Psychonomic Bulletin & Review
1999, 6 (3), 363-378

On the nature of implicit categorization

F. GREGORY ASHBY and ELLIOTT M. WALDRON
University of California, Santa Barbara, California

Current categorization models disagree about whether people make a priori assumptions about the structure of unfamiliar categories. Data from two experiments provided strong evidence that people do not make such assumptions. These results rule out prototype models and many decision bound models of categorization. We review previously published neuropsychological results that favor the assumption that category learning relies on a procedural-memory-based system, rather than on an instance-based system (as is assumed by exemplar models). On the basis of these results, a new category-learning model is proposed that makes no a priori assumptions about category structure and that relies on procedural learning and memory.

There is much recent evidence that human category learning relies on multiple systems (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Smith, Patalano, & Jonides, 1998; Smith, Patalano, Jonides, & Koeppe, 1996). The consensus is that one system is rule or theory based and one involves some form of implicit learning. There is little agreement, however, about the nature of the implicit learning system. One possibility is that the implicit system computes some form of decision function. In most cases, this is equivalent to constructing a decision boundary that separates the contrasting categories (Ashby, 1992a; Ashby & Lee, 1991, 1992; Ashby & Maddox, 1990, 1992, 1993; Maddox & Ashby, 1993). A second possibility, however, is that the implicit system compares the stimulus with the memory traces of past category exemplars (see, e.g., Brooks, 1978; Estes, 1986; Medin & Schaffer, 1978; Nosofsky, 1986) or simply learns to associate responses (or response labels) with different regions of perceptual space (Ashby & Maddox, 1989). In both of these cases, one could still define a decision boundary as the function that separates regions assigned to contrasting categories. Even so, this type of "decision boundary" would only be a mathematical convenience, since it would not correspond to any real computation performed within the brain.

In this article, we test between these two general alternatives. We present data that provide strong evidence that people do not construct decision boundaries or compute decision functions, no matter how complex. Rather, our data are consistent with the general notion that the implicit system accesses exemplar memories or that it gradually learns to associate response labels with clumps of cells in some high-level visual representation area, such as the inferotemporal cortex. We will also review previously published neuropsychological results that favor the latter of these two possibilities. Finally, we identify a class of implicit category-learning models that are consistent with all these results.

PARAMETRIC VERSUS NONPARAMETRIC CLASSIFICATION

To better understand the issues addressed in this article, consider an experiment in which observers are trying to learn two categories of lines that vary in length and orientation. Examples of two separate such experiments are described in Figure 1. Each point in Figures 1A and 1B describes a different stimulus from such an experiment. The "+" signs indicate the lengths and orientations of the exemplars of category *A* and the "o" signs describe the exemplars of category *B*. On each trial, one of these stimuli is sampled randomly and presented to the observer, whose task is to assign it to category *A* or *B*. Feedback about response accuracy is given on every trial. The line in Figure 1A and the curve in Figure 1B are called the optimal decision bounds, because they describe the optimal response strategy—in both cases, accuracy is maximized if the observer responds *A* to any stimulus that falls above the optimal bound and *B* to any stimulus that falls below.

An extensive literature shows that healthy young adults often eventually learn to respond in a nearly optimal fashion in experiments like those shown in Figure 1 (e.g., Ashby & Maddox, 1990, 1992; Maddox & Ashby, 1993). In such cases, however, observers are virtually never able to describe their behavior. For example, in Figure 1A, an explicit analogue of the optimal rule is: Respond *A* if the orientation of the line is greater than the length; otherwise, respond *B*. However, because orientation and length are expressed in different units, this is like comparing apples and oranges. We have collected extensive amounts of data in the tasks shown in Figure 1. After the last experimental session, we typically ask observers for a verbal descrip-

This research was supported by National Science Foundation Grant SBR95-14427. We thank Leola Alfonso-Reese for her invaluable help with the design of the experiments reported here and Barbara Knowlton, Robert Nosofsky, and John Wixted for their helpful suggestions. Correspondence concerning this article should be addressed to F. G. Ashby, Department of Psychology, University of California, Santa Barbara, CA 93106 (e-mail: ashby@psych.ucsb.edu).

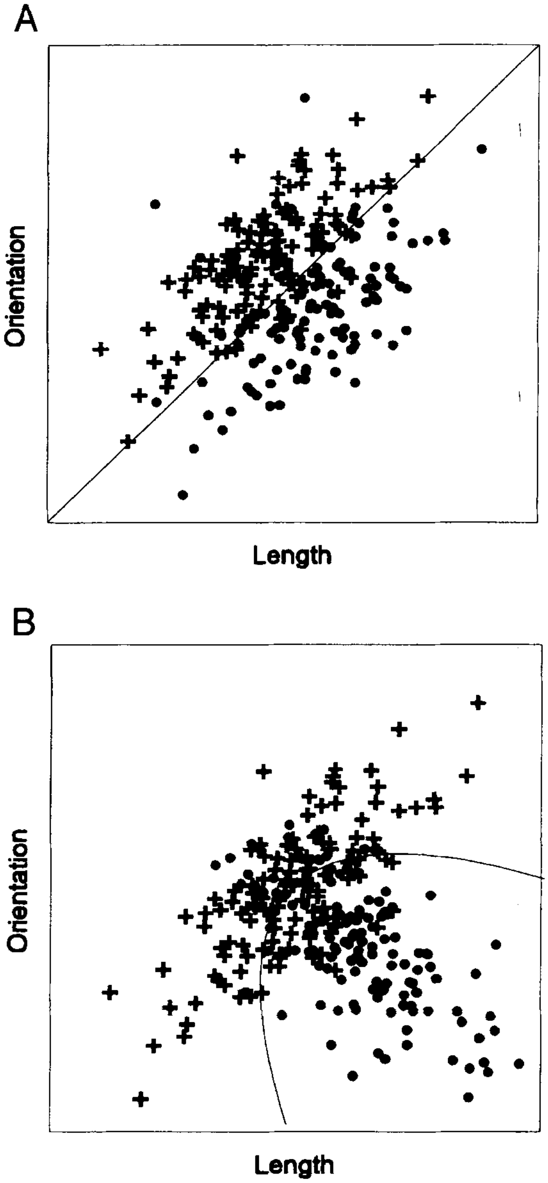

Figure 1. Category structure of two hypothetical experiments. A plus sign indicates an exemplar from category *A*, and a circle indicates an exemplar from category *B*. The solid curves are the decision boundaries that maximize response accuracy (i.e., the optimal boundary). The optimal bound is linear in Figure 1A and quadratic in Figure 1B.

tion of their response strategy. None of the observers in these experiments has ever described the optimal rule, even when his or her performance was well described by this rule. Frequently, observers simply say that their responses were just a "gut reaction." On the basis of this and other evidence, Ashby et al. (1998) argued that people learn the category structures shown in Figure 1 by some form of implicit learning.[1]

An ideal observer in the Figure 1 experiments, who uses the optimal bound perfectly, could learn to do so in several different ways. First, the observer might experiment with many different decision bounds until eventually discovering the one that is optimal. Second, the observer might gradually learn which parts of the (orientation, length) space are associated with category *A* and which parts are associated with category *B*. In this case, the decision bound has no psychological meaning, although it could be defined mathematically as the set of points that partitions the region associated with category *A* from the region associated with category *B*. One of the main goals of this article is to test between these two general alternatives.

The issue of whether the implicit system computes a decision boundary during category learning or just gradually associates response labels with different regions of perceptual space is fundamentally a question of whether a priori assumptions are made about the structure of the categories to be learned. For if a decision boundary is used, some functional form of the boundary must be chosen. For example, in Figure 1A, the optimal bound is linear (i.e., a straight line), and several studies have shown that, in such cases, the responses of practiced observers are well separated by a linear decision boundary (Ashby & Gott, 1988; Ashby & Maddox, 1990). In Figure 1B, the optimal bound is quadratic (i.e., a quadratic curve), and in such cases, the responses of practiced observers are well separated by a quadratic decision boundary (Ashby & Maddox, 1992). To account for data such as these, a model that assumes that category learning is a process of adjusting the parameters of a decision boundary (e.g., either linear or quadratic) must assume that observers sometimes choose a linear form for the decision boundary and sometimes a quadratic form.

How could an observer know which form to choose? Without making extra assumptions about the form of the unknown categories, there is no way for an observer to answer this question. However, if the categories are assumed to be of a certain type, straightforward solutions to this problem are known to exist. For example, if it were known that the exemplars were normally distributed across the various stimulus dimensions within each category, then it is well known that the decision boundary that maximizes categorization accuracy is always linear or quadratic (see, e.g., Ashby, 1992a; Ashby & Gott, 1988). For example, the four categories depicted in Figures 1A and 1B all satisfy this property. In every case, the lengths of the lines in each category are normally distributed, as are the orientations of the lines, and the correlation between length and orientation (if one exists) is linear (as measured, e.g., by the Pearson correlation coefficient). Thus, the optimal bound in Figures 1A and 1B must be linear or quadratic. Furthermore, whether the optimal boundary is linear or quadratic in such cases depends only on the within-category variances on each stimulus dimension and on the correlation between each pair of dimensions. When there are two categories to learn, a linear boundary is optimal if and only if each of these parameters has exactly the same value in each category. In this case, we say that the categories have the same variance–correlation structure. If the categories have a different variance–correlation structure—that is, if any of the variances or correlations have a different value in the two categories—a quadratic boundary is optimal.[2] For example, in Figure 1A, the exemplars in categories *A* and *B* have equal variability in length, they have equal variability in orientation, and the degree to which length and orientation are correlated is the same. As a consequence, the optimal bound in Figure 1A is linear. In Figure 1B, however, the correlations are different in the two categories (i.e., negative in *B* and positive in *A*), so the optimal bound in this experiment is quadratic.

