# The Randomization Procedure in the Study of Categorization of Multidimensional Stimuli by Pigeons

Walter T. Herbranson, Thane Fremouw, and Charles P. Shimp
University of Utah

Pigeons categorized rectangles varying in both height and width in an adaptation of a method used by Ashby and colleagues for the cognitive and neuropsychological analysis of human decision bounds for ill-defined categories. Optimal decision bounds were defined in a stimulus space in which the point $(x,y)$ corresponded to a rectangle with width $x$ and height $y$. Four tasks defined the following 4 optimal bounds: $x = y$, $x = c$, $x = y + d$, and $(x - a)^2 + (y - b)^2 = r^2$, where $a, b, c, d,$ and $r$ were constants given by a task. Estimated decision bounds for individual pigeons conformed approximately to the optimal decision bound in each of the 4 tasks. The new method suggests a way to (a) integrate the disparate literature on ill-defined visual concepts and on optimal performances in nonhuman animals; (b) compare how humans and nonhuman animals categorize ambiguous, multidimensional configural stimuli; (c) model how nonhuman animals categorize naturalistic stimuli; and (d) infer that pigeons' categorizations of naturalistic stimuli may be remarkably close to optimal.

Nonhuman animals may face naturalistic categorization problems that place a premium on nearly optimal performance yet involve ambiguous stimuli that relate only probabilistically to outcomes such as food (Stephens & Krebs, 1986). The present experiments demonstrate a new approach to the question of whether nonhuman animals perform in nearly optimal ways when they confront such visual environments with complex perceptual meaning and ambiguous historical implications.

A theme commonly discussed in evolutionary biology is whether much naturalistic behavior is "optimal." This evolutionary theme has not always been explicitly reflected, however, in the psychology of animal learning and behavior. For much of the 20th century, psychologists studying nonhuman animal learning chose instead to focus on adaptive mechanisms such as the law of effect and associative

learning. The development of probablistic and statistical approaches to learning (Brunswik, 1939; Estes, 1950) helped to lead researchers back to asking how the choice behavior of nonhuman animals was or was not optimal (Graf, Bullock, & Bitterman, 1964; Shimp, 1966, 1969, 1973). At about the same time, behavioral economists began to examine the role of optimality in nonhuman animal behavioral models of economic theory (L. Green & Kagel, 1987; Rachlin, Battalio, Kagel, & Green, 1981; Staddon & Ettinger, 1989). Furthermore, behavioral ecologists began to formalize the role of optimality in foraging and caching behavior (see the review and analysis by Stephens & Krebs, 1986). For these and other reasons, the idea of optimality often appears in discussions of nonhuman animal behavior (Staddon, 1992). We see this as an advance because we feel that, although it is true that the idea of optimality can be grotesquely overextended (Gould & Lewontin, 1979), as was done perhaps to the greatest degree possible by Dr. Pangloss in Voltaire's *Candide,* a carefully qualified use of descriptive notions of optimality can be useful (Dennett, 1987).

It is probably safe to say that behavioral biologists and animal learning theorists agree that optimality is an idea worth examining and that an animal's past experience can often assist it in meeting the natural world's challenge of favoring nearly optimal performance while providing only ambiguous stimuli. They seem to agree further, if only implicitly (Herrnstein & de Villiers, 1980; Tinbergen, 1960), that one way in which an animal might benefit from experience is to learn to see specific visual stimuli as exemplars of general categories. A categorization process could simplify the problem of responding optimally to all the innumerable exemplars of a broad visual category. That is, there may be an inherent link between optimal behavior and categorization. However, this link has gone largely unexamined in research on the behavior of nonhuman animals.

Linking optimality and categorization has been difficult partly because the quantitative structure of naturalistic visual

categories is unknown. We know that animals, especially birds, can learn naturalistic visual concepts such as "fish" or "not fish" that defy analysis in terms of defining features and that instead conform to the idea of "family resemblance" (Herrnstein & de Villiers, 1980; Herrnstein & Loveland, 1964; Jitsumori, 1996; Jitsumori & Yoshihara, 1997; Rosch & Mervis, 1975; Wittgenstein, 1953). However, how to more generally describe the quantitative structure of naturalistic visual concepts has been unclear, and available information is based on methodology that does not immediately lend itself to analysis in terms of quantitative measures of optimal performances. Researchers have not been able to indicate, for example, the degree to which a bird's categorizations of visual stimuli as exemplars of the categories *fish* or *not fish*, or as exemplars of the categories of prey or predator, are optimal. Indicative of this gap between the literatures on optimality and naturalistic concepts is that neither the word *concept* nor *category* appears in the subject index of volumes on foraging and optimal foraging (Commons, Kacelnik, & Shettleworth, 1987; Stephens & Krebs, 1986), and the literature on naturalistic visual concepts was not cited in a fairly recent special issue on behavioral economics of the *Journal of the Experimental Analysis of Behavior* (November 1995). Correspondingly, neither the word *optimality* nor *maximizing* appears in the subject index of volumes on nonhuman animal discrimination and perception (Berkely & Stebbins, 1990; Commons, Herrnstein, & Wagner, 1983; Stebbins & Berkely, 1990).

In short, the extensive literatures on optimality and on naturalistic visual concepts in nonhuman animals are scarcely linked at all. In fact, how even to begin to link these literatures has been unclear. Moreover, except for some notable success in applying signal-detection theory (D. M. Green & Swets, 1974) to stimulus control (Davison & McCarthy, 1988; Nevin, 1981; Wixted, 1993), there have been few attempts to conceptually integrate choice behavior and stimulus control. Thus, Davison and McCarthy (1988) noted that "there has been little regard paid to the role of controlling stimuli in matching research" (p. 216).

Sharply contrasting with this animal literature, the human literature provides well-established method and theory by which one can see how human performance approximates optimality in tasks involving ill-defined visual concepts. Ashby and his colleagues have developed and used to good effect one such method (Ashby & Gott, 1988; Ashby & Lee, 1991; Ashby & Maddox, 1992; Maddox, Filoteo, Delis, & Salmon, 1996). The method, sometimes called the "randomization procedure," can be viewed as a generalization and application of signal-detection theory (D. M. Green & Swets, 1974) and of multidimensional scaling theory (Shepard, 1964) to the notion of ill-defined concepts. At the same time, the method may be viewed as a much generalized probability-learning procedure (Estes, 1964; Estes, Burke, Atkinson, & Frankmann, 1957), with metric and configural properties of stimuli given added attention.

Ashby and his colleagues have used the randomization procedure to study how humans categorize multidimensional stimuli. A standard version of their task requires a subject to categorize stimuli varying in two dimensions,

such as rectangles varying in height and width. Stimuli are drawn from two categories that overlap in the sense that any rectangle can in principle belong to either category. Nearly all stimuli, however, are more likely to belong to one category than to the other. For example, in one typical task, rectangles that are higher than they are wide are more likely to belong to one category, and rectangles that are wider than they are high are more likely to belong to the other category. A subject's task is to assign each stimulus to one of the two categories.

Figure 1 shows a summary of this version of the task. The top panel shows two ill-defined categories, A and B, in the form of two overlapping normal bivariate distributions. The
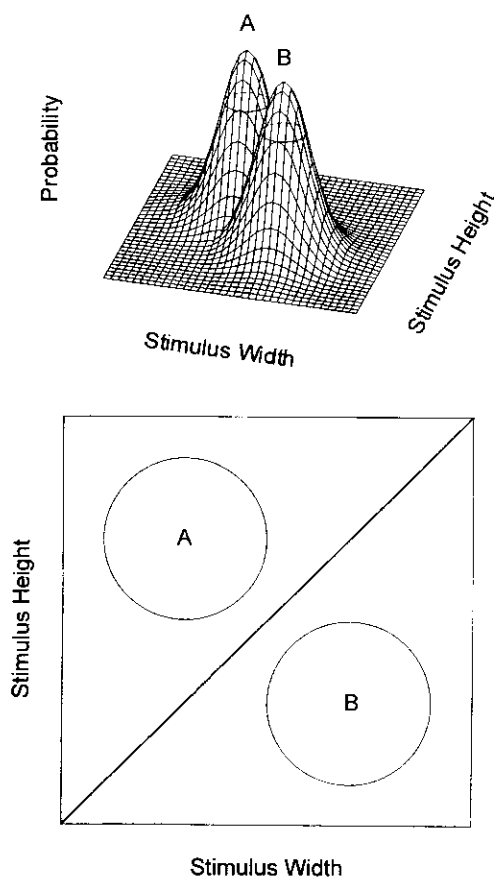


*Figure 1.* Top panel: Bivariate approximately-normal distributions represent likelihoods with which rectangles are sampled from either of two ill-defined (overlapping) categories, A and B. A rectangle is represented in the stimulus space as a point with coordinates equal to the corresponding width and height. A pigeon successively categorizes individual rectangles and is reinforced if a choice corresponds to the category, either A (left key) or B (right key), from which a rectangle was sampled. One arbitrary contour of equal likelihood is shown for each category. Each contour consists of all points corresponding to rectangles equally likely to be sampled from a category. Bottom panel: Arbitrary contours of equal likelihood for each category and the corresponding linear optimal decision bound, $x = y$, according to which a rectangle should be categorized as an A or as a B, depending on whether the rectangle is taller than wide or wider than tall, respectively.

space over which the distributions are defined is typically referred to as the "stimulus space," in which each point represents a particular two-dimensional (2-D) stimulus, such as a rectangle with specified width $x$, and height $y$, and with the third coordinate, $z$, corresponding to the probability with which the rectangle will occur given a particular category. Note that in this basic case, each variable, height, and width in each distribution has the same variance and that the covariance of each distribution is equal to zero. The top panel of Figure 1 also shows two equal-likelihood contours, each of which conveniently summarizes a bivariate normal distribution by showing points corresponding to stimuli that are equally likely to occur given a particular category. The height at which the contours are drawn in Figure 1 is arbitrary; at any height, the collection of equally likely points for a given category forms a circular contour. The bottom panel shows two such equal-likelihood contours, along with the optimal decision bound, the line formed by the points corresponding to rectangles that are equally likely to occur given either category. For the task displayed in Figure 1, the optimal decision bound is a straight line with slope of 1.0, according to which stimuli are categorized optimally when a stimulus on one side is categorized as belonging to Category A and a stimulus on the other side is categorized as belonging to Category B (Ashby & Gott, 1988; Ashby & Lee, 1991; Ashby & Maddox, 1992; Maddox & Ashby, 1993).

The position and shape of an optimal decision bound in this 2-D space depend on the frequencies of occurrence of the stimuli in the two categories, particularly on whether the variances and covariances of the two bivariate distributions are similar. For the present experiments, it is important to note that when the covariances of the two bivariate distributions are equal to zero, the optimal bound is linear or quadratic depending on whether the variances of the two categories are equal or unequal, respectively (Ashby, 1992; Ashby & Maddox, 1998; Maddox & Ashby, 1993).

A person categorizing stimuli on the randomization task produces, over a sequence of trials, an empirical counterpart to the theoretical stimulus space shown in Figure 1; an empirical stimulus space shows which stimuli were sampled from each category and shows how a person categorized each one. Just as one can obtain the optimal decision bound from the theoretical distributions in Figure 1, one can estimate a decision bound from an observer's actual performance. (The details of this estimation procedure are provided later.) These estimated decision bounds are descriptively accurate and demonstrate that humans are remarkably good at learning to approximate optimal decision bounds when they categorize several different types of ambiguous stimuli (Ashby & Gott, 1988; Ashby & Maddox, 1992; Maddox & Ashby, 1993). Human observers have generally approximated optimal decision bounds to an impressive degree in a variety of tasks with linear and even some nonlinear optimal decision bounds (Ashby & Maddox, 1992, 1998).

In summary, the randomization procedure developed and investigated by Ashby, Nosofsky, and many of their colleagues offers attractive possibilities for linking the disparate literature on categorization and on optimization in

nonhuman animals, and our goals in the present two experiments were to exploit some of these possibilities. In particular, we sought to determine whether categorization by nonhuman animals can be described in terms of decision bounds and, if so, whether the bounds approximate optimal solutions to problems that are ill-defined, complex, abstract, and multidimensional. We also sought to explore a suggestion of Ashby and his colleagues that the procedure may serve as a tractable model for studying the role of optimality in naturalistic categorization.

## General Method

### Animals

Eight male White Carneaux pigeons (*Columba livia*) were obtained from the Palmetto Pigeon Plant (Sumter, SC). Each was maintained at approximately 80% of its free-feeding weight, with supplemental grain provided as needed in home cages after daily experimental sessions. Birds were housed individually, in standard pigeon cages with free access to water and grit, in a colony room with a 14:10 light–dark cycle. All experimental sessions took place during the light cycle at approximately the same time 5–6 days a week.

