Psychological Bulletin
2003, Vol. 129, No. 4, 592–613

0033-2909/03/$12.00 DOI: 10.1037/0033-2909.129.4.592

Category Use and Category Learning

Arthur B. Markman
University of Texas at Austin

Brian H. Ross
University of Illinois at Urbana–Champaign

Categorization models based on laboratory research focus on a narrower range of explanatory constructs than appears necessary for explaining the structure of natural categories. This mismatch is caused by the reliance on classification as the basis of laboratory studies. Category representations are formed in the process of interacting with category members. Thus, laboratory studies must explore a range of category uses. The authors review the effects of a variety of category uses on category learning. First, there is an extensive discussion contrasting classification with a predictive inference task that is formally equivalent to classification but leads to a very different pattern of learning. Then, research on the effects of problem solving, communication, and combining inference and classification is reviewed.

A central goal of cognitive psychology is to understand how categories are learned and used. Research on categorization has explored both the structure of people's natural categories and their ability to acquire novel categories through laboratory studies. This work has generated a significant empirical record and a number of important insights into the nature of category representations. Despite this great success, there is a troubling gap between the observations derived from people's natural categories and the models that have been developed on the basis of laboratory studies. In particular, theories of categorization based on laboratory studies do not provide a compelling explanation for the variety of category types, structures, and functions observed in natural categories.

In this article, we suggest that this gap reflects important differences between laboratory methods and the way categories are normally acquired and that the way to bridge the gap is to develop laboratory methods that are analogous to the way natural categories are acquired. Specifically, people's natural categories are acquired in the course of interacting with the categories and with category members. People classify novel items, make predictions about unknown properties, solve problems with categories, make explanations based on them, communicate about them, and form preferences. Each of these tasks leaves its mark on the representation of these categories. Often, category information is learned as a by-product of interactions with the category and is not the central goal of the interaction.

Arthur B. Markman, Department of Psychology, University of Texas at Austin; Brian H. Ross, Department of Psychology, University of Illinois at Urbana–Champaign.

This work was supported by National Science Foundation (NSF) Grant 99-05013 given to Arthur B. Markman and NSF Grant 97-20304 given to Brian H. Ross. The order of authorship was determined alphabetically. Both authors contributed equally to this project. We thank Lawrence Barsalou, Seth Chin-Parker, Bradley Love, Todd Maddox, Douglas Medin, Gregory Murphy, Hunt Stilwell, and Takashi Yamauchi for helpful comments on earlier versions of this article. We also thank Michael Strahan for breaking the single-season sack record.

Correspondence concerning this article should be addressed to Arthur B. Markman, Department of Psychology, University of Texas, Austin, Texas 78712. E-mail: markman@psy.utexas.edu

In contrast to the variety of ways people interact with natural categories, participants in laboratory studies of category acquisition learn almost exclusively through classification. Furthermore, in most laboratory studies, the participants' explicit goal is to learn to classify category members. If laboratory studies of acquisition were extended to include a broader range of tasks, the resulting theories might have greater applicability to natural category structures.

To address this point, we begin with an overview of what we mean by categories. Then, we discuss existing research on a range of important uses of categories. Research on categorization must provide insight into how people represent their knowledge about categories in a way that allows these functions to be carried out. Next, we summarize the factors that have been incorporated into models of categorization derived from laboratory studies. What emerges from this comparison is that theories derived from laboratory studies fall short of explaining the range of functions for which natural categories are used.

Our main thesis is that understanding the richness of use of natural categories requires an extension of laboratory techniques for studying category acquisition. As a framework for thinking about such techniques, we present a view of category learning that is related to research on transfer-appropriate processing in memory research (e.g., C. D. Morris, Bransford, & Franks, 1977). Then, we give a detailed comparison of two empirical techniques, the inductive classification task that has been the basis of much research on categorization and a predictive inference task that is formally identical to the classification task but has been demonstrated to yield very different category representations. Following this overview, we examine two other methods of category acquisition that address additional aspects of category acquisition that the inference–classification comparison cannot address. Although research on how category use affects category learning is in its infancy, it provides a promising opportunity to bridge the gap between observations derived from studies of natural categories and laboratory studies of category acquisition.

Categorization in the World

As a working definition, categories are groups of distinct abstract or concrete items that the cognitive system treats as equivalent for some purpose. Maintaining and using equivalence classes involves mental representations that encode key aspects about category members.[1] Research on categorization focuses on the acquisition and use of these representations. In this section, we discuss research that bears on the ways people use their natural categories. This research makes clear that a critical aspect of categories that must be explained is how representations develop that enable categories to be used.

There are, of course, many other aspects of categorization that have been studied by exploring people's natural categories. For example, research on levels of abstraction in categorization was motivated by observations about people's natural categories (see, e.g., Berlin, 1972; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Furthermore, there is a tendency, when thinking about categories, to focus on simple object categories—often those that can be labeled by count nouns. Observations of natural categories make it clear that the range of categories people possess goes far beyond object categories, including categories of abstract concepts (see, e.g., Malt & Johnson, 1992), substances (see, e.g., Au, 1994), and events (see, e.g., M. W. Morris & Murphy, 1990; Rips & Conrad, 1989), as well as categories that name the roles played by elements in a scene (A. B. Markman & Stilwell, 2001; McRae, Ferretti, & Amyote, 1997).

Category Use

A fundamental observation about natural categories is that they are used for a variety of important cognitive functions. Obviously, one critical aspect of categorization is that people use categories to classify objects. Classification is the ability to determine that a new instance is a member of some known category. Although many laboratory studies involve explicit learning by classifying new instances, many other uses of categories require implicit classification of instances. For example, someone must implicitly classify a newly encountered object in a hotel bathroom as a bar of soap before using it to wash up.

A second critical function is prediction. J. R. Anderson (1990) suggested that category representations are optimized for making predictions. As a reflection of this belief, there have been many studies with children and adults on how categories can be used to predict the values of new features of an item given knowledge about the category to which it belongs. Much of this work has explored how features known to be true of one category can be carried over to other related categories (Gelman & Markman, 1986; Heit & Rubinstein, 1994; E. M. Markman, 1989; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990).

For example, Osherson et al. (1990) demonstrated that people are more likely to attribute some novel property to a member of a general class (e.g., all birds have property X) if they know only that the feature is true of a category that is prototypical of the class (e.g., robins have property X) than if they know only that the feature is true of a category that is not typical of the class (e.g., penguins have property X). Research on inferences made by experts further suggests that they override general principles such as the one observed by Osherson et al. when they have specific causal knowledge about the relationships that members of different categories enter into (Proffitt, Coley, & Medin, 2000). People are also able to use their social categories to make predictions. For example, research on stereotypes demonstrates that people predict the behavior and motivations of a new person they meet on the basis of social categories, including race and profession (see, e.g., Hirschfeld, 1996; Kunda, 1999; Sherman, Lee, Bessenoff, & Frost, 1998). Thus, a range of studies on induction have demonstrated that people are quite willing to use their categories to make predictions about the properties of newly encountered individuals.

Another important function of categories is communication. There is a complex relationship between the words of language and people's categories (Brown, 1958; Malt, Sloman, Gennari, Shi, & Wang, 1999). Brown (1958) pointed out that objects can be named at a variety of levels of abstraction (e.g., poodle, dog, animal) and that speakers must determine what label to use. Furthermore, the particular word that people prefer to use to label an item cannot be predicted with a high degree of accuracy by the labels they use for other items that they perceive to be similar to it (Malt et al., 1999). Despite these complexities, people are able to communicate successfully.

There is also a substantial amount of work on conceptual combination that has demonstrated how people use their concepts productively (Costello & Keane, 2000; Gagne, 2000; Murphy, 1988; Wisniewski, 1997). Simple conceptual combinations occur in the interpretation of adjective–noun phrases like *brown apple* (E. E. Smith, Osherson, Rips, & Keane, 1988). Even for these combinations, much category knowledge needs to be used to add properties to the combination, such as the possibility that a brown apple is rotten or that a wooden spoon is large (Medin & Shoben, 1988). More complex conceptual combinations occur when pairs of nouns are combined. Thus, a *zebra horse* might be a horse with stripes or perhaps a horse that lives near zebras (Wisniewski, 1998; Wisniewski & Love, 1998). People's existing category knowledge constrains their interpretations of these phrases.

Even people's ability to form preferences is influenced by the categories they possess. Research on consumer behavior has explored people's ability to use information about brands of products to form preferences. This work suggests that people have complex beliefs about the characteristics of brands (J. L. Aaker, 1997). Furthermore, they are able to use brands productively, extending existing brands to new products (D. A. Aaker & Keller, 1990; Broniarczyk & Alba, 1994). Extensions that are consistent with their existing beliefs (e.g., Nike squash racquets) are more likely to be viewed positively than are extensions that are inconsistent with their beliefs (e.g., Nike personal computers).

What Must Be Explained?

This review of people's ability to use categories has been selective and has focused on issues that are central to the laboratory methods to be discussed later. There are two main points that emerge from this discussion. First, people are able to use their categories for a wide range of functions. When people are asked to infer a new property of an individual or of a category, they do so on the basis of the information they have acquired about the category. Similarly, when they encounter a novel noun phrase, they are able to interpret it. These observations demonstrate that

[1] The terms category and concept are sometimes used to reflect this distinction (Murphy & Medin, 1985). Category is used to denote the set of items in the world, and concept is used to denote the mental representation that supports this grouping. Although we use these terms consistently in this article, we try to be explicit when we are referring to mental representations to avoid confusion.

Table 1
Sample Category Structures

Category A	Dimension 1	2	3	4	Category B	Dimension 1	2	3	4
Family-resemblance structure									
A1	0	0	0	1	B1	1	1	1	0
A2	0	0	1	0	B2	1	1	0	1
A3	0	1	0	0	B3	1	0	1	1
A4	1	0	0	0	B4	0	1	1	1
A0	0	0	0	0	B0	1	1	1	1
Nonlinearly separable structure									
A1	0	0	0	0	B1	1	1	1	1
A2	0	1	0	1	B2	0	1	0	0
A3	1	0	1	0	B3	1	0	1	1

Note. The category prototypes in the family-resemblance structure (A0 and B0) are not usually presented during training trials.

category knowledge is sufficiently rich and flexible to enable a variety of uses. Thus, people's category knowledge is not encapsulated so that information available in one task cannot be used to perform another.

Second, people are sensitive both to relationships among information within categories and also to distinctions between categories. Making inferences requires people to know how the properties of category members are related. These internal relationships are both statistical and conceptual. At the statistical level, knowing that an object is a member of a particular category and has particular properties changes the likelihood that the object also possesses other properties. Furthermore, people may know how some properties of a category member are causally related to other properties (Rehder & Hastie, 2001). These relationships are useful for predicting the value of missing features.

In addition, people have an understanding of what distinguishes members of one category from those of another. For example, the fact that cobras have a distinctive hood and are poisonous helps to distinguish them from snakes in general but being long and thin and having scales do not. Thus, the former two properties are diagnostic, but the latter two are not. Such diagnostic properties are important for using categories. As one example, part of interpreting a novel noun–noun combination involves highlighting properties of the first (or modifier) noun that are particularly diagnostic of that category (Costello & Keane, 2000). Thus, a *cobra chair* is more likely to be interpreted as a chair in the shape of a cobra's hood than as a chair that is long and thin.

Any theory of concept representation must explain both facets of categorization: that categories can be used for many functions and that people are sensitive both to relationships among properties within a category and also to information that differentiates among categories. Clearly, it has been possible to develop important insights about category structure on the basis of observations of people's natural categories. Nonetheless, there is a strong pressure to explore category acquisition in the laboratory to exert control over the learning setting. Unfortunately, the set of issues that has occupied much of the laboratory research on category acquisition and representation only partially overlaps with the aspects of natural categories described in this section.

Category Representations Developed in Laboratory Studies

Laboratory explorations of category acquisition generally focus on classification. There are two primary methods used in these studies: classification and sorting. In sorting studies, people are shown a set of instances (either as a group or sequentially) and are asked to sort them into a small number of groups. When faced with this method, people often adopt a simple strategy of sorting along a single dimension, perhaps classifying exceptions by their overall similarity to the groups formed (Ahn & Medin, 1992; Medin, Wattenmaker, & Hampson, 1987; Wattenmaker, 1995). Because this strategy probably reflects something about the sorting task itself rather than about mechanisms of category acquisition and use, we do not consider it further.