Ashby and Alfonso-Reese (1995) showed that virtually all of the currently popular models of categorization are equivalent to a process in which the observer estimates the likelihood that the stimulus is a member of each contrasting category. The models differ according to the type of estimator assumed. In statistics, there are two broad classes of estimators. Parametric estimators make strong assumptions about the distribution of the sample data, whereas nonparametric estimators make only weak distributional assumptions. For example, a parametric estimator might assume that the sample data are from a normal distribution. In this case, to estimate the distribution of the data, one needs only estimate the mean and the variance and then insert these estimates into the equation that specifies the normal distribution. This method works well if the data distribution is normal, but of course, if the underlying distribution is skewed, for example, this parametric estimator will be badly biased. In contrast, a popular nonparametric estimator of a data distribution is the relative frequency histogram. The critical distinction is that relative frequency histograms can mimic the shape of any underlying distribution, whereas parametric estimators can only mimic the shape of distributions from the assumed family (e.g., normal).

In accord with this popular distinction in statistics, Ashby and Alfonso-Reese (1995) defined *parametric classifiers* as those classifiers that make strong assumptions about the form of the contrasting categories and *nonparametric classifiers* as those that make almost no assumptions about category form. For example, a parametric classifier might assume that the exemplars have a multivariate normal distribution within each category. In this case, the optimal boundary is always linear or quadratic, so this type of parametric classifier would always use a linear or quadratic decision bound. If some other distributional family were assumed, the optimal boundary might be of some different functional form. The important point, however, is that the decision boundary used by a parametric classifier must always be of the same functional form as that of the optimal boundary, given the distributions assumed by that classifier. As such, parametric classifiers always predict decision bounds of only a limited type (e.g., linear or quadratic). In contrast, nonparametric classifiers can mimic any type of optimal bound. For example, a simple nonparametric classifier might estimate the distribution of the contrasting categories with a relative frequency histogram and then, when a stimulus is presented, give the response associated with the highest histogram. In summary, parametric classifiers can be identified as those that make strong assumptions about the structure of the underlying categories or, *equivalently*, that can mimic only a limited set of decision bounds. Nonparametric classifiers make weak assumptions about category structure and can mimic virtually any decision bound.

Prototype models of categorization assume that the observer gives the response associated with the nearest category prototype (Ashby & Maddox, 1998; Reed, 1972; Smith & Medin, 1981). This strategy is optimal only if the category structures are simple enough that the variance–correlation structure can be ignored.[3] Thus, prototype models make strong assumptions about category structure. Alternatively, it is straightforward to show that the prototype decision strategy is always equivalent to using a linear decision bound (i.e., with two categories; Ashby & Gott, 1988), so for either reason, prototype models are parametric classifiers (Ashby & Alfonso-Reese, 1995). In fact, any model that assumes that the observer operates directly on a decision boundary is equivalent to some parametric classifier (Ashby & Alfonso-Reese, 1995), because to specify a decision bound, or a family of decision bounds, one must specify a functional form. In contrast, Ashby and Alfonso-Reese showed that exemplar models are nonparametric classifiers, because they are equivalent to a process in which the observer estimates the category distributions with a nonparametric estimator (i.e., the Parzen kernel estimator), which is essentially a sophisticated relative frequency histogram. Similarly, models that assume that the observer learns to associate response labels with different regions of perceptual space are equivalent to some nonparametric classifier, because such a process could conceivably mimic any decision bound.

All parametric models of categorization assume that the category distributions are of some specific family. For example, a model that assumes that observers always use a linear or quadratic boundary is equivalent to a parametric classifier that assumes that the category distributions are multivariate normal. Many investigators have suggested the possibility that subjects enter a classification task with the expectation that categories are normally distributed (or at least unimodal and symmetric; e.g., Ashby, 1992a; Ashby & Alfonso-Reese, 1995; Ashby & Maddox, 1992; Flannagan, Fried, & Holyoak, 1986; Fried & Holyoak, 1984; Myung, 1994). The origin of this expectation may arise from the belief that normal distributions provide good approximations to the distributions of many natural perceptual categories, because they assume a dense region of typical members surrounded by a sparse region of less typical members (Rosch & Mervis, 1975). A few studies have investigated the normality assumption empirically. For example, Flannagan et al. found that normal distributions yielded more veridical learning than did bimodal distributions. Furthermore, no transfer effects were obtained for learning normal dis-

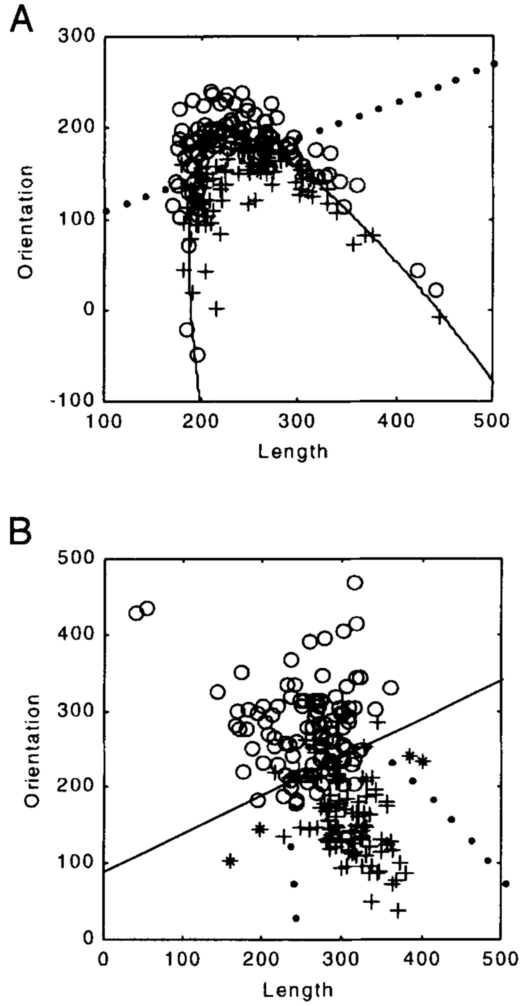

Figure 2. Category structure of Experiments 1 (Figure 2A) and 2 (Figure 2B). A plus sign indicates an exemplar from category *A*, and a circle indicates an exemplar from category *B*. The solid curve is the decision bound that maximizes response accuracy (i.e., the optimal boundary), whereas the boundary that would be used by a parametric classifier that assumed normality is dotted.

tributions, whereas strong transfer effects were observed when the distributions were bimodal. Flannegan et al. suggested that this latter finding indicates that the classification system might use the normal distribution as a prominent null model, along with a repertoire of other, less likely, distributional assumptions.

Furthermore, normality is an attractive choice because, if observers are estimating category means, variances, and correlations, normality is the optimal assumption, because extra assumptions must be added to infer any other distribution (technically, under these conditions, normality is the maximum entropy inference; see Myung, 1994, for details). In addition, it is known that people are extremely sensitive to category means, variances, and correlations (see, e.g., Ashby & Gott, 1988; Ashby & Maddox, 1992, 1993; Fried & Holyoak, 1984; Medin & Schaffer, 1978; Medin, Wattenmaker, & Hampson, 1987), so they certainly have access to enough information to estimate category likelihoods under the assumption of normality. On the other hand, some studies have shown that observers can respond optimally (or nearly so) in tasks in which the categories are not normally distributed (McKinley & Nosofsky, 1995; Neumann, 1977), which might be taken as evidence that humans are nonparametric classifiers.

This article reports the results from two experiments designed to test whether human perceptual categorization is parametric or nonparametric. The idea was to design the contrasting categories in such a way that virtually any parametric classifier would fail to respond optimally. In Experiment 1, the stimulus categories were highly nonnormal, yet they were unimodal and had the same variance–correlation structure. Thus, a parametric classifier that assumed normality would use a decision boundary that was linear, whereas the optimal boundary was highly quadratic. In Experiment 2, the stimulus categories were again unimodal and nonnormal, but now the optimal decision boundary was linear, whereas the parametric decision boundary was highly quadratic (i.e., the variance–correlation structure differed across the two categories).

As in Figure 1, the stimuli in both experiments were lines that varied in length and orientation. The design of the contrasting categories used in the two experiments is illustrated in Figure 2. As in Figure 1, each symbol represents a unique stimulus; the exemplars of category *A* are denoted by the "+" signs, and the exemplars of category *B* are denoted by the "o" signs. In both experiments, the distribution of exemplars in each category was unimodal. In both cases, the decision boundary that maximized response accuracy (i.e., the optimal boundary) is depicted by the solid curve, whereas the boundary that would be used by a parametric classifier that assumed normality is broken. Note that, in both experiments, the categories overlap somewhat, so that perfect accuracy is impossible. In fact, in both experiments, the maximum possible accuracy is 90%. This is the accuracy that would be achieved by an observer who responded *A* to any stimulus falling below the optimal boundary and *B* to any stimulus falling above. An observer using the broken line parametric boundary would achieve 83.2% correct in Experiment 1 and 89% correct in Experiment 2. Because the optimal and parametric classifiers made such similar accuracy predictions in Experiment 2, four transfer stimuli were added to the Experiment 2 design (denoted by the asterisks in Figure 2B). These were presented five times each without feedback during the last experimental session.