### Apparatus

The three experimental chambers had internal dimensions of $38 \times 34.5 \times 50$ cm ($1 \times w \times h$). Each chamber had three clear plastic response keys ($3.5 \times 3.5$ cm) mounted in a horizontal row within a clear Plexiglas viewing window ($17 \times 7$ cm; w $\times$ h). The viewing window itself was mounted within the front wall of the chamber, with its bottom 20 cm above the chamber floor. A 14-in. (16.54-cm) computer monitor (either CTX or Relisys) was 9 cm behind this front wall. Each monitor and experimental chamber was interfaced via digital input/output cards (Metrabyte) either to one of two IBM PS/2 Model 80 computers or to an IBM PS/2 Model 95 computer, which controlled all experimental contingencies and recorded the data. A super MVGA video adapter card (Colorgraphics Communications) was used to keep the monitor in the chamber interfaced to the Model 95 computer blank before and after experimental sessions. A splitter amplifier box (Colorgraphics Communications) was used for the same purpose for the other two boxes. A fan and white noise helped to mask extraneous sounds. A digital sound level meter held at approximately the position of a bird's head in each of the three chambers gave a reading (C scale) varying from approximately 88 to 93 dB.

### Procedure

#### Pretraining

Pigeons were pretrained in sessions consisting successively of habituation to the chamber, magazine training, and autoshaping to attain consistent responding on each key. Stimuli used for autoshaping were 2.4 cm square blocks of various colors (green, red, or blue) appearing on the computer monitor behind the response keys.

#### Stimuli

Stimuli were outlines of rectangles formed by yellow lines approximately 0.3 cm thick on a black background and were centered horizontally and vertically behind the middle response

key. Across trials, rectangles varied both in height and width. Both the height and width of a rectangle could vary from 0.3 to 6.0 cm, in 0.3-cm increments. Thus, there were 20 possible heights and 20 possible widths, for a total of $20 \times 20 = 400$ possible different rectangles. As in the corresponding research with humans, no effort was made to normalize brightness across the different stimuli. On each trial, the computer randomly chose one of the two equally likely stimulus categories, A or B. Each category had a corresponding 2-D distribution, approximately bivariate normal, associated with it. Once the stimulus category for a trial was chosen, height and width were then randomly and independently chosen from the corresponding approximately normal bivariate distribution. Any value could be selected for either dimension of a stimulus, given either category, so that any of the 400 possible stimuli could be selected given either category. The task in this sense was "ill-defined." However, most rectangles were more likely to appear given one category than given the other, so that for most rectangles, a particular response optimized the likelihood of reinforcement.

## Trial Organization

Each of a session's 95 trials consisted sequentially of an orienting cue, stimulus presentation, a categorization response, either reinforcement or a correction procedure, and an intertrial interval. Each trial began with an orienting cue (a $2.4 \times 2.4$ cm green block) presented behind the center key. The first center-key peck to occur after 1 s elapsed turned off this cue. The orienting cue was then replaced with a rectangle randomly chosen as described earlier. A peck to the center key after the rectangle had been displayed for 2 s (Birds 4–8) or 3 s (Birds 1–3) illuminated a solid block ($2.4 \times 2.4$ cm) behind each of the two side keys.[1] The left block was red and the right block was blue. A bird then categorized the rectangle as an exemplar of Category A or Category B, in the sense that a choice of the left key was reinforced if the rectangle had been generated by Category A and a choice of the right key was reinforced if the rectangle had been generated by Category B.

If a choice corresponded in this way to the category from which the rectangle was sampled, mixed grain was presented in a hopper located beneath the Plexiglas window. The hopper was directly below the center key. Hopper presentation time varied across birds from 1.4 to 3.0 s to accurately maintain individual deprivation levels. After reinforcement, there was a 5-s intertrial interval, during which the monitor was blank.

If a response did not correspond to the category from which the stimulus was sampled, a 10-s correction interval began. This interval was signaled by the houselight flashing on and off every 0.5 s. (Otherwise, the houselight was on throughout the entire daily session.) Any subsequent errors caused the correction procedure to continue to recycle, with the same rectangle, until the correct response was given. Only the initial choice response in a trial was recorded and included in data analysis. Any subsequent errors in the correction procedure were excluded.

## Experimental Conditions

The structure of each of the two experiments was the same: each consisted of a first task, a second task, and a replication of the first task. Stable performance in each task was defined as 5 consecutive days over which the percentage of correct responses varied nonmonotonically and within a range of no more than 10%. Conditions were always continued for some time after performance met this criterion, mostly because of our conservative attitude; we did not wish to underestimate the possible impact of long-term drifts in performance in a new procedure.

## Procedure for Estimating Decision Bounds

We had two goals: (a) to determine the practical utility of a decision-bound approach to the problem of how nonhuman animals learn to categorize exemplars of ill-defined categories and (b) to determine whether animals' bounds, if any could be estimated, were approximately optimal. This goal required us to estimate best fitting bounds for each animal for each task. We imposed constraints on this estimation procedure to make the computational task manageable. In the three tasks for which the optimal bound was linear, we limited the search to linear estimations. In the one task for which the optimal bound was nonlinear, in fact a circle, we limited the search to linear and circular bounds. In general, our goal in these initial studies was not to find which of all possible, arbitrary polynomials best described each individual bird's choices in the stimulus space for each task but instead to discover whether a bird's bound in a task approximated the optimal bound for that task.

The method for finding best fitting decision bounds for individual birds worked as follows. For each linear task, each possible linear decision bound that passed through the stimulus space was compared with each bird's performance over the last 5 days of a task. The exhaustive search used a step size of 0.05 for both slope and intercept, so that nearly 1 million potential decision bounds were computed. This step size was the smallest practical value given the computational constraints and given the size of the stimulus space. For the nonlinear (circular) optimal bound in Task 2 of Experiment 2, we similarly conducted an exhaustive search of all possible circular decision bounds, with center $(a, b)$ and radius $r$. The values of all three parameters were varied in increments of 0.25. Possible values for $a$ ranged along the $x$-axis from 0 to 12 for $b$ they ranged along the $y$-axis from 10 to 20, and for $r$ they ranged from 0 to 20.

The best fitting bound was that which accounted for more individual choices than any other bound. That is, it was the bound that maximized the number of individual categorizations of rectangles that were the same as a bird's categorizations. Most searches produced more than one best fitting solution: The estimated decision bound was usually not unique. We provided for each bird in each task the number of best fitting solutions and the envelope of best fitting bounds within which all the others fell. Any estimated best fitting decision bound not explicitly described fell between the extremes represented by this envelope.

## Experiment 1

Humans quickly learn to categorize exemplars in the randomization task such that their estimated decision bounds approximate linear decision bounds (Ashby & Gott, 1988). We therefore began by determining the extent to which pigeons also can master tasks involving different linear optimal decision bounds. We began with tasks similar to those used in initial experiments with human subjects, with our Tasks 1 and 2 loosely patterned after Experiments 1 and 2 in Ashby and Gott (1988).

---

[1] The observation interval was 0 s immediately after initial shaping, and thereafter for the first few birds to be trained on the categorization procedure, this interval was gradually increased to 3 s to ensure adequate exposure duration of the stimuli. As more birds were trained, it became clear that a shorter, 2-s observation period was adequate.

## Task 1 ("Taller Versus Wider" or "Divided Attention," Bound: $x = y$)

We began with a task in which the two categories of rectangles could be discriminated optimally, in terms of plain English, as rectangles that were either "taller than wide" (Category A) or "wider than tall" (Category B). Any rectangle with a height greater than its width was more likely to belong to Category A, and any rectangle with a width greater than its height was more likely to belong to Category B. The variances of the two bivariate, approximately normal distributions that defined the two categories were equal, and their covariances were equal to zero (as in Figure 1, top panel). The optimal decision bound was therefore a straight line with a slope of 1.0 and an intercept of 0 (as in Figure 1, bottom panel). That is, the optimal bound in the stimulus space wa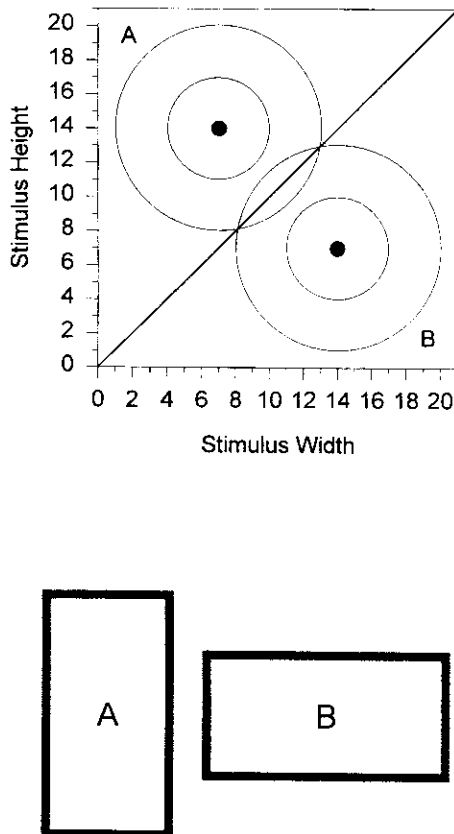s the straight line $x = y$, with $x$ and $y$ being equal to the width and height of a rectangle, respectively. The top panel of Figure 2 shows this optimal bound in the context of the overlap between the two ill-defined categories, with overlap shown in terms of two annuli, located at 1 and 2 $SD$s from the center of each category. The bottom panel of Figure 2 shows the two rectangles corresponding to the centers of the two categories, A and B, respectively.

Simple as this optimal bound is when expressed mathematically, it may nevertheless require relatively complex processing of stimulus information. If the basic perceptual dimensions are height and width, it requires integration of information from these two stimulus dimensions.[2] If an animal either cannot or does not divide attention between the two dimensions, and accordingly does not integrate information, including information about the values of both dimensions of rectangles and about corresponding sampling likelihoods, then categorization cannot be optimal.

## Task 2 ("Wider Versus Narrower" or "Selective Attention," Bound: $x = c$)

If pigeons were able to divide their attention and accurately integrate information across the two dimensions when the optimal bound required them to do so, could they learn also to refrain from integrating information and to ignore one dimension if the optimal bound required them to do so? We arranged a second task to evaluate this possibility. The second task used the same rectangles as the first task, but now only one dimension contained relevant information about category membership. The task required the birds to attend selectively to the relevant width dimension and to ignore information in the randomly varying, irrelevant height dimension. The optimal decision bound was the vertical straight line, $x = c$, where any rectangle with a width less than, or greater than, the value $c$ should be categorized as belonging to Category B or Category A, respectively. (There were no points on the line because it was located at $c = 10.5$.) The top panel of Figure 3 shows a summary of this task, and the bottom panel shows the two rectangles



*Figure 2.* Task 1 of Experiment 1. In this task, an optimal response consisted of categorizing a rectangle as belonging to Category A or B if it was taller than wide or wider than tall, respectively. Top panel: Two contours of equal likelihood are shown for each category, A and B. For each category, the smaller and larger circular contours are 1 and 2 $SD$s from the mean (filled circle) of the corresponding normal bivariate distribution, respectively. The solid line represents the optimal decision bound, $x = y$. Arbitrary screen units along the horizontal and vertical axes correspond to increments of 0.3 cm. Bottom panel: The two exemplars corresponding to the means of Categories A and B, respectively. Line thickness and rectangle size are drawn to scale.

---

[2] It should be noted in passing that in the human literature there is an unresolved issue concerning how physical dimensions of rectangles correspond to underlying psychological dimensions. Krantz and Tversky (1975) suggested, for example, that humans may perceive rectangles in terms of size and shape rather than length and width, although even these suggested psychological dimensions were less than ideal because they interacted and were therefore not independent. Other researchers have continued to suggest that the physical dimensions of length and width adequately represent underlying psychological scales because with proper scaling, they lead to independent psychological dimensions (Borg & Leutner, 1983). Although this issue remains unresolved in the human literature, it is essentially uninvestigated in the nonhuman animal literature. Therefore, throughout this article, we assume for simplicity that the psychological dimensions of rectangles are length and width, but the reader is advised that they may have been some other dimensions, such as size and shape. In any case, as will be explained later, the obtained results make this issue moot for the present experiments.
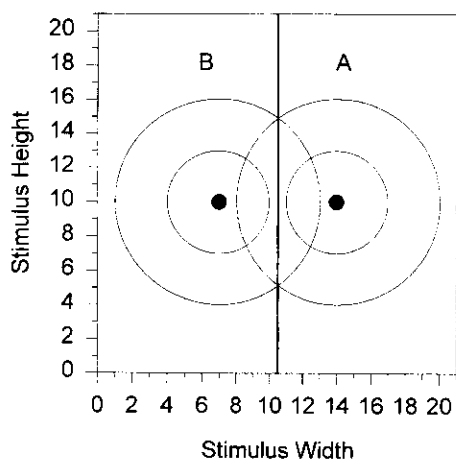
*Figure 3.* Task 2 of Experiment 1. In this task, an optimal response consisted of categorizing a rectangle as belonging to Category A or B if it was wider or narrower than a fixed criterion width, respectively. Top panel: Two contours of equal likelihood are shown for each category, A and B. For each category, the smaller and larger circular contours are 1 and 2 *SD*s from the mean (filled circle) of the corresponding normal bivariate distribution, respectively. The solid line represents the optimal decision bound, $x = c = 10.5$. Arbitrary screen units along the horizontal and vertical axes correspond to 0.3 cm. Bottom panel: The two exemplars corresponding to the means of Categories B and A, respectively. Line thickness and rectangle size are drawn to scale.

corresponding to the centers of the two Categories, B and A, respectively.

