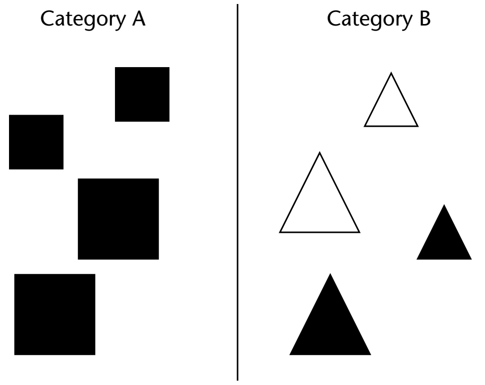

Figure 1. Examples of category A and category B.

Although rules can in principle provide a concise representation of a concept, often more elaborate representations would serve us better. Concept representation needs to be richer than a simple rule because we use concepts for much more than simply classifying objects we encounter. For instance, we often use concepts to support inference (e.g. a child infers members of the category stove can be dangerously hot). Using categories to make inferences is a very important use of concepts. Knowing something is an example of a concept tells us a great deal about the item. For example, if you can classify a politician from the USA as a Republican, you can readily infer the politician's position on a number of issues. The point is that our representations of concepts need to include information beyond what is needed to classify items as examples of the concept. For example, the rule 'if square, then in category A' correctly classifies all members of category A in Figure 1, but it does not capture the knowledge that all category A members are dark. One problem with rule representations of concepts is that potentially useful information is discarded.

The biggest problem with the rule approach to concepts is that most of our everyday categories do not seem to be describable by a tractable rule. To demonstrate this point, Wittgenstein noted that the concept game lacks a defining property. Most games are fun, but Russian roulette is not fun. Most games are competitive, but ring around the roses is not competitive. While most games have characteristics in common, there is not a rule that unifies them all. Rather, we can think of the members of the category game as being organized around a *family resemblance* structure (analogous to how members of your family resemble one another).

A related weakness of the rule account of concepts is that examples of a concept differ in their *typicality*. If all a concept consisted of was a rule that determined membership, then all examples should have equal status. According to the rule account, all that should matter is whether an item satisfies the rule. Our concepts do not seem to have this definitive flavor. For example, some games are better examples of the category game than others. Basketball is a very typical example of the category game. Children play basketball in a playground, it is competitive, there are two teams, each team consists of multiple players, you score points, etc.

Basketball is a typical example of the category of games because it has many characteristics in common with other games. On the other hand, Russian roulette is not a very typical game – it requires a gun and one of the two players dies. Russian roulette does not have many properties in common with other games. In terms of family resemblance structure, we can think of basketball as having a central position and Russian roulette being a distant cousin to the other family members. These findings extend to categories in which a simple classification rule exists. For example, people judge the number three to be a more typical odd number than the number forty-seven even though membership in the category 'odd number' can be defined by a simple rule.

The fact that category membership follows a gradient as opposed to being all or none affords us flexibility in how we apply our concepts. Of course, this flexibility can lead to ambiguity. Consider the concept mother. It is a concept that we are all familiar with that seems straightforward – a mother is a woman who becomes pregnant and gives birth to a child. But what about a woman who adopts a neglected infant and raises it in a nurturing environment? Is the birth mother who neglected the infant a mother? What if a woman is implanted with an embryo from another woman? Court cases over maternity arise because the concept of motherhood is ambiguous. The concept exhibits greater flexibility and productivity than is even indicated above. For example, is it proper to refer to an architect as the mother of a building? All the above examples of the concept mother share a family resemblance structure (i.e. they are organized around some commonalities), but the concept is not rule based. Some examples of the concept mother are better than others.

