Journal of Experimental Psychology:
Learning, Memory, and Cognition
1997, Vol. 23, No. 6, 1362–1377

0278-7393/97/$3.00

Decision Boundaries in One-Dimensional Categorization

Michael L. Kalish and John K. Kruschke
Indiana University Bloomington

Decision-boundary theories of categorization are often difficult to distinguish from exemplar-based theories of categorization. The authors developed a version of the decision-boundary theory, called the single-cutoff model, that can be distinguished from the exemplar theory. The authors present 2 experiments that test this decision-boundary model. The results of both experiments point strongly to the absence of single cutoffs in most participants, and no participant displayed use of the optimal boundary. The range of nonoptimal solutions shown by individual participants was accounted for by an exemplar-based adaptive-learning model. When combined with the results of previous research, this suggests that a comprehensive model of categorization must involve both rules and exemplars, and possibly other representations as well.

In this article, we contrast decision-boundary and exemplar-based models of categorization. The decision-boundary model of categorization (Ashby & Gott, 1988; Ashby & Perrin, 1988; Ashby & Townsend, 1986) holds that people make category membership judgments by evaluating a rule that is based on the dimensional values of a test stimulus. As an example, with a one-dimensional stimulus a single-cutoff rule could be, "If the stimulus is larger than value *X*, then it is a member of Category A; otherwise it is a member of Category B." For multidimensional stimuli, the rule may have no simple verbal gloss, but it can still be described by a decision boundary that is a function of the stimulus dimensions. Additionally, it it not necessary to refer to the actual dimensions of the stimulus in the rule—a different function can be used as a basis for discrimination, such as the relative likelihood of a stimulus coming from each category (Ashby & Townsend, 1986; Nosofsky & Smith, 1992).

The decision-boundary theory also assumes that category membership decisions are made on an all-or-none basis, rather than on a graded probabilistic basis (Ashby & Lee, 1992). Given a stimulus with multiple dimensions, the observer computes the value of the discriminant function. If the function value is greater than a threshold value, the stimulus is assigned a particular category label. Of course, people generally do not respond absolutely deterministically in categorization tasks. Probabilistic responding, the decision-boundary model claims, occurs because perceptual noise sometimes makes a stimulus appear as though it is on the opposite side of the boundary and because errors in memory make the precise value of the threshold uncertain (Ashby & Lee, 1993; Ashby & Maddox, 1992, 1993).

In contrast, exemplar models propose different representations and decision processes (e.g., Nosofsky, 1986). In an exemplar model, when items are presented, each is stored in memory as a point in multidimensional space. When a test stimulus is presented for categorization, its similarity to each item in memory is computed. The relative similarity of the test item to members of the different categories is entered into a probabilistic decision function. The result of this computation is that category membership is decided in a graded fashion.

Thus, decision-boundary and exemplar models differ in two ways. One difference is the assumption of what representations are used during categorization; the exemplar model assumes only the storage of all exemplars, whereas the decision-boundary model assumes the storage of a decision boundary and a discriminant function. The leading decision-boundary theory (the general recognition theory, GRT) also allows that individual exemplars might be represented, but denies that they are accessed during categorization (Ashby, 1992; Ashby & Townsend, 1986; Maddox & Ashby, 1993). The other difference is the type of decision rule the models use. Decision-boundary models are typified by deterministic rules, whereas exemplar models use a graded probabilistic rule. Distinguishing between these models has proved difficult, as the two make formally identical predictions under a number of assumptions (Ashby & Maddox, 1993; Nosofsky & Smith, 1992) and make very similar predictions in many experimental conditions (Estes, 1992). For example, even Nosofsky's (1986) classic data from nonnormally distributed categories could be well fit by a likelihood-based decision-boundary theory.

Support for decision-boundary models falls into two

Michael L. Kalish and John K. Kruschke, Department of Psychology, Indiana University Bloomington.

This research was supported by National Institute of Mental Health (NIMH) Training Grant T32 MH19879-02 and by NIMH FIRST Award 1-R29-MH51572-01. We thank Nathaniel Blair, Michael Erickson, Michael Fragassi, Mark Johansen, Stephan Lewandowsky, and anonymous reviewers for comments on previous versions of this article. We thank Abbey Clawson, Beth Okeon, Laura Owen, Debbie Reas, and Daniel Vote for assistance in administering the experiments.

Correspondence concerning this article should be addressed to Michael L. Kalish, who is now at the Department of Psychology, University of Western Australia, Nedlands, Western Australia 6907 Australia. Electronic mail may be sent via Internet to Michael L. Kalish at kalish@psy.uwa.edu.au. Michael L. Kalish's World Wide Web address is http://boneyard.psy.uwa.edu.au/kalish.html, and John K. Kruschke's World Wide Web address is http://www.indiana.edu/~kruschke/home.html.

primary classes: those experiments that show deterministic responding and those that show a lack of access for individual exemplars. In experiments by Maddox and Ashby (1993), the extent to which responses were deterministic was measured by finding the slope of the *response surfaces* of individual participants. A response surface is the function that describes the probability of giving a particular category label to a stimulus given its dimensional values. The response surface is to be contrasted with the *posterior probability* of the category, given the stimulus that is the actual probability that a stimulus is a member of a given category. In Maddox and Ashby's (1993) study, participants were asked to decide the category membership of stimuli drawn from overlapping Gaussian distributions. When Maddox and Ashby measured the gradient of the response surface across the decision boundary (i.e., the slope of the line tangent to the surface at the decision boundary), they found it to be steeper than the gradient of the posteriors. This runs counter to the predictions of exemplar models, which often are construed to predict a response surface congruent with the posterior probability of category membership, or a *probability-matching* result. Thus, the conclusion was that responding was deterministic. Deterministic responding, as defined above, has been observed when participants were instructed to select criterial stimulus values for segregating two Gaussian distributions (Healy & Kubovy, 1981; Kubovy & Healy, 1977).

An additional prediction of the boundary-based approach is that exemplar-based information (such as the distribution of exemplars within categories), although possibly retained, is not used in categorization (Thomas, 1997; Thomas & Townsend, 1993). Thomas first trained participants to discriminate categories composed of highly correlated two-dimensional stimuli. Once the participants were able to reliably categorize the stimuli, showing that they had learned about the intracategory correlations, they were given a new task. Participants were presented with a stimulus that contained only one of the two relevant dimensions, along with the label of the category to which the stimulus belonged. Participants were then asked to select a value for the missing dimension. Many of the participants gave the mean value of the missing dimension, rather than the value that would have been expected given the strong intracategory correlation of stimulus dimensions. These results suggest that participants could not recall the intracategory structure they had exploited so well in learning the discrimination. This is consistent with the representational assumption of the decision-boundary model, in which only the decision boundary is retrieved from memory, and the information that leads to its adoption (in this case, the intracategory correlations) is not accessible.

Neither of these two lines of research is free of difficulties, however. The steepness of response surface gradients is potentially misleading. Response probabilities may have a steeper gradient than the posterior probability of category membership for at least two reasons. Either responses are deterministic on the basis of noisy perceptual processes as Maddox and Ashby (1993) suggested, or responses are probabilistic on the basis of the posterior probabilities. As long as response probabilities are a function of, rather than identical to, the posteriors, the exemplar model's prediction of probability matching cannot be dismissed. Maddox and Ashby developed a deterministic exemplar model that handled the steep response functions they found.

Similarly, Brooks and his colleagues (Allen & Brooks, 1991; Regehr & Brooks, 1993) have shown exemplar effects in rule use that contradict the forgetting seen by Thomas (1997). Moreover, only a subset of the participants in Thomas's study did not remember intracategory exemplar correlations. Anderson and Fincham (1996) also showed that people can use correlations learned during categorization to make predictions. However, an exemplar model did not fit their data well, especially when extrapolating the correlation. This points out the possibility of individual differences in the relative weighting of boundaries and exemplars, perhaps requiring a hybrid model of categorization (Kruschke & Erickson, 1994; Vandierendonck, 1995).

Much of the difficulty in distinguishing the exemplar and decision-boundary models is methodological. In experiments with a small number of discrete stimuli, the precise placement of the optimal decision boundary is often undefined. To define the optimal boundary, continuous distributions are necessary. However, in experiments with just two Gaussian-distributed categories, it is difficult to distinguish decision-boundary models from exemplar-based models. Both classes of models predict that stimuli near the equiprobability contour (the curve marking the set of stimuli that has an equal likelihood of belonging to each category) will be mislabeled approximately half the time and that mislabeling will diminish logistically with distance from that contour. Exemplar models make this prediction because responses are taken to be a monotonic function of the posteriors. Decision-boundary models made the same prediction because of both (a) the assumed shape of noise distributions that interact with deterministic responding, and (b) the effect that evolution has had in preparing people to cope with approximately Gaussian categories (Ashby & Gott, 1988). It has been shown (e.g., Ashby, 1992) that if the perceiver is responding deterministically in the absence of perceptual noise but with a chosen criterion that is itself subject to random Laplace-distributed noise, then the exemplar and decision-boundary models have an isomorphic functional form, as long as the categories are formed by Gaussian distributions.

The use of the normal distribution for categorization experiments thus makes discrimination of the exemplar and decision-boundary models extremely difficult, if not impossible. Normally distributed exemplars may characterize natural categories, and the assumption that all categories are composed of normal distributions may constrain our range of available decision boundaries (Ashby & Gott, 1988; Ashby & Maddox, 1993), but because of the problems described above, the use of normally distributed exemplars in categorization experiments must be reconsidered.

One way to generate contrasting predictions from the exemplar and decision-boundary models is to use categories composed of mixtures of Gaussians (McKinley & Nosofsky, 1995). A mixture distribution is one that is made up of two or more simple distributions added together. McKinley and

Nosofsky (1995) composed each of their two categories by adding together two Gaussians, each with low variance on one dimension and high variance on the other. The two Gaussians were chosen so that the high-variance dimension of one was the low-variance dimension of the other, and thus each mixture was L shaped. The optimal decision boundary in this case was cubic, rather than quadratic, which is the boundary one would observe if people were constrained to assume that the categories were each normally distributed. Participants were trained with the mixture distributions, and their final response surfaces were compared with the predictions of the normal-distribution decision-boundary model. The results showed that participants were not limited to linear or quadratic decision boundaries, as this boundary model requires. In contrast, the observed cubic response surfaces were well described by an exemplar model.