In addition to this difference between the first and second tasks in terms of selective and divided attention, the tasks were also different in terms of memory requirements. The optimal bound in Task 1 was defined in terms of a comparison between a rectangle's height and width. The optimal bound in Task 2 was defined in terms of a comparison between a rectangle's width and a fixed criterion width, $c$. In terms of these definitions of the tasks, the basis on which optimal categorization depended was therefore literally present in the first task because both height and width were intrinsic physical features of the stimuli, but not in the intervening task, because the criterion $c$ was not part of the displayed stimuli. Optimality in this second task therefore required an animal to behave as though it learned, remembered, and compared the criterion with each rectangle's width. The selective attention task, $x = c$, in this sense

had a memory demand absent from the first, divided attention task.

In summary, for each of two different tasks, in Experiment 1 we examined whether pigeons could learn to categorize multidimensional stimuli in a manner that approximated optimal decision bounds. Our goal here was not to compare performances across different tasks, which is a topic for future research, but to compare a bird's performance with the corresponding optimal bound.

## Method

### Animals and Apparatus

The animals were 4 experimentally naive White Carneaux pigeons (*Columba livia*). The apparatus was the one described in the General Method section.

### Procedure

The procedure conformed to that described in the General Method section. Recall that Task 1 was replicated after the completion of Task 2 to determine the extent to which a once-learned bound could be relearned following the learning of some other bound. Specific details are as follows.

*Task 1 (bound: $x = y$) and its replication.* In Task 1, the heights of Category B rectangles and the widths of Category A rectangles were drawn from identical distributions with means of 7 screen units (2.1 cm) and standard deviations of 3 units (0.9 cm). Similarly, the widths of Category B rectangles and the heights of Category A stimuli had identical means of 14 screen units (3.9 cm) and standard deviations of 3 (0.9 cm).

Categorizing rectangles according to the optimal decision bound in Task 1, a straight line with a slope of 1.0 and an intercept of 0 would result in reinforced choices on an average of approximately 88% of the trials. Expected accuracy could not reach 100% because the overlap between bivariate distributions for the two categories ensured that some stimuli fell on the "wrong" side of the optimal decision bound. In Task 1, 12% of each distribution fell on the wrong side of the optimal bound, that is, opposite the mean of its distribution. Therefore, on the average, 12% of choices to the "better," or optimal, key would not be reinforced if a bird were choosing according to the optimal bound.

*Task 2 (bound: $x = c$).* In Task 2, the widths of Category A stimuli were determined by an approximately normal distribution with a mean of 14 (3.9 cm) and a standard deviation of 3 (0.9 cm), and the widths of Category B stimuli were determined by a distribution with a mean of 7 (2.1 cm) and a standard deviation of 3 (0.9 cm). For both categories, height was drawn from identical distributions with means of 10 (3.0 cm) and standard deviations of 3 (0.9 cm). Height was therefore not diagnostic of the category from which a rectangle was drawn. To review, the width of a stimulus in the second task was generated by the two categories in the same way as in the first task, but height varied randomly, having no impact on category membership, and Category A stimuli were usually wider than Category B stimuli.

The optimal bound for Task 2, a vertical straight line, was located at a point midway along the horizontal axis, corresponding to a width of $c = 10.5$. Behavior conforming to this bound would have produced on the average an optimal level of accuracy of categorization of 82% because 18% of each distribution fell on the "wrong" side of the optimal decision bound. The response keys corresponding to Categories A and B were switched between the first and second tasks to reduce similarity in optimal bounds

between Tasks 1 and 2 (see Figures 2 and 3). Specifically, in Task 2, A and B were switched from the left and right keys, respectively, to right and left. (They were switched back again in the replication of Task 1.) This shift ensured that high levels of accuracy in the second task were attributable to learning the new task, not merely to behavior persisting from the first task. If a bird continued in the second task to use exactly the same decision bound, and the same assignment of left and right keys to categories, the bird would have been correct (its choices would have been reinforced) only 27% of the time. If the bird had used the same decision bound, $x = y$, but had switched assignments of keys to categories, 73% of the choices in the long run would have been correct.

Because the second task was designed to require a bird to compare each rectangle's width with remembered criterion width, $c$, we took steps to remove physical cues in the environment that could serve as a criterion. In particular, stimuli were centered on the video monitor to reduce the effectiveness of a strategy based on the stimulus simply covering a certain location on the screen. This centering presumably also helped to keep subjects from using width cues that might have been available along the edges of the chamber and also kept stimuli as far as possible from the edges of the screen.

## Results

The procedure Ashby and his colleagues have developed has the virtue that it encourages the analysis of behavior of

individual organisms. There are occasions when group averages are illuminating (Estes, 1956), but the analysis of decision bounds apparently is not one of them; averaged decision bounds are particularly likely to misrepresent actual behavior of individual subjects (Ashby, Maddox, & Lee, 1994; Maddox & Ashby, 1993, 1998). Therefore, our focus was on individual performances. To reduce warm-up effects, we excluded performance on the first five trials of each session from all data analyses.

### Acquisition and Overall Accuracy

Table 1 shows that each bird reached the criterion for stable performance within a range across individual birds of 11–28 days in Task 1, 10–31 days in Task 2, and 15–29 days in the replication of Task 1. By the standards of rate of acquisition in some other experiments using complex stimuli and probabilistic reinforcement, this was moderately rapid learning (e.g., see Shimp, Long, & Fremouw, 1996). Training was continued after the arbitrary stability criterion was met merely to verify that stability was relatively permanent once it was established.

The top row of Table 1 shows the percentage of trials on which a bird chose the key more likely to lead to reinforcement (i.e., the "better," or optimal, key). This percentage

Table 1

*Experiment 1: Accuracy and Amount of Training*

| Task | Bird 1 | Bird 2 | Bird 3 | Bird 6[a] | Average |
|---|---|---|---|---|---|
| Task 1: $x = y$ | | | | | |
| % of choices to the optimal or "better" key | 86.3 | 91.5 | 87.3 | 86.4 | 87.9 |
| % of choices accounted for by height | 78.1 | 81.2 | 81.9 | 81.1 | 80.6 |
| % of choices accounted for by width | 80.9 | 82.6 | 77.6 | 79.1 | 80.0 |
| % of choices that were reinforced | 80.9 | 83.8 | 84.5 | 83.1 | 83.1 |
| % of choices expected to be reinforced given optimal performance | 88.0 | 88.0 | 88.0 | 88.0 | 88.0 |
| Day of training on which stability criterion was first met | 18 | 16 | 28 | 11 | 18.3 |
| Total days of training | 91 | 62 | 57 | 31 | 60.3 |
| Task 2: $x = c$ | | | | | |
| % of choices to the optimal or "better" key | 75.8 | 82.2 | 79.8 | 71.3 | 77.3 |
| % of choices accounted for by integration[b] | 72.0 | 70.0 | 70.1 | 63.2 | 68.8 |
| % of choices accounted for by size | 70.6 | 79.0 | 78.0 | 75.2 | 75.7 |
| % of choices that were reinforced | 68.2 | 74.7 | 69.8 | 67.6 | 70.1 |
| % of choices expected to be reinforced given optimal performance | 82.0 | 82.0 | 82.0 | 82.0 | 82.0 |
| Day of training on which stability criterion was first met | 31 | 16 | 10 | 11 | 17.0 |
| Total days of training | 60 | 55 | 55 | 45 | 53.8 |
| Replication of Task 1 | | | | | |
| % of choices to the optimal or "better" key | 86.3 | 84.1 | 83.8 | | 84.7 |
| % of choices accounted for by height | 78.7 | 77.8 | 80.0 | | 78.8 |
| % of choices accounted for by width | 80.9 | 78.7 | 75.6 | | 78.4 |
| % of choices that were reinforced | 82.0 | 80.9 | 81.1 | | 81.3 |
| % of choices expected to be reinforced given optimal performance | 88.0 | 88.0 | 88.0 | | 88.0 |
| Day of training on which stability criterion was first met | 29 | 15 | 23 | | 22.3 |
| Total days of training | 34 | 33 | 41 | | 36.0 |

[a]Data were accidentally not collected from Bird 6 on the replication of Task 1. [b]*Integration* is defined by the decision-bound $x = y$.

was averaged for each task over the last 5 days of training. Accuracy defined this way theoretically could have reached 100% if a bird chose the better key for each stimulus on each trial. Accuracy varied across birds in Task 1 from a low of 86.3% to a high of 91.5%, with a mean of 87.9%, from 83.8% to 86.3%, with a mean of 84.7% in its replication, and from 71.3% to 82.2%, with a mean of 77.3%, in Task 2. There was no reliable difference in accuracy between Task 1 and its replication, demonstrating that a bound once mastered and abandoned could be reestablished.

Although overall accuracy was less than perfect, in every individual case, it also far exceeded the chance level of 50% correct. Note, however, that above-chance performance could be obtained in a variety of ways. Consider Task 1 and its replication. Note that the height and width of stimuli in Category A were on the average different from those in Category B. Specifically, the heights of stimuli in Category A tended to be greater than the heights of stimuli in Category B and vice versa for width. Therefore, it would be possible to exceed chance performance by attending to only a single dimension. To determine how well single-dimension decision bounds accounted for the data, we calculated the percentage of responses accounted for by the best possible single-dimension decision bound for each of width ($x = c$) and height ($y = c$). Table 1 shows that for each bird the optimal decision bound ($x = y$) accounted for performance better than did either single-dimension decision bound.

Alternatively, in Task 2, optimal performance required attention to a single dimension, width. However, above-chance performance could result even if a bird attended equally to height and width, as was required for optimal performance in Task 1. To investigate the possibility that birds did not attend selectively to a single dimension and instead divided attention between two dimensions as in Task 1, we calculated the percentage of responses accounted for by the decision bound from Task 1 ($x = y$) and by the other decision bound that distributed attention equally, $x = -y + d$ (which was investigated further in Experiment 2). Table 1 shows that for each bird, the optimal selective-attention decision bound accounted for performance better than either divided-attention decision bound.

In summary, in Task 1, when attention to both dimensions was required for optimal performance, birds did in fact attend to both dimensions. Alternatively, in Task 2, on which selective attention to a single dimension was optimal, birds selectively attended to that dimension more than to the other. How well they did so can be seen much more clearly in the analysis to which we turn next.

## Individual Estimated Decision Bounds

Figures 4 and 5 show stimulus spaces for 2 individual birds for both Tasks 1 and 2. In each task, these 2 birds were those for which the best fitting lines were best (top panel) and worst (bottom panel), respectively. (It so happened that these were the same 2 birds, Birds 2 and 6, in each task.) The criterion for "best" and "worst" was designed to reflect the extent to which the estimation procedure generated bounds
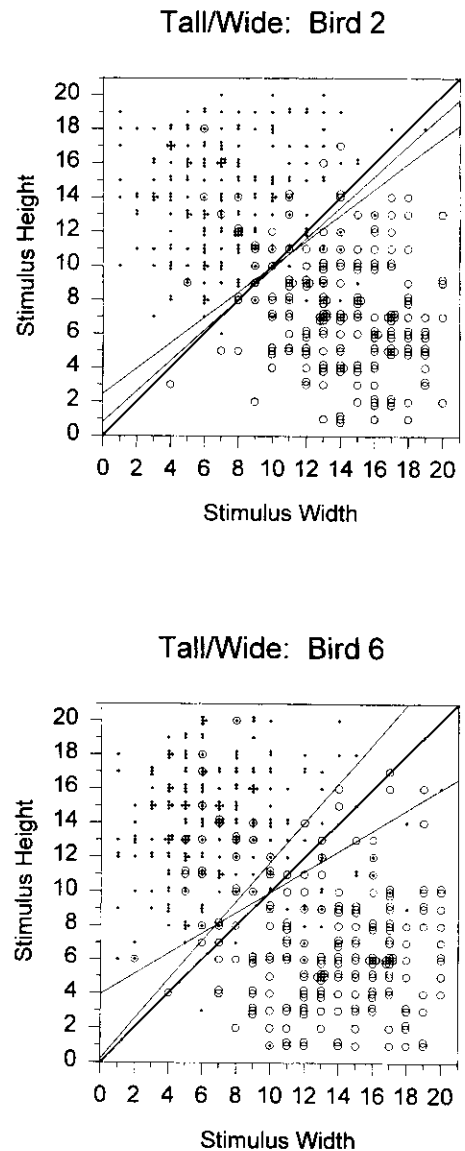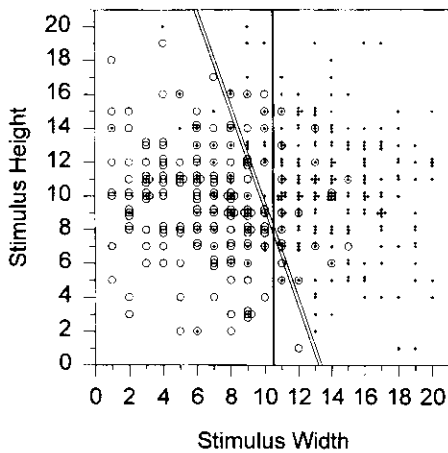


