Environment and Goals Jointly Direct Category Acquisition

Bradley C. Love

University of Texas at Austin

ABSTRACT—*Developing categorization schemes involves discovering structures in the world that support a learner's goals. Existing models of category learning, such as exemplar and prototype models, neglect the role of goals in shaping conceptual organization. Here, a clustering approach is discussed that reflects the joint influences of the environment and goals in directing category acquisition. Clusters are a flexible representational medium that exhibits properties of exemplar, prototype, and rule-based models. Clusters reflect the natural bundles of correlated features present in our environment. The clustering model Supervised and Unsupervised Stratified Incremental Adaptive Network (SUSTAIN) operates by assuming the world has a simple structure and adding complexity (i.e., clusters) when existing clusters fail to satisfy the learner's goals and thus elicit surprise. Although simple, this operation is sufficient to address findings from numerous laboratory and cross-cultural categorization studies.*

KEYWORDS—*categories; clusters; goals; learning; stereotypes*

Judging a person as a friend or foe, a mushroom as edible or poisonous, or a sound as an *l* or *r* are examples of categorization problems. Because people never encounter the same exact stimulus twice, they must develop categorization schemes that capture the useful regularities in their environment.

One challenge for psychological research is to determine how humans acquire and represent categories. Different models simulating theories of category learning have been proposed, but they have not been sufficient to resolve the theoretical debates. For example, the relative merits of exemplar models (in which information about a category is stored as independent episodes or experiences) and prototype models (in which information about a category is stored in a summary format) are still debated, with exemplar models appearing best suited to certain data sets and prototype models for others (see Nosofsky and Zaki, 2002; Smith, 2002). The difficulty in resolving the debate may indicate that both approaches are neglecting a critical variable that modulates human performance.

In this article, I will suggest that the critical variable neglected by both prototype and exemplar approaches is the flexibility with which humans seek and identify structure in their environment and the extent to which this search for regularities is guided by the learner's goals. Neither exemplar nor prototype approaches make room for this flexibility. Irrespective of the nature of the learning problem or the learner's goals, a prototype model represents each category by a single prototype, whereas an exemplar model represents each category as the set of its members.

One alternative to prototype and exemplar representations of category information are clusters—bundles of experiences that group together. Clusters offer a more flexible way of representing information. A category represented by one cluster is a prototype model, whereas a category represented by a cluster for each example is an exemplar model. Cluster models have the ability to represent categories that fall between these two extremes as well. The challenge for a clustering account is to correctly locate human learners along the exemplar–prototype continuum according to the learning environment and the learner's goals. In this article, one such model, Supervised and Unsupervised Stratified Incremental Adaptive Network (SUSTAIN), will be discussed.

FLEXIBLE SEARCH FOR STRUCTURE

There is plenty of evidence to suggest that the key to the psychology of categorization is the flexible search for structure. Since Rosch's (e.g., Rosch & Mervis, 1975) seminal studies of natural object categories, the scholarly consensus has been that, relative to humans' perceptual and conceptual systems, the world comes in natural chunks. That is to say, rather than comprising unrelated groupings of features, the structure of things in the world consists of patterns of correlated features that

Address correspondence to Bradley C. Love, The University of Texas at Austin, 1 University Station A8000, Austin, TX 78712; e-mail: love@psy.utexas.edu.

create discontinuities or clusters. These clusters may provide the basis for cross-cultural agreement in categorization schemes (e.g., Malt, 1995).

Clustering objects in the world by their "external" similarities flexibly captures the natural chunks of our environment. Additionally, forces more "internal" to a learner, such as language, category use, and cultural habits of mind, can influence the discovery of structure. For example, Itzaj Mayan's inclusion of bats in the bird category may serve to highlight commonalities among bats and other birds for speakers of that language, whereas English may serve to highlight commonalties among bats and other mammals. Furthermore, concepts and categories serve multiple functions, and the structure dictated by one goal or function may not be the most useful for some other goal or function (Solomon, Medin, & Lynch, 1999). For example, both veterinarians and chefs have goal-directed interactions with animals; but their very different goals emphasize different aspects of animals, and this should lead to corresponding differences in how these two groups organize their knowledge. Thus, the categorization system must be able to both assimilate structure in the world and discover or even create that structure.

SUSTAIN strikes a balance between these two requirements by assuming categories have a simple structure and incrementally adding complexity as necessary to satisfy the learner's particular goals or needs. Thus, the category structures that SUSTAIN acquires are governed by both the structure of the world and the current task or goal.