PROTOTYPES

The prototype approach to concept learning and representation was developed by Rosch and colleagues to address some of the shortcomings of the rule approach. Prototype models represent information about all the possible properties (i.e. *stimulus dimensions*), instead of focusing on only a few properties like rule models do. The prototype of a category is a summary of all of its members. Mathematically, the prototype is the average or central tendency of all category members. Figure 2 displays the prototypes for two categories, simply named categories A and B. Notice that all the items differ in size and luminance (i.e. there are two stimulus dimensions) and that the prototype is located amidst all of its category members. The prototype for each category has the average value of both the stimulus dimensions of size and luminance for the members of its category. (*See* **Prototype Representations**)

The prototype of a category is used to represent the category. According to the prototype model, a novel item is classified as a member of the category whose prototype it is most similar to. For example, a large bright item would be classified as a member of category B because category B's prototype is large and bright (see Figure 2). The position of the prototype is updated when new examples of the category are encountered. For example, if one encountered a very small and dark item that is a member of category A, then category A's prototype would move slightly towards the bottom left corner in Figure 2. As an outcome of learning, the position of the prototype shifts towards the newest category member in order to take it into account. A prototype can be very useful for determining category membership in domains where there are

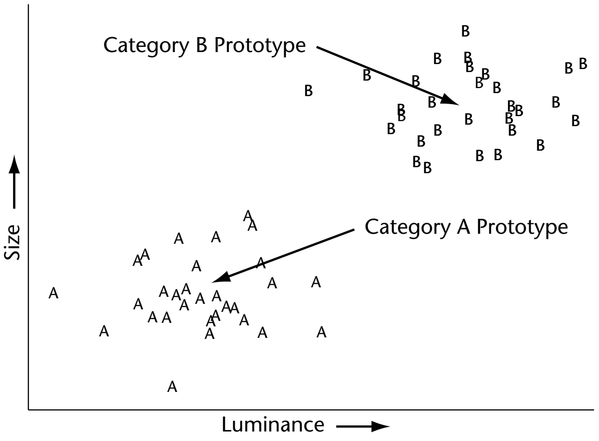

Figure 2. Two categories and their prototypes.

many stimulus dimensions that each provide information useful for determining category membership, but no dimension is definitive. For example, members of a family may tend to be tall, have large noses, a medium complexion, brown eyes, and good muscle tone, but no family member possesses all of these traits. Matching on some subset of these traits would provide evidence for being a family member. (*See* **Multidimensional Scaling**; **Similarity**)

Notice the economy of the prototype approach. Each cloud of examples in Figure 2 can be represented by just the prototype. The prototype is intended to capture the critical structure in the environment without having to encode every detail or example. It is also fairly simple to determine which category a novel item belongs to by determining which category prototype is most similar to the item.

Unlike the rule approach, the prototype model can account for typicality effects. According to the prototype model, the more typical category members should be those members that are most similar to the prototype. In Figure 2, similarity can be viewed in geometric terms – the closer items are together in the plot, the more similar they are. Thus, the most typical items for categories A and B are those that are closest to the appropriate prototype. Accordingly, the prototype approach can explain why robins are more typical birds than penguins. The bird prototype represents the average bird: has wings, has feathers, can fly, can sing, lives in trees, lays eggs, etc. Robins share all of these properties with the prototype, whereas penguins differ in a number of ways (e.g. penguins cannot fly, but do swim). Extending this line of reasoning, the best example of a category should be the prototype, even if the actual prototype has never been viewed (or does not even exist). Indeed, numerous learning studies support this conjecture. After viewing a series of examples of a category, human participants are more likely to categorize the prototype as a category member (even though they never actually viewed the prototype) than they are to categorize an item they have seen before as a category member.

Because the prototype approach does not represent concepts in terms of a logical rule that is either satisfied or not, it can explain how category membership has a graded structure that is not all or none. Some examples of a category are simply better examples than other examples. Also, categories do not need to be defined in terms of logical rules, but are rather defined in terms of family resemblance to the prototype. In other words, members of a category need not share a common defining thread, but rather can have many characteristic threads in common with one another.