Tall/Wide: Bird 2

Tall/Wide: Bird 6

*Figure 4.* Task 1 of Experiment 1. Stimulus spaces for Birds 2 (top) and 6 (bottom) in the divided-attention condition. Points and circles represent choices of Category A or B, respectively. Each panel shows each of a bird's choices over the last 5 days of the task. Bold lines show the optimal decision bound, $x = y$, and lighter lines show the envelope of all best fitting estimated decision bounds. Units along the axes correspond to 0.3 cm.

that accurately described the data. Thus, *best* and *worst* refer to the estimated bound that accounted for the highest and lowest number of choices, respectively. This criterion to determine which individual stimulus spaces to present graphically was designed to reflect the extent to which a decision-bound analysis was applicable to the data and the effectiveness of the estimation procedure, not the accuracy with which birds' decision bounds were optimal. The degree to which estimated decision bounds approximated optimal bounds can be seen in Table 2, which shows the best fitting bounds and the percentage of responses for which they

## Single Dimension: Bird 2
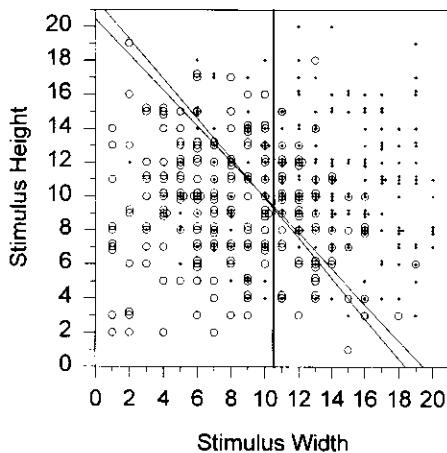


## Single Dimension: Bird 6



*Figure 5.* Task 2 of Experiment 1. Stimulus spaces for Birds 2 (top) and 6 (bottom) in the selective-attention condition. Points and circles represent choices of Category A or B, respectively. Each panel shows each of a bird's choices over the last 5 days of the task. Bold lines show the optimal decision bound, $x = 10.5$, and lighter lines show the envelope of all best fitting estimated decision bounds. Units along the axes correspond to 0.3 cm.

accounted for all birds in each task. Circles and points in Figures 4 and 5 show individual left and right responses, respectively, over the last 5 days of a task. There often were multiple presentations of the same stimulus, so corresponding symbols were shifted slightly to improve clarity in the figures.

*Task 1 (bound: $x = y$).* Table 2 shows that the percentage of choices accounted for by the best fitting lines ranged from 86.4% to 90.8%, so that the estimated straight lines generally effectively summarized what the birds actually

did.[3] Each best fitting bound resembled the optimal bound by having positive slopes, which ranged over birds from 0.60 to 1.65 and which were approximately centered around the optimal value of 1.0. The estimated intercepts ranged from $-3.85$ to 6.20 and again were centered roughly around the optimal value of zero. The estimated decision bounds in these ways at least roughly approximated the optimal bound. Figure 4 shows that in the case of Bird 2, the approximation was truly remarkable.

*Task 2 (bound: $x = c$).* Because the optimal slope for Task 2 was infinite in terms of $x$, we instead report the function in terms of $y$, for which the optimal slope was 0 and the optimal intercept was 10.5. Table 2 shows that three of the four estimated best fitting decision bounds gave reasonably good fits to the optimal bound, with 80.0%–84.9% of the birds' choices being correctly described. Bird 6, the worst case, is an especially interesting exception to which we return later in the *Discussion* section. Estimated slopes for the other 3 birds ranged from $-0.35$ to $-0.05$. The estimated intercepts, excluding Bird 6, ranged from 9.40 to 13.40 and were roughly centered around the optimal value of 10.50. Figure 5 shows that Bird 2's estimated decision bound closely conformed to the optimal bound, and Table 2 shows that Birds 1 and 3 were not far from optimal. Bird 6, however, as we discuss later, apparently performed this task in an entirely different way. It is important that even for Bird 6, a linear function described the data reasonably well.

*Replication of Task 1.* Task 1 was replicated to determine whether a decision bound once mastered can be recovered after a bird has mastered some other bound. Estimated decision bounds (see Table 2) summarized the data about as well as in Task 1, and the percentage of responses correctly described ranged from 84.7% to 87.3%. Estimated slopes varied from 0.50 to 2.70, which were again positive, like the optimal slope of 1.0, but which were slightly more variable than for the original performance. Intercepts varied from $-20.00$ to 5.45, which also were somewhat more variable than in the original task. Nevertheless, Table 2 shows that the estimated bounds were reasonably close to the optimal bound. The replication was generally slightly less successful than the original task in producing behavior similar to the optimal decision bound, but recall that the difference in accuracy was not reliable, so that whatever long-lasting interference was caused by an intervening task was slight.

---

[3] Readers may notice in a few cases that the optimal decision bound accounted for a larger percentage of choices than the "best fitting" bound, as in the case, for example, of Bird 2. This seeming anomaly was an artifact of our methods. When a stimulus fell on the optimal bound, there was no optimal response because either choice was equally likely to be correct. Therefore, choices to such stimuli were excluded from the analysis of accuracy to better reflect performance on trials when an optimal choice existed. On the other hand, it was preferable for best fitting estimated bounds to be based on as many choices as possible, so the searches for these rules were based on all choices. Therefore, the trials on which accuracy and best fitting bounds were based were slightly different, and sampling could produce the anomalous result described above.

Table 2
*Experiment 1: Best Fitting Decision Bounds*

| Task | Bird 1 | Bird 2 | Bird 3 | Bird 6[a] |
|---|---|---|---|---|
| **Task 1: $x = y$** | | | | |
| % of choices accounted for by best fitting straight lines | 87.3 | 90.8 | 87.3 | 86.4 |
| No. of best fitting lines | 15 | 28 | 6 | 39 |
| Parameter values from most extreme best fitting straight lines | | | | |
| Slope (optimal = 1.00) | 1.50 | 0.75 | 0.65 | 0.60 |
| | 1.65 | 0.90 | 0.75 | 1.15 |
| Intercept (optimal = 0.00) | −3.85 | 0.80 | 5.00 | 0.20 |
| | −2.80 | 2.45 | 6.20 | 3.95 |
| **Task 2: $x = c$** | | | | |
| % of choices accounted for by best fitting straight lines | 80.0 | 84.9 | 81.6 | 74.9 |
| No. of best fitting lines | 4 | 5 | 1 | 19 |
| Parameter values from most extreme best fitting straight lines | | | | |
| Slope (optimal = 0.00) | −0.15 | −0.35 | −0.35 | −0.95 |
| | −0.05 | — | — | −0.85 |
| Intercept (optimal = 10.5) | 9.40 | 13.20 | 12.55 | 18.40 |
| | 10.25 | 13.40 | — | 19.45 |
| **Replication of Task 1** | | | | |
| % of choices accounted for by best fitting straight lines | 87.3 | 84.9 | 84.7 | |
| No. of best fitting lines | 5 | 18 | 28 | |
| Parameter values from most extreme best fitting straight lines | | | | |
| Slope (optimal = 1.00) | 1.25 | 2.55 | 0.50 | |
| | — | 2.70 | 0.95 | |
| Intercept (optimal = 0.00) | −4.50 | −20.00 | −0.40 | |
| | −4.30 | −18.10 | 5.45 | |

*Note.* Dashes indicate that the best fitting parameter was unique.
[a] Data were accidentally not collected from Bird 6 on the replication of Task 1.

*Probe stimuli.* Finally, we address a common issue in nonhuman animal research on the nature of concepts. That issue is the question of how animals respond to novel test stimuli. Generalization of performance to novel exemplars is often said to cause problems for theoretical accounts emphasizing control by specific stimulus–response pairs (e.g., see Fremouw, Herbranson, & Shimp, 1998; Honig, 1993; Wright, 1993). This kind of transfer test is not ordinarily conducted in the experiments reported in the human literature on which the present experiments are based. A chief reason why is that the stimulus sets are enormous and most stimuli have therefore seldom if ever been seen before. In the present case, stimulus set size was not as huge as in the corresponding human literature, but the number of stimuli was still large. Therefore, many stimuli were seen only rarely. We can therefore approximate a task with novel test probes by examining accuracy on trials with stimuli only rarely seen. We defined "probe" stimuli in the terminal 5 days of an experiment to be those that a bird had not seen for at least the previous 4 days. Figure 6 shows responses to stimuli that appeared on the final day, but on no previous days within the last 5 days of an experiment. Thus, these stimuli had not been seen by subjects for at least 4 days or 304 trials and possibly had never been seen at all. We calculated accuracy on these trials for the birds for which results are displayed in Figures 4 and 5. For both birds, performance on these probe stimuli was well above chance: In Task 1, Bird 2's accuracy

was 23 optimal out of 26 (88.5%), and for Bird 6 it was 19 of 24 (79.2%). Both accuracies are reliably above chance and in fact resembled corresponding overall performances of 91.5% and 86.4% for Birds 2 and 6, respectively. In Task 2, the outcome was similar, with 16 of 21 (76.2%) and 17 of 24 (70.8%) responses being optimal. Again, these compare well with the corresponding overall performances of 82.2% and 71.3%. Thus, accuracy to rarely seen stimuli was only slightly worse, if at all, than overall accuracy.

## Discussion

Recall our two chief questions: Are the data interpretable in terms of decision bounds? If so, are the estimated decision bounds roughly the same as the optimal bounds? Note that in principle, it would be easy for neither of these questions to be answered: Estimated decision bounds could describe a paltry fraction of the choices, and, even if estimated decision bounds could be found that described most of the choices, there is no reason why they would have to approximate optimal bounds. What, then, are the answers to our two questions?

The answer to the first question is simple: Estimated decision bounds for every bird in each task described a reasonably high percentage of choices. The answer to the second question is correspondingly simple: Estimated decision bounds approximated optimal bounds in 10 of 11
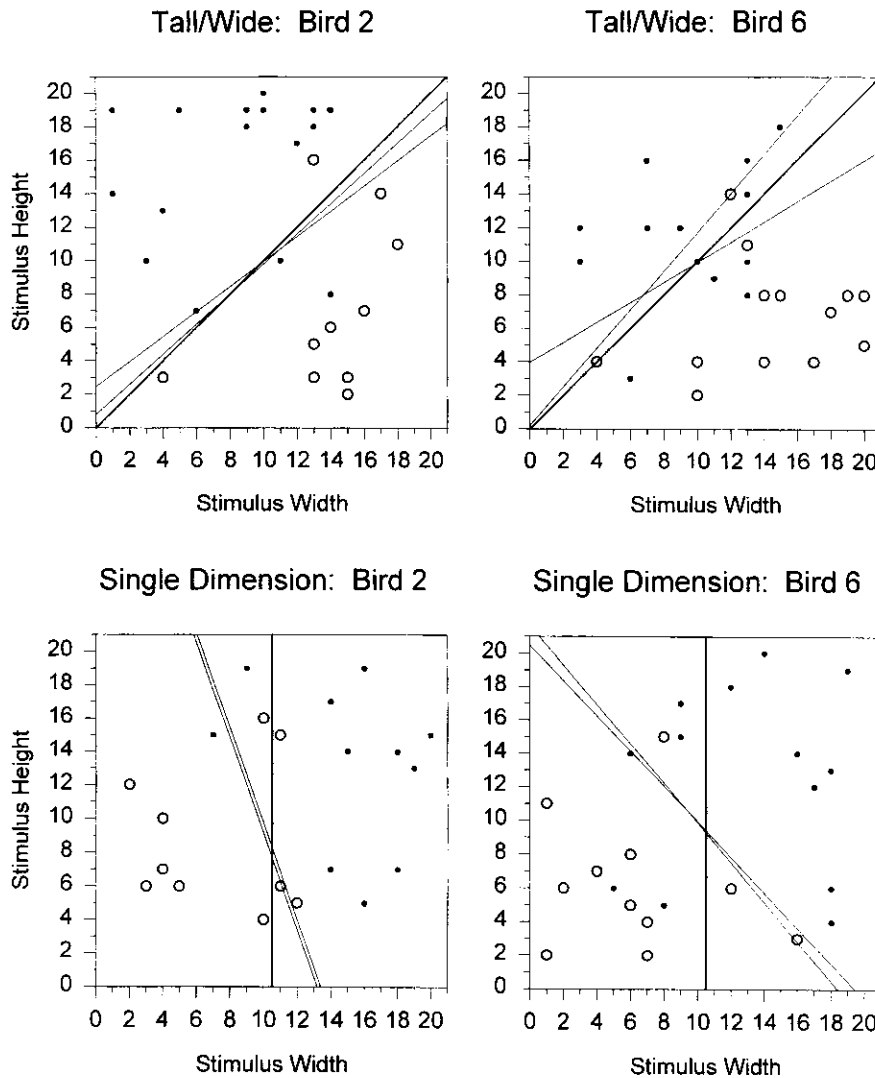
*Figure 6.*    Responses to probe stimuli. Stimulus spaces for Birds 2 (left panels) and 6 (right panels) for Tasks 1 (top) and 2 (bottom). Each panel shows responses to stimuli presented only on the final day of the terminal 5-day periods represented in Figures 4 and 5. Bold lines show the optimal decision bound, and lighter lines show the envelope of all best fitting estimated decision bounds for each bird and condition. Units along the axes correspond to 0.3 cm.

individual cases, and the approximations in several cases were impressive. In the exceptional 11th case, a bound did in fact approximate the data, but the estimated bound did not approximate the optimal bound. In summary, in virtually all cases decision bounds at least roughly described the results, thereby defining descriptive success for a decision-bound approach to describing how nonhuman animals categorize multidimensional exemplars of ill-defined categories.