In classification tasks, people are shown instances sequentially and are asked to classify each instance into one of a small number of categories. During training, participants are given feedback after each trial and are expected to learn the categories by trial and error.[2] Studies of this type generally manipulate the number of dimensions that describe the categories, the category structure, and the base rates of the categories. Table 1 shows two sample category structures. The family-resemblance structure consists of categories in which each instance seen during learning (A1–A4 and B1–B4) shares three feature values with the prototype of its category (A0 and B0). The fourth feature is an *exception feature* and manifests the value typical of the other category. The nonlinearly separable (NLS) structure shown in Table 1 is designed so that there is no clear rule that distinguishes between the categories, and so, the instances cannot be classified on the basis of overall similarity.

After participants learn to classify the instances to some degree of accuracy, they are often given transfer classification in which

[2] There is a growing interest in unsupervised learning in which people are not given feedback about the categories to which items belong (Clapper & Bower, 1994; Love, 2002, in press). To date, there has not been enough research on this topic to provide a clear idea of how these results relate to those of studies of supervised learning.

new instances are presented and are classified. No feedback is usually given on transfer trials. Other transfer tasks such as typicality ratings, recognition memory judgments, and classifications of individual features may be given as well.

Models developed on the basis of people's performance on classification tasks are referred to as categorization models, suggesting that they are meant to provide insight into category representations. Many of these models have been formalized into mathematical or computational models. In this section, we draw broad conclusions about this research, but we do not focus on the details of particular models.

In essence, there are three classes of models that have been used to explain data from classification studies: prototype models, exemplar models, and rule-based models. There is considerable debate about the degree to which these models can explain the extant data, but some combination of these three approaches can explain nearly all of the classification data that has been collected to date (Maddox & Ashby, 1993; Nosofsky & Palmeri, 1997; Nosofsky, Palmeri, & McKinley, 1994; J. D. Smith & Minda, 2000, 2001).[3]

Prototype and exemplar models are similarity-based approaches. On these views, people classify each instance by virtue of its similarity to a stored category representation. In prototype models, the stored category representation is an average exemplar of the category (Posner, Boies, Eichelman, & Taylor, 1969; Reed, 1972; J. D. Smith & Minda, 2001). In exemplar models, individual exemplars are stored along with the label of the category to which they belong. New instances are compared with the stored exemplars and are categorized on the basis of their similarity to individual exemplars (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986).

In the mathematical models that implement prototype and exemplar models, the representations are typically simple, consisting of multidimensional spaces or small numbers of independent features. Similarity among exemplars consists of nearness in the space or degree of feature overlap. Features or dimensions may be weighted independently. There is little provision for relational information or complex representational structure in these models, although it is possible to incorporate information about the degree to which features co-occur by using a suitably structured similarity function (such as a multiplicative function).

According to rule-based models, people try to find some rule that allows all (or most) of the exemplars to be placed into the correct category. If there are exceptions to the rule, then those exceptions may be stored separately (Nosofsky et al., 1994; Palmeri & Nosofsky, 1995).

Only one of the two issues discussed in the previous section has been the source of significant work in laboratory studies of category acquisition. In particular, theories of categorization that are based on classification data typically posit some kind of internal category representation (e.g., exemplars or prototypes) that captures information about features of a category and distinguishes this category from other categories being learned. Rule-plus-exception models also focus on the importance of distinguishing between categories, as the rules that are formed are often focused on finding parsimonious ways to distinguish among the categories being learned.

Finally, the premise of this review of the category-learning literature is that there have traditionally been few laboratory studies that have directly addressed uses of categories other than classification. Thus, it is not clear whether the kinds of representations that people formed in these studies are sufficient for carrying out other functions. Indeed, to the extent that people were forming rules and storing exceptions in some cases, it is not clear how these representations could be used for functions like communication, predictive inference, or preference formation. In the next section, we suggest a way to address this apparent gap between natural categories and laboratory studies.

Category Use and Category Learning

Why is there a gap between the theories developed on the basis of laboratory studies and the observations of the functions and representations of natural categories? It is tempting to ascribe this difference to the sparseness of the laboratory setting in which people are asked to learn a small number of categories composed of simple objects that can be described by a limited number of dimensions. Although this simplicity may indeed lead people to use strategies that they do not outside of the lab, we think the root of the problem is much deeper.

There is a hidden assumption in most research on categorization that the classification task taps a fundamental categorization mechanism. One basis for this assumption is the observation related above that classification is a sufficiently central use of categories such that it lies at the heart of all other functions of categories. An item needs to be classified before category-related knowledge can be used to predict, explain, and so on. A second underlying assumption of laboratory classification work is that categorization involves generating mental representations to support category uses and that any task that makes use of categories involves the same mechanism. On this view, categories learned by classification can be applied straightforwardly to other category uses. This assumption is particularly important because the ability to classify new items has no utility by itself. It is helpful to know the category membership of some new item largely because category representations support more complex reasoning about an item than could be done without knowing the category to which it belongs.

We suggest that the core of the problem is that there is no single category-learning mechanism. Research on memory suggests that what people remember about some item is specific to the way they interact with that item. The concept of transfer-appropriate processing suggests that the way people interact with something affects what they recall about it later (see, e.g., C. D. Morris et al., 1977). In this section, we develop a simple but powerful transfer-appropriate processing view of categorization that can be used to explore the relationship between the way categories are used and what is learned about them.

The cognitive system is conservative (Payne, Bettman, & Johnson, 1993). That is, when people engage in a task, they are unlikely to do more than is necessary to complete that task. In the context of category learning, that means that the information that is necessary to complete a task is much more likely to be learned than is other information that is present during learning but that is not necessary for the successful completion of the task. Furthermore,

[3] There are also iterative clustering models (e.g., J. R. Anderson, 1991; Love et al., in press). These models fall somewhere between exemplar models and prototype models. They retain all of the exemplars seen during training, but they form intermediate groupings in memory.

because people are miserly in their allocation of cognitive effort, they may not learn much additional information once they have developed representations sufficient to complete a task.

This assumption has deep implications for learning. It means that to predict what people learn from a task they perform, it is necessary to analyze the task they are performing and the context in which that task is performed. Furthermore, to understand whether a given task leads to representations that are sufficient to carry out some second task, one must predict what is learned during the first task and then determine whether this information is sufficient for carrying out a second task. Although this statement may seem obvious, it contradicts a deeply held (if implicit) assumption of categorization research, namely, that any learning task, such as classification, leads to category representations that are sufficient for carrying out tasks for which people's categories are used.

The classification task that people are given typically requires them to learn to distinguish among the categories being learned. Thus, features that are diagnostic of the categories are particularly important for classifying new instances. Features that are not diagnostic, either because they are redundant with other diagnostic features or because they do not distinguish between the categories, may not be learned. The particular features of a category that are diagnostic depend on the other categories that are being learned. Relating this analysis to the aspects of natural categories described above, we suggest that the classification task is particularly focused on between-categories information and, thus, that between-categories information is most likely to be acquired by people given a classification task.

The features that are diagnostic of a category in the context of some other category in a classification task may not be the ones that are most useful for making predictive inferences about category members. In a predictive inference setting, the person must predict the value of a missing property given a category label or some of the other feature information, or both. Fortuitously, some of the missing information may be those properties that are most diagnostic. However, in the general case, some amount of the missing information is properties that would not be diagnostic for distinguishing among members of a set of categories. Thus, to carry out predictive inference successfully, it is necessary to have information about the within-category structure of the category: that is, information about how the features of the category relate to each other and to the category label.[4]

This discussion suggests that the study of category acquisition needs to focus on the relationship between the way categories are used and what is learned about them. Only by examining a range of tasks can psychologists begin to understand how category representations are acquired that are rich enough to support the variety of functions for which they are used. Furthermore, there needs to be more emphasis on situations in which categories are learned while processing information about category members because people are rarely in a situation in which their main goal is to learn a category (Brooks, 1999). Indeed, outside of school situations, people rarely have the explicit goal to learn about some item. Thus, laboratory studies need to match the world on this key variable. Furthermore, when a task is proposed, it must be analyzed to determine what information is required to perform it, to understand the degree to which that task is able to use information that is acquired in the process of carrying out other tasks.

Since the late 1990s, there has been a growing emphasis on the relationship between category use and category learning. To illustrate the utility of this approach, we begin with an extensive review of research contrasting categories learned by the standard classification task with categories learned by making predictive inferences. After this presentation, we focus on other means of learning categories to address the integration of category representations from different learning tasks.

Contrasting inference and classification has three benefits. First, there is more work on the influence of these functions on learning than on any other pair of functions. Second, at one level, the classification task and the inference task can be seen as formally equivalent. This formal equivalence is discussed in detail in the next section, but the close similarities of the tasks can be seen quickly using Table 1. In a classification trial, an item (values on all four dimensions) is presented without the category label, and the participant must decide on the label. In an inference trial, the label and the values on three dimensions are presented, and the participant must decide on the value of the fourth dimension. Despite this formal equivalence, the preceding discussion suggests that classification and inference have different information requirements. Thus, there is good reason a priori to believe that the representations that people form when performing these tasks are different.

A third reason for focusing on inference and classification is that both tasks involve a small number of categories and a small number of stimulus dimensions, and so, both tasks are amenable to mathematical modeling. Thus, the models developed on the basis of classification tasks can be extended to inference tasks to gain further insight into these tasks and to provide additional tests of the models.

After the review of research on the inference and classification tasks, we briefly discuss work relating category learning and category use in two other domains: (a) tasks that are combinations of classification and inference and (b) problem solving. These domains allow us to address another important issue relating to category use. One might object that people are not forming category representations when performing these tasks but rather are forming separate representations that are optimized for the learning tasks. The research that has been done on combinations of classification and inference and on problem solving allows us to demonstrate that the information that is acquired while performing one task is used for carrying out a second task.

Inference and Classification

A thorough investigation of the classification and inference tasks requires an explanation of the tasks, a discussion of their similarities and differences, and an investigation of the empirical and modeling results that provide evidence for similarities and differences. These are the goals of this section.

[4] In theory, it is possible to have categories that do not have a category label. Most psychological research focuses on categories that do have labels, and we do not consider the case of categories that do not have labels further in this article.

Overview, Formal Equivalence, and Possible Psychological Differences

Overview

We begin our overview with some terminology because many different labels have been applied to the same properties of categories in different studies. We assume each item (or exemplar) consists of a set of dimensions that can take on one of a number of values. The items can be divided into groups. We use the term *category label* to refer to the symbol that denotes a particular group of items and the term *category feature* to refer to a symbol that denotes the value that an item has along a particular dimension. For example, there may be a set of people who can be separated into groups with category labels such as Democrat or Republican and with dimensions such as their stand on a political issue (e.g., gun control) or their annual income, as well as values along those dimensions (e.g., pro-gun control, annual income of $72,000).

Classification refers to a task in which values for (some of) the dimensions are presented and the category label must be predicted. An example would be to predict whether a person is a Democrat or Republican on the basis of their positions on different political issues such as supporting gun control and the minimum wage. *Inference* refers to a task in which the category label and/or values of some of the category dimensions are presented and the value of another category dimension must be predicted. An example would be to predict a person's position on the minimum wage given that the person is a Democrat and supports gun control (see J. R. Anderson & Fincham, 1996, and Thomas, 1997, for other versions of inference tasks). Of course, inference may also be done with respect to categories in general, as in studies of category-based induction.

As can be seen from this description, the two tasks are quite similar. We first explore the formal similarities and then discuss some possible psychological differences. We illustrate the tasks using the abstract stimulus structure shown in the top part of Table 1, with the two categories, A and B, given in different halves of the table. Each column represents a category dimension, and the numbers (0 or 1) in the table represent a particular value for the dimension (e.g., the first dimension might be color with a 0 representing green and a 1 representing red). Each row (on each half of the table) represents a particular item, such as A1, with the item having values on each of the four category features. We refer to the values of this item's category label and category features as (A, 0, 0, 0, 1). As can be seen with this particular structure, called a family-resemblance structure (Rosch & Mervis, 1975), each category has a value on each dimension that occurs most often for members of that category, and the conjunction of those values is the category prototype given at the bottom of the table. For example, the prototype for Category A, which we call A0, consists of all values of 0 (A, 0, 0, 0, 0). Each exemplar has three values consistent with the prototype of the category to which it belongs and one value consistent with the prototype of the other category. We refer to these latter values as exception features.