The prototype approach, while preferable to the rule approach for the reasons just discussed, does fail to account for important aspects of human concept learning. The main problem with the prototype model is that it does not retain enough information about examples encountered in learning. For instance, prototypes do not store any information about the frequency of each category, yet people are sensitive to frequency information. If an item was about equally similar to the prototype of two different categories and one category was one hundred times larger than the other, people would be more likely to assign the item to the more common category (under most circumstances).

People are also sensitive to the variability along stimulus dimensions. To use Rips' example, a circular object with a 10 cm diameter may be more similar to a US quarter (which is about 2.5 cm in diameter) than to a pizza (which is much larger). Nevertheless, the novel object is more likely to be classified as a pizza than a quarter because quarters display very little variability in their diameters whereas pizzas can vary in size.

Finally, prototypes are not sensitive to the correlations and substructure within a category. For example, a prototype model would not be able to represent that spoons tend to be large and made of wood or small and made of steel. These two subgroups would simply be averaged together into one prototype. This averaging makes some categories unlearnable with a prototype model. One example of such a category structure is shown in Figure 3. Each category consists of two subgroups. Members of category A are either small and dark

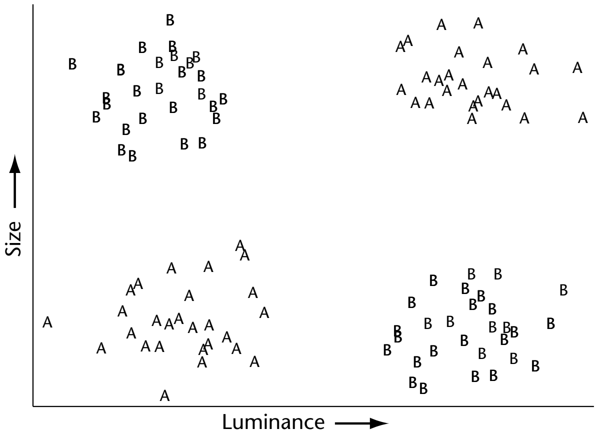

Figure 3. Two categories, each containing two subgroups.

or they are large and light, whereas members of category B are either large and dark or they are small and light. The prototypes for the two categories are both in the centre of the stimulus space (i.e. medium size and medium luminance). Items cannot be classified correctly by which prototype they are most similar to because the prototypes provide little guidance.

In general, prototype models can only be used to learn category structures that are *linearly separable*. A learning problem involving two categories is linearly separable when a line or plane can be drawn that separates all the members of the two categories. The category structure shown in Figure 2 is linearly separable because a diagonal line can be drawn that separates the category A and B members (i.e. the category A members fall on one side of the line and the category B members fall on the other side of the line). Thus, this category structure can be learned with a prototype model. The category structure illustrated in Figure 3 is nonlinear – no single line can be drawn to segregate the category A and B members. Mathematically, a category structure is linearly separable when there exists a weighting of the feature dimensions that yields an additive rule that correctly indicates one category when the sum is below a chosen threshold and the other category when the sum is above the threshold.

The inability of the prototype model to learn nonlinear category structures detracts from its worth as a model of human concept learning because people are not biased against learning nonlinear category structures. Some nonlinear category structures are actually easier to acquire than linear category structures. For example, it seems quite natural that small birds sing, whereas large birds do not sing. Many categories have subtypes within them that we naturally pick out. One way for the prototype model to address this learnability problem is to include complex features that represent the presence of multiple simple features (e.g. large and blue). Unfortunately, this approach quickly becomes unwieldy as the number of stimulus dimensions increases.