Although no studies have looked specifically at the question of whether human category learning is parametric or nonparametric, many studies have used nonnormal categories with the same variance–correlation structure (e.g., Homa, 1978; Homa & Cultice, 1984; Homa, Sterling, & Trepel, 1981; Hyman & Frost, 1975; McKinley & Nosofsky, 1995). McKinley and Nosofsky even reported the results of such a study in which the optimal bound was nonlinear. Observers in this experiment did not

Table 1
Parameters of Bivariate Normal Distributions Used in the First Step of the Stimulus Generation Procedure

	Experiment 1		Experiment 2	
Parameter	Category *A*	Category *B*	Category *A*	Category *B*
μ_x	0	0	0	0
μ_y	0	1	1	−1.897
σ_x^2	1	1	0.25	0.05
σ_y^2	0.1521	0.1521	0.25	0.9
ρ_{xy}	0	0	0	0

use a linear bound, so the results of the McKinley and Nosofsky study support the nonparametric assumption. However, several factors make it difficult to draw strong conclusions from the McKinley and Nosofsky study. First, the McKinley and Nosofsky categories were bimodal, which provided observers with an easy method of discovering the nonnormality. This is especially problematic because Flannagan et al. (1986) showed that people are quite sensitive to bimodality. The categories used in the experiments reported in this article were all unimodal, and the only distributional information that signaled nonnormality was in third and higher moments (e.g., skewness and kurtosis). As will be discussed later in this article, the standard error of estimation of such statistics is so large that, for all practical purposes, they can be considered unestimable. Second, the optimal bound in the McKinley and Nosofsky experiments was neither linear nor quadratic. Thus, the form of the optimal bound could have informed observers that the category distributions were nonnormal.[4] In contrast, in the experiments reported here, the form of the optimal bound provided no information about the nonnormality of the categories. Third, the McKinley and Nosofsky categories were constructed from two subordinate categories that were themselves each normally distributed, and McKinley and Nosofsky acknowledged that they could not rule out the possibility that observers used a parametric classifier of the type that was appropriate, given such a structure (although McKinley & Nosofsky did show that observers had no such explicit knowledge).

GENERAL METHOD

Observers

Five different graduate students at the University of California, Santa Barbara, participated in each experiment. All observers were paid $7 for each 50-min experimental session. All of the observers in Experiment 1 and 3 of the observers in Experiment 2 completed one experimental session on 4 consecutive days. Two of the 5 observers in Experiment 2 completed a session on 5 consecutive days.

Stimuli and Apparatus

The stimuli were lines that varied in length and orientation. In both experiments, the exemplars making up each category were selected by a two-step process. First, random samples were drawn from a bivariate normal distribution, and then each sample was subjected to a nonlinear transformation. The parameters of the bivariate normal distributions are listed in Table 1. Let the vector $[x\ y]'$ denote a random sample from these distributions. In Experiment 1, the value y was transformed to a new value w via the transformation $w = y - 0.7x^2$. In Experiment 2, this transformation was $w = y + 0.6x^2$. Finally, in Experiment 1, the vector $[x\ w]$ was transformed to length and orientation via

$$\begin{bmatrix} \text{length} \\ \text{orientation} \end{bmatrix} = 50 \begin{bmatrix} \cos\left(\frac{\pi}{8}\right) & \sin\left(\frac{\pi}{8}\right) \\ -\sin\left(\frac{\pi}{8}\right) & \cos\left(\frac{\pi}{8}\right) \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} + \begin{bmatrix} 235 \\ 165 \end{bmatrix}.$$

Because this same transformation was used for categories *A* and *B*, and because the original category *A* and *B* variance–correlation structures were equal (see Table 1), the variance–correlation structures for the final *A* and *B* categories were equal (i.e., the two categories were related via a simple translation). In Experiment 2, the vector $[x\ w]$ was transformed to length and orientation via

$$\begin{bmatrix} \text{length} \\ \text{orientation} \end{bmatrix} = 50 \begin{bmatrix} \cos\left(\frac{\pi}{8}\right) & \sin\left(\frac{\pi}{8}\right) \\ -\sin\left(\frac{\pi}{8}\right) & \cos\left(\frac{\pi}{8}\right) \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} + \begin{bmatrix} 285 \\ 235 \end{bmatrix}.$$

After these transformations, the optimal bound was quadratic in Experiment 1 and linear in Experiment 2 (shown in Figure 2), even though the category *A* and *B* variance–correlation structures were identical in Experiment 1 and different in Experiment 2. In addition, the transformations were chosen so that, except for a translation, the optimal bound in Experiment 1 nearly equaled the parametric bound in Experiment 2 and the parametric bound in Experiment 1 equaled the optimal bound in Experiment 2. The distributions were also selected so that an ideal observer using the optimal bound would achieve 90% accuracy in both experiments. An observer using the parametric bound would achieve 83.2% correct in Experiment 1 and 89% correct in Experiment 2. Because the optimal and parametric accuracies were so similar in Experiment 2, four transfer stimuli were added on the last day without feedback (shown as asterisks in Figure 2B). The transfer stimuli were chosen to be between the optimal and the parametric bounds. This would allow an alternative form of evidence, in order to discriminate between the best-fitting decision bounds used by the observers.

Each (line, orientation) pair was converted into a physical stimulus by creating a line representing length in *line* pixels and orientation of ($\pi \times$ *orientation*)/550 radians. For example, the sample (150,160) was used to create a line 150 pixels long, oriented at $160(\pi/550)$ radians. The orientation of stimuli in the optimal quadratic condition varied from 1.18 to 5.833 radians, and the visual angle varied from 3.4° to 7.9°. The orientation of stimuli in the optimal linear condition varied from 1.74 to 5.83 radians, and the visual angle varied from 3.6° to 10.2°. The stimuli were displayed on a Mitsubishi Electric Color Display Monitor Model C-9918NB in a dimly lit room.

Procedure

The observer's task was to assign each presented stimulus to a category by pressing one of two buttons labeled *A* and *B*. Accuracy was stressed more than speed. The display was either response terminated or terminated after 5 sec. After each response, auditory feedback was presented in the form of a sinusoidal tone. A 500-Hz tone indicated a correct response, and a 200-Hz tone indicated an incorrect response. The time between the response and the presentation of the next stimulus was 3 sec. Between each consecutive 50-trial block, the observers were allowed to rest for an amount of time that was observer controlled. An experimental session included 10 blocks of 50 trials (500 trials per session). The observers were instructed that about half the stimuli came from category *A* and half from category *B* and that the categories overlapped so that the best accuracy anyone could achieve would be 90%.

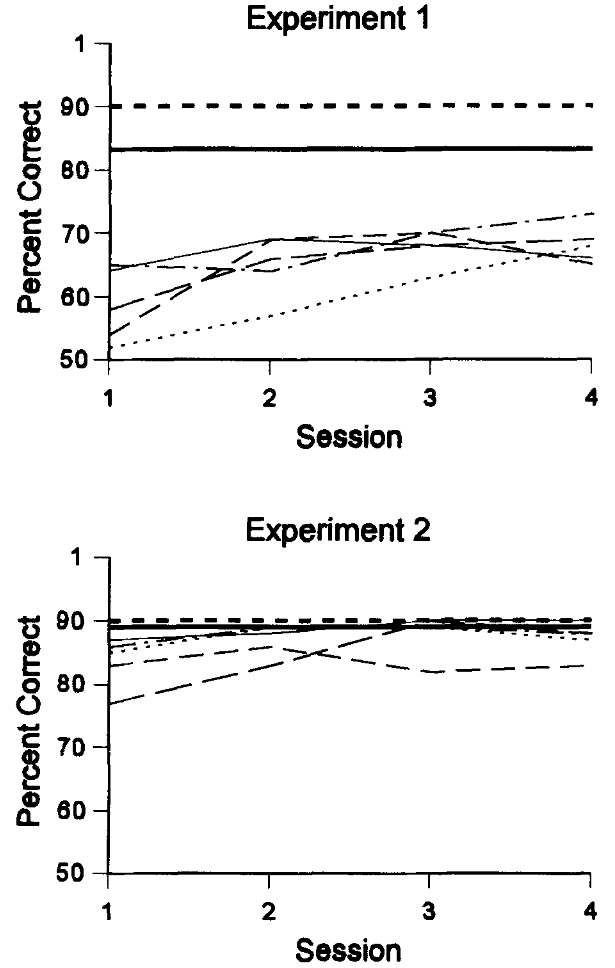

Figure 3. The accuracy of each observer during every experimental session of Experiments 1 and 2. The accuracy of the optimal classifier is denoted by the horizontal broken line, and the accuracy of the parametric classifier is denoted by the horizontal solid line.

RESULTS

Accuracy-Based Analyses

The accuracy of each observer during every experimental session is shown in Figure 3. The horizontal broken line denotes the accuracy of the optimal classifier, and the horizontal solid line denotes the accuracy of the parametric classifier. The observers in Experiment 1 gradually increased their accuracy from about 60% to about 70% over the course of the four experimental sessions, but even at their best, all 5 observers were significantly less accurate than the parametric classifier. In contrast, in Experiment 2, accuracy was high even during the first session, and at least some observers outperformed the parametric classifier at some point in the experiment.