Most of the experiments we consider used the standard incremental learning task described earlier. That is, one item is presented on each trial, and the learner responds with an answer, is given feedback, is given some time to study the item and answer, and proceeds to the next item. For the classification task, this is the procedure that has been used in most experimental studies; the learner is presented with values of all the category features of an item and predicts the category label. For example, the learner might be presented with all the values for Item A1 and have to predict whether it is in category A or B. We can refer to that trial as (?, 0, 0, 0, 1) with the question mark indicating information that was not presented and has to be predicted.

For the inference task, the learner is presented with the category label and the values on all but one of the other category features and predicts the missing value of that category feature. For example, the learner might be asked to predict the value for the second category dimension of A1, given the category label and values of the other dimensions. This trial can be referred to as (A, 0, ?, 0, 1).

Formal Equivalence

This framework makes clear the underlying similarities of the two tasks. The main difference is that in the classification task, the category label is predicted whereas in the inference task, a category-feature value is predicted. Some researchers have proposed that there is no qualitative difference between category labels and category features and that both of these tasks can be viewed as feature prediction (see, e.g., J. R. Anderson, 1990, 1991; though see Murphy, 1993). If the category labels, A and B, are thought of as just another category feature with values 0 and 1, respectively, the underlying similarity of the two tasks is even more striking. Item A1 would be characterized as (0, 0, 0, 0, 1), and its classification trial would be (?, 0, 0, 0, 1), whereas one inference trial would be (0, 0, ?, 0, 1). In both cases, the presented information consists of four of the binary-valued feature values, and the learner has to predict the value of the fifth dimension.[5]

Possible Psychological Differences

Although the tasks can be characterized in a way that suggests formal equivalence, there are reasons to believe that there are significant psychological differences between classification and inference. We first consider some possible differences and then, in the next section, consider empirical examinations of these differences.

To determine what might be learned in each task, we use the transfer-appropriate processing framework to consider how the items are being processed during learning. In classification learning, the goal is to learn to divide items into groups, so the focus is on between-category information. An item is presented, and the learner needs to determine the category to which it should be assigned. Under this task, there is much evidence to suggest that people tend to focus on (often a small number of) features that are diagnostic for this goal (Medin et al., 1987; Nosofsky et al., 1994; A. Tversky, 1977). There may be different strategies people use, including prototypes, exemplars, rules, or some combination such as rules plus exceptions (e.g., J. R. Anderson & Betz, 2001; Nosofsky et al., 1994). However, for all of these strategies, the

[5] As discussed below, when studies comparing inference and classification are run using the family-resemblance structure in Table 1, the prototypes (e.g., A0 and B0) are never shown during classification learning. Similarly, trials in which people have to infer an exception feature are never shown because they would also involve presentation of four values consistent with the prototype of one of the categories.

focus on diagnostic features is central. As already mentioned, in a somewhat different type of classification task, sorting, in which all the items are available and separated into groups, people often use only a single feature for dividing the items even when there is a strong family-resemblance structure (Ahn & Medin, 1992; Wattenmaker, 1995).

In inference learning, the goal is to learn the feature values of the different category members. A partial item is presented along with its category label, and the learner needs to predict the missing value. An important possible difference with classification is that when people are making inferences about classified items, they limit the information being considered to the category given and do not take into account what the other categories are like (Murphy & Ross, 1994). Rather than comparing the stimulus with all items they have seen or with rules for each of the categories, the learners can focus on the single category given to make the inference. For example, given (A, 0, ?, 0, 1), the learner can ask what a Category A item with these three category-feature values would be likely to have for the missing second-feature value. Limiting the focus in inference to the target category may have important psychological implications because the learners are trying to find aspects of categories that facilitate the prediction of missing features. Some evidence is elaborated later showing that inference learners pay particular attention to relationships among category members and often compare category members, leading them to notice commonalities such as the prototypical features in a family-resemblance structure (see, e.g., Lassaline & Murphy, 1996). The important point is that because the learners are focused on the target category that is given along with the partial item, the task may be viewed (from the learners' perspective) as figuring out what the category's members are like, that is, the internal structure of the category. The task leads people to learn what feature values are most relevant for the category, the within-category information, not which feature values are most useful for telling which category the item is in.[6]

We can also examine possible differences by considering the particular stimulus structure in the top half of Table 1. Suppose a learner is given A1 and has to classify it as A or B (?, 0, 0, 0, 1). None of the presented feature values is 100% predictive of the category. If the learner is looking for diagnostic features, it is likely that she or he will construct a simple rule with exceptions (e.g., first feature value 0 means Category A, but two exceptions in A4 and B4) or some disjunctive rule (e.g., Category A if at least two of some three features have the value 0, such as red, square, and large). It is also possible that other disjunctions are learned, such as all the exemplars.

In contrast, the inference-learning task might promote the learning of the prototypes of the two categories. Given the stimulus (A, 0, ?, 0, 1), the learner may focus on Category A and learn what the usual value for the second category feature for this category is. Thus, if the different category features are queried and the learner is given feedback, the learner may learn which are the prototypical values for the category.

The consideration of these two tasks from the perspective of the transfer-appropriate processing framework suggests how they might lead to differences in category representations. In particular, the classification task leads to an emphasis on the diagnostic features, whereas the inference learning emphasizes the internal structure. We now examine the evidence for this analysis.

Evidence for Differences Between Classification and Inference

We have suggested some ways in which the classification and inference tasks might lead to differences in the category representations, so we turn here to the empirical evidence. The first major issue we address is whether the category label is special in any way or if it is treated in qualitatively the same way as any other dimension. If the label is treated as another dimension, then classification and inference are equivalent. Next, we examine studies that have contrasted inference and classification tasks and explore the systematic differences that have been observed between them. We discuss the evidence in four sections: ease of learning, prototype effects, effects of feature diagnosticity, and the ease of abstraction from specific properties.

Is the Category Label Just Another Category Feature?

The view that a category label is just another feature of an exemplar has been argued most forcefully by J. R. Anderson (1990, 1991), who based his influential rational analysis on the assumption that the goal of categorization is to maximize the accuracy of predictive inference. In his model, internal category representations are organized in a manner that maximizes the system's ability to make predictions. Classification is simply a special case of predictive inference where the category label is the dimension that must be predicted. Anderson left open the possibility that the category might receive a different amount of attention than category labels, but the mechanisms for processing features and labels are the same.

There are two general observations that suggest that a category label might not be just a feature of an item (see Yamauchi & Markman, 2000a). First, the labels and features are linked to the category members by different relations. A category label is connected to each category member by a class-inclusion relation (e.g., this object *is a* giraffe). Category features are connected by partonomic relations (e.g., this object *has* a long neck). Second, the scope of the property is quite different. Category labels indicate the whole object, whereas category features indicate parts of objects (Miller & Johnson-Laird, 1976; B. Tversky & Hemenway, 1984). These observations are bolstered by much evidence that category labels are special. We turn to this evidence now.

Developmental research on induction. A great deal of developmental research on induction, much of it conducted by Susan A. Gelman and Ellen M. Markman (e.g., Davidson & Gelman, 1990; Gelman & Markman, 1986), has demonstrated the importance of category membership relative to other category features. Traditionally, children were thought to be limited in their classification ability by a strong reliance on perceptual similarity (Flavell, 1985). When deciding whether a property known to be true of one object was true of a second object, children were thought to focus largely on the perceptual similarity between the two objects.

[6] One way to characterize the difference between classification and inference is by the importance of cue validity (the probability that an object belongs in a category given that it has a particular feature) and category validity (the probability that an object has a particular feature given that it belongs to a category). Classification leads to the acquisition of features that have high cue validity. Inference leads to the acquisition of features that have high category validity.

In contrast to this view, Gelman and E. M. Markman (see, e.g., Davidson & Gelman, 1990; Gelman & Markman, 1986) found that preschool children preferred to base their inductions of new properties on category membership rather than perceptual similarity. In one study, the children saw pictures of two animals and were taught novel properties about them. For example, they saw a flamingo and bat and were told that the flamingo feeds its baby mashed-up food whereas the bat feeds its baby milk. Next, they saw a picture of a new animal, such as a blackbird, that was in the same category as one animal (flamingo) but much more perceptually similar to the other animal (bat), and they were asked whether this new animal (labeled as a bird) would feed its baby mashed-up food or milk. Even 4-year-olds tended to make inferences on the basis of labeled category membership rather than perceptual similarity.

Gelman and Coley (1990) observed similar effects with 2-year-old children, as long as the category label was provided. An important aspect of these results is that the inferences children made on the basis of category membership were selective—when the property was not one that ought to generalize to new category members (e.g., an animal's age), then the children did not generalize from one category to the other (see also Kalish & Gelman, 1992).

In her seminal book, E. M. Markman (1989) provided a number of arguments for the importance of category labels, focusing on the difference between using a noun rather than an adjective when referring to an item. For example, she argued that a noun conveys that the category supports more inferences; is more central to the identity of the object; is relatively enduring, stable, and permanent; and is organized into taxonomies.

The studies contrasting the importance of category labels to inferences have generally examined familiar labels, confounding any effect of the label itself with any prior knowledge effect of that label. Gelman and Heyman (1999) investigated whether the linguistic form itself (noun vs. verbal predicate) is sufficient to lead to differences in the perceived stability of a property. They told 5- and 7-year-old children about four child characters that had some unusual property (e.g., "Rose eats a lot of carrots"). This information was presented as either a noun labeling a category ("She is a carrot-eater") or as a descriptive phrase ("She eats carrots whenever she can"). Children then assessed the stability of this trait for the child character (e.g., "Will Rose eat a lot of carrots when she is grown up?"). Children of both ages thought that the property was more stable when it was presented as a category label than when it was presented as a descriptive phrase.

This evidence demonstrates that category labels are treated differently from other category features even by young children. Children typically use category membership as the basis of inference rather than compelling perceptual similarity. When the same information is presented as a label versus a verbal predicate, children view the label as indicating a more stable property.

Feature similarity versus category membership. Although there is good evidence that categories limit the information considered during inference (Malt, Ross, & Murphy, 1995; Murphy & Ross, 1994; Ross & Murphy, 1996), the difference between category labels and category features has not been the focus of much research in the adult literature. Prior research on inference has not clearly separated the influence of feature similarity from the influence of category membership. Yamauchi and Markman (2000a) examined this difference directly. The logic of their research was based on the well-documented finding that classification is determined largely by the similarity, or featural overlap, between a test stimulus and those items that have already been classified. If the category label is treated as just another feature, then inductive inference is equivalent to classification and should be based on the similarity to earlier items as well. Yamauchi and Markman tested this idea by placing category labels and category features in opposition.

We describe two of Yamauchi and Markman's (2000a) experiments in detail. The stimuli in these studies were a five-feature family-resemblance structure (see Table 1 for a four-feature family-resemblance structure) and were instantiated as pictures of cartoon bugs with dimensions of antenna, head, body, legs, and tail, along with a category label of *monek* or *plaple*. Thus, using the abstract notation from Table 1, the monek prototype was (M, 1, 1, 1, 1, 1), and the plaple prototype had all zeros, so one monek study item would be (M, 1, 1, 1, 1, 0), and one plaple would be (P, 0, 0, 1, 0, 0). The 10 items were given on one sheet of paper with the category label above the item and a line to separate the two categories. The participants were allowed to look at this sheet throughout the experiment and had to answer both classification and inference questions.

The primary manipulation in Yamauchi and Markman's (2000a) Experiment 1 was the similarity of the test item to the prototype. High-similarity items had four features in common with the prototype (and so were equivalent to the study items), medium-similarity items had three features in common with the prototype, and low-similarity items had two features in common with the prototype. For classification questions, it was assumed that the proportion of category responses (called *category-accordance responses*) would decrease with similarity. This decrease would be especially strong for the low-similarity items because they had more features in common with the prototype of the other category; for instance, a monek low-similarity item might be (M, 1, 0, 1, 0, 0), though in classification trials it would be without the label.