The ability to perform approximately optimally in the randomization task is therefore not uniquely human and does not require linguistic competence. Ashby, Alfonso-Reese, Turken, and Waldron (1998) have suggested that linguistic abstractions might play important roles in how humans approximate some optimal solutions. They have noted that language may sometimes help and sometimes

hinder accuracy depending on whether a subject's language habits include a predisposition to think in terms of a bound that is, or is not, the actual optimal bound, respectively. Pigeons presumably do not have an elaborate symbolic system corresponding in detail to natural language, with which they might encode decision bounds. It will therefore be an especially interesting issue for future research to determine whether animals can perform approximately optimally in the randomization procedure in a manner involving abstract strategies similar to "same–different judgments" (Cook, Cavoto, & Cavoto, 1995; Wasserman, Hugart, & Kirkpatrick-Steger, 1995), which, until relatively recently, nonhuman animals other than primates were often believed incapable of learning.

The successful transfer we obtained to novel stimuli

addresses this question of whether nonhuman animals learn abstract strategies. Transfer to novel stimuli is often taken to indicate that an animal has learned a general concept or rule and indeed is virtually the defining feature of demonstrations of "rules" in the nonhuman animal literature. According to this criterion, the present experiment may have demonstrated learning of abstract rules. However, this criterion is not powerful enough to answer the question to the degree one might like: There is a lively and informative dialogue in the human literature over the question of whether a decision-bound approach (Ashby & Lee, 1992; Ashby & Maddox, 1998; Maddox & Ashby, 1993, 1998) or a specific-exemplars approach (McKinley & Nosofsky, 1996) is more suitable for the randomization task. The resolution of this dialogue is far beyond the scope of the present demonstration that the basic randomization procedure can be used effectively with nonhuman animals. Therefore, we feel it would be premature to decide that the pigeons in our experiment learned anything more abstract than memories for specific stimulus–response pairings.

Pigeons' estimated decision bounds approximated optimal bounds even though our procedure was somewhat different from that used by Ashby and his colleagues, in which human subjects have done so well. One potentially important difference was in our use of the character mode, rather than the graphics mode, to generate stimuli, which precluded the enormous number of possible stimuli human subjects confront in the corresponding experiments. The total number of possible stimuli in the present experiments, 400, which is large compared with customary practice in animal research, pales in comparison to the corresponding number, more than 100,000, in experiments by Ashby and his colleagues with human subjects. These and many other inevitable differences between human and avian versions of these experiments might make it surprising that the two versions produce similar results. However, consider the possibility that there has been a selective advantage in the natural world for organisms, both humans and avians, with an ability to perform in nearly optimal ways when sorting multidimensional stimuli into ill-defined categories. This presumably would be the case if multidimensional stimuli were seen as objects, or at least if such stimuli invoked processing mechanisms similar to those invoked by real objects, and if organisms needed to discriminate between ill-defined object categories such as food and predators. The fact that pigeons transfer training from photographs of rooms to the corresponding actual rooms (Cole & Honig, 1994) suggests that the rectangles in the present experiments may have been processed similarly to real objects. If this were so, then the similarity between pigeons and humans in the task developed by Ashby and his colleagues could imply that that task captures at least part of the basic nature of real-world categorization problems and successfully models naturalistic visual categories (Herrnstein & Loveland, 1964; Wittgenstein, 1953).

The combined results of Tasks 1 and 2 address a problem deriving from that fact that researchers do not know for certain of whether the "true" perceptual dimensions of rectangles for pigeons are length and width or shape and size (see Footnote 2). Consider the implications of our combined results. If the true dimensions were length and width, then Tasks 1 and 2 involved information integration and selective attention, respectively. However, if the true dimensions were size and shape, then the reverse would be true: Task 1 would involve selective attention and Task 2 would involve integration. The important combined finding from Tasks 1 and 2 is therefore that whatever the true perceptual dimensions are, both selective attention and integration conditions were conducted and pigeons learned to behave in ways that approximated optimal decision rules in both cases.

We now consider Bird 6 in Task 2, which, it will be recalled, failed to master what we are calling the selective-attention task, with the optimal bound, $x = c$. Recall that Bird 6 produced an estimated decision bound with a slope between $-0.85$ and $-0.95$ and an intercept between $18.40$ and $19.45$ (see Table 2). These values define a decision bound close to the optimal bound in Task 1 of Experiment 2 (described shortly). The estimated decision bound for Bird 6 in the present Task 2, with an optimal bound of $x = c$, instead approximated the decision bound $x = -y + d$. In Experiment 2, we examined this bound $x = -y + d$ from the perspective of optimality. Here, it is instructive to see its role from three other perspectives.

First, note that we can distinguish Bird 6's idiosyncratic behavior in terms of its estimated decision bound, but not in terms of the more common but less diagnostic index of percentage of correct choices, in terms of which Bird 6 was only slightly lower than, and not qualitatively different from, the other birds in the $x = c$ task.

Second, Bird 6's performance shows how important it is to analyze performance of individual birds because it is clear that 3 birds in the $x = c$ task responded one way and Bird 6 responded in a qualitatively different way. This difference would have obscured, and been obscured by, an analysis of group performance. Different birds apparently may solve the same task differently, and these individual differences at this level of task abstractness and complexity might prove useful in diagnosing important differences in how different birds process information.

Third, Bird 6's performance in Task 2 sheds light on the role in the present experiments of an issue that has been fundamental in the history of perception of multidimensional stimuli. The issue is important enough that Bird 6's performance merits a detailed discussion. The issue is the classic one of whether perception is holistic or analytic, that is, whether a complex figure is built up from its component features or is instead perceived as a whole. We consider this question in general and its relevance to Bird 6's performance in particular. The answer to this question in general may depend to some extent on the structure of the perceived stimulus: Nonhuman animals, like humans, might be less able to divide attention between dimensions of some types of multidimensional stimuli than of others. Considerable evidence suggests that compound stimuli do indeed vary in the degree to which their elements are separable or disso-ciable, both with humans (Garner, 1974; Shepard, 1964) and nonhuman animals (Leith & Maki, 1977; Riley & Roitblat, 1978; but also see Cook, Riley, & Brown, 1992). Especially

germane to Bird 6's performance is evidence from human subjects that suggests that rectangular stimuli may be relatively integral (Felfoldy, 1974; however, see Dunn, 1983). Note, however, that the absolute magnitude of an integrality effect can be small (although reliable), can depend on the amount of practice, and derives from the operation of mechanisms that are controversial (Dykes, 1979; Lockhead, 1979).

Even less is known for pigeons than for humans about how rectangles are separable or integral, but indirect evidence in the form of striking similarities between human and pigeon shape perception (Blough, 1984; Cook, Cavoto, & Cavoto, 1996) might be taken to indicate that pigeons perceive rectangles in approximately the same way humans do (see also our earlier discussion on the true perceptual dimensions of rectangles). If rectangles are indeed relatively integral for pigeons, then pigeons might be expected to have found it difficult in Task 2 to attend only to a single dimension. The relevance of the integral–separable issue to the performance of Bird 6 is clear. The estimated decision bound for Bird 6 in Task 2 revealed virtually complete information integration across the two dimensions. Bird 6 showed virtually complete divided attention in the task requiring selective attention. Thus, Bird 6 categorized rectangles in a manner compatible with the possibilities that the rectangles were, at least for Bird 6, powerfully integral or that one of the relevant dimensions was size.

Of course, Task 1 trained Bird 6 to divide attention between the two dimensions, so that Bird 6's continuing in Task 2 to divide attention between dimensions may have been a generalized transfer of past experience. In addition, Tasks 1 and 2 were not equated in terms of the maximum possible percentage of trials on which a maximizing bird's choices would have been reinforced, which also might have been partly responsible for Bird 6's performance in Task 2. Also, the amount of practice apparently affects the integrality effect in humans (Dykes, 1979), and the pigeons in the present experiment were provided with much more practice than were the humans in the experiment by Ashby and Gott (1988), so that the present experiments may not have been comparable in this sense to human experiments on the integral–separable issue. For all these reasons, Bird 6's idiosyncratic decision bound in Task 2 must be interpreted with some caution. Perhaps the most exciting aspect of Bird 6's performance is not so much in the details of that performance per se but in the demonstration of the utility of the randomization procedure as a new tool with which to study the classic problem of holistic versus feature-based perception in nonhuman animals.

As interesting as the integral–separable issue is, it should not be allowed to obscure the result more central to our present goals. Whatever impact the structure of a rectangle had on categorization by pigeons, that effect can often be overcome by reward contingencies. In this sense, our results are consistent with analogous human results (Ashby & Gott, 1988; Dykes, 1979) showing that rectangles can be categorized in a flexible manner, in either a holistic or analytic manner depending on the task, with the residual impact of stimulus structure being relatively small.

## Experiment 2

We have seen so far that pigeons readily learn to categorize exemplars in ways that approximate the two optimal decision bounds: $x = y$ and $x = c$. These two bounds are characterized by a certain quantitative simplicity, but real-world tasks may involve linear bounds with negative slopes or bounds that are nonlinear. In Experiment 2 we therefore arranged such bounds and asked if pigeons can master the corresponding tasks. Tasks 1 and 2 roughly corresponded to experiments in Maddox and Ashby (1993) and Ashby and Maddox (1992), respectively.

### Task 1 ("Larger Versus Smaller," Bound: $x = -y + d$)

In Experiment 1, Task 1 was defined in terms of the integration of features from two dimensions, and Task 2 was defined in terms that required a bird to learn and remember a criterion. In these tasks, each requirement was imposed separately. Some naturalistic tasks may impose both requirements simultaneously. Accordingly, in Task 1 in Experiment 2 we imposed both requirements simultaneously and investigated whether birds could learn to categorize stimuli according to the corresponding optimal decision bound.

Category membership in Task 1 was based on stimulus size. The optimal bound can be stated in plain English as follows: "If volume, or perimeter, or any other linear correlate of size is greater than criterion $d$, then respond B; if it is less than $d$, respond A; otherwise, guess." The top panel of Figure 7 shows this optimal bound, along with the means of the bivariate distributions A and B, and equal-likelihood contours at 1 and 2 SDs from the mean for each category. The bottom panel shows the two rectangles corresponding to the means of the two categories, A and B, respectively.

### Task 2 ("Nonlinear," Bound: $[x - a]^2 + [y - b]^2 = r^2$)

Naturalistic categories may produce nonlinear optimal decision bounds. Ashby and Maddox (1992) have observed that multidimensional naturalistic categories may involve variables with different variances; it might be a rarity for variances of component variables to be equal. For example, consider that for a pigeon to categorize an object either as a pea or as a wheat seed might require a decision bound defined in a stimulus space involving height and width but that there would be no reason to assume that the variances of heights of peas and of wheat seeds, or variances of widths of peas and of wheat seeds, were equal. Recall that unequal variances produce nonlinear optimal decision bounds (Ashby & Maddox, 1992, 1998). Humans can learn at least some nonlinear bounds, such as a bound described by a circle (Ashby & Maddox, 1992). In Task 2 we examined whether birds could learn to categorize rectangles when the optimal, nonlinear decision bound was a circle. The top panel of Figure 8 shows the means of the bivariate distributions for Categories A and B, equal-likelihood contours at 1 and 2 SDs, and the circular optimal decision bound. The bottom
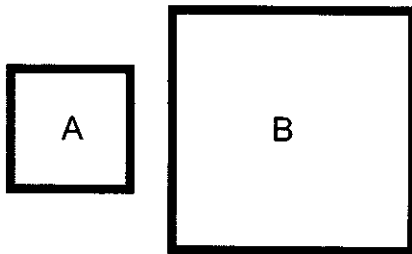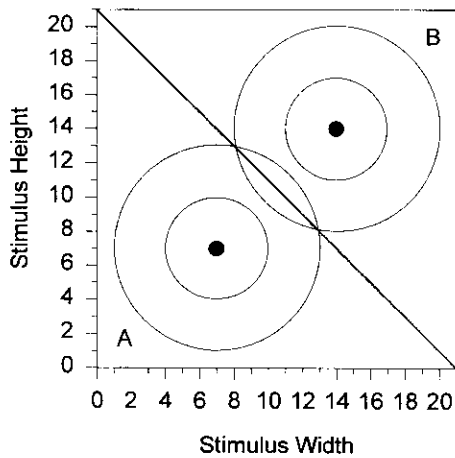
*Figure 7.* Task 1 of Experiment 2. In this task, an optimal response consisted of categorizing a rectangle as belonging to Category A or B if it was smaller or larger than a fixed criterion size, respectively. Top panel: Two contours of equal likelihood are shown for each category, A and B. For each category, the smaller and larger circular contours are 1 and 2 *SD*s from the mean (filled circle) of the corresponding normal bivariate distribution, respectively. The solid line represents the optimal decision bound, $x = -y + d$. Bottom panel: The two exemplars corresponding to the means of Categories A and B, respectively. Line thickness and rectangle size are drawn to scale.

panel shows the two rectangles corresponding to the means of the two categories, A and B, respectively.