An alternative approach to distinguishing the decision-boundary and exemplar models is to alter, not the form of the discriminant function, but the shape of the posterior probability distribution to make it very different from the one that would occur as the result of the combination of a decision boundary and perceptual noise. In principle, any two non-Gaussian distributions will have a ratio that is not logistically distributed. In practice, however, the posteriors must diverge significantly from the logistic for the difference to be detected by an experiment. One way to make the posteriors diverge greatly from the logistic is to change the qualitative form of the ratio of category distributions. Although the ratio of two Gaussians is always either increasing or decreasing, the ratio of overlapping uniform distributions with unequal variance has an extended plateau; it is in essence a two-step function (as shown in Figure 1) instead of a one-step function like the logistic.

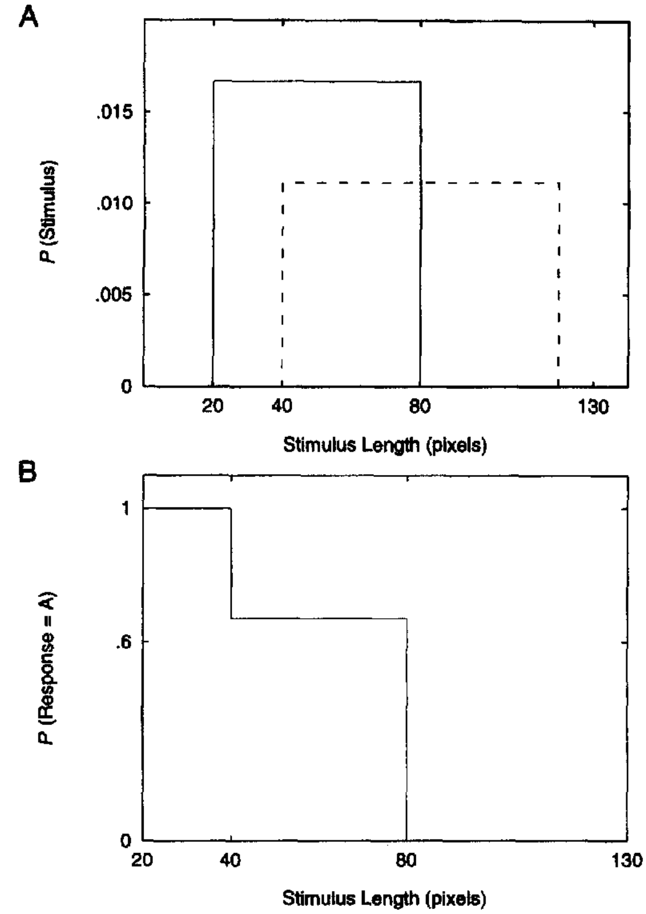

Figure 1. Two overlapping, uniform frequency distributions of different variance, used in Experiment 1. A: The frequency with which different-length stimuli were presented as members of Category A (solid line) and Category B (dashed line). At the extreme left, between stimulus lengths of 20 and 40 pixels, all stimuli were presented as members of Category A. In the overlapping region, from lengths of 40 to 80 pixels, both categories were possible, but stimuli were presented as members of Category A 60% of the time. At the extreme right, from stimulus lengths of 80 to 130 pixels, all stimuli were presented as members of Category B. B: How the uniform distributions led to a plateau of posterior probabilities. P = probability.

The Single-Cutoff Model

The simple verbal rule presented at the beginning of the article was a single-cutoff rule. The GRT predicts, more generally, that as long as one category (A) is more likely than the other (B), then the decision maker will always respond "A." Criterial or perceptual noise will cause the decision maker to sometimes choose the less likely category as a response. In the decision-boundary model, the probability of response "A" given stimulus s is formally

$$P(\text{Response} = \text{A}|s) = \begin{cases} 1 \text{ if } k(s) > c + \text{noise} \\ 0 \text{ otherwise,} \end{cases} \quad (1)$$

where $k(s)$ is the value of the discriminant function for stimulus s, and c is the decision criterion.

The predictions of the decision-boundary model in discriminating overlapping uniform distributions depend largely on the nature of the discriminant function (k). For $P(R = \text{A}|s)$ to increase with increasing s, k must be monotonically increasing. The log of the ratio of the likelihoods of Gaussians is one obvious choice for k and may reflect the basic nature of the environment. However, if k is the log ratio of the likelihoods of two overlapping uniform distributions, then a two-step response surface will result, but response probabilities will be undefined outside the range of the distributions. In one dimension, it is plausible to assume that k is just the psychological magnitude of the stimulus s, rather than being a likelihood ratio. The value of c then becomes the psychological magnitude above which stimuli are labeled "A" and below which stimuli are labeled "B." This model, the *single-cutoff model,* predicts a one-step response surface (as shown in Figure 1) regardless of the densities of the two categories being discriminated. The single-cutoff model incorporates a number of different restricted versions of the GRT, while preserving the basic intuition of what constitutes a rule in a simple categorization problem.

On the other hand, if people encode and retain all category exemplars equally, then the exemplar model can predict a form of matching in which there will be a plateau in the response probabilities.

Experiment 1: Overlapping Uniform Distributions

Participants were trained to distinguish two one-dimensional, overlapping, uniformly distributed categories (Figure 1). Previous research on one-dimensional categorization behavior (Busemeyer & Myung, 1992; Healy & Kubovy, 1981; Kubovy & Healy, 1977) has clearly shown that people can learn to respond as if they are using a decision boundary. The use of normal distributions in earlier research makes distinguishing exemplar and decision-boundary models difficult. In addition, conclusions about the psychological mechanisms responsible for the observed behavior in such studies were compromised by the averaging of response probabilities across subjects. For these reasons, uniform distributions were used in the present experiment to form the categories being learned, and the responses of each participant were considered individually. If participants used a single cutoff, then they should produce a response surface with a steeply sloped sigmoid shape, the gradient of which reflects only criterial and perhaps perceptual noise. If, instead, participants produced a two-step response surface or a response surface too shallow to be reasonably attributed to criterial noise, then the single-cutoff decision-boundary model is potentially disconfirmed.

Method

Participants. Forty-two Indiana University undergraduate students volunteered as part of a psychology course requirement.

Apparatus. Stimuli were presented on a video graphics array (VGA) resolution monitor by a PC-type computer. Each participant sat before a computer in an individual, sound dampened, dimly lit, ventilated cubicle. Responses were entered on a standard keyboard.

Procedure. Participants first read instructions indicating that they would be making probabilistic decisions (i.e., that they could never be completely accurate except by chance) about two kinds (categories) of stimuli that varied on only one dimension.

On each trial, a single horizontal rectangle, 20 pixels (each pixel was approximately 1.24 mm square) high and filled with a high-contrast blue color, was presented in a central screen location, subject to a random horizontal offset of between 0 and 25 pixels. Below the rectangle, a response prompt directed participants to indicate their classification judgment by pressing either the *K* or the *S* key on the keyboard. The computer provided corrective feedback after each trial. If the participant did not respond within 30 s, a message appeared on the screen instructing the participant to go faster, and if the response occurred within 50 ms of the presentation of the stimulus, the participant was warned to wait until the stimulus was presented to make a response.

Of the two categories, A and B, Category A was made up of rectangles drawn from a uniform distribution ranging in length from 20 to 80 pixels in increments of 2 pixels, and Category B was a uniform distribution of rectangles ranging in length from 40 to 130 pixels, also in increments of 2 pixels (see Figure 1). In every block of 180 trials, each of the 30 Category A stimuli appeared three times, and each of the 45 Category B stimuli appeared twice. Thus, stimuli in the overlapping region (between 40 and 80 pixels in length) were presented as Category A stimuli 60% of the time and as Category B stimuli 40% of the time.

Each participant completed four blocks of trials, each with a random ordering of trials. The name (*K* or *S*) given to each category was chosen randomly for each participant.

Results

The data from each participant's last training block were analyzed individually. The goal of the analysis was to determine the shape of the individual response surfaces. To that end, we used a nested model procedure to determine the best fitting empirical response surface. The most complex model surface that we considered allowed for two sigmoidal changes in response probabilities (two steps), each with independent slopes and locations, with the changeover occurring at a freely fitted response level. An example of this sort of surface is shown in Figure 2; the surface represents a form of probability matching in which participants produce probabilistic responses through part of the range of overlap of the two categories. The two-step model has three distinct features. Responses are nearly deterministic at either end of the stimulus scale, whereas responses in an area near the region of category overlap are made with a fixed probability. The transition from deterministic to probabilistic responding is assumed to be graded.

The first simplification of the two-step model is a one-step model in which participants are taken to switch from labeling stimuli as "Category A" to labeling stimuli as "Category B" without an intervening probabilistic response region. Both the point at which this switch occurs and the rate at which it occurs are free to vary in the general one-step model. These parameters are roughly coincident with the criterion and criterial noise parameters of the decision-boundary model. The simplest model tested is a one-step model in which the criterion is fixed to be the optimal criterion. This corresponds to the performance of an optimal categorizer operating under the influence of noise.

We chose to formalize the two-step model as a sigmoidal mixture of two sigmoids. The model begins with the step from deterministic "Category A" responding to probabilistic responding:

$$step_1(x) = s + \frac{(1 - s)}{1 + \exp[g_1(x - x_1)]}, \qquad (2)$$

where s is the probability of responding "A" in the region of category overlap, x_1 is the threshold of the sigmoid, and g_1 is the gain that determines the steepness of the sigmoid. The step from probabilistic responding to deterministic "B" responding is given by the following sigmoid:

$$step_2(x) = \frac{s}{1 + \exp\{g_2[x - (x_2)]\}}, \qquad (3)$$

where x_2 and g_2 are the threshold and gain of the sigmoid, respectively.