The inference questions were of particular interest in these studies. The high- and medium-similarity inference questions should have led to category-accordance responses, as both the category label and the feature values suggested a response in accordance with the prototype of the queried category. The two views of category labels made different predictions for the low similarity items such as (M, ?, 0, 1, 0, 0). This item had two parts in common with the monek prototype (the label and the third feature) and three parts in common with the plaple prototype (the second, fourth, and fifth features). If the label were treated simply as another feature, then there were only two out of five monek properties, so participants should have predicted a 0 value most of the time—as often as they predicted plaple in the classification condition. However, if the category label were special, then the proportion of category-accordance responses should have been much higher for low-similarity items in the inference test than it was in the classification test.

The results supported the idea that the category label is not simply another feature. Although the proportion of category responses for classification and inference were about equal for the high- and medium-similarity test items, the low-similarity items showed many more category-accordance responses for inference than for classification (0.51 vs. 0.23). That is, given a low-similarity test item, such as (M, 1, 0, 1, 0, 0), participants were much more likely to fill in a missing monek-valued feature (e.g., [M, ?, 0, 1, 0, 0]) with the monek prototype value of 1 than they were to fill in a missing category label, (?, 1, 0, 1, 0, 0), with the monek label. Notice that for these low-similarity items, even though more parts were in common with the opposite prototype in the inference test, the participants were more likely to give the feature value that was consistent with the prototype that matched the category label. This result suggests that the label was being treated differently from the other features, though it is possible that the label was simply given more than twice the weight of the other features. The next study provided a test of this idea.

Experiment 4 in Yamauchi and Markman (2000a) provided a strong test of the equivalence of labels and features by varying whether exactly the same difference was referred to by a label or by a feature (part), using a logic similar to that of Gelman and Heyman (1999). In the whole condition of this experiment, the instructions stated that there were poisonous bugs named moneks and nonpoisonous bugs named plaples. For the part condition, the instructions stated that the bugs tagged with monek (plaple) had poisonous (nonpoisonous) needles. Thus, the same feature was described by a class-inclusion relation in one condition and a partonomic relation in the other. If the category label (whole) were different from the category feature (part), then there should have been a greater proportion of category accordance responses to low-similarity items in the whole condition.

Consistent with this prediction, the whole condition showed no influence of similarity, with category-accordance responses of 0.81–0.88 for all three similarity levels, suggesting that the participants were relying only on the label. The part condition did show a drop off with decreasing similarity, but even in this condition, the low-similarity level showed category-accordance responses over 0.50, indicating that both the similarity and label (of the part) were influencing the inference. At the low level of feature match, a significantly higher proportion of category accordance responses was made when the label referred to the whole object (a mean of 0.81) than when it referred to a part (a mean of 0.60). The equivalent responses in the inference condition across the levels of similarity suggests that it is not just that the labels were weighted more than the other features but rather that the participants were relying only on the label.

The results of these studies indicate that people do not treat category labels the same way they treat other category features when making inductive inferences. The developmental research demonstrates the reliance on category membership relative to other category features, and the work with adults supports this distinction. Thus, on the basis of both representational considerations and empirical evidence (from both children and adults), category labels appear to have special importance for inference.

Other Evidence for Differences Between Classification and Inference

The observation that category labels are not treated as features raises the possibility that inference and classification involve different processes. If so, then categories acquired in the process of making inferences may be quite different from those learned by classifying new instances. As discussed above, classification requires a determination of diagnostic properties, between-category information, whereas inference requires knowing the relationship between the category membership of an item and the feature values of that item, within-category information.

This difference between inference and classification can be explored by examining people's performance with different category structures. If classification leads to finding diagnostic dimensions whereas inference leads to finding relationships between feature values and the category label, then category structures that affect the ease of finding diagnostic dimensions or prototypical feature values should differentially affect performance in classification and inference tasks. This general prediction has been explored using data on ease of learning, prototype effects, effects of feature diagnosticity, and the influence of feature variation.

Ease of learning. In this section, we consider how the two types of learning might be differentially affected by differences in category structure. Once again, the top half of Table 1 shows a family-resemblance category structure. This category structure is linearly separable because if the dimension values were drawn out in a multidimensional space, a straight cut through the space could be used to separate the members of the two categories. Classification is not easy for this category structure. As discussed above, it is necessary to use at least three of the dimensions to correctly classify all of the exemplars. Any simple unidimensional rule (e.g., for Category A, Dimension 1 has value 0) correctly classifies only 75% of the exemplars. However, inference learning with a family-resemblance structure should be relatively easy. Within each category, the label is strongly associated with one value on each dimension, so it is not difficult to learn the relationship between the category label and the individual feature values. This analysis suggests that a classification-learning task with this linearly separable structure would be more difficult to perform than would an inference-learning task.

However, the opposite prediction may be made for a comparable NLS structure, like the one shown at the bottom of Table 1. For this structure, the individual exemplars within a category have very little similarity. For example, Exemplars A2 and A3 are both in Category A, but they do not have any feature overlap at all. Classification research demonstrates that this structure can be learned by memorizing exemplars (Medin & Schaffer, 1978; Nosofsky, 1986). Even if people are using rules and exceptions or other disjunctive rules, the linear separability of a structure does not have a large effect on classification learning (Medin & Schwanenflugel, 1981; J. D. Smith, Murray, & Minda, 1997; though see Blair & Homa, 2001). In contrast, inference learners should have significant difficulty with NLS structures because there is no simple relationship between the category label and the values of the features. Thus, the prediction now is that the inference learners would have more difficulty learning NLS structures than would classification learners.

Taken together, these analyses suggest an interaction between the type of learning and the type of category structure: For linearly separable categories, there may be an advantage for inference learning over classification learning, but the advantage reverses for NLS category structures. As is explained below, the absolute levels of performance can be affected by particular factors in each learning paradigm, but the claim here is that the performance in the two paradigms differs greatly with the two category structures.

Although there is no experiment that has explicitly contrasted category structures for these two types of learning, there are experiments with each type of category structure that have looked at the two types of learning. We begin with the experiments contrasting classification and inference learning using linearly separable category structures, such as Yamauchi and Markman

(1998). Classification learners required more blocks of trials (a mean of 13.0 in Experiment 1) to reach a learning criterion of 90% across three consecutive blocks than did inference learners (a mean of 8.5). Using the same structure, A. L. Anderson, Ross, and Chin-Parker (2002) also found a significant advantage for the inference learners (10.6 vs. 13.0 in their Experiment 1, though these learning data were not reported in the article).[7]

The prediction that the inference learners should have more difficulty learning NLS structures than should classification learners was tested by Yamauchi, Love, and Markman (2002), who used slightly different NLS structures than the one shown in Table 1. In both of their studies, inference learners required more blocks to reach a learning criterion than did classification learners (19.9 vs. 12.1 blocks in their Experiment 1; 27.7 vs. 10.4 in their Experiment 2).

Thus, inference learning and classification learning are differentially affected by changes in the category structure. Inference learning, because it focuses people on within-category information, is greatly affected by the extent of the relationship among feature values and on the number of different values to be learned. A family-resemblance category is relatively easy to learn because each item is similar to the underlying prototype, but NLS structures are difficult to learn because each item may be quite different from the prototype and comparisons between category items will yield little overlap. In contrast, classification learning, because it focuses on diagnostic properties, is affected by the number of diagnostic dimensions or specific exemplars that must be considered to learn to classify instances but is less affected than is inference learning by whether the category structure is linearly separable.

Prototype effects from inference learning. Data on ease of learning focus primarily on what sort of information is most useful for classifying instances or making inferences. The broader claim of the transfer-appropriate processing framework, however, is that the information that is most useful for carrying out a task is also most salient in the representation of the category formed by carrying out that task. On this view, categories learned by classification contain information about diagnostic features or exemplars. In contrast, categories learned by inference encode the prototypical values on each dimension. In the next two sections, we examine evidence for these claims, beginning with the prototype effect.

There are three types of evidence suggesting that inference learners acquire the internal structure of the category, such as the prototype: single-feature tests, effects on sorting, and within-category correlations. The first line of evidence comes from a direct test of the individual features of category members (A. L. Anderson et al., 2002). A. L. Anderson et al. (2002) reasoned that if inference learning led to the acquisition of the prototype, then when inference learners were asked to classify single features, they would be more likely to choose the appropriate category than would the classification learners. Thus, the prediction was that inference learners would do better on single-feature classification tests, whereas classification learners would do better on full-feature classification tests. This method also avoided a problem of some earlier findings that inference learners performed better on inference tests and classification learners better on classification tests (see, e.g., Yamauchi & Markman, 1998, Experiment 1). Neither group classified single features during learning.

A. L. Anderson et al.'s (2002) results supported this prediction. For example, in their Experiment 1, inference learners classified single features significantly better than did classification learners (0.84 vs. 0.72) but did worse on the full-feature classification tests (0.62 vs. 0.84). The same interaction was found in their Experiment 2 with a five-dimensional structure and a slightly different set of study items (0.92 vs. 0.70 on the single-feature test; 0.76 vs. 0.88 on the full-feature test).[8]

A second line of evidence that inference learning leads to a prototype representation comes from sorting data. Although it is a common assumption that categories are organized by family-resemblance structure, when participants are asked to sort items, they almost always use a single dimension and ignore any family-resemblance structure (see the review in Lassaline & Murphy, 1996). For example, if they were given the eight nonprototype items in the top of Table 1, A1–A4 and B1–B4, they might sort by the first dimension, putting A4 and B4 with items from the other category. Lassaline and Murphy (1996) argued that a main advantage of family-resemblance structures is that they support inductive inference, but the sorting tasks never make use of this advantage. This view predicts that if learners first had to perform inductive inferences on the items, their subsequent sortings would be more likely to be family-resemblance sorts.

Lassaline and Murphy (1996) tested this prediction using a stimulus structure that was very different from the one in Table 1 to clearly separate family-resemblance and single-dimension sorts (see their article for details). Before sorting, half of the participants made an inference about the relation between properties to promote learning about the internal structure (e.g., for the vehicle category, "If the vehicle has bench seats, what kind of top does it have?"), while the other half judged how often a particular property appeared (e.g., "How many vehicles have a nonconvertible top?"). The questions were constructed so that the two groups had to examine exactly the same properties to provide correct answers. In the subsequent sorting task, participants who were given inference questions made far more family-resemblance sorts than did participants asked the frequency questions (0.54 vs. 0.17 in Experiment 1; 0.54 vs. 0.15 in Experiment 2) and far fewer single-dimension sorts (0.21 vs. 0.50 in Experiment 1; 0.38 vs. 0.69 in Experiment 2). These data suggest that predictive inference in-

[7] The point of this section is to argue that the different category structures have different effects on the two types of learning, but we acknowledge that inference learning need not be easier for all linearly separable category structures. Two factors that can greatly increase the difficulty of inference learning (often with less effect on classification learning) are the number of dimensions and the number of values per dimension. Both of these factors increase the amount that must be learned by inference learners substantially more than the amount that must be learned by classification learners. With sufficient increases in the number of dimensions or dimension values, inference learning can become more difficult than classification learning (A. L. Anderson et al., 2002; Chin-Parker & Ross, 2002a).

[8] One other piece of evidence that is consistent with the different representations is that A. L. Anderson et al.'s (2002) Experiment 2 manipulated whether some items were seen at study or only at test. Consistent with the exemplar view, the classification learners were better able to classify old exemplars than new ones that were equally close to the prototype, 0.88 versus 0.81. In contrast, inference learners showed no effect of old–new, 0.73 versus 0.72.

Table 2
Chin-Parker and Ross (2002b) Category Structure

Category A	Dimension 1	2	3	4	5	Category B	Dimension 1	2	3	4	5
A1	0	0	0	0	1	B1	0	0	1	1	0
A2	0	0	0	1	0	B2	0	0	1	0	1
A3	0	0	1	0	0	B3	0	0	0	1	1
A4	0	1	0	0	0	B4	0	1	1	1	1
A5	1	0	0	0	0	B5	1	0	1	1	1
A0	0	0	0	0	0	B0	0	0	1	1	1

Note. The category prototypes (A0 and B0) were not presented during learning.

creased people's sensitivity to the family-resemblance structure of the categories.

