Journal of Experimental Psychology:
Learning, Memory, and Cognition
1998, Vol. 24, No. 1, 119–143



# Learning Correlations in Categorization Tasks Using Large, Ill-Defined Categories

Robin D. Thomas
Miami University

The experiments revealed whether individual participants are sensitive to exemplar information in the form of within-category correlations between stimulus dimensions after training on large overlapping categories. Participants were trained in 1 of 2 categorization conditions. The sign of the correlation between dimensions differed across conditions, but the categorization rules that best separated the categories were identical. An unannounced attribute-prediction task followed categorization training. Several participants produced predictions consistent with the correct correlation between the dimensions. For other participants, the predictions reflected the correlation only within the region they had associated with the given category, even though the categories overlapped, suggesting that the decision boundary was explicitly represented in memory. Finally, for other participants, no correlational information appeared to be accessible for the prediction task.



One of the central questions in the study of categorization concerns the nature of the mental representation that results when people learn categories from exposure to individual exemplars. An early stance on this issue argued that categories are represented as collections of singly necessary and jointly sufficient object properties (e.g., Bruner, Goodnow, & Austin, 1956; Trabasso & Bower, 1968). This rule-based approach was later supplanted by prototype (Posner & Keele, 1968; Rosch, 1978), by exemplar theories of classification (Estes, 1994; Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986; Shin & Nosofsky, 1992), and by density estimator models (Fried & Holyoak, 1984). The exemplar models have been particularly successful in describing categorization phenomena as exemplified by one of the more known versions of exemplar theory, the generalized context model (GCM; Nosofsky, 1986, a generalization of Medin & Schaffer, 1978). Common to both prototype and exemplar theories is the assumption that the categorization response is determined by a similarity computation of a novel instance to the relevant underlying representation of the category involving the exemplars, either in total (e.g., exemplar theories) or in summary form (e.g., prototype theories). These theories contrast with the rule-based approach in that the latter does not require information about the distribution of exemplars, per se, to be learned, but only about what kind of function separates the categories.

Recently, however, the necessity of storing extensive exemplar information in categorization has been called into question. Nosofsky, Palmeri, and McKinley (1994) resurrected rule-governed categorization that also allows the possibility of storing individual exemplars that are exceptions to the rule. In the rules-plus-exceptions model (RULEX), category learners are assumed to test hypotheses by using simple properties when categorizing objects, starting with one-dimensional rules and then graduating to more complex rules involving conjunctions, disjunctions, and so forth, of those dimensions. The rule that classifies most of the exemplars accurately is retained. However, the final process in this model assumes that individual exceptions to those rules are stored. Participants are assumed to differ with respect to what rule is retained and which exemplars are stored. When individual performance is aggregated, RULEX is successful in accounting for a wide variety of well-known categorization phenomena, such as prototype and specific exemplar effects, sensitivity to correlational information, the difficulty of learning linearly separable versus nonlinearly separable categories, and so forth.

A different line of research into categorization processes arose in response to the criticism that many experiments that investigated formal models of classification used small categories, often containing only three or four exemplars each. Most natural categories are defined on continuous dimensions and are potentially infinite in size. Consequently, Ashby and his colleagues (Ashby & Gott, 1988; Ashby & Maddox, 1992; Ashby & Townsend, 1986; Maddox & Ashby, 1993; Thomas, 1995) have been developing a more ecologically grounded empirical technique to simulate the natural category-learning process. In this technique, categories are experimentally defined to consist of exemplars having values along each stimulus dimension that are probabilistically distributed (e.g., see Figure 1). On a learning trial, an exemplar is sampled according to this distribution and presented to the learner. As a result, the categories are large and overlapping.

Within this empirical setting, Ashby and his colleagues

Some of these results were reported at the 34th Annual Meeting of the Psychonomic Society, Washington, DC, November 1993. Special thanks go to Jim Townsend and Tom Fikes for discussion and guidance and to Rob Nosofsky, Todd Maddox, John Jahnke, Alan Pantle, and an anonymous reviewer for comments on an earlier version of the article.

Correspondence concerning this article should be addressed to Robin D. Thomas, Department of Psychology, Miami University, Benton Hall, Oxford, Ohio 45056. Electronic mail may be sent via Internet to thomasrd@miavx1.acs.muohio.edu.

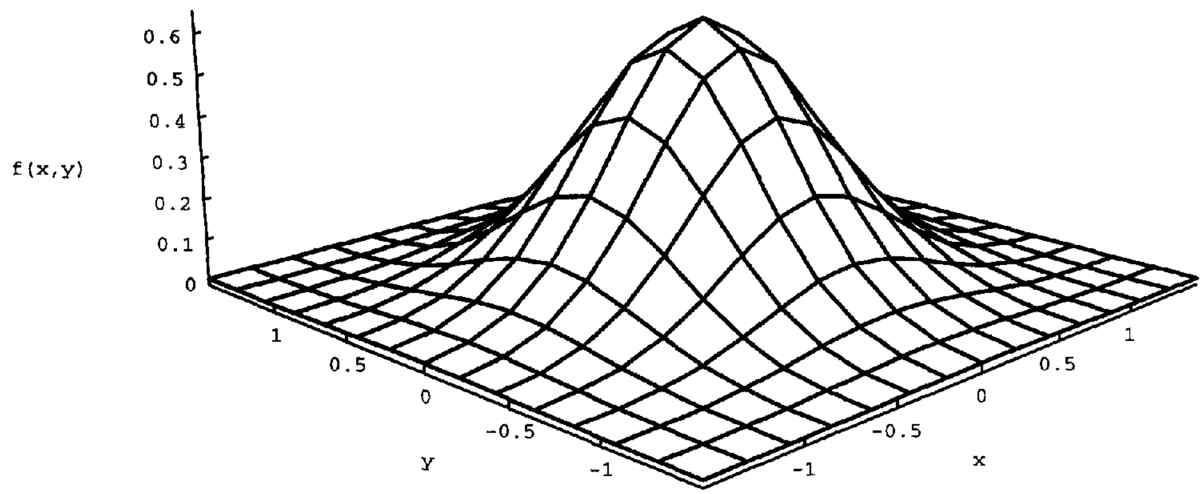



*Figure 1.* A bivariate normal (Gaussian) distribution.

(Ashby & Gott, 1988; Ashby & Maddox, 1992; Maddox & Ashby, 1993) formulated a theory of categorization termed *decision bound theory.* The decision bound theory assumes that experienced participants partition the stimulus space by means of decision bounds into regions that are associated with the appropriate category labels. The categorization process is described as simply locating the given stimulus in the space and emitting the associated response. For example, in the two-category case, if the exemplars vary along two perceptual dimensions, a typical decision bound might be a line or a quadratic curve that the participant would then use to categorize novel stimuli.[1] If the stimulus falls on one side of the bound, the participant says "Category A," if it falls on the other side, the participant says "Category B." The decision-bound approach has been successful in describing categorization behavior using large overlapping category structures (Ashby & Gott, 1988; Ashby & Maddox, 1992; Maddox & Ashby, 1993, but see McKinley & Nosofsky, 1995) and in the small-category case most recently (Maddox & Ashby, 1993). How are these cateogries overlapping when there is a bound though?

From one perspective, the use of a decision bound is a geometrically sophisticated version of the use of rules, albeit not verbalizable rules. To be sure, what exactly constitutes rule-based processing is currently an evolving concept in the literature. From one perspective, any means of assigning exemplars to categories can be called a rule, even the use of similarity aggregated over stored instances. However, current frameworks (e.g., Ashby, Alphonso-Reese, & Turken, 1995; Erickson & Kruschke, 1996) classify categorization processes into generally three categories, verbal rule-based (i.e., using dimensional rules), decision-bound governed, and exemplar-based.[2] The first two classes of models do not require storage and subsequent access of exemplar information for the categorization response, whereas exemplar models such as the GCM do. For the present work, these first two classes are collectively called *rule-based* models. In a strong version of the rule-based models, the participant's memory would be for only the rule or decision bound, and learning may be described as the participant directly operating on the rule or bound (e.g., adjusting its location in perceptual space as a result of feedback, see Maddox & Ashby, 1993, p. 59). *Exemplar-based* refers to the collection of categorization models that assume that the category representation used to classify novel instances involves important aspects of the distribution of category members. For this latter class, the representation may consist of individual exemplars (e.g., the GCM of Nosofsky, 1986) or of summary statistics, such as means, variances, covariances, and so forth, abstracted from the individual exemplars, (e.g., prototype models, Reed, 1972; the density estimator model of Fried & Holyoak, 1984; and other likelihood-based categorization models, Anderson, 1991; Estes, 1986). Of interest, in the infinite-size category case it can be shown that the GCM (as well as the context model of Medin & Schaffer, 1978) is equivalent to a nonparametric density estimator with a probability matching response-

[1] Ashby and Maddox (1992) argued for the special status of quadratic decision bounds because many natural categories seem to be normally distributed and the optimal decision bounds in this situation are quadratic.

[2] Within this taxonomy, it is not clear where traditional prototype models would be placed. It can be shown that one type of decision-bound model, the minimum distance classifier (Ashby & Gott, 1988), is formally equivalent to a prototype model in terms of predicted response probability. However, the two models differ fundamentally in what is assumed to be stored after learning. Decision-bound models assume the bound is learned, whereas the prototype model assumes that the best example (or, in many cases, central tendency) of the exemplars is learned. This difference in representation is detectable when other dependent variables, such as response time, is used (Ashby, Boynton, & Lee, 1994). Because of this difference in assumed category representation, in this article, prototype models are considered "exemplar-based" and not "rule-based," whereas the decision-bound approach is "rule-based." The cited literature does not unambigously classify prototype models in either category.

selection process (Ashby & Alphonso-Reese, 1995). Thus, with large, ill-defined category structures, the difference between exemplar storage and density estimation is fuzzy. The critical distinction, therefore, between rule-based and exemplar-based, as embraced in the present work, concerns whether information about the exemplars, either specific or in summary (i.e., distribution) form, is a significant aspect of what is learned and used during categorization.

One difficulty in determining which categorization process, exemplar-based or rule-based, is supported by the data is that in rigorous quantitative comparisons between the GCM and some appropriately constrained decision-bound models, neither model has enjoyed a consistent advantage over the other in terms of relative fit to the data. Under some conditions, exemplar-based and decision-bound models are mathematically equivalent in terms of predicted response probabilities (Ashby & Maddox, 1993). Given the identifiability problem between these two modeling schemes when response probability is the dependent variable, another methodological strategy (or dependent variable such as response time; Nosofsky & Palmeri, 1997) may be called for to address specific questions of learned category representations. Busemeyer and Myung (1988) made a similar observation regarding the abstraction of schemas in category learning. They argued that to study the prototype abstraction process per se, it may be best to prompt the participant to produce his or her estimate of the prototype on a trial-by-trial basis. Directly interrogating the learned representation may reveal its nature more accurately than examining response probabilities or response times. This method would require, however, that participants could access this representation when so prompted.

The present experiments, therefore, used a categorization training technique that mimicked the ecological properties of categories (i.e., infinite size and overlapping) by defining them as probability distributions on continuously valued stimulus dimensions (Ashby & Gott, 1988). This training was followed by a prediction task in which participants were asked to provide values on one stimulus dimension given values on the other. The prediction task was designed to directly probe participants' memory for category distribution characteristics. The data from this task were the focus of analysis. Specifically, the prediction data should reveal whether any correlation between object dimensions was retained in memory by the participants. If exemplar information was stored during training and later accessed for the prediction task, then the predictions should be consistent with the within-category correlation between the exemplar dimensions, if one exists. However, if only a rule or boundary was learned during categorization, then knowledge of the within-category correlation would not be available to guide the predictions. As described in detail below, the category distributions were chosen specifically to pit a decision-bound representation against an exemplar-based representation with respect to the hypothesized feature predictions.

Using performance from the prediction task as a means of identifying the learned representation from the categorization task assumes that they both tap into the same representation. This assumption would be problematic in the event that decision bounds were used for categorization but residual, yet unnecessary, exemplar information was also stored during learning, which was serendipitously available for the unanticipated task of prediction. Although this dual account lacks parsimony, it is theoretically possible. However, current views of categorization have hypothesized that one of the functions of categorization is prediction of object attributes, thus providing an ecological motivation for such a connection between the two tasks (e.g., Anderson & Fincham, 1996; Lassaline & Murphy, 1996; Malt & Smith, 1984; Murphy & Ross, 1994). Within published accounts of decision bound theory, Maddox and Ashby (1993) have been somewhat inconsistent with respect to what kind of, if any, exemplar information is part of the decision-bound representation, "Exemplar information is not needed; only a response label is retrieved. This does not mean that exemplar information is unavailable, however, because presumably exemplar information is used to construct the decision bound" (pp. 51–52). Later, however, they stated that

This statement can support the motivation underlying our category learning analysis.

> decision bound theory makes no assumptions about categorization at the algorithmic level—that is, about the details of how the participant comes to assign responses to regions. . . . Specifically, it is not necessary for the participant to estimate category means, variances, covariances, or likelihoods, even when responding optimally . . . (p. 52)

and, they described learning in terms of the strong decision-bound perspective, "The decision bound model assumes that the subject operates on the decision bound directly" (p. 59). In addition, Busemeyer and Myung's (1992) rule competition model is cited as a possible learning mechanism for the bound that does not involve storage of exemplars. From this analysis, it appears that the storage of exemplar information is not a fundamental component of the decision-bound theory; it is certainly not required. Performance in the feature prediction task indicating that significant exemplar information was learned during categorization, more than just one or two exemplar exceptions, would not appear to be consistent with this approach.

There has been some previous work on the learning of correlational information in the context of categorization training. Much of this research was concerned with correlations among properties occurring across categories (Billman, 1989; Heit, 1992). With respect to correlations occurring within categories, Malt and Smith (1984) demonstrated that participants were aware of such correlations when the categories were of real-world objects, such as birds. Wattenmaker (1991) also demonstrated participants' sensitivities to correlational information in implicit learning situations. In these studies, the stimuli were composed of discretely valued dimensions or, as in Garner's (1976) terms, of features (e.g., birds that are large do not sing, but small ones do).