Birds, like humans, may face nonlinear bounds in naturalistic settings. A circular bound would appear to be a reasonable first step toward exploring nonlinear bounds in general. Does such a task define an insuperable problem for a nonhuman animal? Or can a pigeon confront such a problem and by one means or another learn to approximate an optimal solution to it?

## Method

### Animals and Apparatus

The animals were 4 experimentally naive White Carneaux pigeons (*Columba livia*) maintained as described in the General Method section. The apparatus was also as described in the General Method section.

## Procedure

The procedure was as described in the General Method section, with additional details as described next. As in Experiment 1, we replicated Task 1 after the completion of Task 2 to determine whether the birds could demonstrate the flexibility required to relearn a rule once learned and then abandoned.

*Task 1 (bound: $x = -y + d$) and its replication.* The heights and widths of rectangles in Category A were independently drawn from identical distributions having means of 7 (2.1 cm) and standard deviations of 3 (0.9 cm). The dimensions of Category B stimuli were similarly selected from identical distributions with means of 14 (3.9 cm) and standard deviations of 3 (0.9 cm). Rectangles from Category A and Category B were therefore, on average, larger and smaller, respectively. Note that a parameter $d = 21$ in this bound is required merely because we used screen units ranging between 1 (0.3 cm) and 20 (6.0 cm) for both $x$ and $y$.

The distributions for stimuli in Task 1 were determined simply by rotating 90° counterclockwise the distributions (and therefore
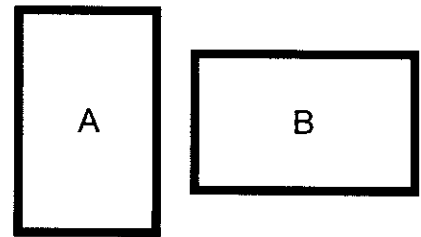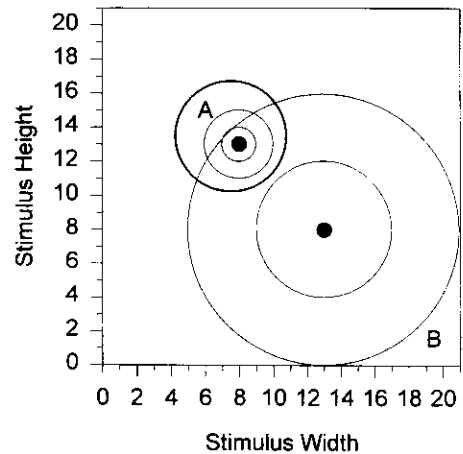




*Figure 8.* Task 2 of Experiment 2. In this task, an optimal response consisted of categorizing a rectangle as belonging to Category A or B if its corresponding point in the stimulus space was or was not in the circular region indicated by the bold circle, respectively. Top panel: Two contours of equal likelihood are shown for each category, A and B. For each category, the smaller and larger contours are 1 and 2 *SD*s from the mean (filled circle) of the corresponding normal bivariate distribution, respectively. The bold circular line represents the optimal decision bound, $(x - 7.5)^2 + (y - 13.5)^2 = 3.25^2$. Units along axes correspond to 0.3 cm. Bottom panel: The two exemplars corresponding to the means of Categories A and B, respectively.

the optimal decision bound) from Task 1 of Experiment 1. The optimal decision bound was therefore the line with a slope of $-1$, defined by length + width = 6 cm. It followed that optimal accuracy in the present Task 1 was the same as in Task 1 of Experiment 1, with an average maximum of 88% correct being theoretically obtainable. Similarly, there was a theoretical optimum of 82% correct if a bird attended only to one dimension and responded according to a decision bound defined by either length or width equal to 10.5, the optimal one-dimensional criterion. For comparison, the nonlinear optimal decision bound from Task 2 described below would have resulted in chance performance, or 50% correct, in Task 1.

*Task 2 (bound: $[x - a]^2 + [y - b]^2 = r^2$).* The nonlinear bound for Task 2 was arranged by sampling stimuli from bivariate distributions with unequal variances. The $x$- and $y$-coordinates of rectangles in Category A were sampled from an approximately normal bivariate distribution with a mean width of 8 (2.4 cm) and a mean height of 13 (3.9 cm). The standard deviation of widths and the standard deviation of heights of rectangles in Category A were the same: 1 screen unit (0.3 cm). For Category B, the widths and heights of rectangles were drawn from an approximately normal distribution with a mean width of 13 (3.9 cm) and a mean height of 8 (2.4 cm). The standard deviation of widths and of heights in the bivariate distribution for Category B were the same: 4 (1.2 cm). Note that the standard deviations for Category B were greater than those for Category A. The optimal decision bound in the stimulus space was a circle, with a center point, $(a, b)$, where $a = 7.5$; $b = 13.5$; and radius, $r = 3.25$. Optimal performance corresponded to an average percentage correct of 96. For comparison, if a subject used the linear bound from Task 1, the average percentage correct would have been only 50%.

## Results

### Acquisition and Overall Accuracy

As in Experiment 1, choices on the first five trials of each session were excluded from data analysis to reduce the impact of warm-up effects.

Table 3 shows that each bird satisfied the stability criterion in an average of 14.5 and 9.5 days for Task 1 and its replication, respectively, and in 16.8 days for Task 2, so that both tasks were learnable in this primitive sense. We again consider how close accuracy was to optimality. The top row of Table 3 shows the percentage of choices that were to the better key averaged over the last 5 days of a task. Mean accuracy across Task 1 and its replication was 90.2 and 80.4 in Task 2. There was not a reliable difference between accuracy in Task 1 and its replication, so that, as in Experiment 1, a previously learned and abandoned rule could be relearned to approximately its earlier level.

As in Experiment 1, overall accuracy was far above the chance level of 50% for each bird in each task. Again, we investigated the possibility that a suboptimal decision bound accounted for above-chance performance. Specifically, as we did for Task 1 of Experiment 1, we calculated the percentage of choices that were accounted for by attending only to height or by attending only to width for Task 1 of Experiment 2. Table 3 shows that for each bird, both single-dimension decision bounds ($x = c$ and $y = c$) accounted for fewer choices than did the optimal decision bound ($x = -y + d$), which requires integration of informa-

tion from both dimensions. This suggests that birds used information from both dimensions of stimuli in making category judgments rather than responding on the basis of a single stimulus component.

Because Task 2 also required birds to attend both to height and width, we again performed a single-dimension analysis. The results, displayed in Table 3, show that each single-dimension decision bound accounted for a relatively low percentage of choices, compared with the nonlinear optimal decision bound. This again suggests that the birds attended to both dimensions, but it says little about whether information from each dimension was used appropriately. That is, one could achieve above-chance performance by using an integration rule that was different from the nonlinear optimal decision bound. Because the optimal decision bound for Task 2 was nonlinear, we investigated this possibility by testing a simpler linear decision bound. Specifically, the decision bound $x = y$ defines the line equidistant from the category means and would result in above-chance, but suboptimal, performance. Table 3 shows the results of this analysis. Note that this decision bound accounts for a comparable (in fact, slightly higher) percentage of responses than the nonlinear optimal decision bound. Thus, at this point, it is not entirely clear whether birds used a nonlinear bound similar to the optimal one or some linear approximation that allowed relatively accurate performance. In the decision-bound analysis in the next section we attempt to clarify this issue.

In summary, the performance of birds in Tasks 1 and 2 of Experiment 2 approximated optimality in terms of their attentional requirements. That is, birds seemed to attend to both height and width in making category judgments, as is required for optimal performance. This brings us to the analysis of individual decision bounds, which provide a clearer picture of the results than analyses of overall accuracy.

### Individual Estimated Decision Bounds

Figures 9 and 10 show the stimulus spaces for the best and worst cases in Tasks 1 and 2, that is, the cases in which the best fitting decision bounds accounted for the highest and the lowest number of choices, respectively. As was the case in Experiment 1, 2 birds (Birds 7 and 8) satisfied these criteria in both Tasks 1 and 2. Table 4 gives the best fitting decision bounds and the percentage of responses accounted for by those lines for each bird in each task.

*Task 1 (bound: $x = -y + d$).* Table 4 shows that the percentage of a bird's categorization responses correctly described by the best fitting estimated lines ranged from 88.0% to 91.1%. The estimated bounds therefore accurately summarized how birds categorized rectangles. The estimated slopes varied from $-2.80$ to $-0.65$. Half the estimated slopes were greater than $-1.00$ and half were smaller, indicating that they were roughly centered around the optimal value of $-1.00$. Table 4 shows also that the estimated intercepts varied from 17.25 to 39.80, again with half below and half above the optimal value of 21.00. In summary, the estimated decision bounds at least roughly

Table 3
*Experiment 2: Accuracy and Amount of Training*

| Task | Bird 4 | Bird 5 | Bird 7 | Bird 8 | Average |
|---|---|---|---|---|---|
| Task 1: $x = -y + d$ | | | | | |
| % of choices to the optimal or "better" key | 90.2 | 89.8 | 91.7 | 87.8 | 89.9 |
| % of choices accounted for by height | 81.3 | 79.6 | 82.7 | 80.7 | 81.1 |
| % of choices accounted for by width | 80.4 | 85.1 | 85.1 | 80.9 | 82.9 |
| % of choices that were reinforced | 82.9 | 84.9 | 85.6 | 82.9 | 84.1 |
| % of choices expected to be reinforced given optimal performance | 88.0 | 88.0 | 88.0 | 88.0 | 88.0 |
| Day of training on which stability criterion was first met | 12 | 9 | 10 | 27 | 14.5 |
| Total days of training | 34 | 37 | 30 | 39 | 35.0 |
| Task 2: $(x - a)^2 + (y - b)^2 = r^2$ | | | | | |
| % of choices to the optimal or "better" key | 78.7 | 81.3 | 86.7 | 74.9 | 80.4 |
| % of choices accounted for by height | 74.0 | 80.2 | 81.3 | 69.8 | 76.3 |
| % of choices accounted for by width | 77.1 | 75.8 | 86.0 | 68.4 | 76.8 |
| % of choices accounted for by integration[a] | 81.2 | 83.7 | 88.3 | 75.2 | 82.1 |
| % of choices that were reinforced | 77.3 | 80.9 | 84.0 | 73.6 | 79.0 |
| % of choices expected to be reinforced given optimal performance | 96.0 | 96.0 | 96.0 | 96.0 | 96.0 |
| Day of training on which stability criterion was first met | 20 | 16 | 15 | 16 | 16.8 |
| Total days of training | 33 | 30 | 31 | 60 | 38.5 |
| Replication of Task 1 | | | | | |
| % of choices to the optimal or "better" key | 90.3 | 88.8 | 92.7 | 89.6 | 90.4 |
| % of choices accounted for by height | 80.0 | 78.4 | 80.9 | 76.0 | 78.8 |
| % of choices accounted for by width | 83.3 | 83.6 | 83.1 | 84.4 | 83.6 |
| % of choices that were reinforced | 85.6 | 84.0 | 84.9 | 85.8 | 85.1 |
| % of choices expected to be reinforced given optimal performance | 88.0 | 88.0 | 88.0 | 88.0 | 88.0 |
| Day of training on which stability criterion was first met | 14 | 7 | 6 | 11 | 9.5 |
| Total days of training | 37 | 33 | 30 | 32 | 33.0 |

[a] *Integration* is defined by the decision-bound $x = y$.

approximated the optimal decision bound. The approximations for Birds 4 and 7 were especially impressive, but Table 4 shows that even the worst case qualitatively conformed to the optimal bound.

*Task 2 (bound: $[x - a]^2 + [y - b]^2 = r^2$).*   Table 4 shows that the percentage of choices accounted for by the best fitting circular decision bounds ranged from 77.3% to 89.1%. Table 4 shows that in all four cases, the estimated radii were greater than the optimal value and that the estimated center points $(a, b)$ were not located at the center point of the optimal decision bound, as they should have been. They were instead shifted in the direction away from the mean of Category B. In general, however, these shifts, although they were in the correct direction, were too big.

To better determine the extent to which performance was optimal, we compared the fits of the nonlinear bounds with fits of straight-line decision bounds. If birds were virtually optimal, one would expect the nonlinear estimated bounds to describe the data better than linear bounds. Table 4 shows that the circular bounds were only slightly better, in the sense that they accounted for a marginally higher percentage of choices, $t(3) = 3.07, p < .05$, one-tailed. The extremely small nature of the improvement gained by the circular bound may suggest that birds learn to approximate optimal solutions to at least some nonlinear tasks essentially by

behaving in accordance with a linear bound similar to the nonlinear one.