The third and final type of evidence that inference learning leads to learning the internal structure of the category examines within-category correlations. These are relations between feature values that do not add to the diagnosticity of the category beyond the predictiveness of the feature values but do indicate structure within the category in terms of which feature values tend to co-occur. For example, the presence of handlebars (whether straight or dropped) and the presence of bicycle-sized tires (whether knobby or slick) can be used to classify bicycles. However, these values are often highly correlated. Mountain bikes have both straight handlebars and knobby tires, whereas road bikes have both dropped handlebars and slick tires. An understanding of the internal structure of the category would require sensitivity to this correlation, but one could successfully classify instances on the basis of the value of either dimension alone.[9]

Chin-Parker and Ross (2002a) examined whether these within-category correlations are learned during inference learning and classification learning. The materials were verbal descriptions of employee files, where each employee had been assigned to work on a particular project. The critical aspect of the design was that two of the four-valued features had values that were perfectly correlated within each category (e.g., the values of education, such as math degree, would always occur with the same values of experience, such as sales experience). For classification learning, an item was presented, and the participant had to respond with the appropriate project. For inference learning, the project and four of the five feature values were presented, and the participant had to choose a value for the missing feature from a category-appropriate choice and a category-inappropriate choice.

There were a variety of tests that all showed roughly the same effects, but the prediction test may be the easiest to understand. For this test, the participant was shown a value for one feature (e.g., math degree) and asked to choose between two values that always occurred with the same category (e.g., sales and advertising experience), one of which had always occurred with math degree and one of which never had. Classification learners were no better than chance, whereas inference learners chose correctly far above chance (71% in Experiment 1; 81% in Experiment 2). These findings suggest that inference learning better supports acquisition of the internal structure of the category than does classification learning.[10]

The three different types of results in this section all suggest that inference learning leads to sensitivity to the internal structure of the category, such as the prototype. First, different category structures (linearly separable and NLS) differentially affected classification and inference learning. Second, inference learners were better able to classify single features than were classification learners. Third, inference learners were able to learn within-category correlations, whereas classification learners were not. In the next section, we turn to evidence that classification learning focuses on the information that is most diagnostic for classifying new instances.

Effects of feature diagnosticity. Two sets of studies have demonstrated that classification learning focuses on diagnostic features. Chin-Parker and Ross (2002b) examined the importance of feature diagnosticity in classification and inference learning with linearly separable categories. Table 2 shows the design used in their two experiments. In both cases, the five-featured bug-like stimuli were in categories with overlapping prototypes, (A, 0, 0, 0, 0, 0) and (B, 0, 0, 1, 1, 1), and there was a family-resemblance structure with all study items having one feature changed from their prototype. If classification were sensitive to the diagnosticity of the features, then the first two dimensions should not have been given much weight because they are not at all diagnostic of the category. If inference learning were not very sensitive to the diagnosticity of features (because it leads to a focus on the internal structure of the category), then the first two features should have been given as much weight as the other features.

In Chin-Parker and Ross's (2002b) Experiment 2, following learning, the participants were shown a number of test items for each category and asked to rate how typical each was of that category. The items varied in their overlap with the prototype and also varied in whether the overlap was of diagnostic features or nondiagnostic features. For example, an item with three features that overlapped with the prototype of A could have the overlap be totally on the diagnostic features—(A, 1, 1, 0, 0, 0), two of the

[9] Although the correlation does provide information about the internal structure, it does not provide additional information to improve the accuracy of the classification. Within-category correlations often signal that there are subcategories, so they are useful for prediction and can be useful for classification at a lower category level.

[10] J. R. Anderson and Fincham (1996) demonstrated that an inference-like learning procedure led to sensitivity to within-category correlations. Murphy and Wisniewski (1989) found evidence that classification learning is insensitive to within-category correlations, though Thomas (1997) obtained a somewhat different result.

diagnostic features and one of the nondiagnostic features—(A, 1, 0, 0, 0, 1), or one diagnostic and two nondiagnostic features—(A, 0, 0, 0, 1, 1). The results for this experiment are given in Table 3, broken down by the diagnostic overlap between the test item and the prototype (expected to be important for classification learning) and the total overlap between the test item and the prototype (expected to be important for inference learning).

As shown in the top half of Table 3, for classification learners, the main influence on ratings was the number of diagnostic features that overlapped with the prototype, with the number of overall features overlapping making no difference at all. That is, within each row, for a given number of diagnostic features, there is very little effect of the total overlap. Thus, classification learners were very sensitive to the diagnosticity of features. In contrast, for inference learners, the typicality ratings were influenced almost equally by the number of diagnostic and nondiagnostic features, as one can see effects within each row of the total feature overlap and smaller effects within each column of diagnostic feature overlap. Chin-Parker and Ross's (2002b) Experiment 1 showed a similar effect in a two-alternative forced choice of which item was a better member of a category—classification learners were very sensitive to the diagnosticity of the features, whereas inference learners were not.

Yamauchi et al. (2002, Experiment 2) obtained a similar result using NLS categories, shown in Table 4, with geometric stimuli and the prototypes (A, 0, 0, 0, 0) and (B, 0, 0, 1, 1). As can be seen there, the NLS structure was a difficult one, with A2 being the prototype of Category B. Of most relevance here is the modeling that gave weights to the different features (see Yamauchi et al., 2002, for details). For the classification learners, the two diagnostic features were weighted marginally more than the two nondiagnostic features, 0.86 versus 0.80, whereas the inference learners weighted the features equally, 0.70 versus 0.70.

The pattern from these studies is clear and consistent with the predictions for the transfer-appropriate processing framework. Inference learning leads people to learn the internal structure of the category, with very little influence of the diagnosticity of the features for distinguishing which category the item is in. Classification learning leads people to learn those features that help to distinguish between the categories.

Table 3
Mean Typicality Ratings (on 1 [Low]–7 [High] Scale) for Chin-Parker and Ross (2002b, Experiment 2) in Terms of Total Overlap With Prototype and Diagnostic Overlap

Number of diagnostic features overlap	Number of total features overlap: 3	4	5
Classification learning			
1	2.93		
2	4.51	4.39	
3	5.88	5.98	5.92
Inference learning			
1	3.66		
2	4.07	4.98	
3	4.63	5.24	6.06

Note. Total overlap of 4 were study items; total overlap of 5 were prototypes.

The influence of feature variation. Features vary. Thus, even if a feature value is described by a particular label, it need not always manifest itself in the same way. Not only do the four legs of dogs differ from those of elephants but the legs of a German shepherd look rather different from those of a spaniel, and people are sensitive to such variation (Solomon & Barsalou, 2001). Although real-world features obviously vary, experimental-world features usually do not. Most of the category-learning experiments have had features that were always instantiated in exactly the same way for every item. For example, the round head of a bug-like item is always exactly the same roundness and size. It is clearly of some interest to examine how people learn experimental categories when there is variation in the instantiation of features because people seem to be able to learn real-world categories with this variation.

This variation is also of interest for this article because inference learning may be much less affected by this variation than is classification learning. This prediction requires some explanation. In inference learning, the label of the category is given along with the partial item. The view presented in this article is that the label is special and leads to focusing on the representation for that category. The value of the missing feature is determined by comparing the partial item with the category representation (whether it be prototype- or exemplar-based). Because the representation of only one category is involved in the comparison, it is much easier to learn the structure of the category and to notice underlying commonalities. For example, suppose that a bug is labeled as a monek and has an irregularly rounded head but is missing a tail (whose value has to be predicted by choosing between a long, bushy tail and a short, thin tail). The learner might compare this bug with the monek representations from previous trials and note that other moneks have had round heads (albeit not quite the same exact roundedness). In addition, the learner might note that most moneks have had longish tails and so choose the long one even if it is the first long, bushy tail he or she has seen. As this new information is added to the category representation, it would tend to reinforce those feature values consistent across the items (such as round-headed and long-tailed) and not reinforce those values specific to each instantiation (such as the bushiness of the long tail). Thus, inference learning would tend to promote a category representation that includes the commonalities among category exemplars even with the feature values instantiated in multiple ways.

Classification learners, in contrast, are faced with a more difficult task as the feature instantiations vary. They cannot focus on a single category because they do not know to which category the presented item belongs. Thus, they must compare the new item with their knowledge of both categories and try to find some way of distinguishing the categories. With multiple instantiations of the features, comparisons with earlier exemplars (or prototypes) would lead to many mismatches compared with when there are single instantiations. For example, a round-head bug with long, bushy tail would have some mismatches with an earlier monek with a round head and long tail if the head were a different round shape and the long tail had been thin. Even finding rules is more difficult, because it is not clear what aspect of the feature one should be considering. For example, if the tails vary in length, width, and

Table 4
Nonlinearly Separable Category Structure From Yamauchi, Love, and Markman (2002, Experiment 2)

	Dimension					Dimension			
Category A	1	2	3	4	Category B	1	2	3	4
A1	0	0	0	0	B1	0	0	1	0
A2	0	0	1	1	B2	1	0	0	1
A3	1	1	0	0	B3	0	1	1	1
A0	0	0	0	0	B0	0	0	1	1

Note. The category prototypes (A0 and B0) were not presented during training trials.

texture, how can one decide that tail length might be useful in the rule?

Two types of results support this theoretical analysis. Yamauchi and Markman (2000b) examined the influence of feature-instantiation variation. In their Experiment 2, which provided the most direct test, the categories to be learned were stylized bugs with the now-familiar family-resemblance structure given in the top of Table 1. In earlier work (Yamauchi & Markman, 1998, Experiment 1), with the usual single instantiation of features, inference learning was easier than classification learning, but 22 of 24 inference participants and 23 of 24 classification participants learned the items to a 90% learning criterion within 30 blocks.

In Yamauchi and Markman (2000b, Experiment 2), however, when each feature value, such as round head, had four different instantiations, the learning differences were huge. Whereas 17 of 24 inference learners reached the learning criterion, only 3 of 24 classification learners did. Thus, although the feature variation does make inference learning a little more difficult, it has a large detrimental effect on classification learning. (Yamauchi & Markman, 2000b, Experiment 1, showed a similar pattern, but the inference and classification conditions had other differences as well.)

A second, less direct piece of evidence for the differential influence of feature variation is found in Rehder and Ross (2001). This research examined *abstract coherent categories,* categories that are defined by systems of relations that interconnect the features of category members without specifying what the specific values of the categories may be. For example, pollution-cleaning devices all have pollutants being cleaned in locations in which those pollutants are found with instruments suitable for removing those pollutants (e.g., sponges for removing oil from the ocean) though the pollutants, locations, and instruments vary across the devices.

Across two experiments, the results indicated a substantial difference in inference and classification learning when the features varied as they do in these abstract coherent categories. When asked to learn the pollution-cleaning device category by inference learning (participants were not told what the underlying nature of the category was, but the items were all called morkels), the learning was trivially easy. Learners were able to infer the appropriate feature value for the coherent morkel category (where the features were related as pollution-cleaning devices) with an average of one error. However, if the same features were rearranged across items so that they did not make a coherent device, then learning was much more difficult, with participants averaging 17.5 errors and 25% of them not reaching the learning criterion. In contrast, when the task was to learn to distinguish morkels from nonmorkels (the incoherent ones with the same features rearranged), the learning was much more difficult, with now 25% of the participants in the coherent condition unable to reach learning criterion. We realize that there are many differences between these two experiments that might have influenced performance, but the large inference/classification difference with these categories defined by an abstract underlying similarity is consistent with the claim that inference provides better support for learning abstract feature values than does classification.

These investigations suggest a major difference in the ability of inference and classification learning to deal with feature variation. Because inference learning focuses on a single category, the effect of feature variation is not very large—comparisons within a category help to bring out the underlying similarities. In contrast, classification learning, because it involves a comparison across all the different items, is much more affected by variations in how features are instantiated.

Summary of evidence. In this section, we have presented evidence that provides support for our processing-framework analysis of classification and inference learning. Classification learning focuses on the diagnostic features, between-categories information, and these are emphasized in the category representation. Inference learning focuses on the prototypical features, the internal within-category information, and these are emphasized in the category representation. This perspective is consistent with the extant evidence that category labels are treated differently from other category features. In addition, this perspective led to a number of predictions concerning the ease of learning (as a function of linearly separable or NLS structure), prototype effects, effects of feature diagnosticity, and the influence of feature variation, and the results are consistent with the predictions.

Implications for Mathematical (Similarity-Based) Models

An important aspect of research on classification is its synergy with mathematical models of categorization. Although there is not yet much work examining model fits of inference tasks, we can briefly review two investigations. Existing classification models do not provide good fits to inference data, but the fits they do provide are instructive.