Figure 3 indicates that, relative to the accuracy of either the optimal or the parametric classifiers, performance was much worse in Experiment 1 than in Experiment 2. In addition, in Experiment 1, accuracy tended to improve across all four sessions, whereas in Experiment 2, it had mostly asymptoted by the end of Session 2. Although not conclusive, such differences favor models that assume a nonparametric classifier, because in general, quadratic bounds (i.e., the optimal bound in Experiment 1) are more difficult to learn than linear bounds (i.e., the optimal bound in Experiment 2; see, e.g., Ashby & Maddox, 1992; Maddox & Ashby, 1993). Because the parametric bound is linear in Experiment 1 and quadratic in Experiment 2, parametric classifiers should find Experiment 1 easier, in which case learning would be faster in Experiment 1 than in Experiment 2, and accuracy would also be higher, relative to the ceilings defined by the solid lines in Figure 3.

Model-Based Analyses

To get a more detailed picture of how observers categorized the stimuli, a number of different models derived from decision bound theory (Ashby, 1992a; Maddox & Ashby, 1993) were fit to each observer's responses. Decision bound theory assumes that each observer partitions the perceptual space into response regions by constructing a decision bound. On each trial, the observer determines which region the percept is in and then emits the associated response. Despite this deterministic decision rule, decision bound models predict probabilistic responding, because of trial-by-trial perceptual and criterial noise. We fit seven different versions of decision bound theory to the data collected in Experiments 1 and 2. All of the models were described in detail by Ashby (1992a).

The goal of the analyses reported in this section is to obtain the best possible description of the data from each individual observer. For example, this analysis will allow us to determine whether the data are better described by the optimal or the parametric boundary. It is important to note, however, that a good fit of any specific model provides only limited information about psychological process. In particular, it is likely that some model that makes very different process assumptions (e.g., a nonparametric classifier, such as an exemplar-based model) might fit as well as the best of these seven decision bound models. With this caveat in mind, we proceed with a description of the seven decision bound models.

1. Optimal classifier. This model assumes that observers use the decision bound that maximizes accuracy (the solid line bounds shown in Figure 2). With the category structures shown in Figure 2, the optimal decision bound is quadratic in Experiment 1 and linear in Experiment 2. The only free parameter of this model is the variance of internal (perceptual and criterial) noise (i.e., σ^2).

2. Parametric classifier. The parametric classifier assumes that observers use the parametric decision bound (the broken line curves in Figure 2). As was described above, the parametric bound is linear in Experiment 1 and quadratic in Experiment 2. This model, which is identical to the optimal classifier model, except for the decision bound, has one free parameter (i.e., σ^2).

3. General linear classifier. The general linear classifier (GLC) assumes that the decision bound is linear. In the present applications, the GLC has three free parameters: the slope and the intercept of the linear decision bound and the variance of internal noise (i.e., σ^2).

4. General quadratic classifier. The general quadratic classifier (GQC) assumes that the decision bound is quadratic. The GQC has seven free parameters in the present application: six parameters that define the quadratic bound and the variance of internal noise.

5 and 6. Unidimensional classifiers. The unidimensional classifiers assume that observers use a unidimensional rule (i.e., a vertical or horizontal decision bound). These models each have two free parameters: the intercept of the decision bound and σ^2.

7. Independent decisions classifier. This model assumes that observers use a conjunctive rule of the form (Ashby, 1992a)

respond A if length $> x_l$ AND if orientation $< y_o$;

otherwise, respond B,

where x_l and y_o are free parameters. Two different versions of this model were created. In one, the noise was assumed to be equal on the two dimensions (so this version had three free parameters: x_l, y_o, and σ^2), and in the other, different noise variances were allowed on the two dimensions (resulting in four free parameters: x_l, y_o, σ_l^2, and σ_o^2). The independent decisions classifier is more similar to the unidimensional classifier than to the GLC, because, with a conjunctive rule, observers never integrate information from the two stimulus dimensions. Rather, they make separate decisions about the two dimensions and then select a response on the basis of the outcomes of these decisions (Ashby & Gott, 1988; Shaw, 1982). In contrast, in the GLC, the stimulus information is integrated (via some linear combination rule), and a response is made on the basis of this integrated value.

Using an iterative maximum likelihood parameter estimation procedure, each of these models was fit separately to the data from the last response block of every observer. To select the best-fitting model, we used the A information criterion (AIC) of Akaike (1974; see, also, Takane & Shibayama, 1992):

$$\text{AIC} = -2L + 2v,$$

where v is the number of free parameters and L is the log likelihood of the data, given the model (see, e.g., Ashby, 1992b, p. 32). The AIC statistic penalizes a model for extra free parameters in such a way that the smaller the AIC, the closer a model is to the "true model," regardless of the number of free parameters. As a result, to find the best model among a given set of competitors, one simply computes an AIC value for each model and chooses the model associated with the smallest AIC. Table 2 lists the AIC scores for the best-fitting version of each model in both experiments. The score of the model that provided the best overall fit for each observer is marked in bold. Figure 4 shows the bounds from each of these best-fitting models, together with a few of the stimuli from the experiment and the optimal bound (dotted line).

In Experiment 1, the parametric classifier fits better than the optimal classifier for every observer. On the other hand, the GQC fits substantially better than the GLC for every observer. In fact, the GQC provides the best overall fit to the data from 4 of the 5 observers (an independent decisions classifier fit slightly better than the GQC for Observer 5). These results strongly indicate that all the observers in Experiment 1 used a nonlinear bound but that the bound they used was suboptimal. This latter result is consistent with the results of Ashby and Maddox (1992), who found that the GQC provided much better fits than the optimal classifier to data from several experiments with normally distributed categories, in which the optimal classifier was quadratic. Even so, Figure 4A shows that, for 3 observers, the best-fitting GQC bounds, although suboptimal, closely approximated the optimal bound.[5] Although the model fits described in Table 2 allow us to reject the assumption that the observers responded optimally, these same fits, together with the response accuracies shown in Figure 3, strongly support the class of nonparametric classifiers over the class of parametric classifiers. If the observers had been using a parametric classifier, the GLC should have fit better than the GQC. Instead, the GQC fit substantially better in every case.

In Experiment 2, the optimal classifier fits much better than the parametric classifier. Of these two models, note that the linear model fits better than the quadratic model in both experiments (i.e., in Experiment 1, the parametric classifier is linear, and in Experiment 2, the optimal classifier is linear). However, the advantage of the linear model over the quadratic model is much larger in Experiment 2 than in Experiment 1. Table 2 also indicates that the GLC fits better than the GQC in four of five cases and that the two best models were the GLC and the independent decisions classifier. The overall good performance of this latter model, however, is due primarily to its success with a single observer (i.e., Observer 4). For example, the median AIC score for the GLC is lower than that for the best independent decisions classifier. The superior performance of the GLC, relative to the GQC, and of the optimal classifier, relative to the parametric classifier, supports the class of nonparametric classifiers over the class of parametric classifiers.

In Experiment 2, a further test was provided by the transfer stimuli, which were positioned in the regions of stimulus space for which the parametric and optimal classifiers predicted contrasting responses. Call the response an observer makes to a transfer stimulus a *transfer response*. Table 3 lists the percentage of transfer responses that were correctly predicted by the best-fitting version of each model (assuming no noise).[6] For example, for Observer 5, the linear bound of the best-fitting GLC correctly partitioned 90% of the responses to the transfer stimuli, but the quadratic bound of the best-fitting GQC