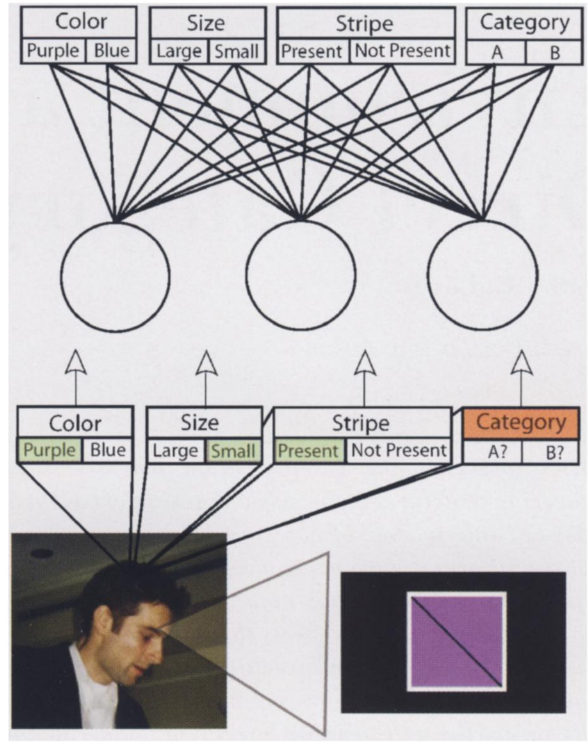

Fig. 1. The basic components of the SUSTAIN model. Information flows from the stimulus encoding at the bottom of the figure (where attention can emphasize select features of an object—e.g., color, shape) to clusters (represented by the three circles) to the output layer at the top. In this case SUSTAIN is being asked to guess the category label, A or B, for the stimulus object. The cluster that is most similar (i.e., closest) to the stimulus determines the model's response by sending a signal to the output units via differentially weighted connections.

SUSTAIN'S OPERATION

SUSTAIN simulates how humans incrementally acquire category knowledge and can be used to generate trial-by-trial predictions of behavior. SUSTAIN represents categories by one or more clusters. Clusters are internal representations that mediate between SUSTAIN's inputs and outputs (see Fig. 1). A stimulus is assigned to the cluster to which it is most similar. This cluster updates itself so that it becomes the central tendency or prototype of the items that strongly activate it. Clustering by similarity allows SUSTAIN to be sensitive to the natural chunks of information present in the environment.

SUSTAIN's cluster-discovery process is also sensitive to the goals of the learner. New clusters are created when existing clusters fail to satisfy the learner's objectives and elicit surprise. A newly created cluster is located so that it is maximally similar to the item eliciting surprise. In supervised learning (i.e., when the learner is alerted to errors) a new cluster is created in response to a prediction error, whereas in unsupervised learning a new cluster is created in response to an unfamiliar stimulus that is not similar to any existing cluster. SUSTAIN's response is based on the cluster that is most activated (i.e., most similar to the current stimulus). Learning rules update this cluster's connection weights and center it amongst its members. Attention is also adjusted so that stimulus dimensions (e.g., color, size) that are most reliable (i.e., items assigned to the same cluster tend to match on those dimensions) become the most influential in determining cluster activations. SUSTAIN's notion of similarity evolves over time according to the clusters it develops and how it allocates attention.

These simple operations are sufficient to capture the joint influence of structure and goals on conceptual organization. Different goals will lead to different goal-related activity, which in turn will influence cluster development. To illustrate this point, consider how SUSTAIN would learn to discriminate between mammals and birds. Here, the goal is simply to appropriately label these animals. After receiving numerous examples of typical mammals (e.g., horse, dog, cow) and birds (e.g., sparrow, robin, finch) SUSTAIN would develop two clusters—one to encode the mammals and one to encode the birds. These clusters capture the natural partitions of the world (i.e., the birds have more in common with each other than they do with mammals and vice versa). However, SUSTAIN's conceptual organization also reflects the goals of the learner.

Conflicts between natural partitions and the structures required to satisfy goals can be seen when an atypical object is encountered, like a bat. SUSTAIN would predict that a bat is a bird based on its properties (bats are small, have wings, and fly).