*Replication of Task 1.*   Table 4 shows that the percentage of choices accounted for by the best fitting straight lines ranged from 89.8% to 92.9%. Estimated slopes varied from $-2.45$ to $-1.15$, with all eight values being smaller than the optimal value of $-1.00$. Estimated intercepts varied from 21.80 to 36.50, with all eight values being greater than the optimal value of 20.00. The estimated decision bounds for Birds 4 and 7 were highly similar to the optimal bound. The birds were clearly able to remaster an ill-defined categorization task after having performed for a while on a different, potentially interfering task.
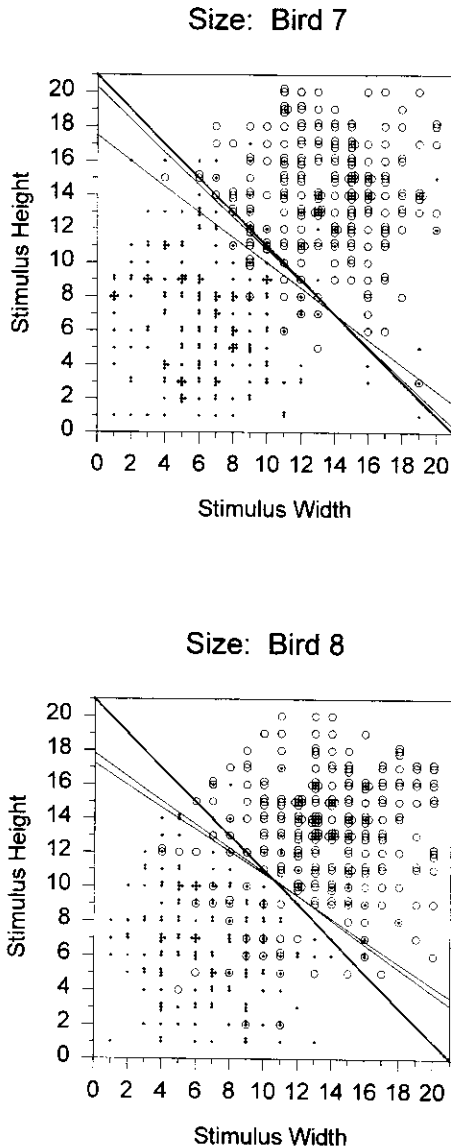
*Probe stimuli.*   As was the case in Experiment 1, we looked for a potential influence of stimulus novelty. Figure 11 displays stimuli that were presented only on the final day of the terminal 5-day periods for Birds 7 and 8. Again, performance on these probe stimuli was impressive: Twenty-four of 26 (92.3%) and 21 of 22 (95.5%) responses to probe stimuli in Task 1 were optimal for Birds 7 and 8, respectively. In both cases, performance on these stimuli was actually better than their overall accuracy rates of 91.8% and 87.8%. This pattern of results can also be seen in Task 2, in which 8 of 12 probe stimuli (66.7%) for Bird 7 and 14 of 16 probe stimuli (87.5%) for Bird 8 were categorized optimally, compared with the corresponding overall values of 86.7% and 73.6%.
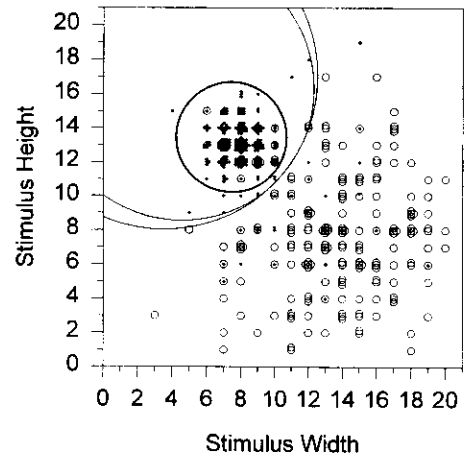
## Discussion

Individual birds learned to perform in tasks with optimal decision bounds that were more complex than those in Experiment 1 in the following ways: In Task 1 of Experiment 2, the optimal bound involved both information integration and memory of a criterion, and in Task 2 the optimal bound was nonlinear.

In summary, it appears that pigeons are capable of learning to place multidimensional stimuli into ill-defined categories in tasks with a considerable range of optimal
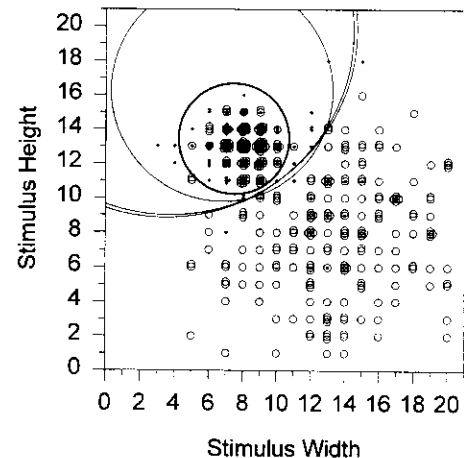
### Size: Bird 7



### Size: Bird 8



### Non-Linear: Bird 7



### Non-Linear: Bird 8



*Figure 10.* Task 2 of Experiment 2. Stimulus spaces for Birds 7 (top) and 8 (bottom). Points and circles represent choices of Category A or B, respectively. Each panel shows each of a bird's choices over the last 5 days of the task. Bold lines show the optimal decision bound, $(x - 7.5)^2 + (y - 13.5)^2 = 3.25^2$, and lighter lines show the envelope of all best fitting estimated decision bounds. Units along axes correspond to 0.3 cm.

decision bounds. This result is what we might expect if the present tasks reflected an ability of pigeons to solve the visual categorization problems they face in the natural world.

## General Discussion

### *How Might the Randomization Procedure Model Naturalistic Categorization?*

One of our goals was to develop a procedure that might serve as a tractable model for the study of naturalistic visual

*Figure 9.* Task 1 of Experiment 2. Stimulus spaces for Birds 7 (top) and 8 (bottom). Points and circles represent choices of Category A or B, respectively. Each panel shows each of a bird's choices over the last 5 days of the task. Bold lines show the optimal decision bound, $x = -y + d$, and lighter lines show the envelope of all best fitting estimated decision bounds. Units along axes correspond to 0.3 cm.

Table 4
*Experiment 2: Best Fitting Decision Bounds*

| Task | Bird 4 | Bird 5 | Bird 7 | Bird 8 |
|---|---|---|---|---|
| Task 1: $x = -y$ | | | | |
| % of choices accounted for by best fitting straight lines | 90.7 | 89.1 | 91.1 | 88.0 |
| No. of best fitting lines | 2 | 108 | 21 | 8 |
| Parameter values from most extreme best fitting straight lines | | | | |
| Slope (optimal = −1.00) | −1.30 | −2.80 | −0.95 | −0.70 |
| | — | −1.15 | −0.75 | −0.65 |
| Intercept (optimal = 21.00) | 22.70 | 21.60 | 17.50 | 17.25 |
| | 22.75 | 39.80 | 20.30 | 17.85 |
| Task 2: $(x - a)^2 + (y - b)^2 = r^2$ | | | | |
| % of choices accounted for by best fitting straight lines | 81.1 | 83.3 | 88.7 | 77.1 |
| No. of best fitting lines | 2 | 9 | 9 | 12 |
| Parameter values from most extreme best fitting straight lines | | | | |
| Slope | 0.70 | 0.70 | 1.05 | 0.70 |
| | — | 0.75 | 1.10 | 1.20 |
| Intercept | 3.90 | 4.50 | −0.30 | 2.20 |
| | 3.95 | 5.25 | 0.45 | 5.55 |
| % of choices accounted for by best fitting circles | 81.8 | 84.4 | 89.1 | 77.3 |
| No. of best fitting circles | 43 | 7 | 7 | 43 |
| Parameters from most extreme best fitting circles | | | | |
| x-coordinate (optimal = 7.50) | 0.25 | 7.75 | 3.50 | 3.50 |
| | 7.00 | 8.25 | 4.50 | 6.75 |
| y-coordinate (optimal = 13.50) | 12.75 | 14.75 | 16.25 | 16.25 |
| | 19.25 | 18.00 | 17.00 | 20.00 |
| Radius (optimal = 3.25) | 5.75 | 6.00 | 7.75 | 6.50 |
| | 15.00 | 8.00 | 9.00 | 11.00 |
| Replication of Task 1 | | | | |
| % of choices accounted for by best fitting straight lines | 90.2 | 89.8 | 92.9 | 90.4 |
| No. of best fitting lines | 5 | 33 | 4 | 22 |
| Parameters from most extreme best fitting straight lines | | | | |
| Slope (optimal = −1.00) | −1.20 | −2.45 | −1.20 | −1.95 |
| | −1.15 | −2.05 | — | −1.65 |
| Intercept (optimal = 21.00) | 22.60 | 32.50 | 21.80 | 25.95 |
| | 23.15 | 36.50 | 21.95 | 29.40 |

*Note.* Dashes indicate that the best fitting parameter was unique.

concepts in nonhuman animals. Consider the several ways in which categories in the randomization task may simulate, at least in a formal sense, naturalistic visual concepts (Ashby & Maddox, 1998). First, presumably like naturalistic concepts, they involve a large number of exemplars. Second, they are ill-defined, "fuzzy," and are characterized better by "family resemblance" than by defining features. That is, both naturalistic concepts and the present categories involve uncertainty because categories overlap in the sense that features usually associated with one category sometimes appear in exemplars of the other category. In fact, the randomization procedure might capture this aspect of naturalistic categories better than does the usual ill-defined concept task (e.g., Medin & Dewey, 1984) because the randomization task includes the all-important case in which the very same stimulus on different occasions belongs to different categories. Gestalt psychologists made much of this situation in the form of reversible images, and these ambiguous images played a critical role in the development of ideas about the nature of family resemblance, fuzzy concepts, and ill-defined concepts (Hanson, 1969; Koffka, 1935; Kohler, 1947; Wittgenstein, 1953). At first glance, it admittedly might seem strange that the very same stimulus might on different occasions belong to different categories. There are occasions, for example, when one would not be inclined to expect that the same stimulus would be categorized differently: A pigeon would seem unlikely to confuse a bit of grain, even a bit of grain swirling in the air, with a peregrine falcon. On further reflection, and on consideration of all the examples and arguments amassed by the gestalt scholars and contextualist epistemologists cited earlier, we are persuaded that there are probably at least some occasions when naturalistic stimuli satisfy this particularly interesting case. A comprehensive model of naturalistic categorization should therefore be able to handle it, and the present task seems admirably suited to do so. Third, the randomization task, with multivariate normal sampling distributions, can model the special case of naturalistic categories having many exemplars located relatively near a category's mean, with progressively fewer being located farther away from the mean. If naturalistic exemplars in other cases do not cluster near prototypes, the randomization procedure might still be able to model the situation with suitable changes in the sampling distributions. This flexibility in the randomization model is sometimes overlooked (Estes, 1994): The procedure does not require the use of overlapping bivariate

## Size: Bird 7

## Size: Bird 8

## Non-Linear: Bird 7

## Non-Linear: Bird 8
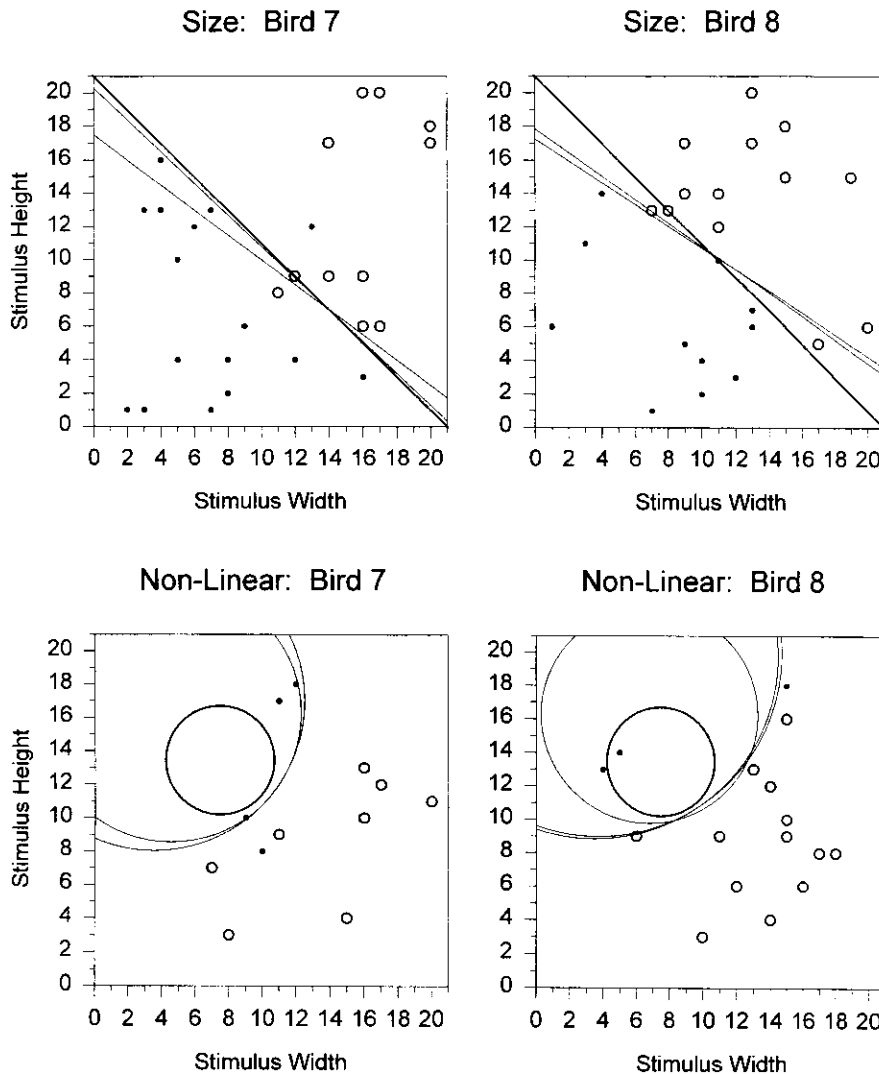
*Figure 11.* Responses to probe stimuli. Stimulus spaces for Birds 7 (left panels) and 8 (right panels) for Tasks 1 (top) and 2 (bottom). Each panel shows responses to stimuli presented only on the final day of the terminal 5-day periods represented in Figures 9 and 10. Bold lines show the optimal decision bound, and lighter lines show the envelope of all best fitting estimated decision bounds for each bird and condition. Units along the axes correspond to 0.3 cm.

normal distributions. Although these distributions are frequently used, as in the present experiments, and although they have desirable properties in mimicking certain naturalistic categories, there is nothing that requires their use. Fourth, exemplars in the randomization task with normal distributions vary more or less continuously along each dimension, at least within a specified range, as is presumably also true for many naturalistic categories, such as faces, with eye-to-eye distances, and eye to mouth distances, varying more or less continuously. Fifth, the randomization task can involve categories with unequal variances: As we discussed in the introduction to Experiment 2, Task 2, it is entirely possible that some naturalistic categories may involve unequal variances.