EXEMPLARS

Exemplar models address many of the shortcomings of the prototype model. Exemplar models store every training example in memory instead of just the prototype (i.e. the summary) of each category. By retaining all of the information from training, exemplar models are sensitive to the frequency, the variability, and the correlations among items. For the learning problem illustrated in Figure 2, an exemplar model would store every training example. New items are classified by how similar they are to all items in memory (not just the prototype). For the category structure illustrated in Figure 2, the pairwise similarity of a novel item and every stored item would be calculated. If the novel item tended to be more similar to the category A members (i.e. the item was small and dark) than the category B members, then the novel item would be classified as a member of category A.

One aspect of exemplar models that seems counterintuitive is their lack of any abstraction in category representation. It seems that humans do learn something more abstract about categories than a list of examples. Surprisingly, exemplar models are capable of displaying abstraction. For instance, exemplar models can correctly predict that humans more strongly endorse the underlying prototype (even if it has not been seen) than an actual item that has been studied (a piece of evidence previously cited in favor of the prototype model). How could this be possible without the prototype actually being stored? It would be impossible if exemplar models simply functioned by retrieving the exemplar in memory that was most similar to the current item and classified the current item in the same category as the retrieved exemplar (this is essentially how processing works in a prototype model, except that a prototype is stored in memory instead of a bunch of exemplars).

Instead, exemplar models engage in more sophisticated processing and calculate the similarity between the current item (the item that is to be classified) and every item in memory. Some exemplars in memory will be very similar to the current item, whereas others will not be very similar. The current item is classified in the category in which the sum of its similarities to all the exemplars is greatest. When a previously unseen prototype is presented to an exemplar model it can be endorsed as a category member more strongly than a previously seen item. The prototype (which is the central tendency of the category) will tend to be somewhat similar to every item in the category, whereas any given non-prototype item will tend to be very similar to some items (especially itself!) in memory, but not so similar to other items. Overall, the prototypical item can display an advantage over an item that has actually been studied. Abstraction in an exemplar model is indirect and results from processing (i.e. calculating and summing pairwise similarities), whereas abstraction in a prototype model is rather direct (i.e. prototypes are stored).

The exemplar model does seem to make some questionable assumptions. For example, exemplar models store every training example which seems excessive. Also, every exemplar is retrieved from memory every time an item is classified. In addition to these assumptions, one worries that the exemplar model does not make strong enough theoretical commitments because it retains all information about training and contains a great deal of flexibility in how it processes information. These issues are currently being resolved by researchers. On the whole, exemplar models seem to be a more viable approach to understanding human concept learning than existing prototype or rule-based approaches, but there is still room for further work. (*See* **Computational Models of Cognition: Constraining**)

NEURAL NETWORK MODELS

Neural network models are intended to learn in a manner analogous to how the brain learns. A neural network consists of layers of neuron-like units that connect to units in other layers. Units can excite and inhibit one another across these connections. An item is represented at the input layer (the first layer) and passes activity to more advanced layers in the network until it reaches the output layer which determines the category the item is a member of (e.g. if the unit in the output layer representing category B is the most activated, then the item is classified as a member of category B). Each unit integrates all the activity originating from the layer below via its connections and passes this summed activity through a transfer function to generate its own output which is passed on to the next layer. Figure 4 illustrates a feedforward neural network with an input, hidden, and output layer. (*See* **Connectionism**)

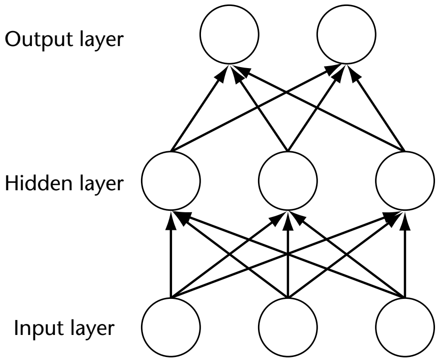

Figure 4. A typical feedforward neural network.