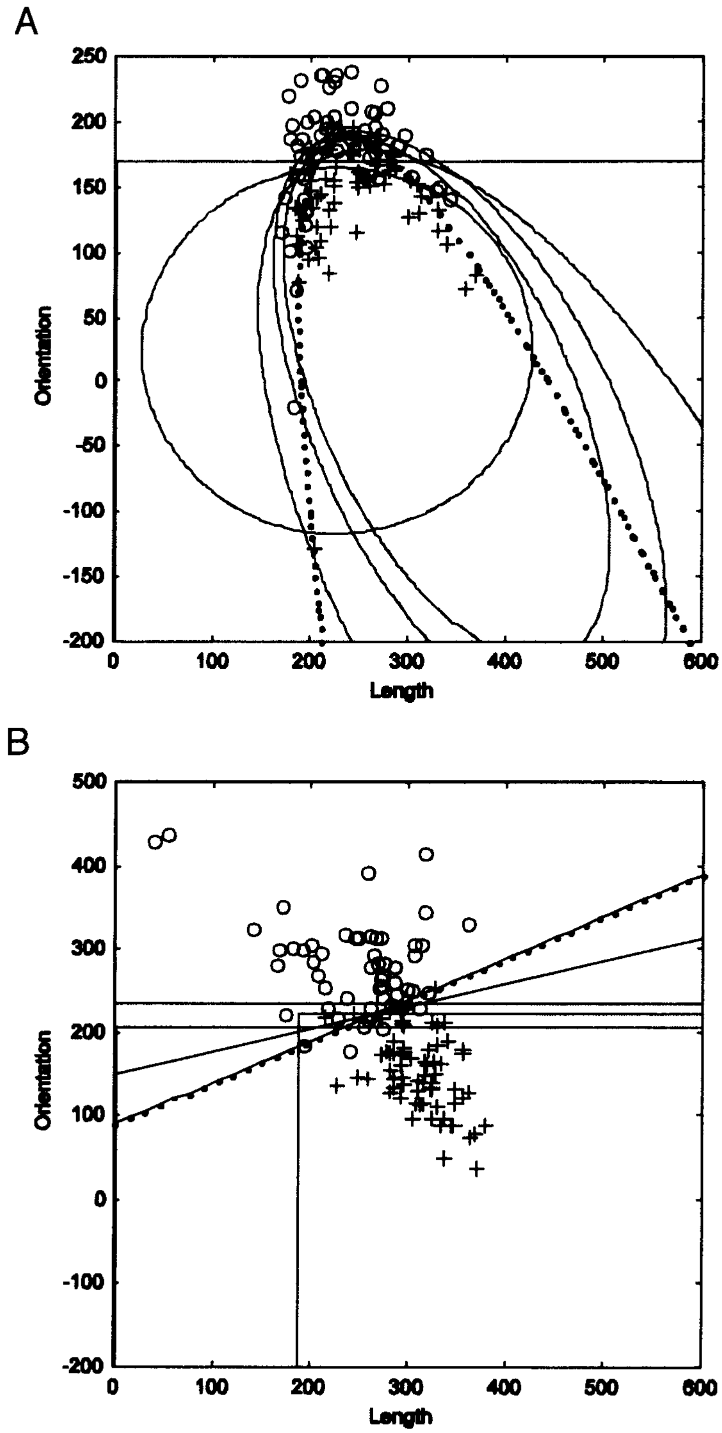

Figure 4. Decision bounds of best-fitting models for Experiments 1 (Figure 4A) and 2 (Figure 4B). Plus signs denote some of the exemplars from category *A*, and small circles denote some of the exemplars from category *B*. The optimal bound is dotted.

correctly partitioned only half of these responses. The transfer responses clearly favor the GLC and the GQC over the best unidimensional and independent decisions classifiers. They also slightly favor the GLC over the GQC. For example, the GLC is the only model that accounts for the transfer responses at least as well as all the other models for every observer.

Taken together, the superiority of the GLC over the GQC, both in goodness of fit and in accounting for the transfer responses, supports a nonparametric account of

Table 2
Goodness-of-Fit (AIC) Scores for Each Model When Fit to the Data From the Last Session of Each Experiment

Experiment	Observer	Optimal	Parametric	GLC	GQC	Best UDC	Best IDC
1	1	696.8	544.3	487.3	**384.7**	509.0	397.7
	2	697.6	550.4	562.4	**439.4**	574.5	452.8
	3	683.4	548.3	518.4	**383.7**	535.4	471.5
	4	686.1	578.1	579.6	**395.5**	597.6	398.0
	5	697.6	596.9	549.9	498.4	551.0	**495.0**
	M	692.3	563.6	539.5	**420.3**	553.5	443.0
2	1	178.6	665.0	**106.8**	114.3	120.2	121.0
	2	**187.9**	667.2	189.1	196.9	199.0	190.5
	3	278.2	666.0	273.0	279.4	273.2	**271.5**
	4	188.0	674.0	156.5	150.6	155.6	**137.6**
	5	356.3	664.4	309.1	314.2	**307.5**	309.6
	M	237.8	667.3	206.9	211.3	211.1	**206.0**

Note—AIC scores of the best-fitting model are in bold. GLC, general linear classifier; GQC, general quadratic classifier; UDC, unidimensional classifier; IDC, independent decisions classifier.

performance in Experiment 2 over a parametric account. Thus, together, the results of Experiments 1 and 2 strongly suggest that human pattern classification is nonparametric, rather than parametric.

POSSIBLE PARAMETRIC ACCOUNTS OF OUR RESULTS

The results from both experiments support the assumption that implicit category learning is a nonparametric process. However, before committing to this notion, it is important to ask whether any parametric categorization scheme could produce the data described above. We can imagine two possibilities. The first is that people always begin with linear bounds, and if there is no linear bound that achieves adequate accuracy, they then try quadratic bounds. According to this hypothesis, the Experiment 1 observers searched through the set of linear bounds, discovered that none achieved adequate accuracy, and then began experimenting with quadratic bounds. In Experiment 2, a linear bound was optimal, so the initial search through the set of linear bounds was successful. Unfortunately, there are several problems with this hypothesis. First, no observer in Experiment 1 exceeded 73.2% correct during his or her last experimental session. Yet the most accurate linear bound (i.e., the parametric bound) achieved 83.2% correct. The analyses described above strongly support the hypothesis that none of the observers in Experiment 1 was using a linear bound by the end of the experiment, even though some linear bound would have significantly improved his or her accuracy. Therefore, it seems unlikely that the observers in Experiment 1 had rejected the entire class of linear bounds because of some dissatisfaction with their response accuracy. Second, Ashby and Maddox (1992) specifically tested this hypothesis in an experiment in which the optimal bound was quadratic. If the hypothesis is correct, the GLC should provide a better account of the data from the first few trials of the first experimental session than the GQC. However, Ashby and Maddox (1992) found that, for every observer, the GQC provided better fits than the GLC to the data from the first 100 trials (although neither model fit very well). Finally, there is a logical problem with this hypothesis—namely, that it would take an inordinate number of trials to reject every possible linear bound. For example, just to decide that the parametric bound in Experiment 1 is suboptimal requires about 222 trials (i.e., using a binomial test, with $\alpha = .05$, $1 - \beta = .80$, and an alternative hypothesis that accuracy equals 85%), and to decide that the parametric bound is the best linear bound would require testing and rejecting many other linear bounds. Thus, it appears that, if observers tried quadratic bounds only after all the possible linear bounds were explicitly rejected, only extremely experienced observers would use quadratic bounds.

A second possible parametric explanation of our data is that observers estimate moments[7] higher than the variance. A parametric classifier can respond optimally in Experiments 1 and 2 only if it recognizes that the category distributions in these studies are nonnormal. Since the distributions are all unimodal and continuous valued, the only way a parametric classifier could infer nonnormality is to estimate moments higher than the variance. For example, in normal distributions, the third central moment is always zero [i.e., $E(\mathbf{X} - \mu)^3 = 0$], whereas it is not

Table 3
Percent Agreement Between Observed Responses to Transfer Stimuli and Responses Predicted by the Best-Fitting Decision Bound Models

Observer	GLC	GQC	Best UDC	Best IDC
1	**95**	**95**	55	30
2	**60**	**60**	30	40
3	**55**	50	30	**55**
4	**100**	**100**	**100**	**100**
5	**90**	50	75	75

Note—The highest percentages are in bold. GLC, general linear classifier; GQC, general quadratic classifier; UDC, unidimensional classifier; IDC, independent decisions classifier.

zero on either dimension in any of the categories used in Experiments 1 and 2. Of cqurse, knowledge of nonnormality is not enough to infer the correct parametric form of the optimal decision boundary, and the unusual method by which we constructed the categories in Experiments 1 and 2 makes it virtually impossible that observers could correctly guess the form of the distributions. Thus, to respond optimally in Experiments 1 and 2, a parametric classifier would have to estimate, not only the first, second, and third moments, but also moments considerably higher than the third. In addition, these estimates would have to be quite accurate. Unfortunately, this latter requirement is virtually impossible to satisfy, because even the best estimators (e.g., maximum likelihood) of higher moments have extremely large standard errors (see, e.g., Kendall & Stuart, 1977; Ratcliff, 1979). For example, whereas the standard error of the sample mean is $\sigma/\sqrt{n}$ (where n is sample size and σ is the population standard deviation), the standard error of the sample variance is $1.414\sigma^2/\sqrt{n}$, and the standard error of the third sample moment is $2.45\sigma^3/\sqrt{n}$ (i.e., these are the standard errors when the parent distribution is normal). Thus, even with perfect memory, accurate estimation of the variance requires such a large sample size that it is probably unrealistic to expect it of human observers, and estimation of the third and higher moments will usually be at least an order of magnitude more difficult.

Because of these arguments, we believe it is extremely unlikely that any plausible model of category learning that assumes, or is equivalent to, a parametric classifier could account for the results of Experiments 1 and 2. The only plausible account of our data is that human category learning is nonparametric. Even so, many alternative categorization models are equivalent to some nonparametric classifier. As was mentioned above, this includes models that assume that observers learn to associate responses with regions of perceptual space (e.g., Ashby & Maddox, 1989) and models that assume that observers access the memory traces of previously seen exemplars from the contrasting categories (i.e., so-called exemplar models; e.g., Brooks, 1978; Estes, 1986; Medin & Schaffer, 1978; Nosofsky, 1986). The critical distinction here is whether implicit category learning uses a form of procedural memory, as is assumed by models of the former type, or an instance-based memory system (e.g., either episodic or semantic), as is assumed by the exemplar models. Our data do not allow a test between these alternatives. However, a number of recent neuropsychological results raise problems for instance-based memory accounts of category learning, so we believe that, at present, the most parsimonious theory is that a major component of human category learning is a procedural-memory-based system that gradually associates response labels with regions in stimulus space. Before developing this idea more formally, we will briefly review the relevant neuropsychological evidence.

