Learning From Examples: The Effect of Different Conceptual Roles

Edward J. Wisniewski
Psychology Department
University of Illinois

Many studies of category learning have emphasized a single role of the concept that is learned --namely, the concept as a mechanism for classifying objects and discriminating them from members of other categories. Recently, researchers have noted that concepts have many purposes besides classification--prediction, communication, explanation, goal attainment, and so on. This paper presents a study that varied the roles of concepts during a classification learning task. Specifically, one group of subjects (the discrimination group) was given standard instructions to learn about pairs of categories. A second group of subjects (the goal group) was given these instructions but also informed about the functions of the categories. The results of the study suggest that the two groups formed different concepts, even though they saw the same examples of the categories. The concepts of the discrimination group were based on those features in the examples that had predictive value--features with high cue and category validity. In contrast, the concepts of the goal group were based on predictive features and features that were important to the function of the category (called "core" features). Relative to the discrimination group, the goal group placed less emphasis on predictiveness. The results are discussed in terms of their implications for standard classification tasks in psychology and explanation-based and similarity-based approaches in machine learning.

INTRODUCTION

In many category learning tasks, the experimenter presents examples of two or more categories, and subjects learn concepts that allow them to discriminate members of one category from those of others. These tasks provide insight into the general nature of people's concepts (e.g., Reed, 1972; Rosch, Simpson, & Miller, 1976; Medin, Wattenmaker, & Michalski, 1987; Nosofsky, Clark, & Shin, 1989).This paradigm is similar to the similarity-based learning paradigm in machine learning. In similarity-based learning, a program examines a number of examples of different categories and creates generalized descriptions (concepts) of those categories. The descriptions enable the program to identify new category members (see Dietterich & Michalski, 1983; Fisher & Langley, 1985, for overviews of these systems).

There are at least two problems with this approach, however. First, it focuses people and programs on forming concepts that emphasize only one of many roles that concepts can have (i.e., classification or discrimination). The importance of the concept is for accurately classifying category members. However, concepts must represent information about a category other than that used to identify its members. Otherwise, why have concepts? In the real world, people don't form concepts solely to use them to identify objects. Object classification is just one of many roles of a concept. Other roles of concepts include using them to attain goals, construct explanations, make predictions, and so on (Schank, Collins, & Hunter, 1986; Matheus, Rendell, Medin, & Goldstone, 1989). Second, the approach focuses people and programs on forming concepts that are based only on information that is explicit in the training examples (but see Stepp & Michalski, 1986). As a result, it ignores the effect of background knowledge on concept formation. In the real world, people inductively learn about categories by integrating background knowledge with the information provided in the examples. Some of this background knowledge includes people's basic, common sense theories of the world (Murphy & Medin, 1985; Medin & Wattenmaker, 1986; Murphy & Wisniewski, in press; Pazzani & Schulenburg, 1989).

In general, by emphasizing only the classification role that concepts play and by ignoring the importance of background knowledge, these learning tasks may obscure the nature of people's concepts and the processes that they use to form them. This paper examines differences in conceptual structure that result when different roles of concepts and different kinds of background knowledge are emphasized. The results of an experiment suggest that: i) people form different concepts from the same set of training examples, when different roles of the concept are emphasized, and ii) people represent information about a category other than that used to identify its members (e.g., information relevant to achieving goals).

EXPERIMENT 1

Undoubtedly, our concepts of objects contain a large amount of information about the functions of objects (Barsalou, 1982). This information is important in achieving goals. As an example, consider one's concept of washing machine and its function "to clean clothes." To achieve the goal of cleaning clothes, we know that clothes, soap, and water must be placed in the machine, it must be turned on, various dials must be set, and so on. While these features are important in achieving a particular goal, any one of the features may not be particularly predictive of category membership. For example, the feature "contains water" is a property of many things and could not be used to classify an object as a washing machine. Yet, the feature is crucial to our understanding of washing machines. This reasoning suggests that when learning about categories, providing people with their functions may induce them to search for features that are relevant to those functions and to incorporate them into their conceptual representations. Providing these functions emphasizes the goal-achievement role of a concept.

In the following experiment, different conceptual roles were emphasized by either informing or not informing subjects about the functions of the categories that they were learning about. In this study, one group of subjects was only instructed to learn the categories (the "classification" group). Because these instructions emphasize the classification role of concepts, we reasoned that subjects might look for features that identify the members of a given category and exclude those of other categories. What will make a feature important is that it have high cue and category validity. Cue validity is the probability that given a particular feature, an object belongs to a category. Category validity is the probability that given an object belonging to the category, it has the feature. So, for example, if the feature of color allows one to discriminate all members of one category from another (a feature with perfect category and cue validity), then that feature will be incorporated into a person's concept. On the other hand, subjects who are informed about the functions of the categories (the "goal" group) should also look for features that are relevant to achieving the category function. So, for example, given the function "used for underwater public transport," subjects might consider the features, "has a propeller" and "emits sonar" as important for achieving the function.