Finally, the two sigmoids are joined by the function

$$mix(x) = \frac{1}{1 + \exp[g_3(x - x_1 + d_T)]}, \qquad (4)$$

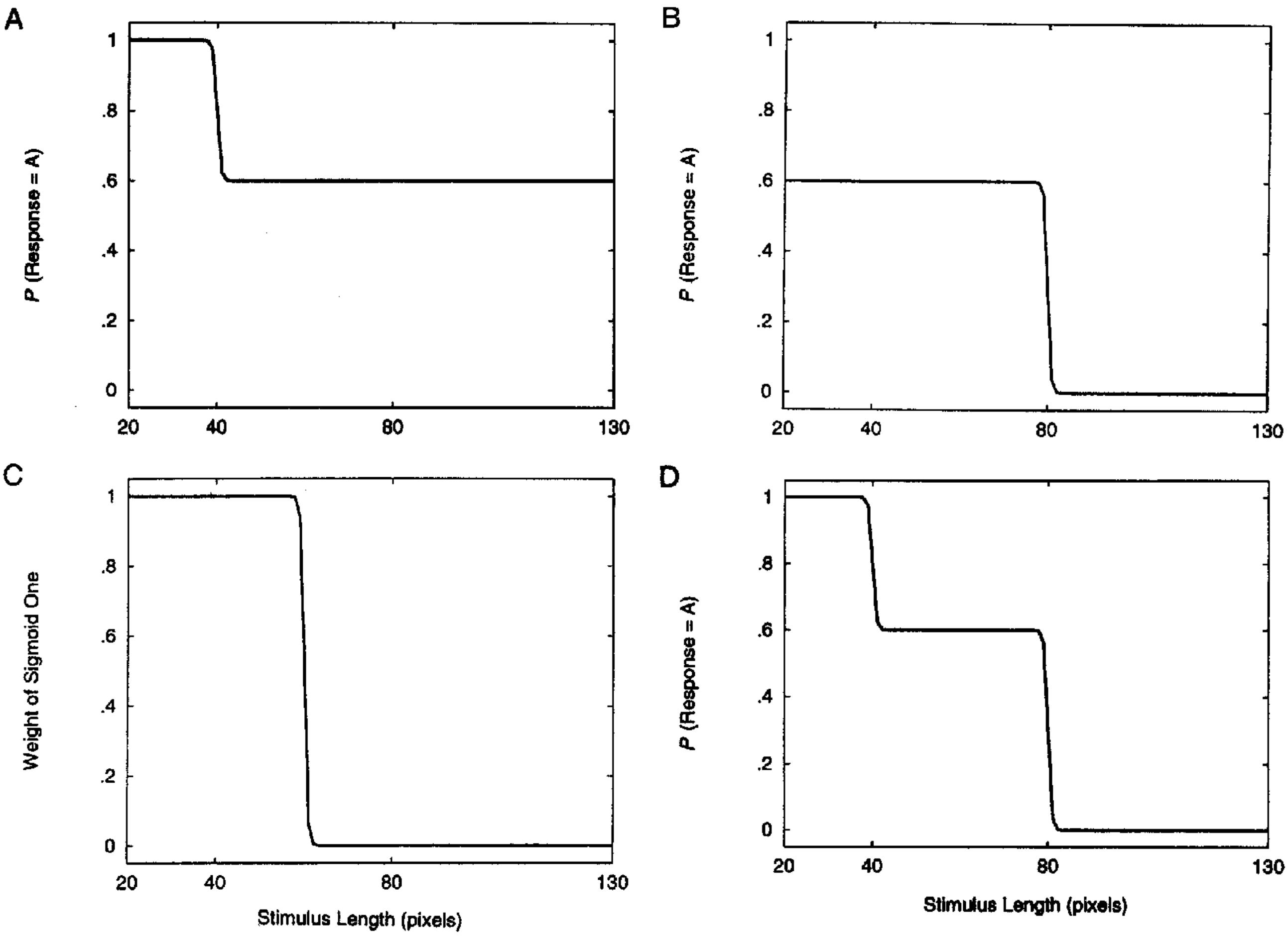

Figure 2. The composition of the general two-step model. A: For Equation 2, the step from deterministic "A" responding to probabilistic responding. B: For Equation 3, the step from probabilistic responding to deterministic "B" responding. C: For Equation 4, the function that determines the relative roles of the first two steps. D: For Equation 5, the two-step model of response probabilities. P = probability.

where d_T is the distance from the threshold of the first sigmoid to the threshold of the mixing sigmoid, and g_3 is the gain. The mixing sigmoid is used to create a single response surface:

$$P(\text{resp} = \text{A}|x) = mix(x)step_1(x) + [1 - mix(x)]step_2(x). \quad (5)$$

The full two-step model has seven free parameters (summarized in Table 1). By setting a number of the parameters appropriately, this two-step model can be reduced to a one-step sigmoid. In particular, if $s = .5$, $g_3 = 0$, $g_2 = g_1$, and $x_2 = x_1$, so that $step_2 = step_1$, then the response surface will be determined by the parameters of $step_1$ alone.

Table 1
Models Used in Describing the Results of Experiment 1

Model	Fixed parameters	Free parameters
Optimal one step	$g_3 = 0, g_2 = g_1, x_1 = x_2 = 80,$ $d_T = 0, s = .5$	g_1
Free threshold one step	$g_3 = 0, g_2 = g_1, x_2 = x_1, s = .5$	x_1, g_1
Two step	none	$g_1, g_2, g_3, x_1, x_2, d_T, s$

Note. g_1, g_2, and g_3 = gains that determine the steepness of the sigmoid; x_1 and x_2 = thresholds of the sigmoid; d_T = distance from the first sigmoid to the mixing sigmoid; s = probability of responding "A" in the region of overlap.

The one-step model can be further constrained such that the threshold is fixed at the statistically optimal criterion ($x_1 = 80$). The three models are summarized in Table 1. The full two-step model along with both the one-step model with free gain (g_1) and criterion (x_1) and the one-step model with the optimal criterion ($x_1 = 80$) were fit to each participant's data.

The stimulus range was divided into 17 bins, with midpoints and frequency indicated in Table 2. Thus, each participant's data consisted of a 17 × 2 frequency table, corresponding to the 17 stimulus bins and the 2 response categories. The marginal frequency of each bin was fixed by the experimental design, so there were 17 degrees of freedom in the data (minus the number of free parameters in each model).

The fit of each model to each participant's data was measured with the likelihood statistic:

$$G^2 = \sum_{i=1}^{34} R_i \log \frac{R_i}{Pr_i},$$

where R is the number of observed responses in each cell, and Pr is the number of responses the model predicts. The parameters of the models were optimized with a gradient descent method.

The optimized models were tested for goodness of fit by a Monte Carlo method. Neither a Pearson chi-square test nor Hosmer and Lemeshow's (1989) test could be used because the observed frequencies of one or the other response were zero in too many adjacent cells. The predicted response probabilities derived from the model were used to generate 100,000 independent participants. The likelihood statistic (G^2) was computed for each participant, and a sampling distribution created. The likelihood statistic value for the actual participant for whom the model was created was then compared with the mean of the sampling distribution. Table 3 shows the fits of the models to participants' data and the critical value of the statistic at the empirically determined $p = .05$ confidence level.

Table 2
Bins Used for Modeling the Results of Both Experiments

Midpoints			
Experiment 1		Experiment 2	
24	73	11	61
34	77	15	65
41	84	19	69
45	94	23	73
49	104	27	77
53	114	41	91
57	124	45	95
61		49	99
65		53	101
69		57	107

Note. In Experiment 1, the first two bins contained 15 stimuli each, whereas the remaining bins contained 10 stimuli each. In Experiment 2, each bin contained 8 observations.

Table 3
Results of the Monte Carlo Simulation for Sigmoid Models in Experiment 1

	Optimal threshold		General threshold		Two step	
Participant	Observed value	Critical value	Observed value	Critical value	Observed value	Critical value
1	12.07	19.83	9.13	21.01	6.84	19.44
2	32.13	29.47	25.47	32.33	17.11	34.61
3	54.03	33.63	25.01	26.84	18.01	27.13
4	41.75	23.78	9.17	28.78	7.73	27.62
5	64.11	27.18	11.13	25.84	7.63	20.82
6	37.11	31.49	20.07	26.06	14.38	23.77
7	32.58	29.37	16.43	25.83	12.08	21.89
8	41.67	28.57	21.34	24.26	16.62	23.49
9	25.13	33.27	19.37	26.64	7.47	26.17
10	58.83	31.23	18.01	27.42	11.11	21.69
11	87.01	24.72	9.57	20.55	5.05	17.99
12	68.79	28.37	18.66	25.45	10.27	22.31
13	48.85	36.93	24.74	28.05	8.67	24.49
14	71.04	32.00	18.37	22.63	16.89	27.07
15	45.77	23.67	18.49	25.35	10.08	24.19
16	65.03	28.18	18.91	26.32	14.79	23.80
17	123.75	31.71	10.48	22.13	9.84	21.27
18	74.12	31.70	11.26	30.30	8.02	22.03
19	58.22	30.12	15.49	25.31	11.12	20.17
20	24.28	26.74	13.89	22.26	9.50	29.54
21	66.98	33.95	11.60	21.28	6.15	25.84
22	49.44	27.99	20.62	29.82	14.06	24.30
23	36.70	31.66	17.52	30.80	17.42	26.62
24	101.40	25.59	17.15	24.52	10.50	17.71
25	18.11	34.60	15.94	26.07	13.09	32.39
26	59.74	28.00	11.51	24.52	8.22	22.39
27	69.99	29.65	17.74	28.13	12.30	25.91
28	43.59	31.10	9.29	29.40	8.18	23.17
29	68.46	27.68	15.28	30.53	8.79	27.44
30	69.29	33.76	49.66	25.47	32.72	28.01
31	45.84	30.38	14.81	25.75	13.64	30.00
32	53.93	23.92	16.15	23.45	9.40	34.07
33	43.52	24.30	13.57	22.99	7.88	24.26
34	107.50	30.99	15.92	29.04	8.29	20.55
35	55.06	29.25	22.07	30.12	15.14	29.36
36	38.44	27.25	17.86	30.78	13.76	24.40
37	56.26	26.43	24.76	29.14	20.79	30.15
38	55.65	29.70	17.56	30.53	13.06	24.79
39	85.72	24.89	10.86	20.90	9.67	21.83
40	87.72	26.75	16.83	22.72	12.60	24.48
41	16.83	22.93	15.56	26.06	9.17	28.08
42	31.99	29.17	17.01	29.08	9.84	26.81

Out of 42 participants, only 1 (Participant 30) could not be fit by the two-step model, because of extreme variability in response probabilities. Both the the general one-step model and the two-step model could fit all the remaining participants, of whom only 5 (Participants 1, 9, 20, 25, and 41) could be fit by the optimal-threshold model.