Yamauchi and Markman (1998) fit two well-known classification models, the *context model* and the *rational model,* to classification and inference transfer data from participants who learned the categories by classification, inference, and a mixture of the two. The context model (Medin & Schaffer, 1978; Nosofsky, 1986) is an exemplar model in which an item is classified into a category when it is similar to the exemplars of that category. The rational model (J. R. Anderson, 1990, 1991) assumes that people form categories to maximize the accuracy of feature prediction by setting up internal partitions (or clusters) of items that are then used to make the feature predictions. In fitting these models, several assumptions were made (see Yamauchi & Markman, 1998, pp. 138–141, for details), one of which was that the category label was treated as just another feature. Although we have already presented evidence that the category label may be special, there is no other obvious way to extend the context model to model an inference task. Furthermore, the rational model explicitly assumes that labels are processed by the same mechanisms as features.

There are two main results of the fit to Yamauchi and Markman's (1998) Experiment 1 (similar results were obtained for their Experiment 2). First, both models did a good job fitting classification transfer for all three learning conditions. Second, neither of the models gave a satisfying account of the inference transfer data (regardless of the learning condition).

The rational model was able to provide a reasonable fit to the data, but the solution it obtained to fit the data contradicted a core assumption of the model's underlying theoretical framework. In particular, the best fitting solution created a separate cluster for each exemplar. This solution is strange, given that the basis of the rational model is that clusters are formed to maximize inductive potential. That is, to account for prediction data, the model must essentially treat each exemplar as independent.

The fits of the context model to the inference transfer data were also poorer than those for the classification data. In particular, the model tended to underpredict performance on inferences of the exception features. The parameters from these model fits are illuminating. The attention weights to all of the features were close to zero except for the weight to the category label. Thus, the context model suggested that people rely almost exclusively on the category label when making inferences. This conclusion is consistent with the idea that category labels are special. However, by focusing exclusively on the category label, the model predicted that people infer the feature value most strongly consistent with the prototype of the category on inference trials. Thus, these fits do not capture any sensitivity people have to the exception features.

Although the context model and the rational model have difficulty accounting for the inference data, the SUSTAIN model (Love, Medin, & Gareckis, in press) does a better job. SUSTAIN is a clustering model that adaptively modifies its architecture during learning. Like the rational model, SUSTAIN represents categories as one or more clusters, but clusters are added when an item has been misclassified (or a feature mispredicted). The model also has attention weights that enable it to focus selectively on dimensions that are diagnostic for the task being performed. The model forms clusters iteratively, so that ease of learning can be modeled.

Love, Markman, and Yamauchi (2000) applied this model to the Yamauchi and Markman (1998) data we just considered as well as to the Yamauchi et al. (2002) data with NLS categories. The qualitative fits generally support the story we have been trying to tell. For the family-resemblance category structure, the model fits led to inference learning being predicted to be faster than classification learning. When fitting inference learning, the model required only one cluster for each category on over 80% of the simulations. This cluster reflected the underlying category prototype, and the model's attention was spread evenly across the dimensions, with a higher weight given to the category label. For classification learning, the modal solution involved three clusters per category, suggesting that the model formed imperfect rules and then stored exceptions to these rules separately.

For the NLS structure given in Table 4, very different results were obtained. In this case, SUSTAIN predicted that classification is easier to learn than inference. The fit for classification learning led to three clusters per category (remember there were three items per category), indicating that each item was memorized. Inference learning led to a very complex clustering, suggesting that trying to use the prototype for predictions led to many failures. The modal solution had nine clusters per category, so that the same exemplar was actually included in more than one cluster. These clusters were relevant to inferences of different dimensions. Thus, the explanation from the fits of SUSTAIN is generally consistent with the idea that inference learning promotes learning the internal structure of the category, which is a prototype for the family-resemblance structure but is difficult to determine for the poorly structured NLS categories. Classification learning promotes learning disjunctive representations (rule plus exceptions or exemplars) for both types of category structures.

The simplicity of the inference-learning task makes it very amenable to mathematical modeling. Nonetheless, current classification models are not able to account for inference learning very well. Future research must address how best to accommodate inference learning and whether it might be best to try to construct more general models of category learning that could learn by both classification and inference learning (such as SUSTAIN). It would be useful to consider models that involve mechanisms for attending both to within-category and between-categories structure, as these two sources of information appear to be differentially emphasized by the inference and classification tasks. These models also need to be applied to both the large set of classification results and the other work on category labels and inference that has been reviewed here.

Conclusions About Inference and Classification Learning

Classification and inference are formally identical, and yet, the evidence reviewed in this section indicates some important differences, consistent with our transfer-appropriate processing view of category learning. We began with an analysis of the information required to perform inference and classification. Inference was expected to focus primarily on within-category relationships, whereas classification was expected to focus on properties that were useful for distinguishing between the categories.

We begin this summary section with a discussion of inference learning. Although much is known about classification learning from the last 25 years of research, psychologists' understanding of inference learning is more limited. There are three general themes that emerge from our review of the existing work.

First, inference learning promotes the learning of prototype representations. The focus on prototypes emerges from two factors. For one, people appear to organize their knowledge of a category around the category label. For another, predictive inferences are well supported by a category representation that encodes information about the most typical relationships among features within a category. When the category cannot be represented well by a single prototype, inference learning still leads to recovery of some underlying structure when possible. For example, in the within-category correlation experiments, inference learning promoted learning the correlations of feature values that consistently occurred together (akin to having multiple prototypes for the category).

This focus on within-category representations leads to the second general conclusion about inference, which is that it is not very sensitive to the contrast categories. The evidence suggests that inference learning focuses on the internal structure of the category. The diagnosticity of the features is not strongly incorporated into the category representation. At first, this lack of sensitivity may seem a problem because people are sometimes sensitive to what other categories are being learned, but there are (at least) two reasons why it may not be a problem: (a) It is important to remember that no category is likely to be learned in exclusively one way. Although people may be learning from making inferences, they also are likely to be learning about the category by classifying items, from explaining some aspect of the item in a situation, and so forth. Thus, people can learn the features that help distinguish cats from dogs without it arising from inference learning. (b) Although inference learning does not lead people to learn which feature values distinguish the categories, people often make inferences about dimensions that have contrasting values across categories. There is some evidence that people prefer to organize categories around a common set of dimensions with contrasting values (Billman, 1996; Billman & Dávila, 2001; Sifonis & Ross, 2002). People are most likely to have to answer inference questions about the values of features on these salient dimensions. Thus, the inference questions that must be answered may lead to parallel representations across a set of contrasting categories.

The third general theme is that inference learning promotes the recognition of abstract commonalities. This idea is suggested both by the research showing it leads to learning the internal structure and by the research showing that variations in feature manifestations have little effect on ease of inference learning. Because inference learning promotes comparisons and does so within a single category, underlying commonalities are more likely to be noticed. Again, consistent with our framework, the processing of the items during learning affects the representation.

One aspect of categories that emerges from the study of inference and classification is the important role of the category labels in the category representation. The differences between the two learning tasks are largely attributable to the difference between category labels and other category features—if there were no difference, the two tasks would be equivalent. The review of cognitive and developmental research points to the clear importance of category labels in the category representations. In many tasks, people classify the item and then use knowledge from that category to decide what to do—infer the feature, solve the problem, explain the behavior. During learning, the category label is a signal that all such items are likely to share some underlying commonalities. Because of the greater inductive potential of category labels than of category features, the category label serves to organize category representations.

Finally, considering the difference between classification and prediction may be useful for studying other puzzling representational phenomena. For example, Sandhofer and Smith (2001) noted that the acquisition of size terms and the acquisition of color terms have very different developmental patterns and proposed that this may be because they are acquired in different circumstances. Color terms are learned in a manner similar to classification learning, in which a child must go from an object to a label. In contrast, size terms are learned in a manner more akin to inference learning, where children learn to find similarities and differences along a dimension by comparing across different objects.

Sandhofer and Smith (2001) tested this explanation by having adults learn terms for novel dimensions using either a color-analogue or size-analogue learning procedure. They found that adults given a task in which they had to match labels with values of single objects showed a learning pattern analogous to children's learning of color terms and that adults given a task in which they compared across items during learning showed a pattern analogous to children's learning of size terms. Thus, the sorts of task differences reviewed in this section may have important implications for a variety of learning phenomena that have not traditionally been conceptualized in terms of categorization. In essence, the work by Sandhofer and Smith is an example of the value of doing a task analysis to understand the acquisition of category representations.

Other Uses, Other Questions

We have argued that to understand category acquisition, it is critical to focus on different ways in which categories are learned and used. We have focused on the distinction between inference and classification because of their formal equivalence, the amount of research on them, and their amenability to mathematical modeling. However, we now review two other aspects of category use to address one other issue. These demonstrations also provide other examples of how category use can be brought into the lab to study category acquisition.

On the basis of the studies reviewed so far, it is possible to conclude that people are not forming categories but rather are generating representations that are useful for the task they have been given. They may attempt to use some of this information when given a new task, but if given extensive practice with this new task, they would ultimately form a distinct representation optimized for it.

This conclusion seems to be antithetical to the spirit of categorization, which assumes that people are forming representations that support people's general interactions with items in the world. To address this point, we focus on two areas of research that have looked at combinations of tasks in the context of category learning. First, there are a number of studies that examine cases in which both inference and classification are given during learning. Second, a series of studies has examined how solving problems with a set of category members affects what is learned about those categories, which has a significant affect on people's subsequent ability to classify new items. To the extent that representations formed in one task influence the learning of the second, it would suggest that people are forming representations of categories rather than representations sufficient for only a single task. In addition, this examination might provide some initial understanding of how categories may be learned by a combination of tasks.

Combining Classification and Inference

Many real-world situations involve a combination of classification and inference. One might see an animal approaching, classify it as a dog, and then use that classification to decide whether to stay or flee. A physician might use the symptoms of a patient to decide on a likely disease the person may have and then use the disease and information about the patient to decide on the best treatment. In situations such as these, one first classifies the item and then uses knowledge about the category (and perhaps the item) to make a further inference. Classification is not the goal but rather is crucial because it allows the classifier to bring to bear relevant category knowledge. This use of the category is important not only because of how the category knowledge may be applied to the task but also because it may feed back and affect the category representation (and, thus, future category-related tasks).

An initial exploration of the combination of inference and classification comes from a study of order effects in the learning tasks (Yamauchi & Markman, 1998, Experiment 2), which provided results that strongly indicate that inference learners given a family-resemblance structure are acquiring prototypes. In this study, two groups of participants learned both classification and inference, but in different orders. Yamauchi and Markman (1998) reasoned that if people acquire a prototype during inference learning, then they should later be able to use this prototype to successfully classify instances. In contrast, if people acquire diagnostic features during classification, then these would not be of much help when the people later make inferences. Thus, the inference-first group should show facilitation in classification learning, whereas the classification-first group should not show facilitation in inference learning. Note that this facilitation in the inference-first group would only be obtained if people performing the second task (classification) make use of the representation constructed during the first task (inference) to the extent that it is helpful.

Yamauchi and Markman's (1998) results support these predictions, with the number of blocks needed for learning to a 90% criterion presented in Table 5. There are two important aspects to these results. First, classification learning was much faster when it followed inference learning than when it was the first learning task (7.8 vs. 12.5 blocks), suggesting a large facilitation from the inference learning. Second, inference learning was not helped by having had classification learning first and was even nonsignificantly slower (9.2 vs. 7.9 blocks).

An additional indication that inference was more likely than classification to lead to a prototype representation comes from the inference tests that followed learning in this experiment. When presented with a partial stimulus that was missing an exception feature, such as A1 (A, 0, 0, 0, ?), the inference-first learners were much more likely to choose the prototype-appropriate value, 0, even though it was contrary to the actual A1 stimulus presented during learning (0.86 vs. 0.54 for the classification-first learners).

Ross (1997, 2000) has conducted a number of experiments combining tasks that promote inference and classification. On each trial, the learner first classifies the item and then uses the classification and the item to make a further inference. After learning to both classify and make the inference, the learner's representation of the category is examined. The main issue is whether the relevance of features for the inference influences how important the features are viewed for later classification. In particular, if there are two features of equal diagnosticity and one is inference relevant and one is inference irrelevant, does the inference-relevant feature become viewed as more central for the classification? The evidence reviewed in this section suggests it does.