After making a prediction error, experiencing surprise, and learning that a bat is actually a mammal, SUSTAIN would create a third cluster to encode the bat. The next time SUSTAIN is asked to classify a bat or another animal similar to a bat, SUSTAIN will predict that the animal is a mammal because the third cluster will be connected to the mammal label from previous training. This example serves to illustrate how SUSTAIN's strategy of starting simple and adding complexity (i.e., new clusters) serves to reflect both the structure of the environment and the learner's goals. In some cultures, such as the Itzaj Maya, bats are considered birds and thus would cluster with birds based on feature similarity, obviating the need for an additional cluster.

The previous example largely focused on classification, which involves predicting a category label from a set of perceptual features. But SUSTAIN is also well suited for inference learning. In making an inference, the category membership of the stimulus is known, but one of its properties is unknown. For instance, a child could know that an animal is a mammal, but be unsure whether it is warm blooded. As in classification, inference predictions are based on the most activated cluster and the connections that transmit cluster activations to the output layer (see Fig. 1).

GOAL INFLUENCE ON STRUCTURE DISCOVERY

One unique aspect of structure discovery in SUSTAIN is that it is intimately coupled with the goals of the learner. Even in the laboratory, small differences in how subjects interact with stimuli can have large effects on how category information is acquired and represented. For example, inference and classification learning focus the learner on different aspects of category structures (see Markman and Ross, 2003, for a review). As previously discussed, inference learning involves predicting the value of a missing feature from the other features and a known category label. Which of the features is predicted on a given trial can vary, but on every trial the category label is unknown.

After feedback on a prediction is received, both inference and classification learning are equivalent in that the learner receives complete information about all features and category membership. Thus, any approach to category learning that seeks to discover structure solely based on the information content of categories without reference to the learner's goals or interactions with the stimuli would predict no differences between these two induction tasks. In reality, inference learning tends to promote a focus on the internal structure of each category (akin to learning about a category independently of contrasting categories), whereas classification learning promotes a focus on information that discriminates between categories. Consequently, for human learners, inference learning is best paired with intercorrelated category structures, such as natural category structures (e.g., animals that fly tend to have feathers and wings), whereas classification learning is better paired with irregular category structures that have less pronounced regularities. Comparisons of inference and classification-learning performance offer clear examples of how goal-directed activity can influence conceptualization even when the content of the categories is identical.

SUSTAIN successfully predicts these performance differences between inference and classification learning because the clusters it discovers are driven not only by the information structure of the stimuli, but also by the learner's goal-directed interactions with the stimulus set. In SUSTAIN, predictions and errors made by the learner are critical to directing cluster creation and development.

SUSTAIN's simulations of these results have yielded successful predictions. Recent work in collaboration with Yasuaki Sakamoto has demonstrated that inference learning leads to more complete knowledge of feature correlations than does classification learning, even for features not directly queried during learning. The importance of these findings for education is underscored by the fact that the participants in this study were students at a local primary school and that the stimuli were classroom materials related to a biology lesson. The results suggest that classroom exercises should stress reasoning from the category label to multiple properties.

The same characteristics that allow SUSTAIN to address differences between inference and classification learning enabled SUSTAIN to predict differences in conceptualization of folk biological categories of freshwater fish by Wisconsin expert fishermen from the majority culture and from the Menominee tribe (Love & Gureckis, 2005). Though both groups had access to the same information about local fish, the Menominees' interactions and culture stressed ecological factors, whereas majority fishermen focused on fishing-related goals (Medin et al, in press). These different goal orientations allowed SUSTAIN to predict differences in knowledge organization despite the fact that both groups had roughly the same knowledge of local fish species.

SUSTAIN captured the data by training on stimulus representations of the fish derived by scaling taxonomic distances (e.g., two species sharing the same genus had a distance of 1). These stimulus representations were intended to parallel experts' knowledge of the fish and were identical for both groups. Unsupervised learning on these representations yielded clusterings loosely organized along both ecological lines and along fishing-related lines. To reflect the central role ecology plays in Menominee culture, Menominee simulations were trained to predict fish habitat (i.e., river, lake, or either). The majority-culture simulations were trained to predict the prestige of a catch, leading to very weak organization along ecological lines and very strong fishing-goal-related organization. Our simulation results (Love and Gureckis, 2005) conform to Medin et al.'s (in press) findings. SUSTAIN's performance is driven by both the statistics of the environment (common across groups) and cultural forces (varying across groups).

COMPARISON TO OTHER APPROACHES

In addition to having parallels with both exemplar and prototype models, SUSTAIN shares a close relationship with rule or hypothesis-testing models (cf. Nosofsky, Palmeri, & McKinley, 1994). These models attempt to classify items by constructing rules and storing exceptions (e.g., "birds fly, except penguins"). SUSTAIN's ability to shift attention to rule-relevant dimensions and to store exceptions in separate clusters allows it to account for data that appear to support hypothesis-testing models. One interesting question is whether clusters with selective attention (i.e., focusing on a limited number of features) are mimicking rules or if a cluster with selective attention is what a "rule" is.