Most recently, Anderson and Fincham (1996) used potentially continuous dimensions and showed that categorization training did lead to participants' learning of correlational structure as revealed in a postcategorization prediction task similar to the one used here. They interpreted their results by using a rational model of categorization (Anderson, 1991), which is a generalization of the GCM of Nosofsky (1986). Even though the dimensions had an underlying continuity, in their experiments, this underlying continuity was not sampled very widely. The categories may also have had less coherence because the exemplars were somewhat uniformly distributed throughout the stimulus range, rather than if they were distributed normally (i.e., Gaussian), clustering around a central tendency, and so forth. This type of coherence has been shown to facilitate the induction of category distributions (Flanagan, Fried, & Holyoak, 1986; Fried & Holyoak, 1984). A novel aspect of the present work, in the context of the research on correlation learning, is the use of hundreds of unique training exemplars drawn from continuously varying and overlapping categories organized around a prototype. In addition, data from the Anderson and Fincham experiments, and from other studies, were averaged over many participants. Ashby, Maddox, and Lee (1994) have argued convincingly that averaging over participants can obscure the true underlying perceptual representation. The present experiments focused on individual participant performance in both the categorization task and the prediction task. In the next section, details of the categorization training procedure and the prediction task are outlined. Following that, the specific hypotheses and motivation for the experiments are given, which contrast the consequences of rule-based and exemplar-based representations for the prediction task.

## Categories as Probability Distributions

All of the categorization tasks used in the experiments reported here were based on the randomization technique developed by Ashby and Gott (1988), designed to mimic the structure of natural categories. Suppose there are two categories to be learned (as in the following experiments) that consist of exemplars composed from two continuously valued dimensions (labeled $X$ and $Y$). Each category is represented by a probability density function such that for each possible stimulus point $(x, y)$, the density function yields the likelihood that an exemplar with those stimulus values is sampled. The form of this distribution is often assumed to be multivariate normal, reflecting the assumption that categories have a best, central example (i.e., the prototype), and category exemplars cluster around this stimulus. Multivariate-normal distributions are characterized by location parameters (mean or prototype value on each dimension), spread parameters (variance on each dimension), and association parameters (correlation or covariance between each pair of dimensions). Within the experimental procedure, on a trial, a category is selected randomly and then an exemplar is sampled from the appropriate distribution. The participant's task is to respond with the correct category label. Note that the categories are continuous and overlapping and, hence, are ill-defined. In the bivariate-normal case, the distribution is a bell-shaped surface (see Figure 1) whose height is the likelihood of sampling that point as an exemplar.

For a better view of the shape of the distribution, one can cut the surface at a constant arbitrary height to yield an equal-likelihood contour. Points on the contour have equal likelihood of being sampled; points inside have higher likelihood; points outside have lower likelihood. Figure 2 illustrates equal-likelihood contours for two different, two-category situations (see Experiment 1). In Figure 2A, Category A has a lower mean on the $X$ dimension than does Category B, but it has a higher mean on the $Y$ dimension. Both categories enjoy a strong positive correlation between

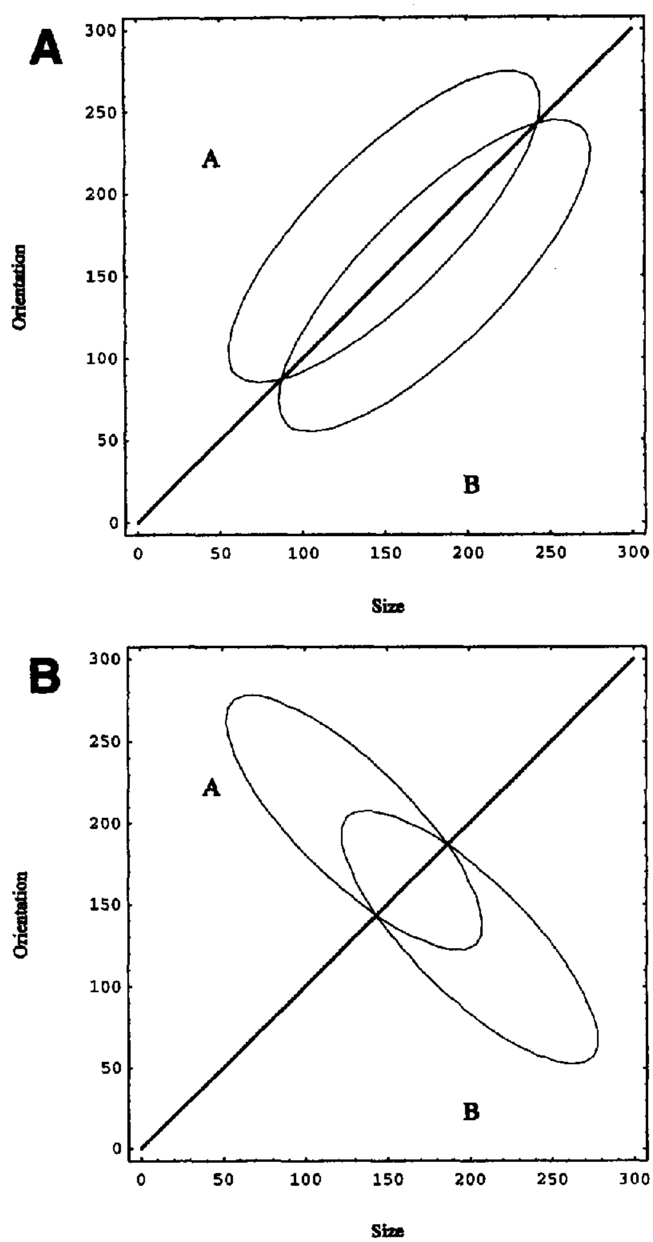



*Figure 2.* Equal-likelihood contours showing the shape of the categories used in Experiment 1 with the optimal bound $y = x$ plotted for (A) positively correlated categories and (B) negatively correlated categories. Points inside the contours have greater likelihood of being sampled as stimuli given the category than as points outside.

the stimulus dimensions as seen in the tilt of the contours. In Figure 2B, both categories have a strong negative within-category correlation. The line drawn in each figure is the computed "optimal bound," which is the bound that if used by the learners, theoretically maximizes probability correct. Technically, in the absence of payoffs and when each category is sampled with equal probability, when classifying a given exemplar, the optimal rule is to choose the category that has the highest likelihood associated with that exemplar (Fukanaga, 1990). The optimal bound, then, corresponds to those points in the $x$-$y$ space that separate regions according to which category yields the greatest likelihood for those points.

The stimuli used in these experiments were circles of varying size with radial lines of varying orientation embedded in them. These dimensions have been shown to be psychologically separable by a number of criteria (e.g., Garner, 1974; Kadlec, 1992; Thomas, 1997) and are well studied in the categorization literature (e.g., Maddox & Ashby, 1993; McKinley & Nosofsky, 1995). Figure 3 shows two example stimuli, the prototypes from two categories used in Experiment 1.

## Predicting Values Along Continuous Dimensions

In addition to categorizing stimuli, the participants were required to predict one dimension's value given a value on the other dimension for a particular category. The categories were (usually, an exception is in Experiment 2) chosen such that there was a strong correlation between the dimensions. This was a visual prediction task in that the given dimension, say size, was displayed as a circle without a line, and the participant's response was to choose an orientation by using the mouse to draw the appropriate radial line. This technique avoids the usual problems of assigning numbers to continua that often plague these types of judgment tasks. The events on a prediction trial were (a) a category was selected, (b) a size or orientation was randomly selected, (c) the category label and either a circle with the selected size or line with the selected orientation was drawn on the screen, and (d) the participant was prompted to choose either the radial line orientation or size by moving the mouse to the appropriate

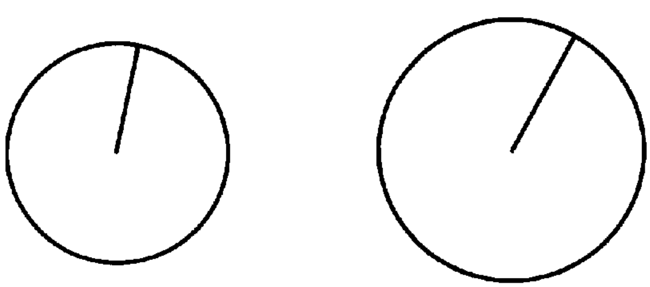


*Figure 3.* Two examples of stimuli used in both experiments. The stimulus on the left is the Category A prototype, and the stimulus on the right is the Category B prototype from the positive correlation condition in Experiment 1. The size is not to scale as a 300-unit radius circle filled a 14-in. (35.56-cm) diagonal computer screen.

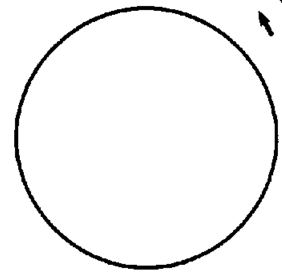



*Figure 4.* The CRT display during a prediction trial. The participant used the mouse to select a size or a line orientation given the prompted category.

place on the screen and pressing the mouse button. This button press would initiate the drawing of a line or circle conforming to the mouse position. A typical prediction-trial display is shown in Figure 4.

## Motivation for the Experiments

The following two experiments were designed to assess whether the participants had retained and could subsequently access information about the exemplar distributions during categorization. In each experiment, a categorization-training task was followed by a surprise prediction task. Category-distribution parameters were chosen so that the optimal bound was the same across conditions within an experiment but that the underlying correlation between the stimulus dimensions was vastly different across the conditions (e.g., a positive correlation in one and a negative correlation in the other). If only the boundary was learned during training, then for the prediction trials, the participant-generated sizes and orientations would not differ across category conditions because the optimal bound was identical for both conditions. Exactly how rule users perform this task is not presently specified. It may be that their predictions are consistent with a maximal separation of the categories or that they are randomly distributed within the region associated with the category.

If exemplar information is stored during categorization training and is later accessible, then the prediction responses should be at least qualitatively consistent with the underlying correlation. For example, if the learner estimates the category densities perfectly (as the GCM model predicts in the limit), then, when faced with a prediction trial, he or she may perform a regression across the stored exemplars to produce the objective expected value (i.e., normative value) for the queried dimension, conditional on the value of the dimension provided by the experimenter (see Anderson & Fincham, 1996, for possible models of this task that incorporate a similar idea). Actual performance in the prediction task, even when relying on exemplars, will no doubt differ from the normative functions due to sampling error and noise in the memory representation. However, the predictions should be in qualitative agreement with the correlation.

Note, however, that the prediction data cannot directly test what type of process was actually used to categorize the stimuli. It may be that, even if the predictions do conform to the normative function, the participant still used a rule-like categorization process but stored exemplars as well (see above discussion). However, parsimony may favor an exemplar-based process underlying both tasks in this case. The recent trend in the literature has been to relate performance in a variety of tasks by using a common underlying theoretical framework (Ashby & Lee, 1991; Ashby & Maddox, 1994; Nosofsky, 1986; Thomas, 1996). On the other hand, if the predictions fail to reveal the underlying correlation, strong doubt is cast against an exemplar-based categorization process. One would have to provide an explanation for why this information would be available to classify a novel stimulus but not be available immediately to draw inferences concerning properties of exemplars within a category. The literature regarding correlation learning reviewed above would suggest this scenario is unlikely to occur. Nevertheless, from a logical point of view, the prediction task can only speak to what information was both stored during categorization training and subsequently accessed later and not necessarily to what process was actually used to categorize the stimulus.

As a forecast of the findings, participants in Experiment 1 were able to accurately predict values along one dimension given a value on the other when the within-category correlation was positive (see Figure 1A). When this correlation was negative (see Figure 1B), participants appeared to selectively attend to orientation during categorization and, generally, were less able to reproduce the correlation during prediction. Unexpectedly, a few participants in this condition exhibited a peculiar result, suggesting that the location

Table 1
*Category Parameters and Optimal Decision Bounds Used in Experiments 1 and 2*

| Parameter | Category A | Category B |
|---|---|---|
| | Experiment 1 | |
| Positive correlation | | |
| $\mu_x$ | 150 | 180 |
| $\mu_y$ | 180 | 150 |
| $\sigma_x$ | 40 | 40 |
| $\sigma_y$ | 40 | 40 |
| $\rho_{xy}$ | +.80 | +.80 |
| Normative prediction functions | $y = .8x + 60$ | $y = .8x + 6$ |
| | $x = .8y + 6$ | $x = .8y + 60$ |
| Negative correlation | | |
| $\mu_x$ | 130 | 200 |
| $\mu_y$ | 200 | 130 |
| $\sigma_x$ | 30 | 30 |
| $\sigma_y$ | 30 | 30 |
| $\rho_{xy}$ | −.80 | −.80 |
| Normative prediction functions | $y = -.8x + 304$ | $y = -.8x + 290$ |
| | $x = -.8y + 290$ | $x = -.8y + 304$ |
| | Experiment 2 | |
| Gaussian correlation | | |
| $\mu_x$ | 120 | 220 |
| $\mu_y$ | 120 | 120 |
| $\sigma_x$ | 45 | 45 |
| $\sigma_y$ | 45 | 45 |
| $\rho_{xy}$ | +.95 | −.95 |
| Normative prediction functions | $y = .95x + 6$ | $y = -.95x + 319.5$ |
| | $x = .95y + 6$ | $x = -.95y + 324$ |
| Uniform correlation | $f_a = .81u_1 + .03u_2 + .13u_3 + .04u_4$ | $f_b = .04u_1 + .13u_2 + .03u_3 + .81u_4$ |
| Normative prediction functions | | |
| ($x < 165$) | $y = 103.0$ | $y = 204.6$ |
| ($x > 165$) | $y = 204.6$ | $y = 103.0$ |
| ($y < 167$) | $x = 103.1$ | $x = 226.9$ |
| ($y < 167$) | $x = 209.1$ | $x = 120.9$ |

*Note.* For Experiment 1, the optimal bound was $y = x$, and for Experiment 2, $xy - 167.4x - 165y + 27621 = 0$. $\mu$ = mean; $\sigma$ = standard deviation; $\rho$ = correlation; $u_i$ is a uniform density over regions demarcated by optimal decision bound starting at lower left and moving clockwise (see Figure 10B). $X$ = size, $Y$ = orientation.

of the boundary that best separated the categories may have interacted with their memory for exemplars distributed nearby (as shown later in Figure 7). That is, these participants did give accurate predictions but only for given sizes or orientation within their learned categorization regions, a finding not presently accounted for by exemplar-based theories (see General Discussion). Experiment 2 was a replication of Experiment 1 but with category distributions that would require attention to be distributed across both dimensions. The underlying correlation within categories differed across the two conditions in Experiment 2, but the same decision boundary described optimal performance just as was the case for Experiment 1. Evidence for exemplar storage was obtained again, as well as the unexpected interaction between the boundary and the category exemplars that was found in Experiment 1.