The connections between units are altered as a result of learning in order to minimize the prediction error (i.e. the weights are altered in order to correctly classify items). Sophisticated learning algorithms dictate how the weights should be altered as a result of learning. Neural networks with only an input and output layer share many of the limitations of the prototype model – they can only learn linearly separable functions (i.e. simple category structures). More complicated neural networks with a hidden layer (and nonlinear transfer functions) can learn just about any category structure. However, neural networks of this variety are not very good models of human concept learning because they tend to learn problems quickly that people learn slowly and vice versa.

Neural network models that are conceptually related to rule, prototype, and exemplar models have been successful as models of human concept learning. For example, the ALCOVE model replaces the hidden layer in Figure 4 with encoded exemplars. In other words, units in the hidden layer are added as exemplars are encountered. This exemplar neural network model, which combines an exemplar representation of concepts with the powerful learning algorithms of neural networks, does a good job of accounting for aspects of human concept learning. The SUSTAIN model is a neural network model that combines aspects of both exemplar and prototype models. SUSTAIN initially begins like a prototype model, but it can store exemplars (which themselves can later evolve into prototypes) when prediction errors occur. For the problem illustrated in Figure 3, SUSTAIN would form four prototypes that correspond to the four clusters of items. The ability to store multiple prototypes per category allows SUSTAIN to avoid the problems that plague prototype models. Both ALCOVE and SUSTAIN also incorporate rule-like dynamics. These models learn to attend to the most relevant stimulus dimensions and neglect the less meaningful dimensions, much like how rule models tend to focus on a limited number of stimulus dimensions (e.g. if it is *large*, then it is in category A).

CONCLUSIONS AND FUTURE DIRECTIONS

From this brief review of concept learning models we saw that the progression from rule models to prototype models to exemplar models was marked by a shift towards more concrete representations (i.e. more information about the training examples is retained), greater fluidity (i.e. category boundaries are not seen as rigid), and more sophisticated processing at decision time (exemplar models are the quintessential case – all abstraction is done after the training examples are encoded). Although all three approaches have their shortcomings, they all reflect some aspects of human concept learning. The successful neural network models of concept learning retain characteristics of all three approaches. Like the rule approach, these neural network models acknowledge the utility of strategically focusing on a subset of stimulus dimensions. If a stimulus dimension is irrelevant to a learning problem, the models will ignore the dimension and not be distracted by it. Like prototype models, some of these neural network models form abstractions which can assist generalization and reduce storage requirements. Like exemplar models, these neural network models are quite fluid, can encode individual exemplars, and engage in sophisticated processing at decision time.

One important aspect of concept learning that these models do not address is the influence of prior knowledge. Our prior knowledge exhibits strong influences on what we learn from a series of examples. For example, even if all the blue cars on a mechanic's lot have transmission problems and none of the red cars do, the mechanic would never predict that blue cars in general have transmission problems. Certainly, the mechanic would not paint a car red in the hope of repairing it. The mechanic's prior knowledge and theories of how cars function preclude this association. Instead, the mechanic is oriented towards more fruitful solutions. One important challenge for concept learning models is to illuminate how prior knowledge affects our interpretation of examples. Conversely, more work is needed in understanding how examples we encounter affect our theories of the world.

Further Reading

Lakoff G (1987) *Women, Fire, and Dangerous Things: What Categories Tell Us About the Nature of Thought*. Chicago, IL: University of Chicago Press.

Medin DL (1998) Concepts and conceptual structure. In: Thagard P (ed.) *Mind Readings*, pp. 93–126. Cambridge, MA: MIT Press.

Mervis CB and Rosch E (1981) Categorization of natural objects. In: Rosenzweig MR and Porter LW (eds) *Annual Review of Psychology* **32**: 89–115.

Rumelhart D (1989) The architecture of mind: a connectionist approach. In: Posner MI (ed.) *Foundations of Cognitive Science*, pp. 133–159. Cambridge, MA: MIT Press.

Wisniewski EJ (in press) Concepts and categorization. In: Medin DL (ed.) *The Steven's Handbook of Experimental Psychology*. New York, NY: John Wiley and Sons.