The gains of the best fitting one-step models were inspected to determine the steepness of the response surfaces. Gains ranged from 0.036 to 0.190, with a mean of 0.100 and a standard deviation of 0.038. These gains can be compared with those predicted by a deterministic response rule, under certain assumptions. If we assume, as Ashby (1992) did, that high-contrast self-terminated displays create zero perceptual noise and that perceived length scales linearly with actual length, then we can compare the gain of the best fitting sigmoid,

$$P(\text{resp} = \text{A}|x) = \frac{1}{1 + \exp[-g(x + a)]},$$

with the gain that would be caused by criterial noise alone,

$$P(\text{resp} = \text{A}|x) = \frac{1}{1 + \exp[(-x + a)/\sigma]}.$$

The standard deviation of the criterial noise, σ, is equal to one over the fitted gain, $1/-g$. Thus, the minimum and maximum criterial noise, as estimated from the gains of the best fitting single sigmoid model, are 5.26 and 27.78 pixels (6.8 and 36.1 mm), respectively. Even this minimum value is easily discriminated, and the maximum one represents an extremely high level of noise.

Although testing the goodness of fit of each single model required the Monte Carlo technique, testing the difference in fit between models did not. Nested models can be compared simply by comparing the difference between the fit of the model that has more free parameters (q) and the fit of the model that has fewer parameters (p), by using the following formula:

$$\chi^2_{(q-p)} = -2(G^2_q - G^2_p).$$

The fits of the optimal-threshold and general one-step models were compared for the 5 participants for whom the optimal-threshold model provided a statistically unrejectable description. For 4 participants (Participants 1, 9, 20, and 25) the general one-step model fit significantly better than the model with a threshold at the optimal location. The remaining participant (Participant 41) was not better fit by the free-threshold model, $\chi^2(4, N = 180) = 2.53, p > .1$, because the participant produced an extremely shallow response function. However, this participant, along with 18 of the remaining 41 participants, was significantly better fit by the two-step model.

Overall then, 23 participants showed one-step response functions with nonoptimal thresholds, as determined by the lack of improvement in fit provided by the addition of a second step in the response function. The responses of a typical participant from this group are shown in Figure 3. The response functions of these participants had slopes too shallow to be accounted for by perceptual noise. Another 18 participants showed response functions that were more complex than simply one step, as was evidenced by the superiority of the two-step model. Figure 4 shows the responses of 4 typical members of this group of participants. Not every member of this group showed responses that had a visually identifiable plateau. Finally, even the complete two-step model was unable to adequately describe the responses of 1 participant who thus could not be fit by any model in this class.

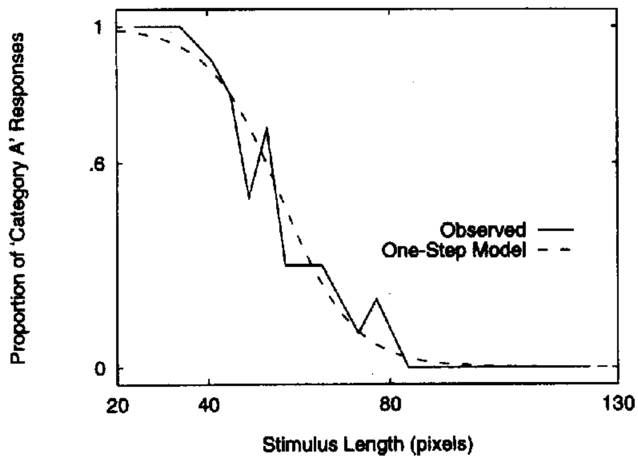

Figure 3. Results for Participant 11 in Experiment 1. This participant was fit best by the general one-step model. The graph shows the proportion of responses for Category A as a function of stimulus size. Within the overlap region from 40 to 80 pixels, the true proportion of Category A stimuli was .6.

Discussion

The message of these results is that no participant produced a response surface clearly compatible with any version of the decision-boundary model in which perceivers use single-cutoff decision rules, such as would result from the use of any strictly increasing discriminant function. Although the majority of participants were indeed well described by a single sigmoid, the slopes of their response surfaces were generally extremely shallow. This could be due to very high levels of criterial noise, but in the absence of a compelling explanation for such noise—such as a theory of criterion change during learning—these data argue against a deterministic response selection mechanism. Instead, the shallow sigmoid likely results from probabilistic responding throughout the range of overlap of the two categories, reflecting the slow change from one deterministic region to the next.

It is also of interest to note that only 1 participant showed a criterion that was above the optimal threshold, and no participant had a criterion below the region of overlap. In untrained participants, thresholds would be placed randomly across the range of stimuli. The lack of thresholds outside the overlap region suggests that participants were not simply guessing about each stimulus in the last block of trials. Rather, participants might have been sensitive to the locations of the boundaries between the overlapping and deterministic ranges, and may not have used this knowledge optimally.

A number of participants produced rates of probabilistic responding in the overlapping range that were consistent with a two-step model. These participants showed sensitivity to both the beginning and end of the probabilistic range, information that would not be available to an individual who had access only to a single criterion value or a summary description of the two categories as Gaussian distributions.

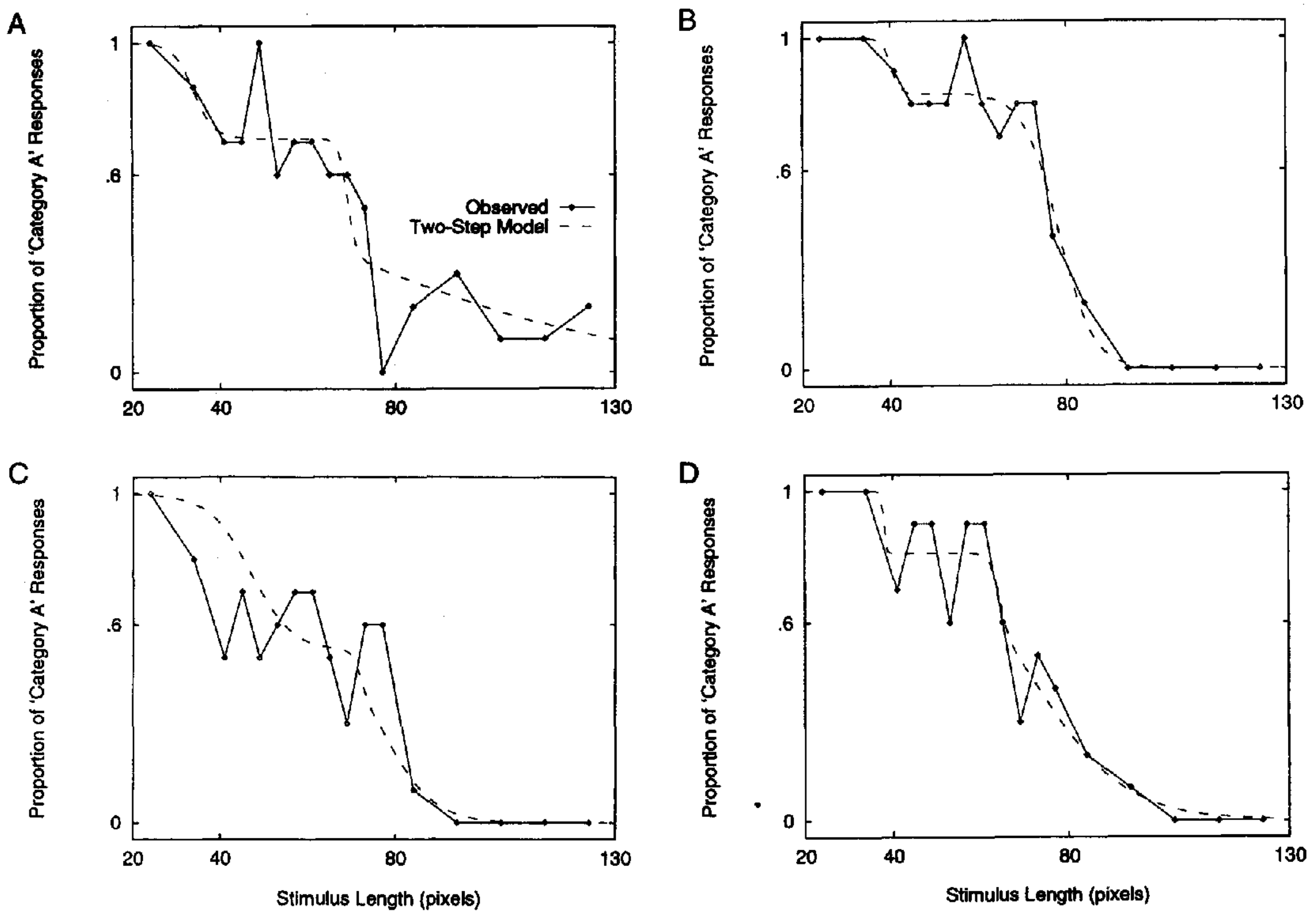

Figure 4. Results for Participants 2, 9, 13, and 42 in Experiment 1. These participants were fit best by the two-step model. A and B show the proportion of responses for Category A as a function of stimulus size. Within the overlap region from 40 to 80 pixels, the true proportion of Category A stimuli was .6.