## Experiment 1

Experiment 1 contrasted two different categorization training conditions, one in which the correlation between dimension was large and positive within both categories learned and one in which this correlation was large and negative. For both conditions, the rule or decision bound that would maximize accuracy was the same bound. In both conditions, the category distributions were bivariate Gaussian (normal). The equal-likelihood contours for the category distributions appear in Figure 2. The actual parameters defining the categories appear in Table 1, together with the optimal bounds and objective (normative) prediction functions (see Experiment 2 for an explanation of the category parameters used there). The $x$-dimension refers to the circle size, and the $y$-dimension refers to line orientation.

In the positively correlated condition, the correlation between the size and orientation dimension was +.80 for both categories. The optimal decision bound is the line $y = x$, so that if a stimulus falls above this line, the participant should respond "A" and, if it falls below, he or she should respond "B." In the negatively correlated condition, the correlation between size and orientation was −.80 in both categories. Observe that the optimal bound in this condition is also the line $y = x$. The amount of overlap between the categories was the same in both conditions so that the optimal probability of correct classification was approximately 88% in both cases.

Recall that if the participant stored distribution information during categorization training and could later access this information for the prediction task, then the predictions should show a strong positive trend in the positively correlated condition and a strong negative trend in the negatively correlated condition. If the participant stored only decision bounds, then the prediction functions, whose form is unspecified, should be the same for both conditions because the optimal boundaries for the two conditions are the same.

### Method

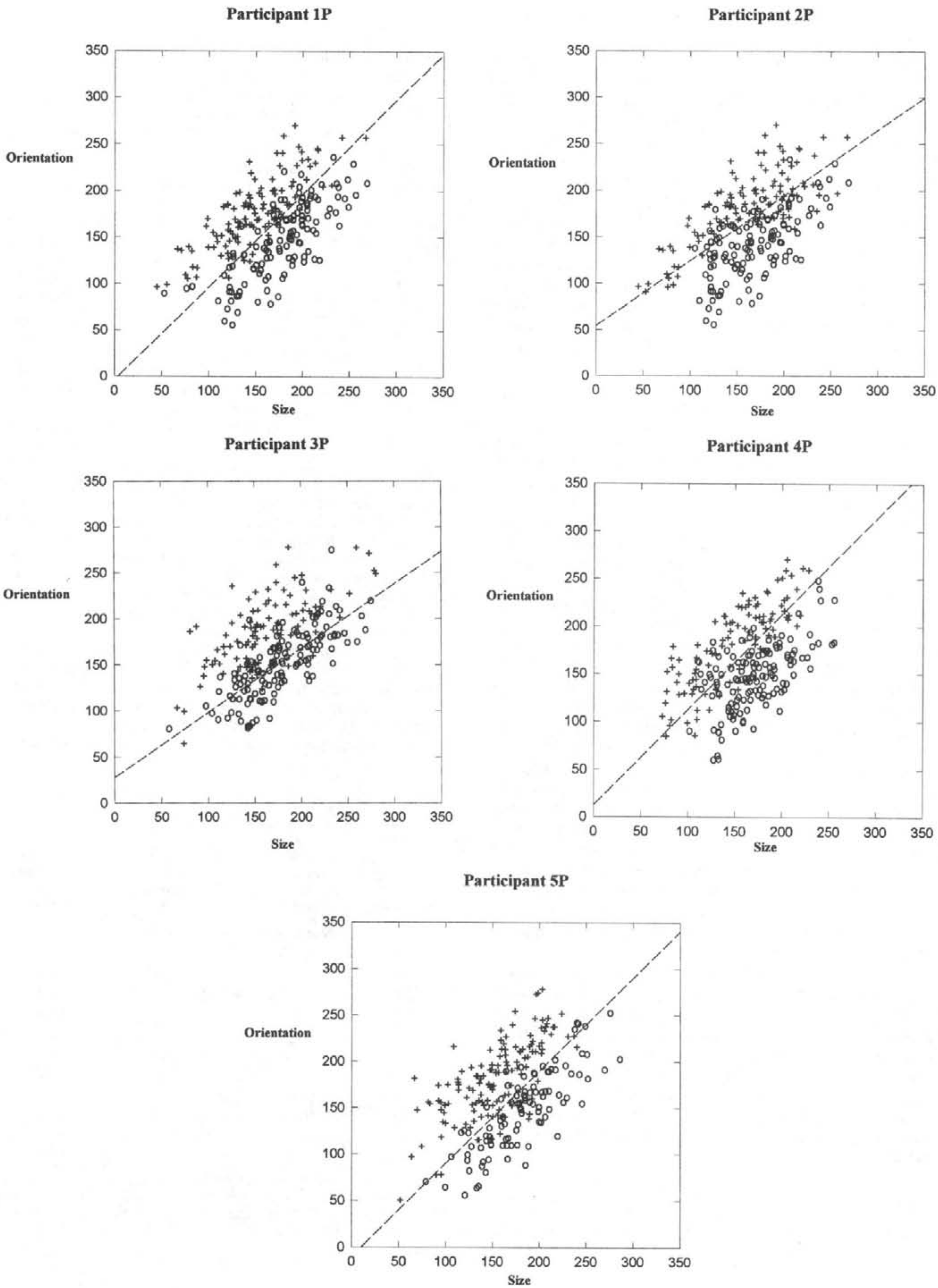
*Participants.* Ten graduate students served as participants, 5 in each condition. They were paid $5 per session plus a bonus based on performance in the categorization task. None of the participants were aware of the theoretical questions being addressed.

*Materials and procedure.* All of the experiments were conducted on a CompuAdd 386 with enhanced VGA video (640 × 400). There were three sessions total. First, there were two sessions of 960 categorization trials (in four blocks of 240 trials each), using the generalized randomization technique. On a categorization trial, one of the categories was randomly selected (with equal probability) and an orientation value and size value were sampled according to the category's distribution parameters. These values were then used to construct the stimulus. The orientation units were such that 0 corresponded to a horizontal line pointing to the right and 200 corresponded to a line pointing straight up (an angle = $\pi/2$). The size units were such that a circle with a radius of 300 filled the screen (14 cm = 300 size units). Sizes below 30 units (1.4 cm) were not used because these were deemed too small for reliable perception. Sizes sampled above 300 were also discarded. To maintain sampling symmetry, angles of less than 30 units or greater than 300 were not used either. The probability of sampling a size or angle outside of the restricted region was less than 2%. The learners had unlimited time to input their responses by using two keys on the computer keyboard; this was not a timed task. Subsequent to participants' response, one of the words *CORRECT!* or *INCORRECT* was displayed on the center of the screen. Participants then used the space bar to initiate the next trial. Participants were encouraged to take breaks often. In the third session, one block of 240 categorization trials was followed by an unannounced, two blocks of prediction trials, in which participants predicted one dimension given a value on the other, as described above (120 trials each block). There was no feedback during the prediction trials. The given dimensional values were sampled with uniform probability from the possible range. Each category was sampled with equal probability as well. The participants were not informed of the prediction blocks prior to completing the final categorization block. During the experiment it was feared that participants might selectively attend to one of the dimensions in the negatively correlated case because the difference in predicted percentage correct for a best one-dimension bound versus the optimal bound

Table 2
*Categorization Response Characteristics*

| Participant | Last session % correct | Best fitting linear discriminant bound | % responses accounted for by bound |
|---|---|---|---|
| | Negatively correlated condition | | |
| 1N | 89.0 | $y = 3.2x - 378.9$ | 93.3 |
| 2N | 86.7 | $y = 0.1x + 145.2$ | 93.3 |
| 3N | 87.9 | $y = 0.6x + 70.8$ | 93.3 |
| 4N | 82.9 | $y = 0.6x + 75.9$ | 91.3 |
| 5N | 82.9 | $y = 0.4x + 101.5$ | 92.5 |
| | Positively correlated condition | | |
| 1P | 83.3 | $y = x - 3.7$ | 87.5 |
| 2P | 75.0 | $y = 0.7x + 54.5$ | 80.8 |
| 3P | 76.3 | $y = 0.7x + 27.9$ | 81.3 |
| 4P | 77.5 | $y = x + 12.5$ | 79.2 |
| 5P | 79.6 | $y = x - 10.1$ | 85.0 |

*Note.* Optimal percentage correct is 88%. The optimal bound is $y = x$. There are different participants in the different groups even though they are numbered the same. N = participant in the negatively correlated condition; P = participant in the positively correlated condition.


Participant 1P
Participant 2P
Participant 3P
Participant 4P
Participant 5P
Orientation
Size
0
50
100
150
200
250
300
350

was only 3%. Forty categorization transfer trials were included at the end to assess this possibility in the negatively correlated condition. The stimuli for the transfer trials were randomly selected from the lower left and upper right regions in the space but away from the presented exemplars. Because these trials occurred after all of the other trials, they did not have any effect on the prior predictions. Their inclusion, after the fact, was designed to establish whether learners were selectively attending to only one of the dimensions during categorization training.

### *Results and Discussion*

Each participant's categorization response characteristics for the final session are summarized in Table 2. The column labeled "Best fitting linear decision bounds" contains the equation for the linear bound found from linear discriminant analysis of the participants' responses (Morrison, 1976). These bounds are linear discriminant functions that maximally separate the participants' categorization responses. For example, an "A" response was correctly predicted if it was above the bound, and a "B" response was correctly predicted if it was below. The number of correct predictions was maximized by the discriminant bound. How well this bound accounts for the responses is shown in the last column (if the linear bound perfectly described performance, this percentage would be 100). Figures 5 and 6 plot these responses as points in the two-dimensional stimulus space for the positively correlated condition and negatively correlated condition, respectively. The symbols denote the learners' responses, not the correct category labels for the stimuli.

From Table 2, it is evident that the overall difficulty of the positively correlated case, as reflected in the last session's percent correct, was greater than that of the negatively correlated condition, even though amount of distributional overlap was held constant. No doubt the requirement to integrate information across the two dimensions in the former condition led to the greater difficulty. Any deviation of the participant's actual boundary from the optimal one in the positively correlated case drastically reduced accuracy. In contrast, accuracy was not reduced as much in the negatively correlated case when participants deviated from optimality.

Selective attention did occur in the negative correlation case in that the best fitting decision bounds in the positive case were closer to the optimal bound ($y = x$) than were the best fitting bounds in the negative correlation condition. Another interesting feature to note is that when the distribution of categorization responses was examined, the category responses for the negatively correlated group did not overlap (Figure 6). Responding for these learners was deterministic, replicating similar findings of Ashby and his colleagues (Ashby & Gott, 1988; Ashby & Maddox, 1992; Maddox & Ashby, 1993). However, when the responses of the learners in the positive condition were examined, there was much greater overlap (Figure 5). This is also as shown by the lower percentages accounted for by the best fitting bound. The differences between the response characteristics of the two groups suggests the possibility of different categorization strategies, although task difficulty may have lead to greater perceptual noise in the positive condition.

Table 3 contains the obtained slopes for the best fitting simple linear regressions to the prediction data. There are four regressions per participant (two categories, two regressions each). In two cases in the negatively correlated condition, the predictions unexpectedly separated into two clusters (see Figure 7). Participants 2N and 5N (i.e., Participants 2 and 5 in the negatively correlated condition) produced predictions consistent with the correlation between dimensions only when the given size or orientation was in their response region for that category. Hence, to capture the correct nature of their predictions, a linear spline model with a discontinuity was fit to these data (Montgomery & Peck, 1982). Essentially, this involves fitting two simple linear regression models to the predictions, one model for predictor values less than some point and another model for given values greater than that point (the technical details of this procedure are discussed in the results of Experiment 2, where hypothesis testing with this model was required). The slopes of the spline pieces for prediction functions that involve dimensional values typical of the category are the ones reported in Table 3 for Participants 2N and 5N and are the ones used in the statistical analysis below. An example of a participant's prediction data that were described by a spline regression is shown in Figure 7. Recall that high orientations are typical of Category A exemplars. As can be seen in Figure 7B, for orientations greater than 151, the fitted regression line to size predictions given Category A does accurately capture the correlation between the dimensions. For these learners, the availability of exemplar information during the prediction task appeared to be contingent on whether the given dimensional value satisfied their categorization rule. The categories in this paradigm overlapped by over 12%, so the learners did see Category A's exemplars in the B region and vice versa. Even though the percentage of overlap was moderately low, there were 2,160 categorization trials, so that the number (260) of "exceptions" to the boundary was significant. However, they apparently did not store those exceptions during training. The importance of this result along with similar results obtained from Experiment 2 are highlighted in the General Discussion.

The main effect of condition (positive versus negative correlation) on the predictions can be seen by examining the slopes for the estimated regression lines within each category across the two conditions. The regression slopes for the positively correlated condition were greater than those in the negatively correlated condition: for Category A, $t(8) = 8.1, p < .001$; for Category B, $t(8) = 3.98, p < .01$. Figure 8 shows prediction responses with the best fitting regression lines that were typical for learners in the positive correlation

*Figure 5 (opposite).* Categorization responses from the last session of training of the 5 participants in the positive correlation condition (P). $N = 240$; $\rho = .80$; pluses indicate Category A, and circles indicate Category B.