NEUROPSYCHOLOGICAL EVIDENCE AGAINST INSTANCE-BASED ACCOUNTS OF HUMAN CATEGORY LEARNING

As was mentioned above, exemplar theory assumes that people assign objects to categories by accessing memory traces (perhaps subconsciously) of exemplars from the relevant categories (see, e.g., Brooks, 1978; Estes, 1986; Medin & Schaffer, 1978; Nosofsky, 1986). As such, exemplar theory hypothesizes that categorization depends on instance-based or exemplar-based memory. Thus, if exemplar theory is correct, people with an impaired ability to store (or consolidate) the memory of previously seen exemplars should also be impaired in category learning.

Although this seems a straightforward prediction, a complication arises because exemplar theorists have not taken a strong position about the details of their hypothesized instance-based memory. The critical issue seems to be whether the associated trial-dependent context is stored along with the instance (i.e., the exemplar). Context-rich memory of an instance is frequently called *episodic memory*, whereas memory of an instance without the associated context is often called *semantic memory* (Tulving, 1972).

There is strong evidence that episodic memory is relatively unimportant in normal human category learning. For example, patients with medial temporal lobe amnesia, who have impaired episodic memory, have been found to perform normally on a variety of different category-learning tasks[8] (Knowlton, Ramus, & Squire, 1992; Knowlton & Squire, 1993; Kolodny, 1994). To account for this result, some exemplar theorists have argued that exemplar memory is intact in amnesic patients and that their only problem is that they have lost conscious access to those memories (e.g., Higham & Vokey, 1994), or that "amnesiacs may have more difficulty than normals in discriminating among distinct exemplars in memory" (Nosofsky & Zaki, 1998, p. 249). However, if either of these hypotheses were correct, it would seem that medial temporal lobe amnesiacs should have equal trouble with the recall of *all* exemplar memories, whereas the data show that anterograde amnesia is more likely and generally more severe than retrograde amnesia (see, e.g., Zola, 1997).

This latter result is consistent with a common view of medial temporal lobe amnesia in which damage to hippocampal structures impairs memory consolidation, rather than retrieval or the ability to discriminate among distinct memory traces (e.g., Gluck & Myers, 1997; McClelland, McNaughton, & O'Reilly, 1995; Polster, Nadel, & Schacter, 1991; Squire & Alvarez, 1995).

Another problem for episodic-memory-based accounts of category learning comes from several neuropsychological studies that have demonstrated a double dissociation between category learning and recognition memory. There is strong agreement that recognition memory requires intact episodic memory, so if category learning also uses episodic memory, performance on these two tasks should be highly correlated. In contrast to this prediction, as was mentioned above, amnesic patients, with impaired episodic memory, perform normally on a variety of different category-learning tasks, even though their recognition memory is severely impaired (Knowlton et al., 1992; Knowlton & Squire, 1993; Kolodny, 1994; Squire & Knowlton, 1995). The opposite dissociation has also been shown. Knowlton, Mangels, and Squire (1996) reported that patients with Parkinson's disease have normal recognition memory but have impaired category learning. Filoteo, Maddox, and Davis (1998) also reported category-learning deficits in Parkinson's disease patients. To date, there are no exemplar-based accounts of this latter dissociation. A number of studies have established a related double dissociation with nonhuman animals (e.g., Malamut, Saunders, & Mishkin, 1984; McDonald & White, 1993, 1994; Packard, Hirsch, & White, 1989; Packard & McGaugh, 1992; Packard & White, 1991).

A possibility that is more difficult to refute is that the instance-based memory assumed by exemplar theory is a context-free semantic memory. For example, according to this hypothesis, double dissociations between category learning and recognition memory are possible because recognition memory tasks require intact episodic memory, whereas category-learning tasks only require intact semantic memory. Therefore, if semantic memory is intact in medial temporal lobe amnesia, recognition memory would be impaired, but not category learning. It also makes sense that, to learn the structure of a category, it is not necessary to store all the contextual cues associated with the presentation of the category exemplars.

Despite its appeal, there are also problems with this semantic-memory-based account of category learning. For example, a number of studies have reported that medial temporal lobe damage often results in both episodic and semantic memory deficits (e.g., Ostergaard, 1987; Shimamura & Squire, 1987, 1991), so medial temporal lobe amnesiacs with intact semantic memory and impaired episodic memory may be rare. Recently, however, some researchers have argued that damage restricted to the hippocampus proper impairs context-rich episodic memory, but not context-free semantic memory, whereas more widespread medial temporal lobe damage that includes the parahippocampal region impairs both forms of memory (Eichenbaum, 1997; Vargha-Khadem et al., 1997). If this hypothesis is correct, then, according to semantic-memory-based accounts of category learning, amnesic patients with damage restricted to the hippocampus should be impaired in recognition memory, but not in category learning, whereas amnesic patients with damage to the hippocampus and the parahippocampal region should be impaired in both tasks. Few published studies have examined this issue specifically. However, several studies have reported results from patients with massive medial temporal lobe damage, which included damage to the parahippocampal region, who learned normally in complex categorization tasks (Filoteo, Maddox, Davis, & Hopkins, 1996; Squire & Knowlton, 1995). In fact, the Filoteo et al. (1996) study used a task that was virtually identical to the one shown in Figure 1B. Thus, although more data are needed to answer this question completely, the early results raise problems for the hypothesis that an intact context-free semantic memory system is crucial for normal category learning.

As presently formulated, it is difficult to see how instance-based theories of category learning (e.g., exemplar theory) can account for these results. It may be possible, however, that a reformulation or elaboration of some instance-based theory might prove more successful. For example, Nosofsky and Zaki (1998) provided an important first step in such a reformulation. Thus, the point of this section is not to argue that the neuropsychological evidence falsifies exemplar models. Such a conclusion would be premature. Rather, the purpose of this section is to argue that the neuropsychological data provide a significant challenge to current exemplar models. For this reason, we believe it is prudent to investigate alternative nonparametric theories of category learning. In particular, the neuropsychological data are easily and intuitively accounted for by the notion that category learning involves a process of associating responses with regions of perceptual space. According to this idea, category learning is a form of procedural learning that uses procedural memories, not instance-based memories. In contrast, recognition memory does require accessing exemplar memory traces. Thus, amnesic patients, with impaired episodic memory but intact procedural learning, are impaired in recognition memory but relatively normal in category learning. At the same time, Parkinson's patients, with impaired procedural learning but relatively normal episodic memory, are impaired in category learning and relatively normal in recognition memory.

A NONPARAMETRIC PROCEDURAL-LEARNING-BASED MODEL

Recently, there has been a surge of interest in the neural mechanisms and processes that mediate human category learning (e.g., Ashby et al., 1998; Knowlton, Mangels, & Squire, 1996; Smith et al., 1996; Squire, 1992). Much of the evidence indicates an important role for the striatum—a region of the basal ganglia that includes the caudate nucleus and the putamen. For example, there are recent reports that patients with striatal dysfunction (including those with either Parkinson's or Huntington's disease) are impaired in category learning (e.g., Filoteo et al., 1998; Knowlton, Mangels, & Squire, 1996; Knowlton, Squire, et al., 1996). In one such report, Filoteo et al. (1998) found that Parkinson's patients were impaired (as were Huntington's disease patients) relative to age-matched controls, in a study using a design that was essentially the same as the one shown in Figure 1B (i.e., normally distributed categories, quadratic optimal bound). In addition, lesions of the caudate nucleus in rats and monkeys have been

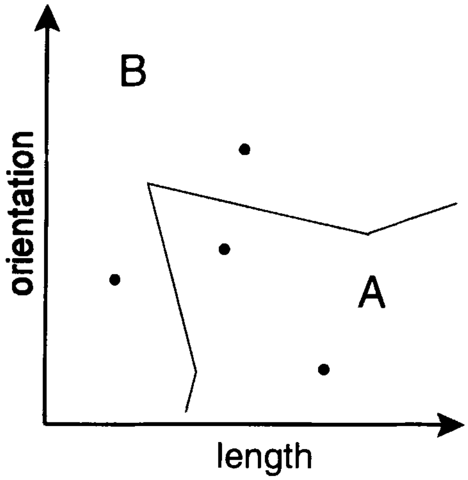

Figure 5. A simplified version of the striatal pattern classifier. The dots represent four different striatal units, and the solid line is the "boundary" that separates the regions of perceptual space assigned to the two category responses.

shown to disrupt simple forms of category learning (see, e.g., Buerger, Gross, & Rocha-Miranda, 1974; Divac, Rosvold, & Szwarcbart, 1967; McDonald & White, 1993, 1994; Packard, Hirsch, & White, 1989; Packard & McGaugh, 1992; Wang, Ainger, & Mishkin, 1991, cited by Petri & Mishkin, 1994). There is also substantial evidence that the striatum plays a key role in procedural learning (e.g., Jahanshahi, Brown, & Marsden, 1992; Mishkin, Malamut, & Bachevalier, 1984; Saint-Cyr, Taylor, & Lang, 1988; Willingham, Nissen, & Bullemer, 1989), so it is plausible that the striatum is mediating the procedural learning that we have argued is operating during categorization.