Method

Subjects. Subjects were 24 students at the University of Illinois who participated in the experiment as part of course credit.

Materials. Three category pairs were used. Each category was given a two-syllable nonsense name such as mornek or donker. Four types of features were used to construct the training examples of a category. A feature could be classified as predictive or non-predictive and core or superficial. A predictive feature was one with high cue and category validity (see definitions above). All of the examples that contained a predictive feature were members of the category (high cue validity) and 80% of the category members had the feature (high category validity). Non-predictive features occurred equally often in the examples of both categories of a given pair. Specifically, they occurred in 80% of the members of both categories. Besides being predictive or non-predictive, a feature was either core or superficial. A core feature was one that was especially relevant to the category's function. For example, the function of one category was "for killing bugs." A core feature of the category was "contains poison." A superficial feature was one that was not particularly relevant to the category's function. For example, the feature "manufactured in Florida" is not particularly relevant to the function "for killing bugs." Table 1 lists the features that were used to construct the examples of the category pairs, along with their functions. The structure of each category contained three superficial features (two predictive and one non-predictive) and two core features (one predictive and the other non-predictive). One of the superficial-nonpredictive features of a category was the core-nonpredictive feature of its contrast category (see Table 1).

The training examples that were presented to subjects consisted of all possible combinations of four of the five features from each category. In addition, two "random" features were added to each example. These features appeared equally often with examples of both categories in a pair. To make the random features, two stimulus dimensions with two values each were selected for each category pair. For example, the two dimensions selected for the donker/oostap category pair were "color" (with the values gray and blue) and "developed by" (with the values Japanese company and American company). A value from each dimension was then randomly selected for each example. There were 15 training examples (consisting of six features) from each category. Each exemplar was represented as an alphabetized list of features on an index card. Subjects in the function group also saw the function of each category typed in bold-face above each learning example.

There were 10 test examples from each category, divided into four types. The superficial-core* type contained the two superficial-predictive features of the category and the two core features of the (other) contrast category. Four test examples of this type were constructed by adding a different combination of two random features to the core and superficial features. There were two test examples of the core type, consisting of the two core features of the category and different combinations of two random features. There were two test examples of the superficial type, consisting of the two superficial-predictive features of the category and different combinations of two random features. Finally, there were two test examples of the superficial-core type. Each of these examples consisted of the three superficial features and two core features of the category and one random feature. Table 2 shows an example of each type for the mornek category. Notice that none of the test examples are ambiguous in that they all contain more predictive features of one category than the other.

Procedure. Subjects read instructions telling them that they were to learn various categories of objects by reading descriptions of individual objects. The instructions also mentioned that they would see new examples of the objects and that they would have to decide which kind of object each example was. The task was introduced by way of example. Subjects were told to imagine that they did not know what cars and airplanes were. They would be shown some examples of cars and airplanes and had to figure why the examples of car were cars and why the examples of airplane were airplanes. Then, they would see new examples and have to decide whether they were cars or airplanes. In addition, the function group was told that knowing what an object is used for (its function) is helpful in learning about that object. They were also told that they would see descriptions of each category's function.

For practice, subjects first learned two categories (unrelated to the experimental categories) to give them an idea of the difficulty of the task. Next, the experimental categories were taught. The randomized examples of the two contrast categories were presented on index cards, and subjects

Table 1. Features used to construct examples of the category pairs and their functions.

Mornek	Plapel
Function: for killing bugs	Function: for wallpapering
sprayed on plants **C-P**	sprayed on walls **C-P**
contains poison **C-NP**	contains poison **S-NP**
contains a sticky substance **S-NP**	contains a sticky substance **C-NP**
stored in a garage **S-P**	stored in a basement **S-P**
manufactured in Florida **S-P**	manufactured in Ohio **S-P**
Ardon:	**Carpel:**
Function: warns people about leaking gas	Function: used to suck up water
has a red flashing light **C-P**	has a rubber hose **C-P**
makes a loud noise **C-NP**	makes a loud noise **S-NP**
has a vacuum chamber **S-NP**	has a vacuum chamber **C-NP**
box-shaped **S-P**	barrel-shaped **S-P**
turned on by a dial **S-P**	turned on by a button **S-P**
Donker:	**Oostap:**
Function: reads printed text to people	Function: underwater public transport
emits word sounds **C-P**	emits sonar **C-P**
contains speakers **C-NP**	contains speakers **S-NP**
has a propeller **S-NP**	has a propeller **C-NP**
made of platinum **S-P**	made of steel **S-P**
costs thousands of dollars **S-P**	costs millions of dollars **S-P**