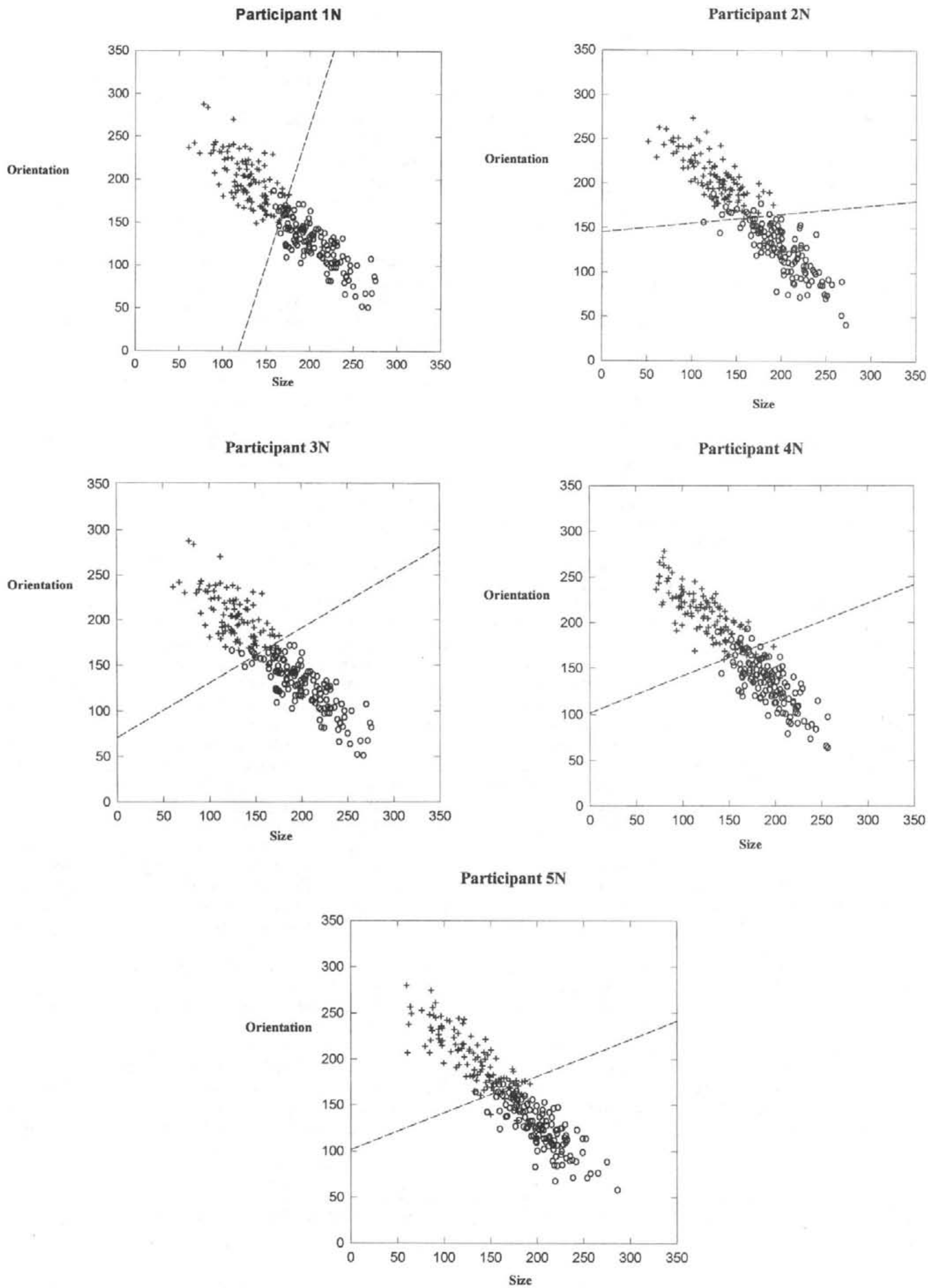

Participant 1N
Participant 2N
Participant 3N
Participant 4N
Participant 5N
Orientation
Size

Table 3
*Regression Slopes for Prediction Trials*

| Participant | Category A: *Y* given *X* | Category A: *X* given *Y* | Category B: *Y* given *X* | Category B: *X* given *Y* |
|---|---|---|---|---|
| | Negatively correlated condition | | | |
| 1N | −0.61 | 0.06 | 0.02 | −0.03 |
| 2N | −0.42 | −0.91* (−.009) | −0.62* (.28) | −1.00* (−.02) |
| 3N | −0.08 | −0.16 | 0.01 | −0.10 |
| 4N | −0.26 | −0.76 | 0.51 | 0.49 |
| 5N | −0.80* (−.48) | −0.65 | −0.55 | −0.83 |
| *M* | −0.43 | −0.48 | −0.13 | −0.28 |
| | Positively correlated condition | | | |
| 1P | 0.82 | 0.43 | 0.97 | 0.85 |
| 2P | 0.58 | 0.63 | 1.10 | 0.73 |
| 3P | 0.58 | 1.10 | 0.67 | 1.10 |
| 4P | 0.74 | 0.71 | 0.52 | 0.50 |
| 5P | 0.92 | 0.63 | 0.75 | 0.58 |
| *M* | 0.73 | 0.70 | 0.80 | 0.76 |

*Note.* Asterisk indicates a spline regression. The first slope tabled is that of the regression line within the category region prompted as defined by the participant's responses. This slope was used for the statistical analyses. The second number in parentheses is the regression slope describing the predictions for given dimensional values associated with the "other" category. *Y* = orientation; *X* = size; N = participant in the negatively correlated condition; P = participant in the positively correlated condition.

condition. Participants in this condition were very good at reproducing the correct correlation, as can be seen in the strong positive slopes of the estimated regression lines.[3]

When the transfer trials in the negative correlation condition were visually examined, it appeared that participants were attending primarily to orientation. This selective attention effect may have greatly attenuated performance in the prediction task. However, because the best fitting slopes were somewhat positive, some integration across dimensions is suggested. In three of the five cases, the prediction responses were flat, like those of Participant 3N shown in Figure 9 (Figure 9A shows size given orientation, and Figure 9B shows orientation given size). This suggests that attention to both dimensions during categorization may be required for the storage of correlational information.

In summary, category learners in the positive correlation condition were overwhelmingly able to reproduce the underlying correlation, as reflected in their postcategorization predictions. In contrast, perhaps because of the effects of selective attention, the learners in the negative correlation condition were less able to use the underlying correlation to generate predictions, although evidence exists that some exemplar information was stored during categorization. Finally, 2 participants' predictions reflected memory for exemplar information only in the region they associated with the given category. The sharpness of the break in the cluster of predictions for these 2 participants is remarkable especially considering that the categories overlapped considerably and that this overlap was gradual given the continuous nature of the category distributions.[4]

## Experiment 2

One alternative explanation for the excellent prediction results in the positive correlation case might be to argue that

[3] The fact that the positive relationship was learned better than the negative relationship bears some resemblance to results from multiple-cue learning in decision-making research (e.g., Naylor & Clark, 1968). However, caution in interpreting the direction of the relationship between size and orientation in the present context must be observed. This is because orientation is arbitrarily labeled as *positively increasing* if the line is rotating counterclockwise. Unless the participants are aware of this mathematical convention, one line orientation being deemed as greater than another will not make sense to them. The greater learning in the positive-correlation condition when compared with the negative-correlation condition is no doubt due to the fact that attention to both dimensions is required in the former but not as much so in the latter.

[4] One concern may be that the low frequency of experience with category exemplars accounts for the failure to predict in the other category response region. However, sampling frequency for exemplars gradually declines as one moves away from the mean of the category because of the continuous nature of the underlying distributions. The sharpness of the break in the clustering of the predictions strongly suggests something special about the category boundary beyond mere frequency of experience.

*Figure 6 (opposite).* Categorization responses from the last session of training of the 5 participants in the negative correlation condition (N). $N = 240$; $\rho = -.80$; pluses indicate Category A, and circles indicate Category B.

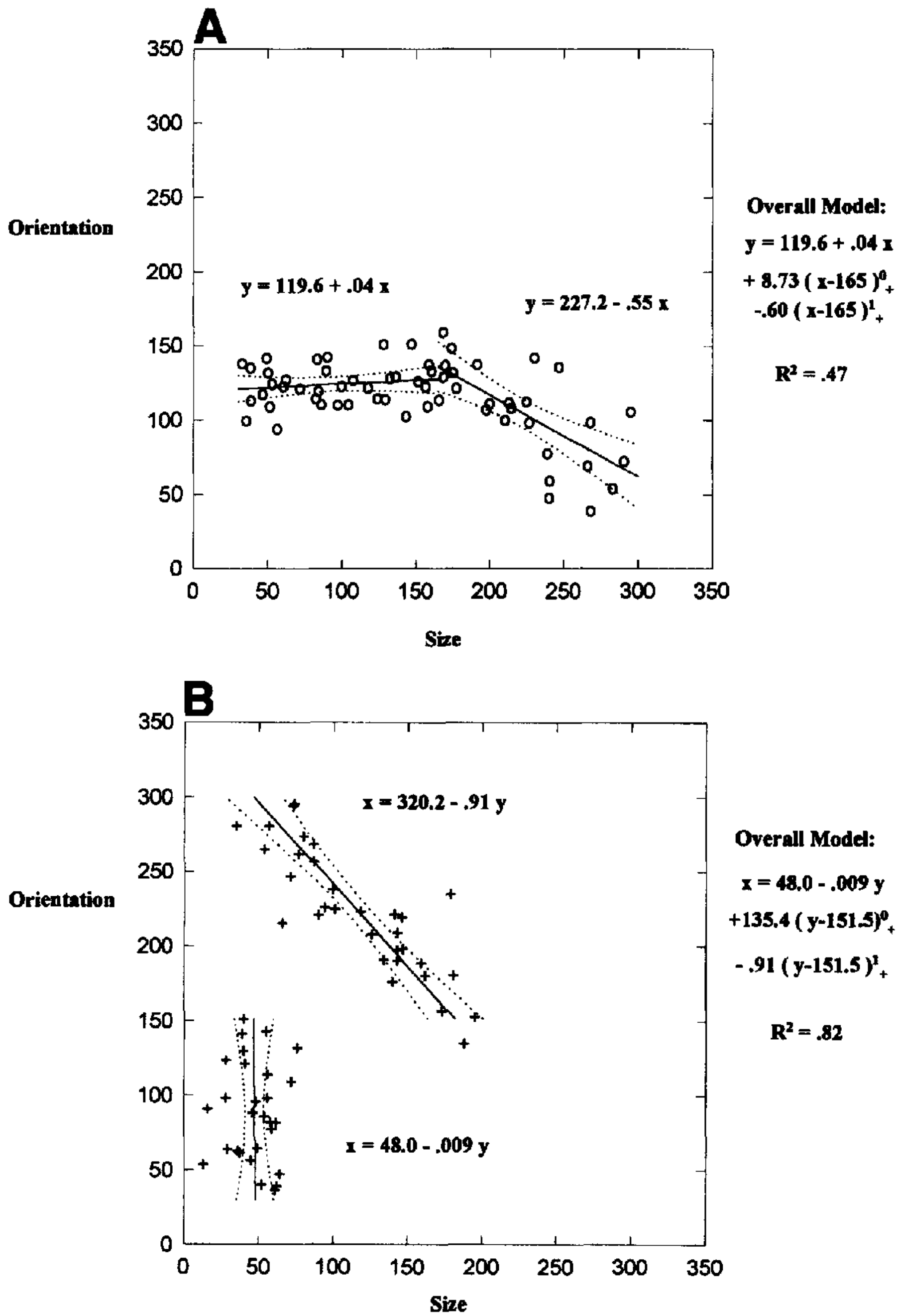



*Figure 7.* Prediction responses for two of the participants, (A) Participant 5N and (B) Participant 2N, in the negative correlation condition, together with the best fitting regression lines showing the interaction between exemplar information and the form of the rule. Pluses indicate Category A, and circles indicate Category B.

the participants' predictions were merely reflecting the shape of the optimal bound ($y = x$). That is, the steep positive regression functions for those participants' predictions are simply a copy of the optimal boundary. Thus, it was this boundary that was stored in memory and subsequently accessed from memory. The orthogonal choice of correlations (positive versus negative) was made particularly to refute that possible explanation. However, the operation of selective attention in the negative correlation case makes ruling out this alternative explanation a little more problematic. For this reason, Experiment 2 was essentially a replication of Experiment 1 but with category distribution parameters chosen so that learning the optimal bound would require participants to distribute attention across both stimulus dimensions.

Again, there were two categorization conditions. In the Gaussian condition, two bivariate Gaussian (normal) distributions were chosen to generate the category exemplars. The Gaussian distributions led to an optimal bound that can be described logically as an exclusive–or rule (see Figure 10). That is, if the size and orientation are small or if the size and orientation are big, then the rule would classify the exemplar as an "A," otherwise it would be a "B." In the uniform condition, two probability mixtures of uniform distributions were chosen so that the optimal bound would be the same as in the Gaussian case. Recall that optimal responding occurs when, for a given exemplar, the responder provides the category label that has the highest likelihood. In Figure 10B, throughout the lower left and upper right quadrants, Category A has the highest likelihood. In the other two quadrants, Category B is most likely. Therefore, the optimal boundary in this case emerges by construction of the densities. Also, the mixtures were chosen so that the distributional overlap in the uniform case would be the same as the amount of overlap for the Gaussian distributions (see Table 1 for category parameter values and Figure 10 for equal-likelihood contours in the Gaussian condition and the density function in the uniform condition).

To understand the category distributions in the uniform condition, consider when the exemplar to be classified is drawn from Category A. For this exemplar, there is an 81% chance that its size is (uniformly) between 30 and 165 and that the line orientation is (uniformly) between 30 and 167.4; there is a 3% chance that the size is between 30 and 165 and that the angle is greater than 167.4 but less than 300; there is a 13% chance that the size is greater than 165 but less than 300 and the angle is greater than 167.4 but less than 300; finally, there is a 4% chance that both the size is greater than 165 and the angle is greater than 167.4. These four percentages (81, 3, 13, and 4) are equal to the percentages

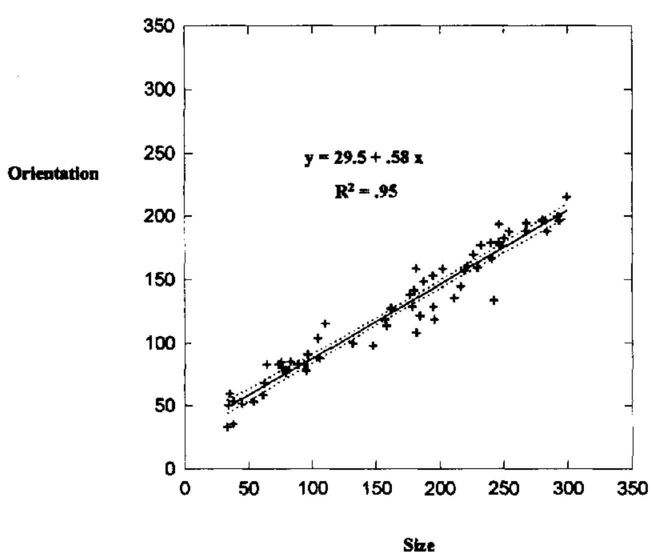



*Figure 8.* Prediction responses for orientation given size judgments in Category A for Participant 3P in the positive correlation condition, together with the best fitting regression lines. The accuracy in the predictions are typical of learners in this condition.