An example of this paradigm and results (from Ross, 1997, Experiment 1) clarifies this point. Learners were told that there were two fictional diseases, *terrigitis* and *buragamo*, and, for each disease, there were four symptoms that were perfectly predictive. Suppose that terrigitis could be predicted by fever, dizziness, abdominal pain, or itchy eyes. In addition, each disease could be treated with two fictional drug treatments, which were different for the two diseases. For example, patients with terrigitis might be treated with either *lamohillin* or *pexlophene.* On each trial, a patient's symptoms were presented, consisting of two disease-predictive symptoms (e.g., fever and dizziness) and one symptom that was not predictive of either disease (e.g., swollen tongue). The learner classified this patient as having one of the two diseases and was given feedback. Then, the learner was asked to decide which of the two disease-appropriate treatments should be given. The learner was then given feedback on this treatment choice.

Although there were four category-predictive symptoms, the design was constructed so that only half were perfectly predictive of a treatment, whereas the other half were not at all predictive of the treatment. Thus, fever might be predictive not only of terrigitis but also that terrigitis should be treated with lamohillin, whereas dizziness might be predictive of terrigitis, but given this symptom, the two drugs lamohillin or pexlophene would be used equally often. We refer to symptoms that are predictive of both the disease and treatment (e.g., fever) as *relevant-use symptoms* and symptoms that are only predictive of the disease (e.g., dizziness), as *irrelevant-use symptoms*.

After participants learned to make correct diagnoses and treatment decisions consistently, they were given tests to examine their representation, with the main interest being whether relevant-use and irrelevant-use symptoms were thought to be equally important for classification. That is, although the two types of symptoms were equally predictive of the disease category, the question was whether the relevant-use symptom, because of its importance for the treatment decision, might come to be viewed as more central to the disease.

One way to test this hypothesis is through single-symptom classification—if all that is known is that a patient has a particular symptom, which disease does one think the patient is most likely to have? Although learners were good at choosing the correct disease for the irrelevant-use symptoms (when given dizziness, they would classify the patient as having terrigitis 80% of the time), they were almost perfect for relevant-use symptoms, such as fever (96%). (Other dependent measures also showed this relevant-use symptom advantage.) The important point here is that the relevant-use symptoms, compared with the irrelevant-use symptoms, were not only viewed as more predictive of the treatment (which they were) but were also viewed as more predictive of the disease (which they were not). The category representation for the

Table 5
Number of Learning Blocks for Each Learning Task in Yamauchi and Markman (1998, Experiment 2)

Learning order	Inference	Classification	Total
Inference-first	7.9	7.8	15.7
Classification-first	9.2	12.5	21.7

Note. Each block contained eight trials. For classification-first learning order, classification learning preceded inference learning.

symptoms was a joint function of their diagnosticity for the disease classification and their importance for the treatment decision.

This effect of category use on later classifications was quite robust. Similar (though smaller) advantages were found in Ross (1997) when the classification was probabilistic (i.e., the symptoms were not perfectly predictive) and when the cover story was changed so that participants were predicting not treatments but something unrelated to the biology of the disease (to ensure that the effect was not due just to the special relation between diseases and treatments). There was a relevant-use symptom advantage when the dependent measure was generation of symptoms for each disease and even when learners were asked to judge how frequently each symptom occurred (i.e., relevant-use symptoms were judged as having been seen more during the experiment than irrelevant-use symptoms, though they were presented equally often). Ross, Gelman, and Rosengren (2002) showed similar effects with children classifying fictional creatures and making inferences about their actions.

The studies by Ross (1997, 2000) used an *interleaved* learning paradigm in which the learner makes a classification and inference decision about each item. This paradigm was chosen because it often occurs in real-world situations, but it is not the only way in which categories are learned.[11] Another common and very different means of learning, called *postclassification,* is when learners first learn to classify and only later learn to use the categories to make an inference. When the disease/symptom experiment was conducted with the postclassification learning paradigm, the results were very similar to those with the interleaved paradigm (e.g., 0.76 for irrelevant use vs. 0.90 for relevant use symptoms; Ross, 2000). This effect did not depend on learners having to continue classification during the treatment decisions—if learners were presented with the patient symptoms and told the disease (as in the inference-learning paradigm of Yamauchi & Markman, 1998), the relevant-use symptom advantage still occurred (0.58 vs. 0.80). Only when the treatment learning was performed without any mention of the disease did this advantage disappear.

In summary, results make clear that when people are learning to both classify and use the classification to make an inference, the features relevant to both the classification and the inference are viewed as more central to the classification than those features relevant only to the classification. That is, the use of the category affects the category representation, which in turn influences a variety of later category-related judgments, including classification.

We note three points from this research. First, as mentioned earlier, these findings indicate that people were not forming separate representations from the different tasks. In the disease experiments, we saw the effects not just for the interleaved learning but even for the postclassification paradigm when classification was learned before the inference task was introduced. (This paradigm is discussed further in the next section.) The Yamauchi and Markman (1998) results, with a very different procedure and materials, show how inference learning can affect later classifications. Second, these findings provide some further support for the framework because the analysis of the task was necessary to predict the effects of the two tasks. This was seen most clearly in the inference-first facilitation (and the lack of facilitation in the classification-first condition). Third, the disease studies show how the perspective that has been applied to the simple inference versus classification contrast may also apply to more complex situations in which categories are used. In addition, it should be noted that there are parallel results in real-world settings in which the uses of categories have influenced the category representation, although in the real-world settings there is no control over the type of uses made of the categories (see, e.g., Boster & Johnson, 1989; Medin, Lynch, Coley, & Atran, 1997; Proffitt et al., 2000).

Problem Solving

Research on the relationship between problem solving and category learning has also demonstrated how multiple tasks all influence the same representation. The research just discussed on combining classification and inference makes clear that judgments after classification may influence the category representation, but this research has two limitations. First, the inference judgments were very similar to classification judgments (as evidenced by our claims of formal similarity earlier). Thus, it is possible that category representations are only influenced by tasks that have a similar formal structure to the classification task. Second, the prior work showed that postclassification judgments can affect the weighting of features but did not address whether the category uses may have a wider influence on category learning. The research on problem solving addresses both of these limitations, and it supports a test of some additional issues. We begin with a demonstration of problem-solving influences on category representations and then turn to research that examines how this learning occurs and what is being learned.

Ross (1997, Experiments 6 & 7) showed the influence of category-based problem solving on the category representation, including knowledge used to classify, with both mathematical problem solving and "spy decoding" paradigms. In this section, we focus on the spy decoding task. Learners were told they were clerks in an intelligence-gathering operation and that their mission was to receive spies' coded messages, consisting of letters and numbers. Participants used the letters to figure out which spy had sent the message (classification) and then used that spy's special decoding formula on the numbers to get a preliminary decoding to pass on to the supervisor (problem solving). The decoding provided an opportunity for the learners to notice some of the relations among the numbers. The crucial question was whether these number relations would be incorporated into the category representation and used to classify later messages. Most importantly, to be described shortly in detail, the critical relations among the numbers were true of all the messages. Thus, if learners incorporated different relations into the category representations of the different spies, it must have been due to the use because only during the use were the relations treated differently for the categories.

Figure 1 provides an example of the coded messages, decoding formulas, and one of the test conditions (the results were consistent across the test conditions). The learner received a coded message, such as PD286742, and had to decide if it was sent by Spy A or Spy B using the letters. In this experiment, the classification was simple (PD or DP was Spy A, SF or FS was Spy B). Learners were given feedback on their classification and were then asked to use

[11] Note that if category learning involves only classification learning, there are a very limited number of learning situations, but if category learning involves both classification learning and learning to use the category, there are many uses and many ways in which the classification and learning can be related.

Decoding formula (refer to position of number in the coded message)

Spy A:	2nd + (6th × 3rd) + 1st + 5th
Spy B:	6th + (2nd/5th) + 4th + 3rd

Sample study messages		6th × 3rd	2nd/5th
Spy A:	PD286742	12	2
Spy B:	SF266632	12	2

Sample no-letter tests	6th × 3rd	2nd/5th
--384233	12	2.67
--483145	15	2

Figure 1. Sample materials for spy code experiments from Ross (1997, 1999).

the Spy A decoding formula, shown in Figure 1, to decode the numbers in the message. Applying that formula to this message led to a preliminary decoding of 26 ($8 + 2 \times 6 + 2 + 4$). All Spy A messages had the product of the sixth and third numbers equaling 12, and it was thought that learners might notice this relation and incorporate it into the category representation. Similarly, all Spy B messages had a consistent number relation, that the quotient of the second divided by the fifth was always equal to 2, so that this might be incorporated into the Spy B representation. As mentioned, all the coded messages had both number relations (see Figure 1); however, each message was decoded in one way, dependent on the spy, so only one relation was used in each decoding.

At test, the learners had to classify a message consisting only of numbers, but in these number sets, only one of the number relations held (i.e., either the product of the third and sixth numbers was 12 or the quotient of the second divided by the fifth numbers was 2). If learners had incorporated the number relations into the category representations, then we expected that they would classify these test items on the basis of the number relation. They did so for 0.79 of these no-letter tests. Thus, although the number relations during study were not predictive of the category (because all messages had both number relations), the use of the number relations did lead learners to incorporate them selectively into the category representations and use them for later classifications.

What Is Necessary for This Learning?

Ross (1999) extended this work in two ways using the spy decoding materials. First, this effect of using the category on later classification occurred with a postclassification paradigm, just as it did with the disease/symptom materials (0.68 on the no-letter tests). Second, and directly addressing the issue of distinct representations, the crucial aspect for finding this effect was that the learner activated the category representation during the category use. If learners during the decoding did not know which spy's message was being decoded, then there was no influence of the decoding on later classification judgments (0.48 on the no-letter tests). In contrast, if the category representation was activated during the decoding, then the effect returned. For example, if learners were encouraged to fully process the message during decoding, especially to note the letters in the message (which predict classification), then the effect of category use was found (0.69) even though the processing was incidental to the classification (that is, the learners were told to read the message aloud and write it down, not to classify). These findings provide strong support for the hypothesis that the critical aspect for category use to influence later category judgments is that the category representation be activated during the learning of the use.

What Is Learned?

These experiments show that the use of categories in problem solving can lead to a change in the category representations, but the features and relations that were learned were very simple and observable. In many categories, especially nonobject categories such as problem-solving ones, the commonalities that tie the category members together are often not simple or directly observable but rather involve relations among features and often even require some interpretation and abstraction of features (see, e.g., Chi, Feltovich, & Glaser 1981). Does the use of categories allow the learning of these more abstract features and relations?

The earlier experiments with the code paradigm did not require learning abstract relations (e.g., second divided by fifth equaled 2). Ross and Warren (2002) found that more abstract relations could be learned, such as whether the intermediate product of a subtraction (third minus sixth) was negative. For the no-letter tests, new items (with new negative numbers resulting from the intermediate product of the subtraction) were still likely to be classified as indicating the spy that the subtraction formula had been used with (ranging across experiments from 0.64 to 0.76). Thus, problem-solving category uses can lead to changes in the representation even when the features and relations are abstract ones, rather than specific, observable ones.

In summary, this research supports the claim that different category-related learning tasks are leading to a common category representation, not to distinct representations. There is much work on expertise in problem-solving domains (see, e.g., Chi et al., 1981) showing that experience in the domain leads to a change in the category structure and, as a result, to a shift from classification based on superficial properties to classification based on deeper properties of the domain. However, as mentioned in the previous section, the examination of expert performance does not allow one to determine which aspects of the experiences led to the change. In the experiments presented here, we can determine that the problem solving influenced the category representation, including knowledge for classification judgments. When people learn to classify and make use of the classification to solve a problem, the knowledge they use to solve the problem is incorporated into the category representation.

Implications for the Study of Categorization

In this article, we have argued that a full understanding of category learning and representation requires expanding the investigation of category learning beyond classification. Category acquisition occurs in the course of using categories for different functions. The particular information that is acquired about a category member in the context of carrying out a particular task depends on the information that is required to carry out that task successfully. This transfer-appropriate processing view of category acquisition is useful for helping to understand a number of recent lines of research that explore how people learn category representations. In addition, this view helps to provide a bridge between category learning and many other research areas in cognition. Categories are critical for a wide variety of tasks, and one would hope that category-learning research can provide some further ideas of category learning and representation to these areas in which category use is so important.