C-P: core predictive
C-NP: core non-predictive
S-P: superficial predictive
S-NP: superficial non-predictive

Table 2. Examples of the test items for the mornek category.

superficial-core*	core
stored in the garage **S-P**	contains poison **C-NP**
manufactured in Florida **S-P**	sprayed on plants **C-P**
contains a sticky substance **C-NP***	best if used within 5 years **R**
sprayed on walls **C-P***	came in a 32-ounce container **R**
best if used within 1 year **R**	
superficial	**superficial-core**
stored in a garage **S-P**	stored in a garage **S-P**
manufactured in Florida **S-P**	manufactured in Florida **S-P**
best if used within 1 year **R**	contains a sticky substance **S-NP**
came in a 16-ounce container **R**	contains poison **C-NP**
	sprayed on plants **C-P**
	best if used within 1 year **R**

S-P: superficial predictive
C-P: core predictive
C-P*: core predictive (of contrast category)
R: random
S-NP: superficial non-predictive
C-NP: core non-predictive
C-NP*: core non-predictive (of contrast category)

were given 12 sec to study each item. A timer beeped every 12 sec to mark the time. Subjects learned about three pairs of contrast categories. All possible orders of the pairs were presented and order was counterbalanced across subjects. After subjects had learned all three category pairs, they classified the 60 test examples (10 from each category), presented in random order. Each item consisted of a list of features followed by a rating scale, and each was presented on a separate sheet. The rating scale asked subjects to indicate which of the two contrast categories they thought the object was in and to indicate their confidence on a 1 to 7 scale. A 1 was the most confident value for one category of the pair and a 7 was the most confident value for the other. Subjects were instructed to use the other numbers of the scale to indicate intermediate degrees of confidence. It was assumed that certainty rating is closely related to the typicality of the example to the category (see Murphy & Wisniewski, in press). The whole experiment took about 45 minutes to complete.

Results and Discussion

The mean confidence ratings for the types of test examples are shown in Table 3.
Recall that none of the test examples were ambiguous in that they all contained more predictive features of one category than its contrast category. In the table, the higher the rating, the more confident subjects were that a test example was a member of this more predictable category. All of the statistical analyses reported are t-tests for paired comparisons.

Table 3. Confidence ratings for test examples.

	Function	Discrimination
superficial-core*	4.00	5.02
core	6.16	5.93
superficial	6.04	6.36
superficial-core	6.43	6.54

There were several important findings. First, compared to the discrimination group, the goal group was more confident that the superficial-core* items belonged to the less predictable category. These items contained only one predictive feature of this category (a core feature) compared to two predictive features of its contrast category (both superficial). The mean certainty score was 4.00 for the goal group compared to 5.02 for the discrimination group. In fact, the goal group rated all 24 of the superficial-core* items lower than the discrimination group. This difference was highly significant, $t(23) = 11.96$, $p < .001$, in the item analysis. The rating difference was also considerably larger than any other differences between item types among the two groups. Second, the discrimination group rated the superficial items higher than they rated the core items (6.36 versus 5.93). This difference was significant, $t(11) = 2.33$, $p < .05$, in the subject analysis. In contrast, the goal group rated the core items slightly higher than they rated the superficial items (6.16 versus 6.04), although this difference was not significant. Finally, the discrimination group rated the superficial items higher than the goal group (6.36 versus 6.04), a difference that was significant, $t(11) = 2.48$, $p <.05$, in the item analysis. In contrast, the goal group rated the core items slightly higher than the discrimination group (6.16 versus 5.93), although this difference was only marginally significant, $t(11) = 1.84$, $p < .10$, in the item analysis.

Taken together, the results suggest that the goal and discrimination groups formed different concepts, despite being presented with the same examples of a category. The discrimination group, given standard instructions to learn categories, formed concepts that were based on predictive features. Two results support this conclusion. First, the discrimination group rated the superficial items as better examples of their category than the core items. The superficial items contained more predictive features than the core items (two versus one). Second, given items containing two predictive features of one category and one predictive feature of its contrast category (the superficial-core* items), subjects rated the items as better members of the predictive category.