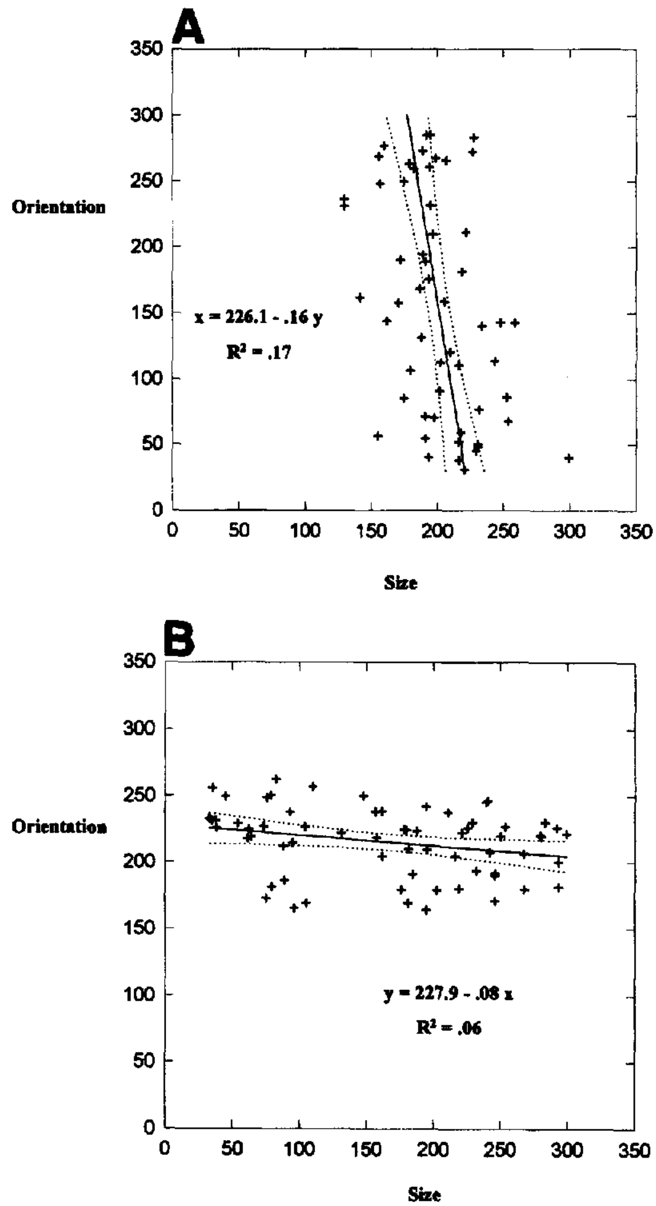



*Figure 9.* Prediction responses of Participant 3N for Category A in the negative correlation condition, together with the best fitting regression lines showing that the form of the category rule determined prediction responses. A: Predictions for size given orientation. B: Predictions for orientation given size.

under the distribution in those regions, respectively, for Category A in the Gaussian condition. A similar uniform mixture describes Category B in the uniform condition. The limits of 30 and 300 correspond to the cutoff values for stimulus construction (see Experiment 1, *Materials and procedure*).

If participants store and can subsequently access exemplar information for the prediction task, participants in the Gaussian-category condition should produce strong positive slopes for Category A predictions and strong negative slopes for Category B predictions. However, if only the exclu-

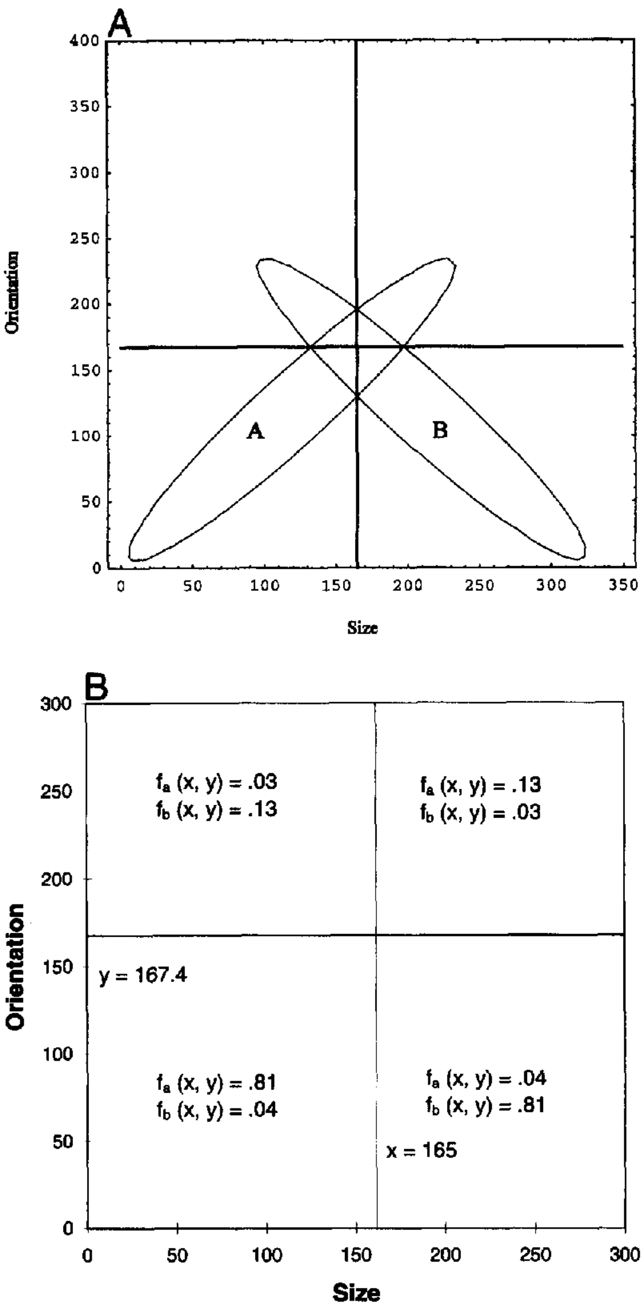



*Figure 10.* Illustrations of the category distributions used in Experiment 1. A: Equal likelihood contours of the categories in the Gaussian condition with the optimal bound plotted ($x = 165$ and $y = 167.4$). B: Schematic showing the category structures used for the uniform condition. Each distribution is a probability mixture of four uniform distributions on the four square regions given by the plotted optimal bound. An optimal responder would classify the stimulus as an "A" if the size were less than 165 and the orientation less than 167.4 or if the size were larger than 165 and the orientation greater than 167.4; otherwise, the response would be "B."

sive–or rule is remembered, then the prediction data would be likely to correspond to step function with a knot (i.e., a jump discontinuity) at the vertical bound for predictions of orientation given size and a knot at the horizontal bound for predictions of size given orientation. Participants were tested in the uniform condition to reject the hypothesis that the shape of the decision bound might induce an illusory correlation between the stimulus dimensions. The prediction functions for this control group would most likely correspond to the above-described step function.

## Method

*Participants.* Five graduate students served in the Gaussian condition, and 4 served in the uniform condition. These participants did not serve in Experiment 1, and none of these participants were aware of the theoretical questions being addressed. Participants were paid \$5 per session plus bonus based on performance in the categorization task.

*Materials and procedure.* The details of the stimuli and apparatus can be found in the *Method* section of Experiment 1. For each participant, there were three categorization sessions of 800 trials each. The generalized randomization technique was used and extended to the uniform mixture distribution case. On the fourth session, 200 categorization trials were followed by 200 prediction trials. Participants were unaware of the upcoming prediction task until after categorization on the 4th day.

## Results and Discussion

Table 4 contains the categorization response characteristics for the last session for Experiment 2. Figures 11 and 12 display the categorization responses for the last session. Visual inspection of the data, participant accuracy, and the large percentage of responses accounted for by the optimal bound indicate that participants did learn the optimal bound in both conditions. This is important because the optimal bound (i.e., the exclusive–or rule) requires attention to both dimensions. Selective attention to the size dimension was a problem in Experiment 1 in comparing correlation learning across the two conditions. Clearly, participants were attending to both dimensions in both the Gaussian and uniform conditions. Responding in both conditions appeared to be fairly deterministic as well in that there was little or no overlap in the responses. This was especially true for the uniform-category condition.

To determine whether the within-category correlations were learned, two types of regression models were fit to each prediction function, a simple regression model (e.g., $y = ax + b$) or a linear spline model. The latter spline model was fit to assess the possibility that only the exclusive–or rule was stored in memory, and thus, a step function might best describe the prediction data. For example, if Category A were under consideration and the participant's task was to predict orientation given size, for sizes less than 165 units, a low (constant) orientation (specifically, $y = 103$; see Table 1) is the best prediction, but for sizes greater than 165, a high

Table 4
*Categorization Response Characteristics*

| Participant | Last 300 trials % correct | % responses accounted for by optimal bound |
|---|---|---|
| Gaussian distribution condition | | |
| 1G | 87.7 | 88.0 |
| 2G | 84.8 | 90.0 |
| 3G | 93.7 | 94.0 |
| 4G | 85.0 | 85.7 |
| 5G | 93.0 | 92.3 |
| Uniform mixture distribution condition | | |
| 1U | 90.3 | 93.7 |
| 2U | 88.0 | 96.0 |
| 3U | 88.0 | 95.0 |
| 4U | 79.0 | 83.0 |

*Note.* Optimal bound for both conditions: $xy = 167.4x - 165y + 27{,}621 = 0$. G = participant in the Gaussian distribution condition; U = participant in the uniform mixture distribution condition.



orientation (204.6) is the best prediction.[5] Formally, the linear spline model is a regression with a knot, $x_0$, so that for example, for a "predict orientation given size" function,

$$\text{for } x \leq x_0,\ y = b + ax,$$

$$\text{for } x < x_0,\ y = b' + a'x. \qquad (1)$$

Recall that the spline is functionally two separate regressions, one for data below the knot and one for data above. This spline model can be written as one regression model,

$$y = b + ax + c\,(x - x_0)^0_+ + d(x - x_0)^1_+, \qquad (2)$$

where $(u)_+ = u$ if $u > 0$ and 0 if $u < 0$ (see, e.g., Montgomery & Peck, 1982). The separate splines in Equation 1, as functions of these arguments are

$$\text{for } x \leq x_0,\ y = b + ax,$$

$$\text{for } x > x_0,\ y = (b + c - dx_0) + (a + d)x. \qquad (3)$$

Viewed this way, the simple regression model is a special case of the spline model in Equation 2, where $c$ and $d$ are fixed at zero. Thus, we can use an $F$ statistic for nested models to determine if the more general model leads to a significant improvement in fit, with a penalty for having extra parameters. If the more general spline model fails to significantly improve fit and the regression in the simple linear model is significant, then there is evidence of correlation learning.

For the nested model comparison, the $F$ statistic is based on the sum of squared error for each regression model,

$$F_{obs} = \frac{(SSE_r - SSE_g)/(df_r - df_g)}{SSE_g/df_g}, \qquad (4)$$

where $SSE_r$, $df_r$ and $SSE_g$, $df_g$ refer to the sum of squared error and degrees of freedom of the simple regression (restricted) and of the spline regression (general), respectively. Under the null hypothesis that the restricted model is correct, $F_{obs}$ follows an approximate $F$ distribution with $df_r - df_g$ numerator degrees of freedom and $df_g$ denominator degrees of freedom (e.g., Khuri & Cornell, 1987).[6]

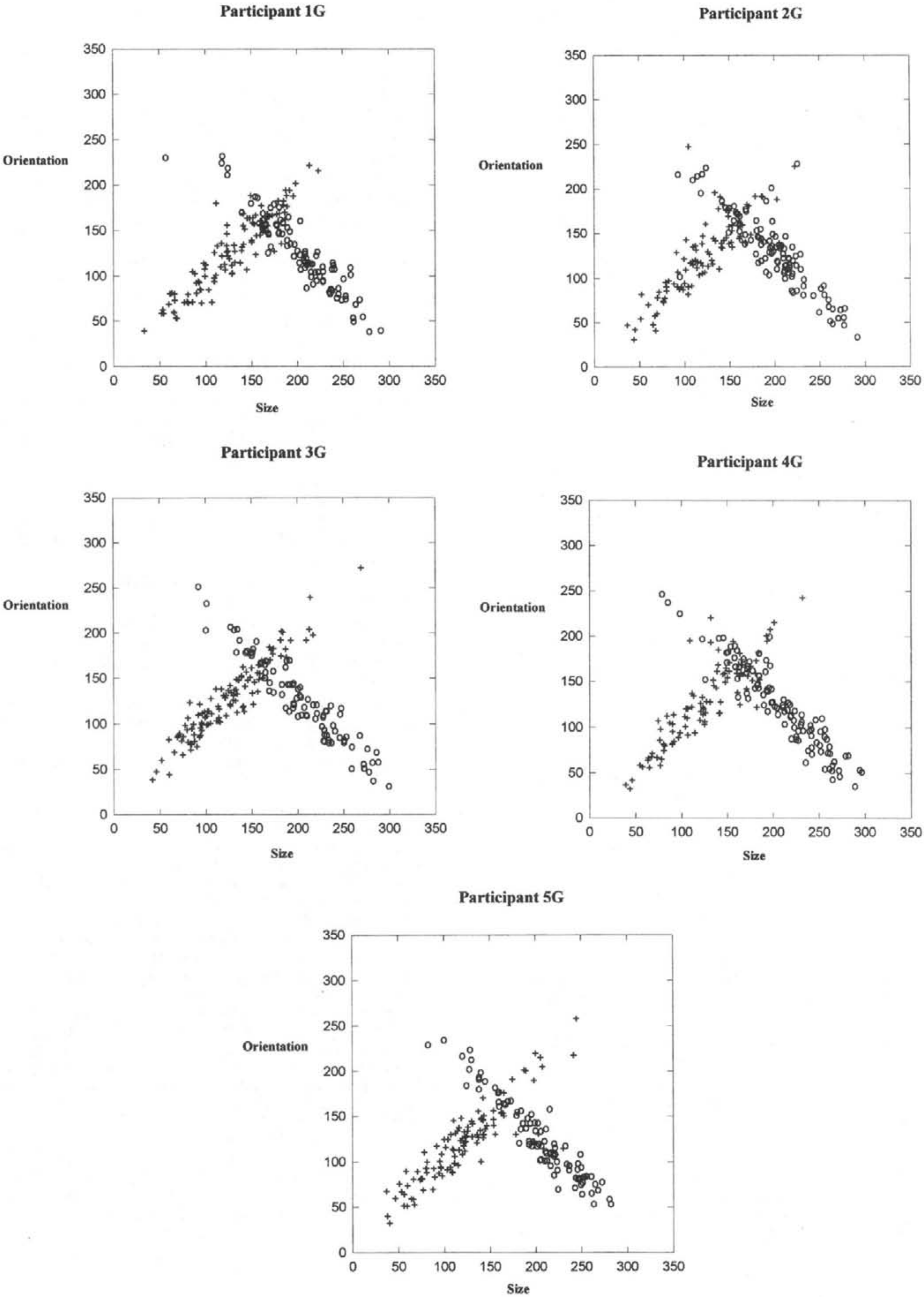
In the Gaussian case, out of the 20 total prediction functions (4 for each of the 5 participants: 2 categories, 2 functions for each category), the more general spline model led to a relative improvement in fit in 14 cases ($p < .05$).[7] In the remaining 6 cases, the simple regression model was significant (i.e., the slope was different from 0 in the predicted direction, all $ps < .05$). In the uniform condition, the spline model outperformed the simple model in all but one case. The best fitting regression models, according to the $F$ test, for each participant are in Table 5. An overall summary of the estimated regression slopes appears in Table 6. This table gives the individual and average slope values for both conditions for both types of prediction trials in the upper and lower regions of the stimulus space (i.e., above and below the bound for orientation).