The finding of nonoptimal thresholds, shallow sigmoids, and level probabilistic response regions all argue against the notion that responses were due simply to the selection of a noisy threshold of a steep monotonic discriminant function.

Experiment 2: Training With Overlapping Skewed Distributions and Testing Without Feedback

The interpretation of the results of the first experiment depended on an analysis of the last training block. It is possible that participants were still adjusting their responses during that block, and if so the nested modeling would be thrown into question for the following reason. Notice that the measure of fit used to find the best model for each participant assumes the independence of each observation. Because participants were receiving corrective feedback during their final block, this assumption might not be justified. Further, the decision-boundary model assumes that responses are generated on the basis of a stationary decision boundary. A decision-boundary model that used corrective feedback to change the location of the boundary could very well produce shallow response surfaces, such as those seen in Experiment 1. To control for these possibilities, in Experiment 2 participants were given a transfer block of trials following the training trials; there was no feedback during the transfer phase. In addition, whereas in Experiment 1 the categories overlapped only by 40 pixels and differed in probability by a factor of only one half within that range, in Experiment 2 the range of overlap was increased. If participants were probability matching, then this extended range would make it easier to discriminate two-step from shallow one-step response surfaces. The ratio of the two categories within that region also was increased to encourage use of a single criterion point, giving the decision-boundary model a better chance of success.

Method

Participants. Forty-seven students participated for partial credit in an introductory psychology course at Indiana University.

Design and procedure. Two categories (A and B) were defined. Category A was composed of filled rectangles drawn from two adjoining uniform distributions, as shown in Figure 5. The first, A-only region ranged from 10 pixels (13 mm) to 19 pixels in increments of 1 pixel, and the second, overlap region ranged from 20 to 92 pixels in increments of 8 pixels. Category B was also defined by two distributions: The first was uniform from 20 to 92 pixels in increments of 8 pixels, and the second was uniform from 100 to 120 pixels (124 mm to 149 mm) in increments of 1 pixel.

There were 120 trials in each of five blocks. In every block, each of the 10 stimuli from the A-only region was displayed twice, each of the 10 stimuli from the overlap region was shown four times, and each of the 30 stimuli from Category B was shown twice. Thus, there were 60 stimuli from each category shown in every block, and the rate of Category A to Category B stimuli between 20 and 92 pixels (the region of overlap) was 2:1.

The instructions were the same as in Experiment 1. The only other change was the addition of a transfer phase at the end of the training trials. Transfer stimuli were bars ranging from 10 to 28, 40 to 78, and 90 to 108 pixels in 2-pixel intervals. Each of the 80 test stimuli was presented twice, for a total of 160 transfer trials. Transfer trials had no corrective feedback.

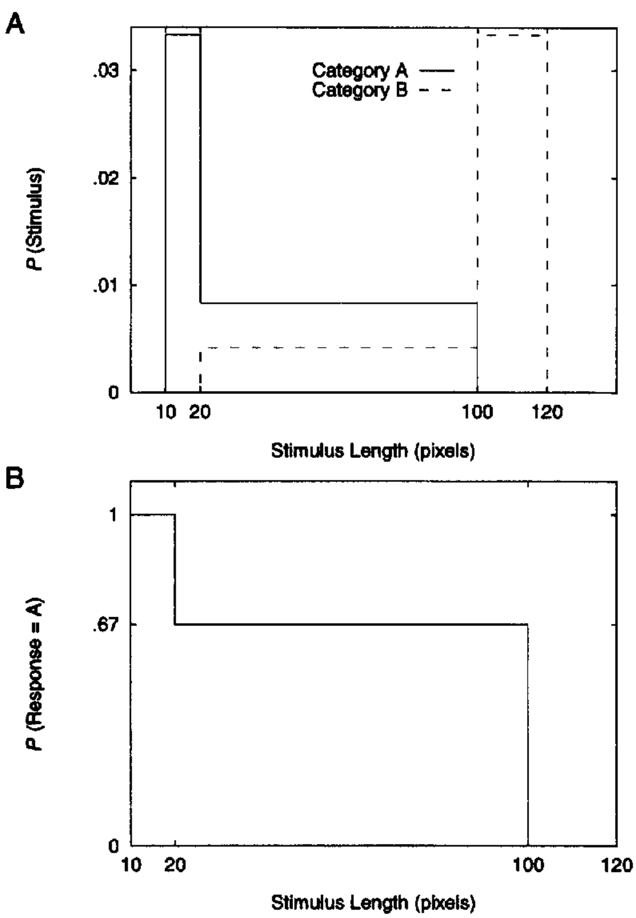

Figure 5. Category distributions used in Experiment 2. A and B show the probability (*P*) of Category A (solid line) and Category B (dashed line) as a function of stimulus length. At the extreme left, between stimulus lengths of 10 and 20 pixels, only Category A was possible. In the overlapping region, from lengths of 20 to 100 pixels, both categories were possible, with Category A occurring 67% of the time. At the extreme right, from stimulus lengths of 100 to 120 pixels, only Category B was possible.

Results and Discussion

The same models as in Experiment 1 were tested, with the optimal-threshold one-step model having a threshold at 100 instead of 80 pixels. Models were fit to the 160 trials in the transfer block. Trials were again binned according to stimulus length. There were 20 bins, each with 8 trials. The bins are shown in Table 2.

Following Monte Carlo simulation of the models, only 1 out of the 47 participants (Participant 25, whose responses were nondeterministic at the extremes of the stimulus range) could not be described by the two-step model. Observed and critical values of the likelihood test statistic are shown in Table 4. Only 1 participant's data were adequately described by the optimal-threshold model, although they were significantly better described by the free-threshold model. This participant (Participant 13) barely produced any change in response probabilities across stimulus lengths and produced highly variable response probabilities.

Comparison between models, based on differences in model fit, revealed that all of the remaining 45 participants were better fit by the general-threshold one-step model than by the optimal-threshold model. Of the 45 participants, 18 were fit better still by the two-step model.

Figure 6 shows a typical one-step participant (37), and Figure 7 shows 4 typical two-step participants.

The criterial noise necessary to account for the best fitting single sigmoid models was calculated for each participant, as in Experiment 1. The best fitting gains ranged from 0.021 to 0.276, with a mean of 0.121 and a standard deviation of 0.063. This equates to criterial noise standard deviations ranging from 3.62 pixels (4.7 mm) to 47.62 pixels (61.9 mm).

It appears, therefore, that participants were largely unable to select a criterial stimulus value and to respond deterministically on its basis. Instead, responses tended to be probabilistic over the entire range in which the two categories overlapped (often rising rather than showing a sustained plateau per se). On the other hand, 6 participants (22, 24, 31, 33, 37, and 40) were fit by one-step models with relatively high gains (greater than .2—i.e., with criterial noise standard deviations of less than 7 mm). These participants did seem to pick criterion values. None of these models had a criterion close to the optimal one, however, as indicated by the improvement given by the free-threshold model. As in Experiment 1, all observed thresholds were within the overlap region.

These results again point to a discrepancy between the participants' responses and the predictions of the single-cutoff decision-boundary model. Participants showed changes in response probabilities at both ends of the probabilistic region, indicating sensitivity to changing contingencies rather than simple storage of a single decision criterion. Although evidence for two-step response surfaces is relatively weak, the evidence against a one-step surface (the single-cutoff model's prediction) is strong.

Exemplar Model

Decision-boundary theory is based on the selection and application of a criterion. Observers are supposed to handle

Table 4
Observed and Critical Values of G^2 for the Results of Experiment 2

Participant	Optimal threshold		General threshold		Two step	
	Observed	Critical	Observed	Critical	Observed	Critical
1	42.74	34.29	12.60	20.06	9.30	20.86
2	39.88	35.20	21.57	34.55	14.61	28.84
3	91.08	30.74	19.13	23.77	13.74	22.40
4	68.72	38.85	11.65	37.28	4.00	31.11
5	96.25	31.48	22.15	21.05	7.35	19.44
6	60.82	26.79	21.99	27.96	14.89	24.54
7	51.50	33.23	13.91	15.28	3.04	13.87
8	72.66	35.47	18.43	31.32	15.11	29.73
9	99.96	38.85	15.42	29.34	11.08	29.78
10	84.66	32.85	9.54	29.66	5.13	21.09
11	84.04	31.03	23.92	25.55	17.27	22.45
12	61.26	28.48	20.88	28.67	15.74	30.29
13	33.22	38.85	28.73	32.54	26.13	31.34
14	112.86	33.61	20.02	29.45	12.15	27.45
15	52.91	28.28	10.14	28.80	6.43	24.59
16	122.44	33.42	12.38	20.01	6.47	23.12
18	54.20	33.76	17.41	29.25	14.80	27.53
19	48.80	33.17	15.73	21.77	9.63	22.30
20	78.90	30.76	14.52	23.00	11.76	28.09
21	54.60	29.57	10.09	21.07	7.17	18.54
22	143.94	36.68	4.62	14.11	3.54	13.02
23	64.04	35.25	49.99	25.85	23.42	29.18
24	48.14	30.09	2.95	11.83	1.62	12.92
25	43.29	30.82	24.78	21.21	20.09	19.32
26	59.30	28.11	31.64	28.66	16.92	33.32
27	70.41	29.91	19.63	27.27	16.58	24.93
28	44.70	25.12	16.65	27.57	9.33	28.03
29	53.35	33.72	16.23	35.39	8.70	26.25
30	33.66	29.94	10.65	21.58	8.36	24.07
31	148.38	33.92	4.37	14.33	2.28	11.38
32	84.36	33.47	3.43	21.61	2.40	19.20
33	112.73	33.79	6.97	17.98	2.17	15.43
34	59.63	32.21	18.43	23.99	13.23	21.47
35	98.47	35.27	17.24	20.23	11.95	17.89
36	74.67	29.90	21.45	22.18	12.97	22.56
37	67.34	32.53	5.48	13.03	4.15	12.74
38	43.55	28.45	25.05	20.24	9.08	19.00
39	73.97	35.58	5.51	17.47	2.69	20.18
40	31.59	23.33	6.53	14.59	5.08	10.20
41	103.12	30.07	17.31	29.70	11.07	22.12
42	22.20	11.98	7.17	11.88	6.86	16.10
43	57.50	32.97	13.88	29.39	9.01	36.18
44	73.46	38.85	18.45	27.08	17.60	34.82
45	57.03	39.30	22.40	31.95	9.92	27.41
46	93.44	30.15	15.53	25.38	8.76	24.92
47	45.57	30.26	19.73	25.85	14.25	21.61

Note. G^2 is the likelihood test statistic.

base rates optimally (Maddox, 1995). This is not to suggest that observers estimate base rates accurately, but that their estimates are used optimally. In the experiments here, this would mean that participants would respond "A" deterministically whenever it was perceived as being more likely to be correct, and otherwise would switch to deterministic "B" responding. But to the extent that participants can be described as using single criteria at all, the criteria are not placed optimally and, in addition, the transition from "A" to "B" responses covers a wide range of physical stimulus values. The success of the two-step model showed that some participants were producing stable, probabilistic responding when faced with nondeterministic stimulus–category mappings. Both of these results suggest that some process other than, or in addition to, decision boundaries must have been at work.