On the other hand, the goal group, provided with information about the category function, formed concepts that emphasized both predictive features and features that were relevant to the function (the core features). Clearly, the core features were important in the conceptual representations of the function group. Compared with the discrimination group, subjects in this group rated the superficial-core* items as more typical of the <u>less</u> predictable category. Presumably, subjects had represented the core features of these items as important information about the function of the less predictable category. Furthermore, the goal group rated the core items as slightly better examples of the category than the superficial items--even though the superficial items contained more predictive features and one of core features of each core item was nonpredictive.

GENERAL DISCUSSION

The strategies used by the subjects in this experiment are somewhat similar to explanation-based learning or EBL (Dejong & Mooney, 1986). When provided with category functions, people incorporated features from the training examples into their concepts that were relevant to those functions. Most likely, people used their intuitive theories of the world to determine the relevant features (see also Murphy & Medin, 1985; Medin & Wattenmaker, 1987). For example, when informed that the category members were used "for killing bugs," people may have used their basic knowledge of the world that poison kills animate things and that insects inhabit plants and often destroy them, to determine that the features "contains poison" and "sprayed on plants" are important aspects of the category members. People's use of theories is similar to an EBL program that uses its domain theory to explain why a training instance is an example of some high-level concept (often functional in nature). On the other hand, people's strategies were different from EBL in several ways. First, many EBL programs construct a <u>deductive</u> proof (explanation) for why a single example fits a high-level concept. They then inductively generalize this explanation to form a new concept. Features are not included in the new concept if they are not part of the explanation. In contrast, subjects in the goal group probably constructed <u>plausible</u> explanations (as opposed to deductive proofs) for why category members fit their functions. They also incorporated non-explanatory (but predictive) features into their concepts. Current EBL programs discard such features.

The results of this study have implications for several areas of cognitive science. In psychology, the findings suggest that the standard classification learning paradigm provides a limited view on the nature of concepts and the processes that are used to construct them (see also Schank, Collins, & Hunter, 1986). Researchers have generally focused on one purpose or role of concepts--namely, their role as representations that allow one to classify or discriminate objects.
As with the classification paradigm in psychology, many similarity-based approaches in machine learning primarily focus on this role. Yet, concepts have a number of different purposes (Matheus, et al. provide a useful taxonomy). It is important to study concept acquisition and use in contexts that highlight these purposes. In fact, different purposes of a concept may interact with its purpose as a classification mechanism. Thus, standard classification tasks may be further limited in their emphasis on the classification role to the exclusion of other roles. To take a simple example, it might be the case that some combination of simple features (i.e., a higher-order feature) achieves an object's function and that the combination is also predictive of category membership. In a standard classification task, finding such a combination might be extremely difficult, given the problem of combinatorial explosion. However, the search for such a combination can be constrained by knowing the function of the category.

REFERENCES

Dejong, G., & Mooney, R. J. (1986). Explanation-based learning: An alternative view. Machine Learning, 1, 145-176.

Fisher, D., & Langley, P. (1985). Methods of conceptual clustering and their relation to numerical taxonomy. University of California Irvine Technical Report 85-26.

Matheus, C. J., Rendell, L. R., Medin, D. L., & Goldstone, R. L. (1989).Purpose and conceptual functions: A framework for concept representation and learning in humans and machines. In The Seventh Annual Conference of the Society for the Study of Artificial Intelligence and Simulation of Behavior.

Medin, D. L., & Wattenmaker, W. D. (1987). Category cohesiveness, theories, and cognitive archeology. In U. Neisser (Ed.), Concepts and conceptual development: The ecological and intellectual factors in categorization. Cambridge: Cambridge University Press.

Medin, D. L., Wattenmaker, W. D., & Michalski, R. S. (1987). Constraints in inductive learning: An experimental study comparing human and machine performance. Cognitive Science, 11, 319-359.

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

Murphy, G. L., & Wisniewski, E. J. (in press) Feature correlations in conceptual representations. In G. Tiberghien (Ed.) Advances in Cognitive Science, vol 2, John Wiley & Sons.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 282-304.

Pazzani, M. J., & Schulenburg, D. (1989). The influence of prior theories on the ease of concept acquisition, to appear in The Proceedings of the Eleventh Annual Conference on Cognitive Science.

Reed. S. K. (1972). Pattern recognition and classification. Cognitive Psychology, 3, 382-407.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 2, 491-502.

Schank, R. C., Collins, G. C., & Hunter, L. E. (1986). Transcending inductive category formation in learning. Behavioral and Brain Sciences, 9, 639-686.

Stepp, R. E., & Michalski, R. S. (1986). Conceptual clustering: Inventing goal-oriented classifications of structured objects. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, (Eds.), Machine Learning, vol 2, Los Altos, CA: Morgan Kaufmann Publishers

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, Relational properties, and concept naturalness. Cognitive Psychology, 18, 158-194.