The prediction results that conformed best to the true underlying correlation were provided by Participant 3G. When prompted with either dimension, this participant provided judgments of the value of the other dimension that were clearly consistent with the correct correlation, as can be seen in Figure 13 by the tight regression very nearly identical to the normative prediction function. In contrast, Figure 14 shows prediction responses from Participant 5G (size given orientation in Figure 14A and orientation given size in Figure 14B, both for Category A), which would be expected if only the category rule was available to guide the predictions. That is, the predictions are homogeneous throughout the square of the exclusive–or rule and break at the middle boundary. Finally, as in the first experiment, several participants appeared to be sensitive to both the

[5] Generally, the location of the knot was allowed to vary across subjects and prediction functions according to a clustering criterion. The data for one type of prediction in a given category had to split into two clusters such that there was a greater-than-50 unit difference on the predicted dimension across the clusters, and the knot was placed at the midpoint. In all cases that did not fit the clustering criterion, the knot corresponded to the optimal boundary location on that dimension. However, in some cases, outlier responses were screened. At most, this screening resulted in the loss of only five observations from approximately 50 responses per prediction condition.

[6] Technically, the statistical test requires independent observations. When individual data are analyzed, in which repeated observations from a single participant are collected, this assumption is almost certainly not satisfied. However, in nearly all modeling analyses of individual observer data, statistical tests based on maximum likelihood or sum of squared error, this assumption has to be made. In any case, the results should be interpreted with this caveat in mind.

[7] The superior fit of the spline model to Participant 3G's size-given-orientation responses for Category B was obviously an artifact of the extremely low standard error of estimate of the regression. This learner clearly stored the category correlation information in all cases.


Participant 1G
Orientation
Size
Participant 2G
Orientation
Size
Participant 3G
Orientation
Size
Participant 4G
Orientation
Size
Participant 5G
Orientation
Size
0
50
100
150
200
250
300
350

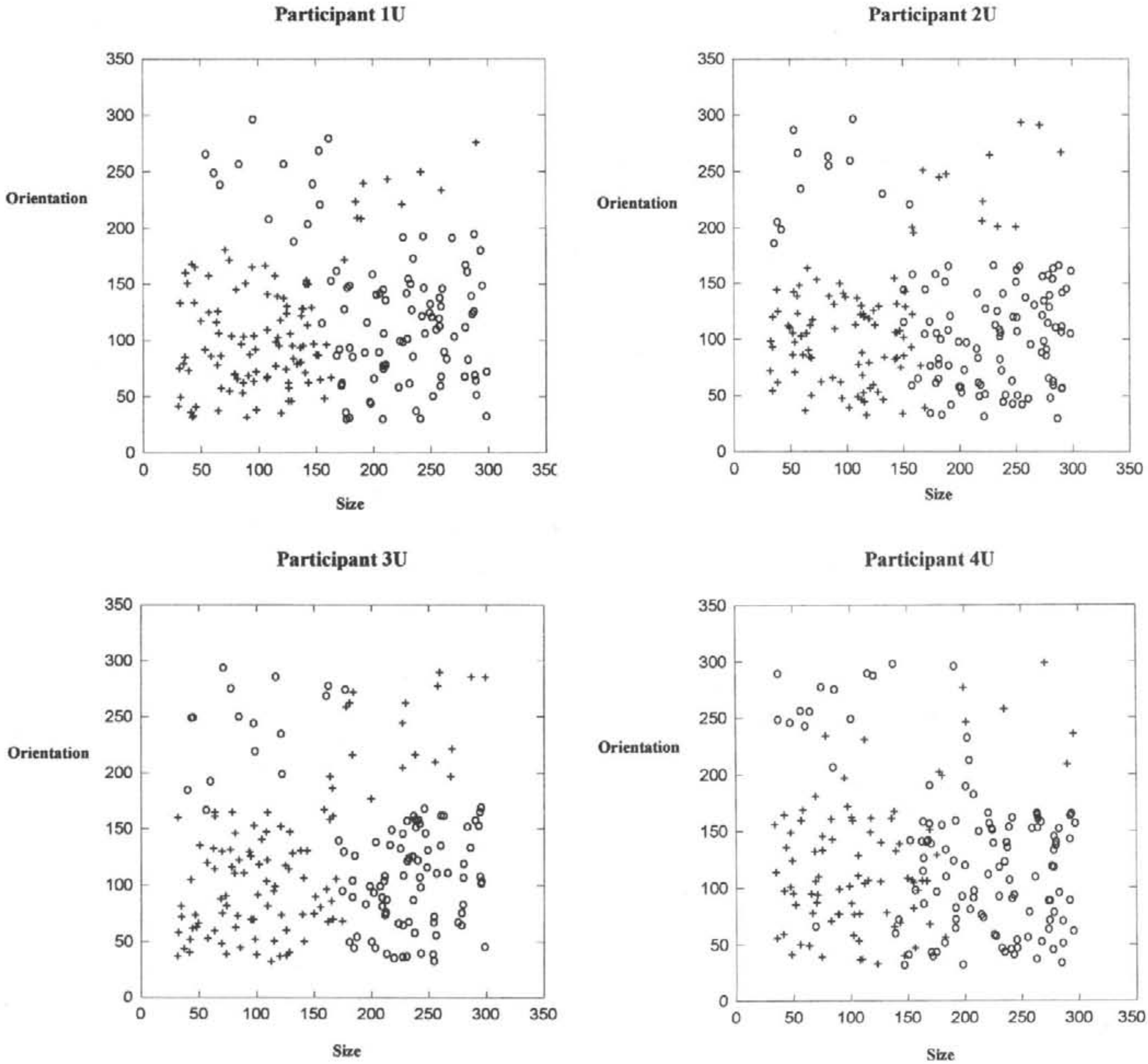



*Figure 12.* Categorization responses from the last session of training of the 5 participants in the uniform condition (U). $N = 200$; pluses indicate Category A, and circles indicate Category B.

correlation and the location of the decision bound that separated the two categories. Telling examples of this behavior are illustrated in Figure 15, showing the prediction responses for Participants 1G, 2G, and 4G. For each of these learners, within the category region containing most of the exemplars, the lower regions, regression slopes of the predictions were consistent with the correlation, but the responses clearly broke off at the category boundary.

The prediction responses in the uniform condition were much more homogeneous across learners (see the small standard deviations in Table 6). Figure 16 shows the data from Participant 3U, which were typical for this condition.

*Figure 11 (opposite).* Categorization responses from the last session of training of the 5 participants in the Gaussian condition (G). $N = 200$; $\rho = \pm.95$; pluses indicate Category A, and circles indicate Category B.

Table 5
*Best Fitting Regression Models for Experiment 2 Predictions*

| Participant and measure | Category A | | Category B | |
|---|---|---|---|---|
| | *Y* given *X* | *X* given *Y* | *Y* given *X* | *X* given *Y* |
| **Gaussian condition** | | | | |
| 1G | | | | |
| Model | $y = 12.4 + .78x$ | $x = 1.45 + .81y + 117.0(y - 165)^0_+ - .49(y - 165)^1_+$ | $y = 267.7 - .75x$ | $x = 278.6 - .11y - 214.6(y - 165)^0_+ + .11(y - 165)^1_+$ |
| $R^2$ | .86 | .88 | .88 | .98 |
| 2G | | | | |
| Model | $y = -79.0 + 1.37x + 77.34(x - 165)^0_+ - 1.2(x - 165)^1_+$ | $x = 93.9 + .07y + 73.9(y - 165)^0_+ + .05(y - 165)^1_+$ | $y = 411.2 - 1.7x - 105.7(x - 165)^0_+ + 1.3(x - 165)^1_+$ | $x = 324.9 - .64y - 132.9(y - 185)^0_+ + .54(y - 185)^1_+$ |
| $R^2$ | .78 | .82 | .77 | .92 |
| 3G | | | | |
| Model | $y = -40.8 + 1.1x$ | $x = 25.7 + .95y$ | $y = 322.6 - .79x$ | $x = 321.4 - .65y - 24.5(y - 165)^0_+ - .46(y - 165)+1$ |
| $R^2$ | .94 | .97 | .92 | .96 |
| 4G | | | | |
| Model | $y = 50.9 + .31x + 69.4(x - 165)^0_+ + .17(x - 165)^1_+$ | $x = 3.2 + .93y + 109.0(y - 200)^0_+ - 1.2(y - 200)^1_+$ | $y = 285.6 - .51x - 56.2(x - 165)^0_+ - .24(x - 165)^1_+$ | $x = 324.5 - .91y$ |
| $R^2$ | .94 | .96 | .82 | .80 |
| 5G | | | | |
| Model | $y = 51.9 + .22x + 188.2(x - 163)^0_+ - .14(x - 163)^1_+$ | $x = 96.2 - .06y + 166.5(y - 165)^0_+ + .14(y - 165)^1_+$ | $y = 351.0 - .40x - 189.6(x - 163)^0_+ + .03(x - 163)^1_+$ | $x = 261.2 - .09y - 189.6(y - 165)^0_+ + .45(y - 165)^1_+$ |
| $R^2$ | .91 | .95 | .97 | .94 |
| **Uniform condition** | | | | |
| 1U | | | | |
| Model | $y = 41.7 + .41x + 207.1(x - 172.5)^0_+ - .2(x - 172.5)^1_+$ | $x = 89.9 + .17y + 122.8(y - 190)^0_+ + .08(y - 190)^1_+$ | $y = 268.6 - .34x - 114.5(x - 141)^0_+ + .28(x - 141)^1_+$ | $x = 294.1 - .36y - 120.7(y - 198)^0_+ + .25(y - 198)^1_+$ |
| $R^2$ | .97 | .91 | .96 | .94 |
| 2U | | | | |
| Model | $y = 55.9 + .21x + 173.3(x - 150)^0_+ - .29(x - 150)^1_+$ | $x = 69.4 + .07y + 169.2(y - 179)^0_+ - .23(y - 179)^1_+$ | $y = 291.5 - .37x - 131.7(x - 152)^0_+ + .22(x - 152)^1_+$ | $x = 277.8 - .29y - 131.8(y - 180)^0_+ + .31(y - 180)^1_+$ |
| $R^2$ | .95 | .97 | .94 | .92 |
| 3U | | | | |
| Model | $y = 109.6 + .09x + 103.3(x - 171)^0_+ - .09(x - 171)^1_+$ | $x = 97.8 - .04y + 203.2(y - 175)^0_+ - .02(y - 175)^1_+$ | $y = 270.0 - .12x - 113.5(x - 178)^0_+ + .08(x - 178)^1_+$ | $x = 301.8 - .14y - 189.4(y - 175)^0_+ + .20(y - 175)^1_+$ |
| $R^2$ | .96 | .99 | .96 | .98 |
| 4U | | | | |
| Model | $y = 96.1 + .02x + 291.8(x - 168)^0_+ - .60(x - 168)^1_+$ | $x = 144.3 - .19y + 208.9(y - 194)^0_+ - .29(x - 194)^1_+$ | $y = 340.5 - .12x - 235.9(x - 158)^0_+ + .08(x - 158)^1_+$ | $x = 350.7 - .91y$ |
| $R^2$ | .99 | .92 | .98 | .82 |

*Note.* $X$ = size; $Y$ = orientation; G = participant in the Gaussian condition; U = participant in the uniform condition.

The prediction responses display the steplike function consistent with the exclusive–or rule and not consistent with exemplar storage and access.