There are a number of mechanisms that might be involved in one-dimensional categorization besides single perceptual thresholds. Participants might have been using prototypes, ideal values, or decision regions. They might have been estimating the densities of the two category distributions, computing the ratio of likelihoods, and comparing that value to a (very) noisy criterion. It is also possible that participants were storing individual exemplars in memory and consulting those stored exemplars when making their responses. ALCOVE (Kruschke, 1992), a connectionist implementation of an exemplar model with error-driven learning, has been successful in explaining the results of a wide range of categorization experiments (Choi, McDaniel, & Busemeyer, 1993; Kruschke, 1992, 1993; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Nosofsky & Kruschke, 1992; Nosofsky, Kruschke, & McKinley, 1992). Whereas the experiments presented here were designed to test the single-cutoff model, they were not necessarily optimal for testing ALCOVE. Nonetheless, the ability of an exemplar model to fit these data is still of interest.

The ALCOVE formalism of exemplar-based learning, as used here, has three free parameters. Each exemplar representation has a *specificity*, c, which governs the extent to which nearby stimuli activate the representation. The exemplar representations have modifiable connections to a node that represents the perceived category. Those connections are changed with a characteristic *learning rate*, λ. Finally, the perceived category representation is used in a graded fashion to map onto response probabilities. That mapping is controlled by a *scaling parameter*, ϕ, which determines the extent to which the most active category dominates choice

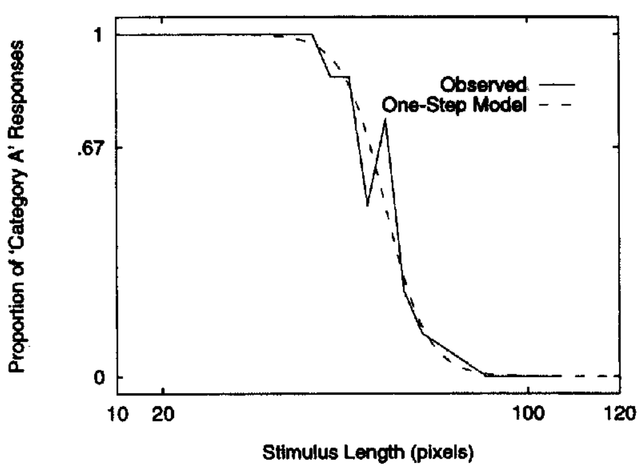

Figure 6. Results for Participant 37 in Experiment 2. This participant was fit best by the general one-step model. The graph shows the proportion of responses for Category A as a function of stimulus size. Within the overlap region from 20 to 100 pixels, the true proportion of Category A stimuli was .67.

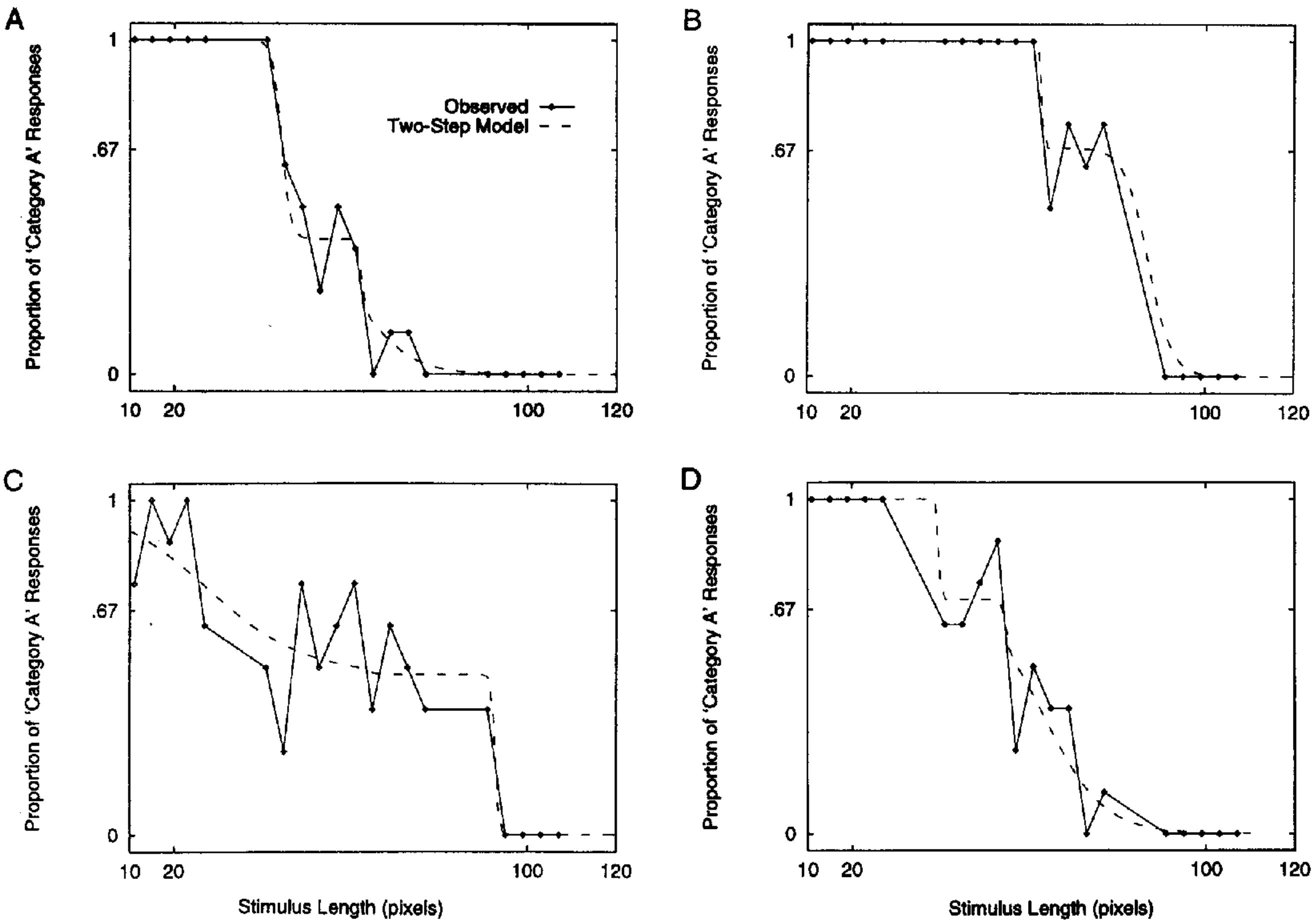

Figure 7. Results for Participants 16, 7, 26, and 46 in Experiment 2. These participants were fit best by the two-step model. A and B show the proportion of responses for Category A as a function of stimulus length. Within the overlap region from 20 to 100 pixels, the true proportion of Category A stimuli was .67.

probabilities. Because stimuli vary on only one dimension, the original fourth parameter of ALCOVE, which governs the allocation of attention between stimulus dimensions, is not relevant here.

Formally, the activation of exemplar (or *hidden*) unit j is given by

$$a_j^{\text{hid}} = \exp(-c|h_j - a^{\text{in}}|), \tag{6}$$

where h_j is the location of the exemplar unit in psychological space and a^{in} is the location of the stimulus in the same space, and where c is the specificity constant.

Category node activation is then a weighted sum of exemplar node activations:

$$a_k^{\text{out}} = \sum_j w_{kj} a_j^{\text{hid}}. \tag{7}$$

The weights (w_{kj}) are modifiable by

$$\Delta w_{kj} = \lambda(t_k - a_k^{\text{out}})a_j^{\text{hid}}, \tag{8}$$

where λ is the learning rate and t_k is the "humble" teacher value, defined as

$$t_k = \begin{cases} \max(+1, a_k^{\text{out}}) \text{ if the stimulus is in category } k \\ \min(0, a_k^{\text{out}}) \text{ otherwise.} \end{cases} \tag{9}$$

The actual response probabilities predicted by the model are derived from the category node activations. The probability of responding "Category A" is given by an exponentiated version of the Luce choice rule:

$$P(\text{A}) = \frac{\exp(\phi a_{\text{A}}^{\text{out}})}{\exp(\phi a_{\text{A}}^{\text{out}}) + \exp(\phi a_{\text{B}}^{\text{out}})}, \tag{10}$$

where ϕ is the scaling parameter.

ALCOVE was fit to both data sets with a simulated annealing procedure, and the resulting models were then tested with the Monte Carlo method described above. The fit of ALCOVE was computed using the likelihood metric.