Recent studies from Sakamoto and Love (2004) favor the latter interpretation. One key difference between clusters and rules is that clusters preserve some information about the stimulus set outside the scope of relevant rules. Experiments showed that human learners are sensitive to the same information SUSTAIN picks up in its clusters.

Rational approaches to understanding category learning do not attempt to explain how people categorize items; rather they focus on why certain types of problems are harder to master based on formal analysis of the information structure of categories. Although rational analyses are certainly of use, goal influences on learning suggest that these analyses need to be expanded to incorporate other factors. For example, Anderson's (1991) rational model is a clustering model like SUSTAIN, but it cannot explain how a learner's interactions with a set of stimuli (e.g., whether the learner is engaged in classification or inference learning) shape conceptual organization.

Within a given task or goal set, SUSTAIN is in accord with rational explanations that stress the importance of complexity in determining the difficulty of acquiring a concept (cf. Feldman, 2003). Complexity in SUSTAIN is related to the number of clusters required to master a concept, with more difficult problems requiring the creation of more clusters (all else being equal). Indeed, for simple problems SUSTAIN can behave like a prototype model, whereas for difficult problems SUSTAIN can behave like an exemplar model.

In response to the difficulties exemplar, prototype, and rule-based models have had in accounting for select results, researchers have proposed multiple-system models that combine the outputs of disparate constituent models. One open question is whether a flexible-cluster approach will prove sufficient, obviating the need for multiple-systems explanations. A further challenge for the cluster approach is addressing emerging findings from the cognitive neuroscience of category learning and developing theories that bridge brain and behavior.

CONCLUSIONS

People's search for regularities in the world is governed by both the structure of their environment and their goals. To capture this joint influence of environment and goals on conceptual organization, SUSTAIN starts simple and adds complexity as needed. In particular, SUSTAIN creates new clusters to individuate events that are surprising, such as a prediction errors in supervised learning or unfamiliar stimuli in unsupervised learning. This straightforward approach is sufficient to address numerous aspects of human categorization, including those explained by exemplar, prototype, and rule-based approaches. Importantly, SUSTAIN's sensitivity to both environment and goals allows it to capture cultural influences on conceptual organization and differences in performance across different induction tasks. These same operations also address phenomena from the stereotype and schema literatures concerning how people encode experience in terms of existing conceptual structures (Sakamoto & Love, 2004).

Recommended Reading

Ahn, W., Goldstone, R.L., Love, B.C., Markman, A.B., & Wolff, P. (2005). Categorization inside and outside of the laboratory. Washington, DC: American Psychological Association.

Hahn, U., & Ramscar, M. (2001). Similarity and Categorization. New York: Oxford University Press.

Love, B.C., Medin, D.L, & Gureckis, T.M. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review*, *111*, 309–332.

Acknowledgments—This work was supported by Air Force Office of Scientific Research grant FA9550-04-1-0226 and National Science Foundation CAREER Grant 0349101 to the author.

REFERENCES

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

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, *12*, 227–232.

Love, B.C., & Gureckis, T.M. (2005). Modeling learning under the influence of culture. In W. Ahn, R.L., Goldstone, B.C., Love, A.B., Markman, & P. Wolff (Eds.), Categorization inside and outside the laboratory (pp. 229–248). Washington, DC: American Psychological Association.

Markman, A.B., & Ross, B.H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592–615.

Malt, B.C. (1995). Category coherence in cross-cultural perspective. *Cognitive Psychology*, *29*, 85–148.

Medin, D.L., Ross, N., Atran, S., Cox, D., Waukau, H.J., Coley, K., Proffitt, J.B., & Blok, S. (in press). The role of culture in the folk biology of freshwater fish. *Cognition*.

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

Nosofsky, R.M., & Zaki, S.F. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 924–940.

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

Sakamoto, Y., & Love, B.C. (2004). Schematic Influences on Category Learning and Recognition Memory. *Journal of Experimental Psychology: General*, *133*, 534–553.

Smith, J. (2002). Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science*, *13*, 437–442.

Solomon, K.O., Medin, D.L., & Lynch, E.B. (1999). Concepts do more than categorize. *Trends in Cognitive Science*, *3*, 99–105.