Two types of statistical analyses were performed to assess whether the correlation in each condition had any effect on the prediction responses. First, average slope values across subjects for each type of prediction for both categories and conditions (Table 6) were compared with 0. The slopes used for this test were limited to those in the lower regions of the stimulus space in which most of the exemplars fell when the spline model fit the data best (the first numbers in each column of Table 6). Recall that in the Gaussian case, the learners may have stored exemplar information, but the location of the bound was stored as well (the latter having

Table 6
*Individual and Average Regression Slopes for Predictions From Experiment 2*

| Participant | Category A: *Y* given *X* | Category A: *X* given *Y* | Category B: *Y* given *X* | Category B: *X* given *Y* |
|---|---|---|---|---|
| | | Gaussian condition | | |
| 1G | .78 | .81, .32 | −.75 | −.11, 0.0 |
| 2G | 1.37, .17 | .07, .12 | −1.7, −.4 | −.64, −0.1 |
| 3G | 1.1 | .95 | −.79 | −.65, −1.11 |
| 4G | .31, .48 | .93, −.27 | −.51, −.75 | −.91 |
| 5G | .22, .08 | −.06, .08 | −.4, −.37 | −.09, .36 |
| *M* (*SD*) | .756, .522 | .54, .24 | −.83, −.612 | −.48, −.352 |
| | (.495, .424) | (.493, .45) | (.512, .208) | (.363, .629) |
| | | Uniform condition | | |
| 1U | .41, .21 | .17, .25 | −.34, −.06 | −.36, −.11 |
| 2U | .21, −.08 | .07, −.16 | −.37, −.15 | −.29, .02 |
| 3U | .09, 0.0 | −.04, −.06 | −.12, −.04 | −.14, .06 |
| 4U | .02, −.58 | −.19, −.48 | −.12, −.04 | −.91 |
| *M* (*SD*) | .183, −.112 | .002, −.112 | −.238, −.073 | −.425, −.235 |
| | (.171, .335) | (.154, .3) | (.136, .053) | (.336, .456) |

*Note.* The first number in a cell is the slope of the regression line in the lower region of the stimulus space (below participant's boundary for orientation). The second number is the slope for the predictions above this bound. Only one number appears when the simple linear model fit best. The reported $t$ tests for the slope values used the first number in each column. $X$ = size; $Y$ = orientation; G = participant in the Gaussian condition; U = participant in the uniform condition.

led to the significant improvement of the spline model). The exemplar information may emerge in the slope of the lower regressions where the bulk of the exemplars fell. An analysis of the predictions in the Gaussian condition showed that for Category A, the average slope (.756) for the orientation given angle regressions was greater than 0 by a one-sample $t$ test, $t(4) = 3.41$, $p < .025$. The slope for size given orientation was also greater than 0, $t(4) = 2.45$, $p < .05$. Both prediction slopes for Category B were less than 0, with orientation given size, $t(4) = -3.61$, $p < .025$, and size given orientation, $t(4) = -2.96$, $p < .025$. In the uniform condition, for Category A, neither slope differed from 0, with orientation given size, $t(3) = 2.13$, $p = .061$, and size given orientation, $t(3) = 0.036$, $p = .49$. For Category B predictions, however, both prediction slopes were less than 0: orientation given size, $t(3) = -3.48$, $p < .025$; size given orientation, $t(3) = -2.52$, $p < .05$. This may have been due to the very small standard deviation of Category B regression slopes in this condition.

The Category B average slopes in the uniform condition suggest a possible illusory correlation emerging that was due to the shape of the decision bound. Those statistical tests, though, should be interpreted with caution given the few data points (5 in the Gaussian comparisons and 4 in the uniform cases). The large number of prediction trials for each participant allows us to look at individual data. Thus, to assess perception of correlation, each regression slope for each individual participant was tested for significance in the predicted direction in both conditions. Recall that for the 6 cases in the Gaussian condition in which a simple linear regression best accounted for the data (see above and, e.g., Participant 1G orientation-given-size predictions in Category A), the slope differed from 0 in the correct direction, greater than 0 for Category A and less than 0 for Category B (all $ps < .01$). In 9 of the 14 cases in which the spline model fit significantly better in the Gaussian condition, the slopes were different from 0 and in the correct direction (i.e., positive for Category A and negative for Category B, for both slopes in the spline, above and below the knot, $p < .01$). These results can be contrasted with those in the uniform condition. Of the cases for which the spline fit better, only one spline slope was significantly different from 0. In the one case in which the simple linear model fit best (Participant 4U, Category B, size given orientation) the slope was less than 0 ($p < .05$). However, the bulk of the individual participant data in the uniform condition clearly indicate that the shape of the optimal bound did not lead to the perception of an illusory correlation between dimensions within each category.

In summary, Experiment 2 indicates that exemplar information is retained for most, but not for all, participants when attention to both dimensions is required. Also, the shape of the decision bound did not produce an illusory correlation between dimensions. Finally, the location of the bound appeared to interact with the stored exemplar information for some participants in the Gaussian condition.

## General Discussion

By using large, ill-defined categories composed of stimuli defined on continuously valued dimensions, exemplar information was available for the prediction task for many of the participants. This was evidenced by the different slope characteristics of the regression lines fitted to the participants' predictions across category correlation conditions in both experiments. The prediction responses from partici-

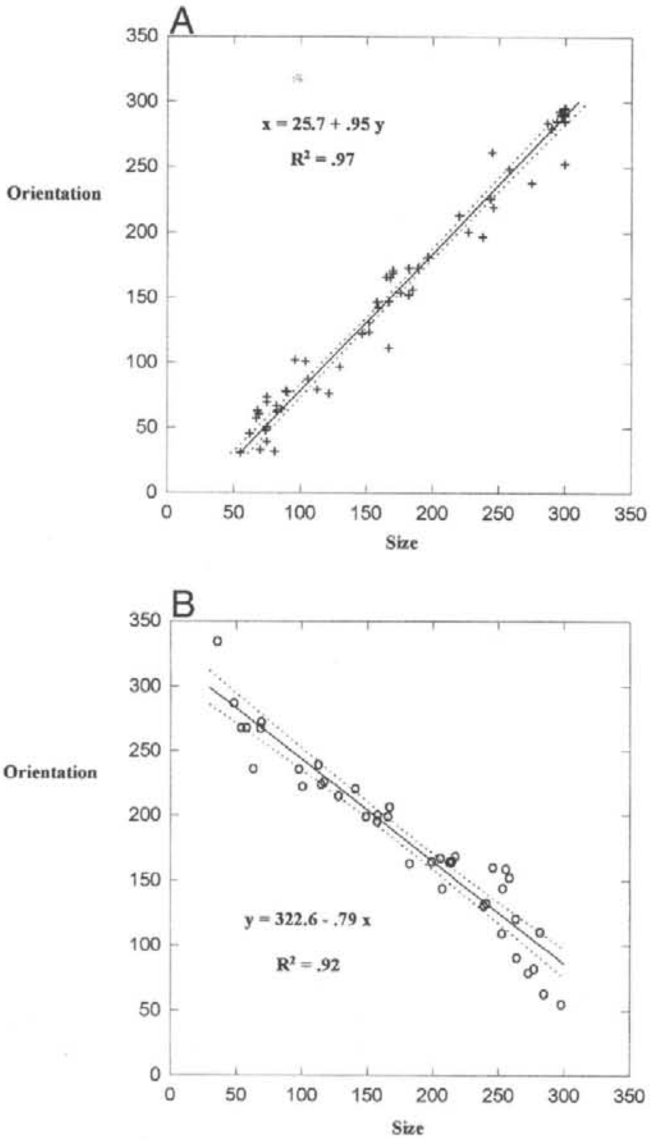



*Figure 13.* Prediction responses for Participant 3G in the Gaussian condition, together with the best fitting regression line showing a high degree of sensitivity to the underlying correlation. A: Size given orientation, Category A. B: Orientation given size, Category B.

pants for whom the stimulus dimensions were positively correlated within each category (Experiment 1) clearly reflected the learning of that correlation. Unfortunately, a selective attention effect in the contrasting (negative correlation) condition of Experiment 1 prevented many of these participants from producing analogous results. One interesting consequence of this failure is to suggest that selective attention to a specific dimension during categorization may prevent information about the other dimension from being stored in memory. Unexpectedly, two of the learners' predictions in this latter group displayed an interaction between exemplar information and the category boundary location (Participants 2N and 5N, Figure 7). For these participants, the correct regression was reproduced only within the response region they had associated with the given category. From Experiment 2, in the Gaussian condition, the prediction results of many participants were consistent with the underlying correlation but, again, not always, and sometimes only within the learner's response region for the given category. We can, however, rule out the hypothesis that the shape of the decision bound produced an illusory correlation for these learners. Participants in the uniform condition, in which the optimal category decision bound was identical to that of the Gaussian categories, did not generate the regressions indicating a perceived (false) correlation within the categories. Thus, for most learners, an exemplar-based categorization strategy provides the most parsimonious account of performance in both tasks. For those few participants failing to learn the category correlations, a rule-based approach is strongly suggested.

The unexpected but important finding that, for some, the exemplar distribution and location of the decision rule interacted during the prediction task (e.g., Participants 2N and 5N of Experiment 1 and Participants 1G, 2G, and 4G in Experiment 2) suggests two possible conclusions. One is that categorization may not be governed entirely by either exemplars or rules but by some mixture of both processes. Pure exemplar models (e.g., the GCM) would have difficulty predicting this phenomenon without incorporating the forgetting of exemplars on error trials. Failing to store exemplars from error trials would be a major change in the structure of these models. This change might lead current models to more poorly predict categorization performance, especially for asymptotic learning. For example, the GCM currently predicts near optimal categorization responding after extensive training (Ashby & Alphonso-Reese, 1995; Ashby & Maddox, 1993), an empirical result that has been obtained in many studies (Maddox & Ashby, 1993; McKinley & Nosofsky, 1995), including the present one. To do this, however, the model needs to estimate the category distributions very nearly accurately. This estimation includes exemplars from the "wrong" category region. A second interpretation of these data may simply be that participants form and use the decision boundary for categorization but residually store some exemplar information. The challenge for this interpretation would be to explain the differential forgetting of exemplars when they appeared in the wrong category region.

The mixture approach has received some consideration in categorization research and in other learning domains. In language acquisition research, the peculiar three-stage learning of past-tense verbs (Ervin, 1964) has led some investigators to call for both rule and exemplar memory modules to account for the data (Pinker, 1991; Prasada & Pinker, 1993). In addition to RULEX (Nosofsky et al., 1994), another modeling scheme incorporating both processes within a connectionist framework is currently being investigated (Erickson & Kruschke, 1996; Kruschke & Erickson, 1994). Essentially, the model is a probability mixture of two modules: a rule-based module, modeled as a set of linear-threshold perceptrons aligned to the dimensional axes of the input space, and an exemplar module, modeled as a set of

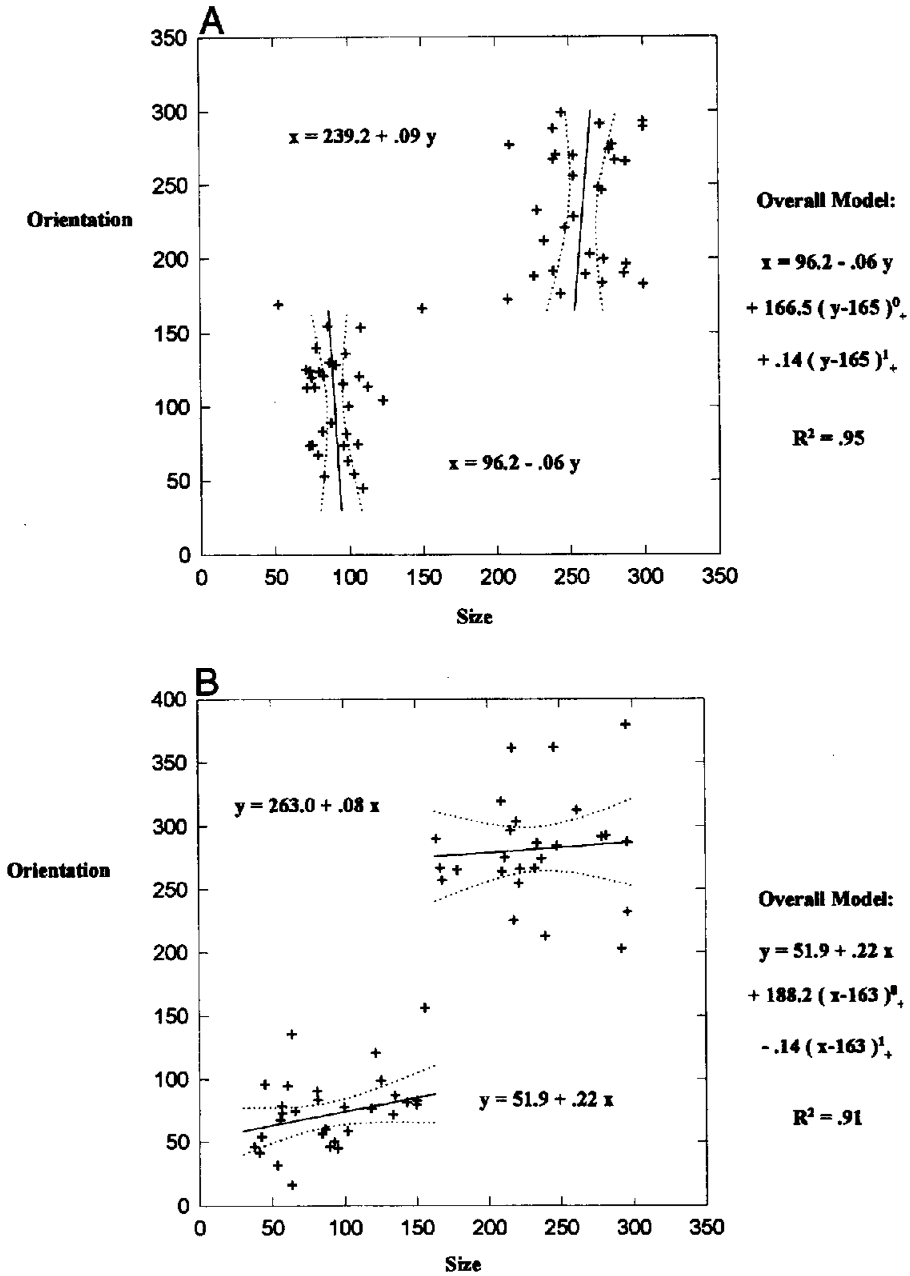
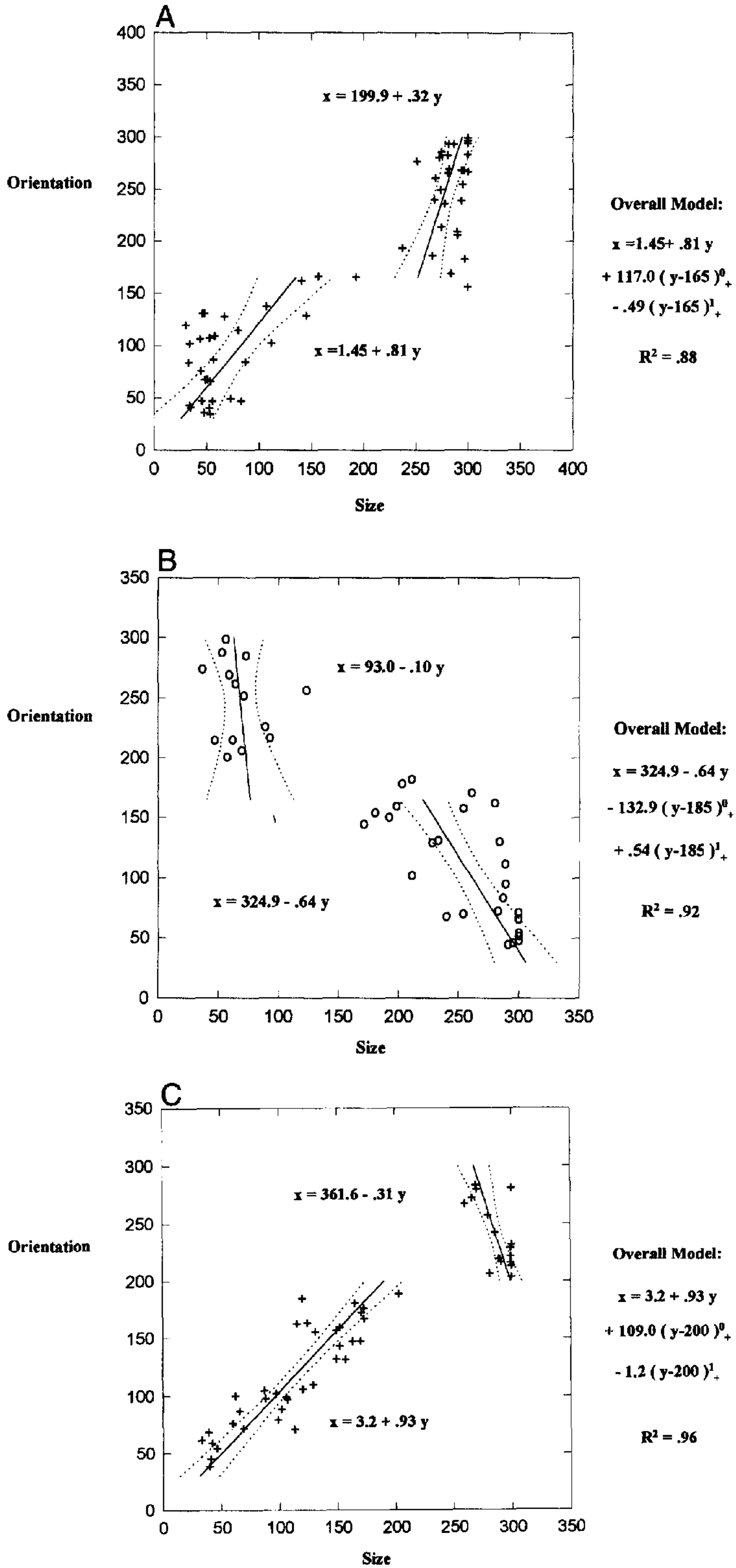