Fit to Experiment 1

The exemplar units for ALCOVE comprised a covering map from 0 to 148 pixels with a unit placed every 2 pixels, providing one unit for each possible exemplar. Data were put into 17 bins, as in the initial analysis (Table 2). ALCOVE was able to fit the data of all 42 participants. The fits are shown in Table 5. Even Participant 39, who was not fit by any sigmoid model, was well fit by ALCOVE. Direct comparison of the fits of ALCOVE and the one- and two-step logistic models is in principle possible with a penalized error metric, such as the Akaike information criterion (AIC). However, in this case use of the AIC would be misleading. Although the AIC is a measure of relative goodness of fit, we are not arguing that ALCOVE is necessarily a better model in all cases. The fit of ALCOVE to these data is indicative of the extent to which exemplar-based representations need to be considered *in conjunction* with other representations in category learning. These data do not reject an exemplar-only model, although they do reject one sort of rule-only model—at least for some participants.

Table 5
Results of the Monte Carlo Simulation for the ALCOVE Model in Experiment 1

Participant	Observed	Expected (p = .5)	Critical (p = .05)
1	12.85	11.57	18.29
2	19.91	19.00	31.48
3	24.11	17.26	25.50
4	11.97	18.44	31.79
5	16.70	17.88	27.92
6	20.28	18.69	27.02
7	13.88	16.97	23.82
8	17.03	15.20	20.90
9	7.49	13.68	17.30
10	18.57	12.95	19.22
11	14.59	16.45	26.69
12	23.66	17.36	26.24
13	12.95	17.92	24.79
14	12.71	15.51	29.06
15	10.16	17.13	24.06
16	21.85	17.87	29.76
17	9.62	13.69	20.71
18	8.86	14.95	21.77
19	11.05	17.71	29.90
20	13.88	18.56	26.19
21	10.59	13.12	22.03
22	20.09	15.90	20.30
23	12.78	18.86	27.32
24	5.91	10.37	20.77
25	14.12	19.19	26.09
26	13.05	15.84	24.36
27	14.55	17.66	26.20
28	9.66	17.48	22.73
29	6.93	12.63	15.88
30	14.69	18.44	24.06
31	13.35	18.92	30.70
32	18.86	18.96	23.88
33	9.84	12.48	23.26
34	23.35	17.33	24.86
35	13.27	18.69	28.39
36	16.85	17.93	28.12
37	19.46	16.82	24.71
38	18.07	17.66	25.03
39	12.82	15.72	21.27
40	17.03	13.30	21.50
41	17.42	18.02	26.08
42	15.65	16.04	25.92

Fit to Experiment 2

The covering map consisted of exemplar nodes placed at every pixel, from 10 to 119 pixels, giving one exemplar for each possible exemplar. The data were again put in bins, as in the earlier analysis (Table 2). ALCOVE fit all but Participants 20 and 23. Participant 20 had a noisy transition from "A" to "B" responding, which ALCOVE could not match, whereas Participant 23 showed a clear transition from "A" to "B" responding but made several "B" responses to extreme Category A stimuli, which ALCOVE could not account for. Fits for all participants are shown in Table 6.

The categories used in this experiment were chosen to provide a large sample of responses to stimuli within the probabilistic region. The results from the experiment showed, however, that participants had large individual differences in their response surfaces. This might suggest that different participants used different strategies to distinguish the categories. In fact, however, with its fixed exemplar-based representations, ALCOVE was able to fit all of the different kinds of observed response patterns.

Four participants in particular are good illustrations of the range of responses ALCOVE was able to predict. Participant 9 from Experiment 1 and Participants 22, 40, and 32 from Experiment 2 (shown in Figure 8) produced (a) a two-step response profile, (b) a sharp response transition near the left of the probabilistic region, (c) a sharp response transition near the right of the probabilistic region, and (d) a gradual transition across the region, respectively. Parameters and fits for those participants are shown in Table 7. These values differ in cases by an order of magnitude or more, and show that an exemplar model such as ALCOVE can capture large individual differences.

Although ALCOVE can capture the individual differences seen in these data, this should not give the impression that it can account for any possible response surface. Because it is an exemplar model, it would have great difficulty fitting a response surface with a response plateau *below* that of the true probabilities throughout the overlap region. For example, if the true probability of Category A is .6 in the overlap region, but some participant showed a response plateau of .4, then ALCOVE would fail to fit that participant. Conversely, that same participant would be easily fit by the two-step logistic model or a hypothetical two-boundary model. In fact, a decision-boundary model based on participants' likelihood estimates might be able to fit any observed response function, as long as the likelihood estimates can be fit freely from the data. The exemplar model provides a parsimonious way of accounting for most of the participants' responses; this does not mean it provides a complete, or even the most accurate, account.

Table 6
Results of the Monte Carlo Simulation for the ALCOVE Model in Experiment 2

Participant	Observed	Expected ($p = .5$)	Critical ($p = .05$)
1	12.45	13.54	23.80
2	17.47	22.08	34.78
3	14.92	17.12	27.85
4	24.00	20.33	32.45
5	12.61	14.45	22.02
6	16.23	19.12	24.15
7	14.95	11.35	18.95
8	16.77	19.90	25.03
9	14.94	20.87	27.31
10	16.99	18.83	33.39
11	24.10	18.33	26.83
12	20.76	21.67	29.68
13	29.61	22.05	37.26
14	20.16	16.25	22.60
15	14.40	19.66	32.50
16	9.04	10.91	14.14
18	18.76	22.04	29.07
19	14.73	14.68	18.30
20	26.29	14.07	22.81
21	16.42	12.78	21.80
22	3.99	8.19	15.02
23	25.73	19.09	23.28
24	3.79	8.75	16.18
25	24.27	17.38	27.62
26	21.59	21.53	33.08
27	24.37	19.99	27.33
28	15.05	16.71	24.29
29	16.61	22.30	30.69
30	15.32	17.56	27.33
31	12.88	11.93	22.61
32	3.29	12.35	20.25
33	16.44	10.65	21.06
34	22.65	16.82	23.18
35	17.53	12.99	20.79
36	20.01	17.60	24.75
37	8.66	12.41	18.93
38	11.86	10.37	16.58
39	4.34	10.27	17.36
40	8.03	8.31	17.52
41	12.83	17.98	24.45
42	10.43	15.60	21.87
43	12.35	21.61	31.22
44	18.04	22.05	32.83
45	23.94	17.54	30.32
46	16.89	14.53	22.73
47	17.10	19.37	26.29

General Discussion

By responding probabilistically to stimuli of indeterminate category origin, participants showed a sensitivity to the relative likelihood of category membership. This sensitivity has been established for many years by numerous experiments involving partial reinforcement, in the context of both discrimination and categorization (Brunswik, 1941, 1956; Grant, Hake, & Horseth, 1951). The results of the experiments presented here suggest that this same sensitivity may play a role in classifications of overlapping categories. These experiments do not imply that decision boundaries play no role in classification of overlapping categories. Rather, simple rules alone cannot provide a complete description. For that, it might be necessary to consider the role of individual exemplars in memory, learning mechanisms for the boundary, or both.

Exemplar and decision-boundary models make similar predictions in a large number of experimental conditions. Much of the data taken to support the exemplar model can also be interpreted within the decision-boundary framework. The exemplar model seems to give a more convincing account of performance with complex categories (McKinley & Nosofsky, 1995), although Maddox and Ashby (1993) showed that a decision-boundary model could fit data from discrimination of nonnormally distributed categories better than a deterministic exemplar model could. In the simple experiments presented here, there is further evidence separating the exemplar model from a class of decision-boundary models. Many of the participants were not fit by a single-cutoff model, but an exemplar model, ALCOVE, was able to fit the data of almost every participant. For the general decision-boundary model to fit all of these same data, qualitative differences would have to exist in different participants. Some participants would have to assume normal distributions, and others would have to assume uniform distributions. The precise nature of the distributional assumptions and constraints on the facilities of participants to estimate the parameters of their chosen distributions are beyond the scope of this article to determine. Although these experiments were not designed to test the exemplar model as rigorously as the single-cutoff decision-boundary model, these results still mark the differences between these two models of categorization.

What is it about the models that makes them different? There are three basic areas in which ALCOVE and the decision-boundary model differ, which we discuss in turn.

Adaptive Learning

ALCOVE is adaptive, with its internal parameters adjusted to the degree to which there is error in categorization. In contrast, the decision boundary is adjusted constantly to fit the statistics of the training data so that at all times, the boundary is (or ought to be) the optimal one given the training data. This distinction is not an in-principle one. It is possible to imagine algorithms for updating the decision boundary based on mismatches between the model's responses and the training data (e.g., Busemeyer & Myung, 1992). If this strategy were followed, then some of the results of this experiment (especially the finding that individuals showed probabilistic responding over wide ranges of stimuli) could be handled by a decision-boundary model. However, those participants best fit by a two-step function in these experiments and those described by nonquadratic discriminant functions in other experiments (Kalish, in press; McKinley & Nosofsky, 1995) would require different, substantial revisions to the model.