Neuroanatomical data suggest that the striatum, which is the input structure within the basal ganglia, is particularly well suited for such pattern association. First, the striatum receives projections from virtually all areas of the neocortex, including extrastriate visual cortical areas (Saint-Cyr, Ungerleider, & Desimone, 1990). These projections are known to be both diffuse and highly convergent, in the sense that many cortical afferents converge on relatively few striatal units and any single cortical afferent makes contact with many striatal units. Second, the striatum is an area with a high degree of synaptic plasticity, much of which is mediated by dopaminergic projections from the substantia nigra (pars compacta) that fire selectively in the presence of reward (Schultz, Apicella, & Ljungberg, 1993; Stein & Belluzi, 1989; Wickens, 1993). Furthermore, the basal ganglia is known to have prominent projections to the prefrontal cortex and motor output areas (i.e., via the thalamus; see, e.g., Alexander, DeLong, & Strick, 1986). On the basis of the above evidence, it has been suggested that the basal ganglia functions to associate a particular pattern of cortical activation with a motor response (e.g., Rolls, 1994; Wickens, 1993). Together, the neuropsychological and anatomical data suggest that the cortical-striatal-cortical system may be a good candidate for the neural substrate of perceptual classification (see Ashby et al., 1998, for a much more thorough discussion of these data).

A relatively simple model of visual pattern classification emerges from a consideration of this architecture. First, it is assumed that stimuli are represented in a *perceptual space* somewhere in higher level visual areas, such as the inferotemporal cortex. Because of the convergence of afferents from the cortex to the basal ganglia, it is proposed that a low-resolution map of perceptual space is represented among the striatal units. The information loss from this mapping could account for the suboptimal performance sometimes observed with complex category bounds (e.g., as in Experiment 2 of McKinley & Nosofsky, 1995). Through learning, the striatal units become associated with one of the category labels, so that, after learning is complete, a category response label is associated with each of a number of different regions of perceptual space. In effect, the striatum has associated a response with clumps of cells in the visual cortex. We call this model the *striatal pattern classifier* (SPC).

A simplified version of the SPC is illustrated in Figure 5. The two-dimensional length–orientation space depicts the perceptual representation of the lines used in our experiments. Thus, each point in this space is associated with a distinct cell in some extrastriate visual area (i.e., the cell maximally stimulated when a particular stimulus is shown). The four large dots represent four different striatal units. Each striatal unit is associated with one of the two category responses, which creates four distinct regions in perceptual space. In Figure 5, two of those regions are associated with category *A*, and two are associated with category *B*.

A number of authors have proposed models that are highly similar to the SPC. This includes the grid model of Ashby and Maddox (1989), the covering version of Kruschke's (1992) ALCOVE model (but not the more widely used exemplar-based version of ALCOVE), and Anderson's (1991) rational model. In all of these models, a low-resolution grid is mapped onto perceptual space, and a decision is made on the basis of which grid points are activated by the stimulus.

Note that, with only two striatal units, the SPC always predicts linear decision boundaries. In fact, by moving the

Table 4
Goodness-of-Fit (*SSE*) Scores When the General Quadratic Classifier, General Linear Classifier, and Striatal Pattern Classifier Are Fit to the Data From the Last Session of Experiment 1

Observer	GQC	GLC	SPC
1	52.3	56.5	52.6
2	67.8	74.7	68.9
3	79.4	83.0	79.7
4	59.4	67.9	61.9
5	76.9	78.6	77.3
M	67.2	72.1	68.1

Note—GQC, general quadratic classifier; GLC, general linear classifier; SPC, striatal pattern classifier.

two units around, the SPC can reproduce *any* linear bound. Thus, with two striatal units, the SPC is equivalent to the GLC (both models have three free parameters). As such, a two-unit version of the SPC can account for the data of Experiment 2 as well as can the GLC. Because the bound separating any two striatal units is linear, for any finite number of striatal units, the SPC predicts that the boundary separating the category *A* and category *B* regions of perceptual space is piecewise linear. Therefore, the SPC is never equivalent to the GQC. However, by varying the number and response assignments of the striatal units, the SPC "boundary" can provide an arbitrarily close approximation to the GQC boundary (or to virtually any decision bound). Because of this, one important difference between the SPC and the GQC is that the SPC is a nonparametric classifier, whereas, as was already mentioned, the GQC is parametric. However, because of its ability to mimic the GQC, some version of the SPC can account for the Experiment 1 data as well as can the GQC.[9]

The model described in Figure 5 is incomplete. To account for learning data, algorithms must be incorporated that specify how the striatal units become associated with the various categories (one possibility was proposed by Ashby & Maddox, 1989), how the number of units is selected, and how the visual units are mapped onto the striatal units. For example, these latter problems might be solved by Kohonen learning (e.g., Haykin, 1994; Kohonen, 1982). Once such details are added, the SPC would provide a model of the implicit system of the recent competition between verbal and implicit systems (COVIS) model of category learning (Ashby et al., 1998). COVIS assumes there are multiple category-learning systems but that the two most important are an explicit hypothesis- or theory-testing system (i.e., the verbal system) and an implicit procedural-learning-based system. Ashby et al. (1998) hypothesized, however, that an exemplar-based system and a perceptual priming system might also contribute to category learning under certain specialized conditions.[10]

The SPC is a procedural-learning-based account of categorization that we propose is valid under conditions in which the observer is not using some explicit rule.[11] It is a procedural-learning-based model because the striatal units learn to associate percepts with actions (i.e., category responses) in an incremental and implicit fashion (in a manner specified by COVIS). Note that instance- or exemplar-based memories are never accessed in this model and the hippocampus is never activated (or any other medial temporal lobe structure). For these reasons, when incorporated into COVIS, the SPC accounts for the neuropsychological data discussed in the previous section in a natural and intuitive fashion. According to this account, medial temporal lobe amnesic patients are relatively normal in category learning because instance-based memories are generally not accessed during categorization (but see note 10). Parkinson's patients are frequently impaired because the loss of dopamine in the striatum impairs the incremental learning process through which the striatal units become associated with category responses. A much more thorough discussion of these issues can be found in Ashby et al. (1998), who also show that the predictions of COVIS are generally consistent with category-learning data from other special neuropsychological populations (Huntington's disease, major depression, patients with frontal lesions, elderly people, children, and nonhuman animals).

CONCLUSIONS

The validity of categorization models can be evaluated either by goodness-of-fit testing or by testing the axioms that are used to build the models. Although the categorization literature has been dominated by the former approach, this article attempts to use the latter. Specifically, we attempted to determine whether the process through which people learn the structure of new categories is parametric or nonparametric. Ashby and Alfonso-Reese (1995) showed that current categorization models sharply disagree about this assumption.

The data from the two experiments strongly supported the assumption that human pattern classification is nonparametric. In fact, it would be extremely difficult for a parametric model to account for the results presented here. Thus, our results rule out prototype models and many decision bound models. They also rule out any model that assumes that observers compute a decision function of any specific functional form. On the other hand, our results do not allow us to discriminate among a variety of different nonparametric models. One way to reduce the set of candidate models further is to examine a different basic assumption that divides the nonparametric models into two classes. In this article, we focused on the assumption that category learning relies on an instance-based memory system (as is assumed by exemplar models) versus the assumption that it relies on a procedural-memory-based system. Although more work needs to be done on this problem, the present evidence favors procedural memory over instance-based memory. On the basis of these results, we proposed a category-learning model, called the SPC, that is nonparametric and relies on procedural learning and memory.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, **9**, 357-381.

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

Ashby, F. G. (1992a). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449-483). Hillsdale, NJ: Erlbaum.

Ashby, F. G. (1992b). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, **39**, 216-233.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, **105**, 442-481.

Ashby, F. G., & Gott, R. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, **120**, 150-172.

Ashby, F. G., & Lee, W. W. (1992). On the relationship among identification, similarity and categorization: Reply to Nosofsky and Smith (1992). *Journal of Experimental Psychology: General*, **121**, 385-393.

Ashby, F. G., & Maddox, W. T. (1989, November). *Toward a theory of natural categorization*. Paper presented at the 30th Annual Meeting of the Psychonomic Society, Atlanta.

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 598-612.

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 50-71.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.

Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Handbook of perception and cognition: Measurement, judgment, and decision making* (pp. 251-301). San Diego: Academic Press.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169-211). Hillsdale, NJ: Erlbaum.

Buerger, A. A., Gross, C. G., & Rocha-Miranda, C. E. (1974). Effects of ventral putamen lesions on discrimination learning by monkeys. *Journal of Comparative & Physiological Psychology*, **86**, 440-446.

Divac, I., Rosvold, H. E., & Szwarcbart, M. K. (1967). Behavioral effects of selective ablation of the caudate nucleus. *Journal of Comparative & Physiological Psychology*, **63**, 184-190.

Eichenbaum, H. (1997). How does the brain organize memories? *Science*, **277**, 330-332.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, **127**, 107-140.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, **18**, 500-549.

Filoteo, J. V., Maddox, W. T., & Davis, J. (1998, April). *Probabilistic category learning in patients with amnesia, Huntington's disease, or Parkinson's disease: The role of the hippocampus and basal ganglia*. Paper presented at the Fifth Annual Meeting of the Cognitive Neuroscience Society, San Francisco.

Filoteo, J. V., Maddox, W. T., Davis, J., & Hopkins, R. O. (1996, April). *Quantitative modeling of category learning in a patient with medial temporal lobe damage: Beyond accuracy scores.* Paper presented at the Third Annual Meeting of the Cognitive Neuroscience Society, San Francisco.

Flannagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 241-256.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 234-257.

Gluck, M. A., & Myers, C. E. (1997). Psychobiological models of hippocampal function in learning and memory. *Annual Review of Neuroscience*, **48**, 481-514.

Haykin, S. S. (1994). *Neural networks: A comprehensive foundation.* New York: Macmillan.