Much of this article has been taken up with a comparison of classification and inference learning. Even though these two tasks can be viewed as formally equivalent, an analysis of the information requirements for these tasks suggests that they are quite different. Consistent with this analysis, we found very different results for the two types of learning. Classification focuses learners on diagnostic properties that distinguish among the categories being learned. In contrast, inference focuses within a category on the relationship between the category label and the feature values, as well as on relationships among feature values.

It is most straightforward to contrast inference and classification because of their similarity. However, research on other uses of categories has enabled us to address other important aspects of categorization that have heretofore been missing from laboratory studies. The studies of combined inference and classification and of problem solving suggest that people are forming true category representations rather than independent representational structures that are suitable only for the task for which they were developed.

Suggestions

So, how should research on the relationship between category use and category learning proceed? Given the amount of research necessary to clarify theoretical issues, psychologists are just beginning to understand how to examine this relationship. Nonetheless, a general plan for research is beginning to take shape. First, it should be obvious after this review that laboratory studies must include a wider range of category-learning tasks. As we discuss below, there is a danger that different theories will emerge for every task, but our belief is that categorization research is hampered by an overly restricted set of learning tasks and that the potential danger of proliferating theories of tasks is more than outweighed by the salutary effects of considering the role of use on learning.

Second, for any study on category learning, it is important to do a task analysis to gain some understanding of the information required to perform the task. The information required depends both on the processing (use) and the information provided to the learner, so it is important to carefully consider the particular requirements of each experimental setting. This same caveat applies even to category learning by classification. For example, A. B. Markman and Maddox (2003) suggested that the reason people given classification tasks have difficulty learning categories when the features have multiple manifestations (as in the work by Yamauchi & Markman, 2000b) is that the family-resemblance category structure with exception features leads them to seek rules that distinguish between the categories. The number of possible rules increases with the number of feature manifestations. Markman and Maddox suggested that a family-resemblance structure with nondiagnostic feature values on some dimensions (rather than exception features) might lead to holistic processing. Because these categories have a family-resemblance structure, holistic processing based on similarity should be straightforward. Consistent with this task analysis, people learned categories with multiple feature manifestations easily when the category structure had nondiagnostic features instead of exception features.

Third, experimenters need to keep improving the task analyses by expanding the aspects of natural categories that are considered. In this article, the task analyses have focused on two issues: the relative attention to relationships among properties within a category versus between categories and the ability to use category representations for multiple tasks. Tasks vary with respect to their emphasis on these factors. More importantly, new investigations will suggest new foci and provide finer distinctions. For example, there are a number of within-category relationships that may well have different effects on the learning.

This research program will require shifting between developing task analyses and exploring new uses of categories to understand what is learned. By starting with the framework suggested in this article, research on category use need not be unconstrained. Instead, the range of tasks and aspects of category structure can be gradually expanded from the base provided by existing research.

Concluding Comments

We conclude this article with a call to action and three warnings. The call to action is that laboratory research on categorization must explore a range of category uses to provide insight into the richness of category representations observed outside of the lab. Mismatches between current models and observations of natural categories are due largely to the nearly exclusive focus of laboratory research on classification tasks (see Schank, Collins, & Hunter, 1986, for an earlier warning about this problem).

Although we believe that this change in investigations of category learning is critical, we must make some caveats. First, as mentioned above, this is the beginning of an ambitious research program. Over the past 25 years, there have been many studies teasing apart specific issues in classification learning. We are suggesting an approach to category learning of which classification is just a part.

Second, although researchers have learned quite a bit about inference learning and its differences from classification in a relatively short time, our guess is that inference may be easier to study than many other category uses. Indeed, one reason why Yamauchi and Markman (1998) contrasted inference and classification learning was because the tasks were formally identical and yet seemed to differ in their emphasis on within-category and between-categories structure. Many other category uses differ from classification in more complex ways and thus involve more detailed task analysis (as in the work on problem solving described above). In addition, although we believe that most category uses involve a central conceptual representation (as we showed with classification combined with inference and problem solving), we also recognize that specialized uses of categories may lead to specialized representations. Learning about these cases will help to better understand the limits of general category knowledge.

Third, this multifaceted approach to studying category acquisition must be undertaken conservatively. Science is a search for generalizations. If every task leads to the acquisition of somewhat different information, there is a risk that every model of categorization developed will only apply to the particular task on which the model is based. Thus, it is important to carry out a careful task analysis on the laboratory procedures to understand the degree to which they are likely to yield representations that are compatible with the performance of other tasks. It is possible, of course, that researchers will find an infinite number of variations in the structure of category representations that result from the myriad of ways that people interact with categories. However, we view this as a position to be driven to rather than one that should be adopted at the outset.

In our view, the danger of this kind of conceptual relativism is more apparent than real because there is a core set of tasks for which categories are frequently used. Clearly, classification is important. The other uses discussed in this article—inference, problem solving—are also important. There are a few other areas (e.g., preference formation, communication, and causal reasoning) for which less research has been done (Ahn, 1999; A. B. Markman & Makin, 1998; Zhang & Markman, 1998). Research on category acquisition should focus on this set of tasks to understand how they affect category representations. As these studies accrue, psychologists may also begin to see generalizations across the tasks and better understand how they interrelate. Finally, a better understanding of the individual tasks will allow researchers to investigate category learning from multiple interleaved uses, which is the typical manner in which categories are acquired.

References

Aaker, D. A., & Keller, K. L. (1990). Consumer evaluations of brand extensions. *Journal of Marketing, 54,* 27–41.

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research, 34,* 347–356.

Ahn, W. K. (1999). Effect of causal structure on category construction. *Memory & Cognition, 27,* 1008–1023.

Ahn, W. K., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16,* 81–122.

Anderson, A. L., Ross, B. H., & Chin-Parker, S. (2002). A further investigation of category learning by inference. *Memory & Cognition, 30,* 119–128.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

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

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin and Review, 8,* 629–647.

Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 259–277.

Au, T. K. (1994). Developing an intuitive understanding of substance kinds. *Cognitive Psychology, 27,* 71–111.

Berlin, B. (1972). Speculations on the growth of ethnobotanical nomenclature. *Language in Society, 1,* 51–86.

Billman, D. (1996). Structural biases in concept learning: Influences from multiple functions. In D. Medin (Ed.), *The psychology of learning and motivation* (Vol. 36, pp. 283–321). San Diego, CA: Academic Press.

Billman, D., & Dávila, D. (2001). Consistent contrast aids concept learning. *Memory & Cognition, 29,* 1022–1035.

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

Boster, J. S., & Johnson, J. C. (1989). Form or function: A comparison of expert and novice judgments of similarity among fish. *American Anthropologist, 92,* 171–179.

Broniarczyk, S. M., & Alba, J. W. (1994). The importance of the brand in brand extension. *Journal of Marketing Research, 31,* 214–228.

Brooks, L. R. (1999). Orientation to meaning suppresses explicit knowledge of surface cues in category learning. *Abstracts of the Psychonomic Society, 4,* 51.

Brown, R. (1958). How shall a thing be called? *Psychological Review, 65,* 14–21.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

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

Chin-Parker, S., & Ross, B. H. (2002b). *The effect of contrast categories in classification and inference learning.* Manuscript submitted for publication.

Clapper, J. P., & Bower, G. H. (1994). Category invention in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 443–460.

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science, 24,* 299–349.

Davidson, N. S., & Gelman, S. A. (1990). Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development, 5,* 151–176.

Flavell, J. H. (1985). *Cognitive development* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Gagne, C. L. (2000). Relation-based combinations versus property-based combinations: A test of the CARIN theory and the dual-process theory of conceptual combination. *Journal of Memory and Language, 42,* 365–389.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26,* 796–804.

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science, 10,* 489–493.

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

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

Hirschfeld, L. A. (1996). *Race in the making.* Cambridge, MA: MIT Press.

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: Multiple classification and children's category-based inductions. *Child Development, 63,* 1536–1557.

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

Kunda, Z. (1999). *Social cognition: Making sense of people.* Cambridge, MA: MIT Press.

Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic Bulletin and Review, 3,* 95–99.

Love, B. C. (2002). Comparing supervised and unsupervised learning. *Psychonomic Bulletin and Review, 9,* 829–835.

Love, B. C. (in press). The multifaceted nature of unsupervised category learning. *Psychonomic Bulletin and Review.*

Love, B. C., Markman, A. B., & Yamauchi, T. (2000, August). *Modeling classification and inference learning.* Paper presented at the Annual meeting of the American Association for Artificial Intelligence, Austin, TX.

Love, B. C., Medin, D. L., & Gureckis, T. (in press). SUSTAIN: A network model of human category learning. *Psychological Review.*

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

Malt, B. C., & Johnson, E. J. (1992). Do artifact concepts have cores? *Journal of Memory and Language, 31,* 195–217.

Malt, B. C., Ross, B. H., & Murphy, G. L. (1995). Predicting features for members of natural categories when categorization is uncertain. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 646–661.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity of the linguistic categorization of artifacts. *Journal of Memory and Language, 40,* 230–262.

Markman, A. B., & Maddox, W. T. (2003). Classification of exemplars with single and multiple feature manifestations: The effects of relevant dimension variation and category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 107–117.

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

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence, 13,* 329–358.

Markman, E. M. (1989). *Categorization and naming in children.* Cambridge, MA: MIT Press.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes, 12,* 137–176.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology, 32,* 49–96.

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

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

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

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness and category construction. *Cognitive Psychology, 19,* 242–279.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and Perception.* Cambridge, MA: Harvard University Press.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Morris, M. W., & Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition, 18,* 407–418.

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

Murphy, G. L. (1993). A rational theory of concepts. In G. Nakamura, R. Taraban, & D. L. Medin (Eds.), *The psychology of learning and motivation* (Vol. 29, pp. 327–359). Orlando, FL: Academic Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–315.

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology, 27,* 148–193.

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 572–586.

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

Nosofsky, R. M., & Palmeri, T. J. (1997). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Perception and Psychophysics, 59,* 1027–1048.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–97.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200.

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

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology, 79,* 1–16.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 811–828.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382–407.

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.

Rehder, B., & Ross, B. H. (2001). Abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1261–1275.

Rips, L. J., & Conrad, F. G. (1989). Folk psychology of mental activities. *Psychological Review, 96,* 187–207.

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

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

Ross, B. H. (1997). The use of categories affects classification. *Journal of Memory and Language, 37,* 240–267.

Ross, B. H. (1999). Postclassification category use: The effects of learning to use categories after learning to classify. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 743–757.

Ross, B. H. (2000). The effects of category use on learned categories. *Memory & Cognition, 28,* 51–63.

Ross, B. H., Gelman, S. A., & Rosengren, K. (2002). *Children's category-based inferences affect classification.* Manuscript submitted for publication.

Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 736–753.

Ross, B. H., & Warren, J. L. (2002). Learning abstract relations from using categories. *Memory & Cognition, 30,* 657–665.

Sandhofer, C. M., & Smith, L. B. (2001). Why children learn color and size words so differently: Evidence from adults' learning of artificial terms. *Journal of Experimental Psychology: General, 130,* 600–620.

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

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75,* 589–606.

Sifonis, C. M., & Ross, B. H. (2002). Alignment effects on learning multiple, use-relevant classification systems. *Memory & Cognition, 30,* 1148–1159.

Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12,* 485–527.

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

Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category:

The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 984–1002.
Smith, J. D., Murray, M. J., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 659–680.
Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology, 43,* 129–169.
Thomas, R. D. (1997). Learning correlations in categorization tasks using large, ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 119–143.
Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.
Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113,* 169–191.
Wattenmaker, W. D. (1995). Knowledge structures and linear separability. Integrating information in object and social categorization. *Cognitive Psychology, 28,* 274–328.
Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin and Review, 4,* 167–183.
Wisniewski, E. J. (1998). Property instantiation in conceptual combination. *Memory & Cognition, 26,* 1330–1347.
Wisniewski, E. J., & Love, B. C. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language, 38,* 177–202.
Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 585–593.
Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39,* 124–148.
Yamauchi, T., & Markman, A. B. (2000a). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 776–795.
Yamauchi, T., & Markman, A. B. (2000b). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. *Memory & Cognition, 28,* 64–78.
Zhang, S., & Markman, A. B. (1998). Overcoming the early entrant advantage: The role of alignable and nonalignable differences. *Journal of Marketing Research, 35,* 413–426.

Received March 4, 2002
Revision received November 26, 2002
Accepted December 18, 2002 ■