*Figure 14.* Prediction responses (Category A) for Participant 5G who apparently had access to only the exclusive–or rule representing the categories. A: Size-given-orientation responses. B: Orientation-given-size responses.

exemplar nodes, each activated by a limited region of the input space according to the similarity of the stimulus to the given exemplar node. The exemplar module is an extant connectionist instantiation of the GCM of Nosofsky, termed *attention learning covering map* (ALCOVE; Kruschke, 1992). The relative contribution of the two modules are determined by a competitive gating mechanism (Jacobs, Jordan, Nowlan, & Hinton, 1991). The goal of the model was to account for both rule-governed behavior and the influence of high-frequency exceptions to the rule as occurs in language learning.

The data reported here suggest how a mixture model might work in the prediction task. Perhaps, to generate correct predictions, the exemplar module is consulted as long as the rule stored in the rule module is satisfied by the value of the given dimension, giving the rule module "veto" power in the prediction task. Consider the task of predicting orientation given size for Category A in the negatively


A
Orientation
Size
x = 199.9 + .32 y
x =1.45 + .81 y
Overall Model:
x =1.45+ .81 y
+ 117.0 ( y-165 )$^0_+$
- .49 ( y-165 )$^1_+$
$R^2$ = .88
B
Orientation
Size
x = 93.0 - .10 y
x = 324.9 - .64 y
Overall Model:
x = 324.9 - .64 y
- 132.9 ( y-185 )$^0_+$
+ .54 ( y-185 )$^1_+$
$R^2$ = .92
C
Orientation
Size
x = 361.6 - .31 y
x = 3.2 + .93 y
Overall Model:
x = 3.2 + .93 y
+ 109.0 ( y-200 )$^0_+$
- 1.2 ( y-200 )$^1_+$
$R^2$ = .96

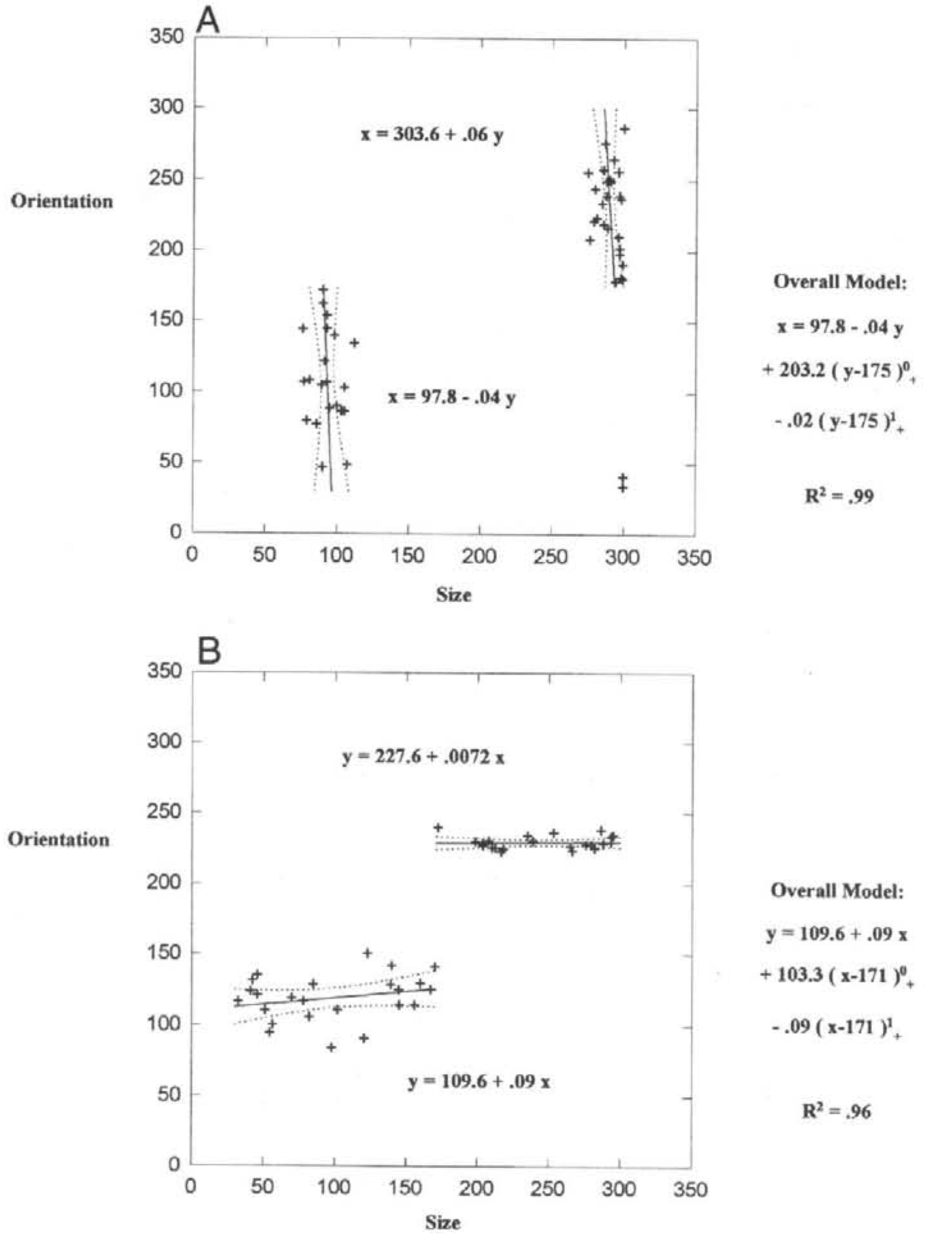



*Figure 16.* Prediction responses for a typical learner in the uniform condition (Participant 3U). A: Size given orientation, Category A. B: Orientation given size, Category A.



correlated case of Experiment 1. The participant might give the correct orientation values up to the point at which the size provided by the experimenter fell in his or her response region for Category B. When the size is consistent with Category B and not with Category A, the participant might provide an orientation that is either random or, more interesting, most unlike typical Category B orientations (e.g., Participant 2N). This finding provides compelling evidence that the decision bound argued to exist by Ashby and his colleagues (e.g., Maddox & Ashby, 1993) is part of

*Figure 15 (opposite).* A: Prediction responses for Participant 1G, B: prediction responses for Participant 2G, and C: prediction responses for Participant 4G, exhibiting an interaction between the stored exemplars and the category rule. Pluses indicate Category A trials, and circles indicate Category B trials.

the category representation of these learners. In previous research, evidence that the location of the category decision bound was stored in memory had to be inferred from the learner's distribution of responses. In fact, one could not rule out the possibility that the resulting boundary was an artifact of an exemplar similarity computation giving rise to an equi-similarity or "equivocality contour" (Ashby & Maddox, 1993). To my knowledge, the results reported here are the first to demonstrate that the boundary appears to be explicitly represented in memory for some learners as the result of the categorization learning process.

One other consequence of these data is the importance of individual differences in categorization strategies. Some learners may be naturally predisposed to use rules, whereas others are more apt to use exemplar-similarity processes. In addition, the experimental conditions, stimulus conditions, or both, may dictate one processing mode over another (such as the availability of a simple rule or integrality–separability of the stimulus dimensions). The best approach to understanding categorization learning may not be to seek a universally correct model but to try to identify and understand the relationships between learner characteristics, experimental conditions, and the choice of categorization strategy.

In summary, exemplar information in the form of learned within-category correlations is available to use in an attribute prediction task for most category learners. However, this information was not available to a few participants who appeared to have memory only for the categorization decision boundary. Perhaps, most important, evidence for the existence of a decision bound in the representation of the stimulus space was observed for a number of participants because their prediction data were influenced by its location.

## References

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

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., & Alphonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology, 39,* 216–233.

Ashby, F. G., Alphonso-Reese, L. A., & Turken, A. W. (1995, November). *Competition between verbal and implicit rules of category learning.* Paper presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles, CA.

Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorization response times with multidimensional stimuli. *Perception & Psychophysics, 55,* 11–27.

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

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., & Maddox, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology, 38,* 423–466.

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity choice model. *Psychological Science, 5,* 144–151.

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

Billman, D. (1989). Systems of correlations in rule and category learning: Use of the structured input in learning syntactic categories. *Language and Cognitive Processes, 4,* 127–155.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking.* New York: Wiley.

Busemeyer, J. R., & Myung, I. J. (1988). A new method for investigating prototype learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 3–11.

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.

Erickson, M. A., & Kruschke, J. K. (1996). *Rules and exemplars in category learning.* (Indiana University Cognitive Science Research Report #183). Bloomington, Indiana: Indiana University.

Ervin, S. M. (1964). Imitation and structural change in children's language. In Lenneberg, E. G. (Ed.), *New directions in the study of language* (pp. 163–189). Cambridge, MA: MIT Press.

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

Estes, W. K. (1994). *Classification and cognition.* Oxford, England: Oxford University Press.

Flanagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, and 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, and Cognition, 10,* 234–257.

Fukanaga, K. (1990). *Introduction to statistical pattern recognition* (2nd ed.). San Diego, CA: Academic Press.

Garner, W. R. (1974). *The processing of information and structure.* New York: Wiley.

Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology, 24,* 341–380.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 2,* 79–87.

Kadlec, H. (1992). *Information processing of stimulus modules and dimensions in visual perception: Independence and interactions.* Unpublished doctoral dissertation, Purdue University, West Lafayette, IN.

Khuri, A., & Cornell, J. A. (1987). *Response surfaces: Designs and analyses.* New York: Marcel Dekken.

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

Kruschke, J. K., & Erickson, M. A. (1994). Learning of rules that have high-frequency exceptions: New empirical data and a hybrid connectionist model. In *Proceedings of the Sixteenth*

*Annual Conference of the Cognitive Science Society* (pp. 514–519). Hillsdale, NJ: Erlbaum.

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

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

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior, 23,* 250–269.

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.

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

Montgomery, D. C., & Peck, A. E. (1982). *Introduction to linear regression analysis.* New York: Wiley.

Morrison, D. F. (1976). *Multivariate statistical methods.* New York: McGraw-Hill.

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

Naylor, J. C., & Clark, R. D. (1968). Intuitive inference strategies in interval learning tasks as a function of validity magnitude and sign. *Organizational Behavior & Human Performance, 3,* 378–399.

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

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

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

Pinker, S. (1991). Rules of language. *Science, 253,* 530–535.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes, 8,* 1–56.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 383–407.

Rosch, E. (1978). Principles of categorization. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General, 121,* 278–304.

Thomas, R. D. (1995). Gaussian general recognition theory and perceptual independence. *Psychological Review, 102,* 192–200.

Thomas, R. D. (1996). Separability and independence of dimensions within the same–different judgment task. *Journal of Mathematical Psychology, 40,* 318–341.

Thomas, R. D. (1997). *Identifying separable dimensions: Classification and same–different judgments.* Manuscript submitted for publication.

Trabasso, T., & Bower, G. H. (1968). *Attention in learning.* New York: Wiley.

Wattenmaker, W. D. (1991). Learning modes, feature correlations, and memory-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 908–923.

Received February 7, 1996
Revision received June 9, 1997
Accepted June 9, 1997 ■