Higham, P. A., & Vokey, J. R. (1994). Recourse to stored exemplars is not necessarily explicit: A comment on Knowlton, Ramus, and Squire (1992). *Psychological Science*, **5**, 59.

Homa, D. (1978). Abstraction of ill-defined form. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 407-416.

Homa, D., & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 83-94.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 418-439.

Hyman, R., & Frost, N. H. (1975). Gradients and schema in pattern recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 630-654). New York: Academic Press.

Jahanshahi, M., Brown, R. G., & Marsden, C. (1992). The effect of withdrawal of dopaminergic medication on simple and choice reaction time and the use of advance information in Parkinson's disease. *Journal of Neurology, Neurosurgery, & Psychiatry*, **55**, 1168-1176.

Kendall, M. G., & Stuart, A. (1977). *The advanced theory of statistics.* New York: Macmillan.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, **273**, 1399-1402.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, **3**, 172-179.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category level knowledge. *Science*, **262**, 1747-1749.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, **1**, 106-120.

Knowlton, B. J., Squire, L. R., Paulsen, J. S., Swerdlow, N. R., Swenson, M., & Butters, N. (1996). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology*, **10**, 538-548.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Proceedings of the 6th International Conference on Pattern Recognition* (pp. 114-128). Silver Spring, MD: IEEE Computer Society Press.

Kolodny, J. A. (1994). Memory processes in classification learning: An investigation of amnesic performance in categorization of dot patterns and artistic styles. *Psychological Science*, **5**, 164-169.

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

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

Malamut, B. L., Saunders, R. C., & Mishkin, M. (1984). Monkeys with combined amygdalo-hippocampal lesions succeed in object discrimination learning despite 24-hour intertrial intervals. *Behavioral Neuroscience*, **98**, 759-769.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, **102**, 419-457.

McDonald, R. J., & White, N. M. (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience*, **107**, 3-22.

McDonald, R. J., & White, N. M. (1994). Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. *Behavioral & Neural Biology*, **61**, 260-270.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 128-148.

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

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

Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: Two neural systems. In G. Lynch, J. L. McGaugh, & N. M. Weinberger (Eds.), *Neurobiology of human learning and memory* (pp. 65-77). New York: Guilford.

Myung, I. J. (1994). Maximum entropy interpretation of decision bound and context models of categorization. *Journal of Mathematical Psychology*, **38**, 335-365.

Neumann, P. G. (1977). Visual prototype formation with discontinuous representation of dimensions of variability. *Memory & Cognition*, **5**, 187-197.

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

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesiacs and normals: An exemplar-based interpretation. *Psychological Science*, **9**, 247-255.

Ostergaard, A. L. (1987). Episodic, semantic and procedural memory in a case of amnesia at an early age. *Neuropsychologia*, **25**, 341-357.

Packard, M. G., Hirsch, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience*, **9**, 1465-1472.

Packard, M. G., & McGaugh, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behavioral Neuroscience*, **106**, 439-446.

Packard, M. G., & White, N. M. (1991). Dissociation of hippocampus and caudate nucleus memory systems by post-training intracerebral injection of dopamine agonists. *Behavioral Neuroscience*, **105**, 295-306.

Petri, H. L., & Mishkin, M. (1994). Behaviorism, cognitivism and the neuropsychology of memory. *American Scientist*, **82**, 30-37.

Polster, M. R., Nadel, L., & Schacter, D. L. (1991). Cognitive neuroscience analyses of memory: A historical perspective. *Journal of Cognitive Neuroscience*, **3**, 95-116.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446-461.

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

Rolls, E. T. (1994). Neurophysiology and cognitive functions of the striatum. *Reviews in Neurology*, **150**, 648-660.

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

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostriatal dysfunction in man. *Brain*, **111**, 941-959.

Saint-Cyr, J. A., Ungerleider, L. G., & Desimone, R. (1990). Organization of visual cortical inputs to the striatum and subsequent outputs to the pallido-nigral complex in the monkey. *Journal of Comparative Neurology*, **298**, 129-156.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience*, **13**, 900-913.

Shaw, M. L. (1982). Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology*, **14**, 353-409.

Shimamura, A. P., & Squire, L. R. (1987). A neuropsychological study of fact memory and source amnesia. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 464-473.

Shimamura, A. P., & Squire, L. R. (1991). The relationship between fact and source memory: Findings from amnesic patients and normal subjects. *Psychobiology*, **19**, 1-10.

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

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, **65**, 167-196.

Smith, E. E., Patalano, A. L., Jonides, J., & Koeppe, R. A. (1996, November). *PET evidence for different categorization mechanisms*. Paper presented at the 37th Annual Meeting of the Psychonomic Society, Chicago.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, **99**, 143-145.

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: A neurobiological perspective. *Current Opinion in Neurobiology*, **5**, 169-177.

Squire, L. R., & Knowlton, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences*, **92**, 12470-12474.

Stein, L., & Belluzi, J. (1989). Cellular investigations of behavioral reinforcement. *Neuroscience & Behavioral Reviews*, **13**, 69-80.

Takane, Y., & Shibayama, T. (1992). Structures in stimulus identification data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 335-362). Hillsdale, NJ: Erlbaum.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York: Academic Press.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, **277**, 376-380.

Wickens, J. (1993). *A theory of the striatum*. New York: Pergamon.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 1047-1060.

Zola, S. (1997). Amnesia: Neuroanatomic and clinical aspects. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral neurology and neuropsychology* (pp. 447-461). New York: McGraw-Hill.

NOTES

1. The critical feature that makes the task implicit is not the use of normally distributed categories, but the precise nature of the optimal bound. If the optimal bound were linear and orthogonal to either of the stimulus dimensions, the task would be explicit. For example, if the optimal bound is orthogonal to the length dimension, the optimal rule is: Respond *A* if the line is short, and *B* if it is long.

2. Of course, even if the population variances and correlations all have identical values in the two categories, the sample variances and correlations will not be exactly equal. Presumably, an observer who assumes that the category distributions are normal would use a statistical criterion for equality of variance–correlation structure.

3. It is important to note that we are not claiming that prototype theory makes any optimality assumptions. Rather, the claim is only that the prototype model is *mathematically equivalent* to a model that makes strong distributional assumptions and that assumes that the observer responds optimally (under the constraints imposed by those distributional assumptions). For a more thorough discussion of this point, see Ashby and Alfonso-Reese (1995).

4. For example, consider a parametric classifier that assumes that the optimal bound is quadratic. Since the optimal bound in the McKinley and Nosofsky (1995) experiments was more complex than a quadratic bound, such a classifier might discover that no quadratic bound was adequate. This would inform the classifier that the category distributions were nonnormal.

5. Although it appears in Figure 4 that the best-fitting bound for one observer (i.e., Observer 5) is unidimensional, Table 2 indicates that the IDC fit the Observer 5 data substantially better than the best unidimensional classifier. The best-fitting version of the IDC for this observer had a different noise variance on each dimension, and with a large

enough variance on the length dimension, the vertical segment of the IDC bound improves goodness of fit (relative to a unidimensional classifier with a horizontal bound), even though all the stimuli from both categories fell on its right-hand side. This same scenario occurred for 1 observer in Experiment 2 (i.e., see Figure 4B).

6. The transfer data were not used in the model-fitting procedure (since no feedback was provided to transfer responses). Although the GLC can never provide a better absolute fit to the training data than the GQC (since the GLC is a special case of the GQC), there is no logical reason that the GLC bound that best fits the training data cannot predict the transfer responses better than does the GQC bound that best fits the training data.

7. The *n*th central moment of a random variable **X** is defined as $E(\mathbf{X} - \mu)^n$, where E denotes expected value and μ is the distribution mean.

8. One study reported that amnesic patients performed as well as controls during the first 50 trials of category learning, but thereafter showed a deficit (Knowlton, Squire, & Gluck, 1994). This study used 14 highly distinct stimuli, so it is possible that the amnesic deficit occurred because the control participants began memorizing the responses to some of the stimuli. This hypothesis is supported by the results of a study that used randomly configured dot patterns as stimuli (Kolodny, 1994). With confusable stimuli of this type, memorization is a more difficult strategy. In the Kolodny (1994) study, amnesiacs and controls each categorized several hundred dot patterns, yet there was no accuracy difference between the two groups, even during the last test block.

9. For example, we fit a version of the SPC with four striatal units to the data from the last experimental session of each observer in Experiment 1. This version of the model (shown in Figure 5) has the same number of free parameters as the GQC (i.e., seven). The fits of the GQC and SPC were virtually identical (see Table 4), which indicates that a version of the SPC with at most four grid points can account for the data from Experiments 1 and 2 about as well as can the best decision bound models.

10. For example, when there are only a few exemplars in each category, observers might actively memorize responses. Such a strategy would almost certainly depend heavily on medial temporal lobe structures.

11. COVIS assumes that the separate explicit and implicit category-learning systems compete throughout training (Ashby et al., 1998). Of the decision bound models tested in this article, the unidimensional and independent decisions classifiers assume that observers always use explicit rules (since unidimensional and independent decisions rules are easy to verbalize), whereas the optimal and parametric classifiers and the GLC and GQC all assume that observers use implicit rules (since these rules are almost always extremely difficult to verbalize). The good fits of the GQC in Experiment 1 and of the GLC in Experiment 2 indicate that most (but perhaps not all) of the observers were responding implicitly by the end of the experiment.

(Manuscript received February 11, 1998;
revision accepted for publication December 10, 1998.)