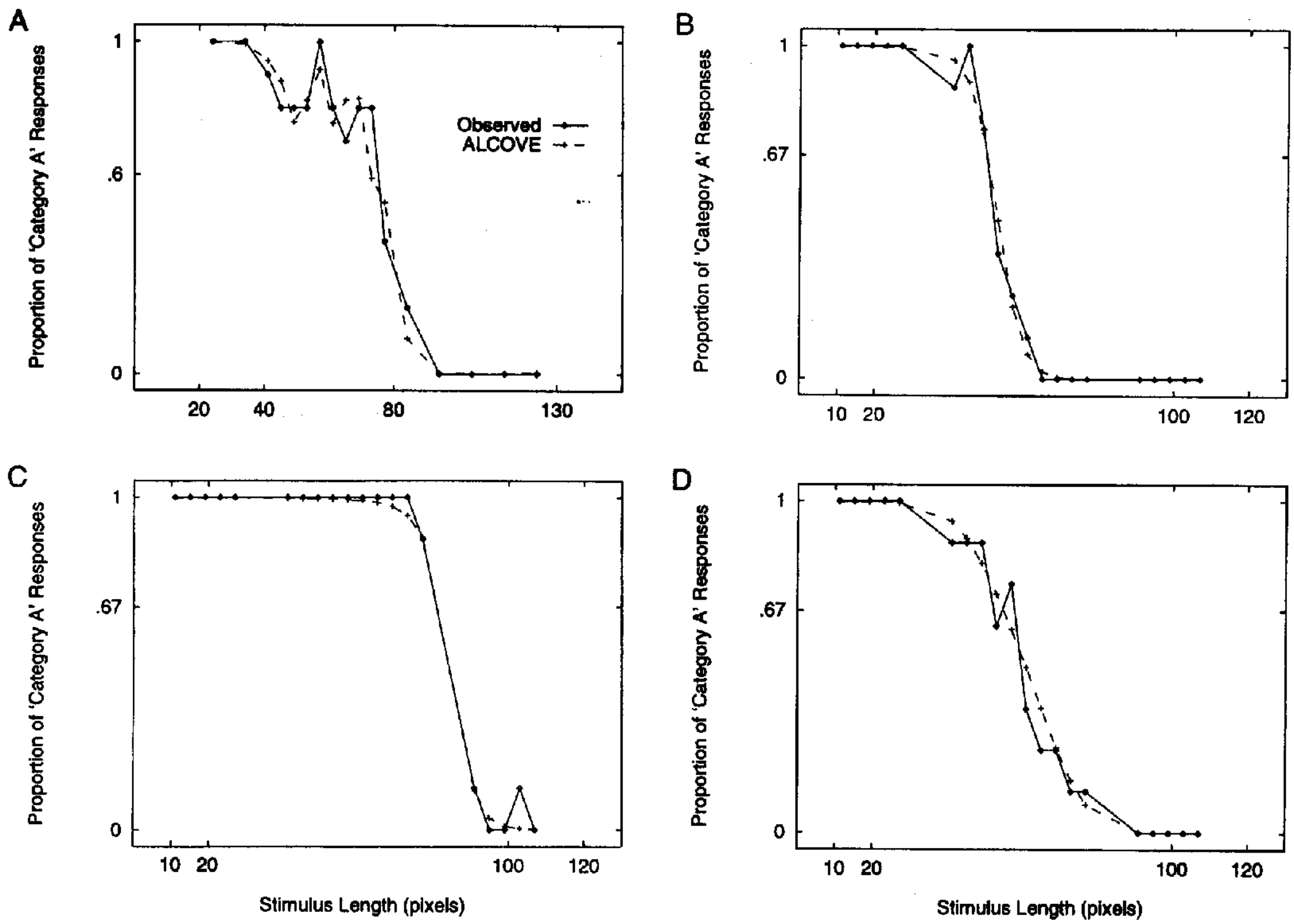

Figure 8. Four representative participants from the two experiments. A: Participant 9 from Experiment 1 with a two-step response surface. B–D: Participants 22, 40, and 32 from Experiment 2 showing, respectively, a sharp change from "B" to "A" occurring near the left side of the probabilistic region, a sharp transition near the right side of the probabilistic region, and a gradual transition across the entire region. ALCOVE fits all four of these response-surface types.

Extrapolation

The experiments presented here and in most other studies test interpolation, in which participants make responses to stimuli lying within the outer limits of those seen during training. Performance on extrapolation, however, is predicted to be quite different by the two models. Exemplar models, including ALCOVE, tend to predict that as the distance from the training stimuli increases, the categorical preferences tend to weaken. Thus, very extreme stimuli do not always receive the same, deterministic responses that less extreme stimuli might. A decision-boundary model, on the other hand, would predict that responses should become absolutely deterministic and very rapid for distant stimuli, as there would be no chance of perceptual noise bringing the perceived stimulus to the opposite side of the boundary (Ashby & Lee, 1993; Busemeyer, DeLosh, Choi, & McDaniel, 1992; DeLosh, Busemeyer, & McDaniel, 1997). Evaluation of exemplar models in extrapolation tasks has tended to show that they do *not* fit the data very well (Anderson & Fincham, 1996; Busemeyer et al., 1992; DeLosh et al., 1997; Erickson & Kruschke, in press).

Table 7
Parameters and Fits of ALCOVE for 4 Representative Participants of the Two Experiments

Experiment	Participant	ϕ	Learning rate	Specificity	G^2
1	9	4.383	0.080	0.067	7.50
2	22	25.059	0.005	0.002	3.97
2	40	3.518	0.028	0.019	7.00
2	32	5.864	0.025	0.021	3.28

Note. ϕ = the scaling parameter; G^2 = the likelihood test statistic.

Exceptions

Exceptions are individual stimuli of one category that are surrounded by stimuli from another category. Decision-boundary models do not naturally handle exceptions. When limits are placed on the form of the boundary—even very weak limits such as that the boundary should be continuous—the model is prevented from accommodating the exceptions. In contrast, exemplar models can easily cope with exceptions (Kruschke & Erickson, 1994; Nosofsky, Palmeri, & McKinley, 1993). In fact, the ALCOVE formalism can sometimes handle exceptions too well, relative to instances of the rule. For example, Kruschke and Erickson (1994) reported that when people are trained to discriminate categories, most of whose members can be divided by a one-dimensional rule with several remaining exceptions, some people learn the *rule* instances *faster* than they learn the exceptions. To accommodate this result, Kruschke and Erickson proposed a hybrid model in which rule-based and exemplar-based representations competed to categorize each stimulus.

It is clear that exemplars, summary representations, and rules all differ (albeit sometimes only subtly) in the way they can be used to make categorical judgments. Although these experiments did rule out a single-cutoff rule-only model of categorization, they did not, of themselves, require a hybrid approach. However, although the exemplar-only model (ALCOVE) did fit most of the participants' responses and although there are rule-only models that can fit each individual participant's responses, the qualitative differences among participants might be indicative of the operation of two competing categorization strategies. Some participants may have been responding on the basis of perceived similarity of the target stimulus with the stored members of each category, whereas others might have been actively searching for a single criterial stimulus value. A comprehensive model of category learning might need to include both decision boundaries and exemplar representations.

References

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3–19.

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

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

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

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., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Advances in psychology: Foundations in perceptual theory* (Vol. 99, pp. 369–399). Amsterdam: North-Holland.

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and 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., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review, 95,* 124–150.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Brunswik, E. (1941). Organismic achievement and environmental probability. *Psychological Review, 50,* 255–272.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Busemeyer, J. R., DeLosh, E., Choi, S., & McDaniel, M. (1992, November). *Extrapolation: The sine qua non of abstraction.* Paper presented at the 33rd Annual Meeting of the Psychonomic Society, St. Louis, MO.

Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory and human performance. *Journal of Experimental Psychology: General, 121,* 177–194.

Choi, S., McDaniel, M. A., & Busemeyer, J. R. (1993). Incorporating prior biases in network models of conceptual rule learning. *Memory & Cognition, 21,* 413–423.

DeLosh, E., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 968–986.

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

Estes, W. K. (1992). Models of categorization and category learning. In D. L. Medin, R. Taraban, & G. Nakamura (Eds.), *Psychology of learning and motivation: Special volume on categorization* (pp. 15–56). San Diego, CA: Academic Press.

Grant, D. A., Hake, H. W., & Horseth, J. P. (1951). Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *Journal of Experimental Psychology, 42,* 1–5.

Healy, A. F., & Kubovy, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 344–354.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Kalish, M. (in press). Connectionism and an ecological approach to category learning. In Z. Dienes (Ed.), *Connectionism and human learning.* London: Oxford University Press.

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

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science, 5,* 3–36.

Kruschke, J. K., & Erickson, M. A. (1994). Learning of rules that have high-frequency exceptions: New empirical data and a hybrid connectionist model. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 514–519). Hillsdale, NJ: Erlbaum.

Kubovy, M., & Healy, A. F. (1977). The decision rule in probabilistic categorization: What it is and how it is learned. *Journal of Experimental Psychology: General, 106,* 427–446.

Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 288–301.

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

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 and Performance, 21,* 128–148.

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

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22,* 352–369.

Nosofsky, R. M., & Kruschke, J. K. (1992). Investigations of an exemplar-based connectionist model of category learning. In D. L. Medin (Ed.), *Psychology of learning and motivation* (Vol. 28, pp. 207–250). San Diego, CA: Academic Press.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 211–233.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1993). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Nosofsky, R. M., & Smith, J. E. K. (1992). Similarity, identification, and categorization: Comment on Ashby and Lee (1991). *Journal of Experimental Psychology: General, 121,* 237–245.

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122,* 92–114.

Thomas, R. D. (1997). *Learning correlations in categorization tasks.* Manuscript submitted for publication.

Thomas, R. D., & Townsend, J. T. (1993, November). *Learning distributional information in a categorization task.* Paper presented at the 34th Annual Meeting of the Psychonomic Society. St. Louis, MO.

Vandierendonck, A. (1995). A parallel rule activation and rule synthesis model for generalization in category learning. *Psychonomic Bulletin and Review, 2,* 442–459.

Received June 12, 1996
Revision received March 11, 1997
Accepted March 11, 1997 ■

The 1997 Research Awards in Experimental Psychology

The awards program of the Division of Experimental Psychology of the American Psychological Association recognizes work by new investigators in all areas of experimental psychology. There is a separate award named for each of the five *JEPs*, and each year an outstanding young investigator is selected for each award. The selection is based on the quality of that person's work and its consistency with the primary subject-matter domain of that *JEP* for which the award is named. In addition, the individual selected normally is targeted for consideration by being a recent author (single, senior, or junior) of an outstanding article that was either published or accepted for publication in that *JEP*.

Kim Kirkpatrick-Steger
New Investigator Award in Experimental Psychology
Animal Behavior Processes

Jennifer Stolz
New Investigator Award in Experimental Psychology
Human Perception and Performance

Neil Mulligan
New Investigator Award in Experimental Psychology
Learning, Memory, and Cognition

Jeffrey Andre
New Investigator Award in Experimental Psychology
Applied

Akira Miyake and Priti Shah
New Investigator Award in Experimental Psychology
General