In addition to these formal similarities between the present task and naturalistic tasks, we can now see that there may be processing similarities as well. Different naturalistic concepts may involve either attending selectively to a single dimension or dividing attention across dimensions, and we have seen that the present task may also establish either of these different attentional processes in nonhuman animals. Finally, we reemphasize a possible naturalistic implication of the fact that pigeons and humans can perform in nearly optimal ways in a variety of different decision-bound tasks. Why might individual humans and individual pigeons, having such different neuroanatomical structures and processes and such different environmental histories, perform similarly? One obvious possibility is that the procedure taps into how these different species have evolved to solve, in approximately optimal ways, the same naturalistic problem,

which is to sort incalculably many, complex, and overlapping visual stimuli into a manageably small number of categories. In short, if we assume that the procedure taps into a naturalistic problem common to birds and humans, then the similarity in performance is readily understandable.

We hasten to add that we do not claim that a bird categorizing a rectangle into ill-defined categories of overlapping bivariate normal distributions is necessarily in every way like a bird categorizing an ambiguous naturalistic stimulus into ill-defined categories of, say, prey or predator. In particular, the stimuli used here were static and 2-D, whereas many naturalistic stimuli may be dynamic and may have many dimensions. Nevertheless, we would like to suggest that the several ways in which the present task resembles naturalistic situations recommends the task as a tractable model and useful heuristic for the study of naturalistic categories.

## Optimal Categorization of Naturalistic Stimuli

If the present method resembles the situation birds face when they categorize naturalistic stimuli, then we can infer that birds may do a remarkably good job at approximating optimal solutions in visual naturalistic categorization. That this may be so has not been apparent from the existing literatures, either on naturalistic visual concepts, because that literature is based on methods without any applicable analyses in terms of optimality (e.g., see Commons et al., 1983; Jitsumori & Yoshihara, 1997; Vaughan & Green, 1984), or on optimization in naturalistic contexts, because that literature is based on methods without any applicable analyses in terms of categorization of multidimensional visual stimuli (Commons et al., 1987; Stephens & Krebs, 1986).

For many years, data have suggested that pigeons can and often do approximate optimal solutions to diverse tasks involving probabilistic reinforcement (Graf et al., 1964; Shimp, 1966, 1969, 1973; Shimp et al., 1996; Silberberg & Ziriax, 1985, but see Vaughan, 1987; Hineline, Silberberg, Ziriax, Timberlake, & Vaughan, 1987). The present results extend those previous findings to situations in which configural properties of stimuli play critical roles in determining reward payoff and suggest that optimal solutions are also approximated in corresponding naturalistic situations involving probabilistic choice and complex, configural stimuli. The possibility that naturalistic performance is optimal has long been considered (see the review by Stephens & Krebs, 1986), and the present results add credibility to this possibility for situations involving complex, configural visual discriminations.

To the extent to which birds approximate optimal solutions to naturalistic categorization problems, they appear to conform more to a tradition of "rational" choice (Shelly & Bryan, 1964; Von Neumann & Morgenstern, 1947) than to that of "irrational" (Kahneman, Slovic, & Tversky, 1982) or "minimally rational" choice (Cherniak, 1986). Humans have been shown to make such poor use of available statistical information (such as base-rate information) in certain simulated naturalistic tasks that it is now common to emphasize an irrational quality in human judgment. Categorization of behavior in probabilistic tasks as rational or irrational is fraught with many conceptual and empirical difficulties (Gigerenzer et al., 1989). Nevertheless, it would seem that the present results from pigeons, as well as the previous results from humans in the experiments on which the present experiments were based, conform about as well as they could to the notion of rational choice, and new computational perspectives increase the plausibility of such a notion (Gigerenzer & Goldstein, 1996).

## Potential for Future Research

What are some other possible uses of a nonhuman animal version of the randomization procedure? We can now ask new questions. Are there any decision bounds avians cannot learn to approximate, and, if so, what are they? What bounds are easier or harder to learn, and what effect does learning one bound have on learning another? Will the dialogue between exemplar theorists and decision-bound theorists be resolved in the same manner as with humans? Will exemplar theory account for performance of avians on some decision-bound tasks and not on others? It now is possible to use stimuli other than the rectangles used here and thereby to begin systematically to determine what stimuli are relatively integral or separable (as in Ashby & Gott, 1988; Ashby & Maddox, 1998). Such results should help to clarify how holistic or feature-based perception by nonhuman animals affects naturalistic detection of different types of prey or predators. More generally, it would be of interest to know how a decision-bound approach applies to tasks involving stimuli more nearly naturalistic than the present rectangles, including human facial expressions (Jitsumori & Yoshihara, 1997). Would it be easier for pigeons to learn to approximate optimal bounds based on stimuli that looked like birds, grain, predators, and so on? Finally, the human version of the procedure is proving useful in neuropsychological analyses of cognitive deficits in patients with Parkinson's disease (Maddox et al., 1996), so it should now be possible for an animal model to help to unlock the neurobiological bases of this and other disorders. The search for an animal model of highly flexible human performances has been a high-priority goal for some behavioral neuroscientists, and the present results further encourage such a search because they provide additional converging evidence that in numerous visual tasks, pigeons and humans behave in strikingly similar fashions.

Without listing more than a fraction of the tasks for which there are such striking similarities between avian and human performances, we can note that Cook et al. (1996, p. 165), who used multidimensional stimuli to study visual organization, found "strikingly human-like" results, as did Fremouw et al. (1998) in research on avian shifts of attention between local and global levels of perceptual analysis. Shimp and Friedrich (1993) found evidence for a spatial attentional system in avians similar in several ways to that in humans. Blough (1984) found that confusion matrices summarizing pigeons' discrimination among letters of the alphabet, together with other data, "suggests that some common pro-

cesses determine letter recognition in pigeons and people" (p. 281). Herrnstein and Loveland's (1964) pioneering study showed that pigeons can quickly learn visual concepts based on naturalistic objects meaningful to humans. It should prove fascinating and useful to discover why pigeons and humans, with their dissimilar visual systems and neuroanatomical structures, behave so similarly in these and other visual tasks.

## References

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

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

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

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

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. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making: Handbook of perception and cognition* (pp. 251–301). San Diego, CA: Academic Press.

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,* 144–150.

Berkely, M. A., & Stebbins, W. C. (Eds.). (1990). *Comparative perception: Basic mechanisms.* New York: Wiley.

Blough, D. S. (1984). Form recognition in pigeons. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 277–289). Hillsdale, NJ: Erlbaum.

Borg, I., & Leutner, D. (1983). Dimensional models for the perception of rectangles. *Perception & Psychophysics, 34,* 257–267.

Brunswik, E. (1939). Probability as a determiner of rat behavior. *Journal of Experimental Psychology, 25,* 175–197.

Cherniak, C. (1986). *Minimal rationality.* Cambridge, MA: Bradford.

Cole, P. D., & Honig, W. K. (1994). Transfer of a discrimination by pigeons (*Columba livia*) between pictured locations and the represented environments. *Journal of Comparative Psychology, 108,* 189–198.

Commons, M. L., Herrnstein, R. J., & Wagner, A. R. (Eds.). (1983). *Quantitative analyses of behavior: Discrimination processes* (Vol. 4). Cambridge, MA: Ballinger.

Commons, M. L., Kacelnik, A., & Shettleworth, S. J. (Eds.). (1987). *Quantitative analyses of behavior: Foraging* (Vol. 6). Hillsdale, NJ: Erlbaum.

Cook, R. G., & Cavoto, K. K., & Cavoto, B. R. (1995). Same-different texture discrimination and concept learning in pigeons.

*Journal of Experimental Psychology: Animal Behavior Processes, 21,* 253–260.

Cook, R. G., Cavoto, K. K., & Cavoto, B. R. (1996). Mechanisms of multidimensional grouping, fusion, and search in avian texture discrimination. *Animal Learning and Behavior, 24,* 150–167.

Cook, R. G., Riley, D. A., & Brown, M. F. (1992). Spatial and configural factors in compound stimulus processing by pigeons. *Animal Learning and Behavior, 20,* 41–55.

Davison, M., & McCarthy, D. (1988). *The matching law.* Hillsdale, NJ: Erlbaum.

Dennett, D. C. (1987). *The intentional stance.* Cambridge, MA: Bradford.

Dunn, J. C. (1983). Spatial metrics of integral and separable dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 242–257.

Dykes, J. R., Jr. (1979). A demonstration of selection of analyzers for integral dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 734–745.

Estes, W. K. (1950). Towards a statistical theory of learning. *Psychological Review, 57,* 94–107.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53,* 134–140.

Estes, W. K. (1964). Probability learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 89–128). New York: Academic Press.

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

Estes, W. K., Burke, C. J., Atkinson, R. C., & Frankmann, J. P. (1957). Probabilistic discrimination learning. *Journal of Experimental Psychology, 54,* 233–239.

Felfoldy, G. L. (1974). Repetition effects in choice reaction time to multidimensional stimuli. *Perception & Psychophysics, 15,* 453–459.

Fremouw, T., Herbranson, W. T., & Shimp, C. P. (1998). Priming of attention to local or global levels of visual analysis. *Journal of Experimental Psychology: Animal Behavior Processes, 24,* 278–239.

Garner, W. R. (1974). *The processing of information and structure.* Hillsdale, NJ: Erlbaum.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Kruger, L. (1989). *The empire of chance.* Cambridge, England: Cambridge University Press.

Gould, S. J., & Lewontin, R. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society, B205,* 581–598.

Graf, B., Bullock, D. H., & Bitterman, M. E. (1964). Further experiments on probability-matching in the pigeon. *Journal of the Experimental Analysis of Behavior, 7,* 151–157.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics.* Huntington, New York: Krieger.

Green, L., & Kagel, J. H. (Eds.). (1987). *Advances in behavioral economics* (Vol. 1). Norwood, NJ: Ablex.

Hanson, N. R. (1969). *Perception and discovery.* San Francisco, CA: Freeman, Cooper.

Herrnstein, R. J., & de Villiers, P. A. (1980). Fish as a natural category for people and pigeons. *Psychology of Learning and Motivation, 14,* 59–95.

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concepts in the pigeon. *Science, 146,* 549–551.

Hineline, P. N., Silberberg, A., Ziriax, J. M., Timberlake, W., & Vaughan, W., Jr. (1987). Commentary prompted by Vaughan's

reply to Silberberg and Ziriax. *Journal of the Experimental Analysis of Behavior, 48*, 341–346.

Honig, W. K. (1993). The stimulus revisited: My, how you've grown. In T. R. Zentall (Ed.), *Animal cognition: A tribute to Donald A. Riley* (pp. 19–33). Hillsdale, NJ: Erlbaum.

Jitsumori, M. (1996). A prototype effect and categorization of artificial polymorphous stimuli in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 22*, 405–419.

Jitsumori, M., & Yoshihara, M. (1997). Categorical discrimination of human facial expressions by pigeons: A test of the linear feature model. *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology, 50B*, 253–268.

Kahneman, D., Slovic, P., & Tversky, S. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Harcourt, Brace & World.

Kohler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology.* New York: Liveright.

Krantz, D. H., & Tversky, A. (1975). Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology, 12*, 4–34.

Leith, C. R., & Maki, W. S., Jr. (1977). Effects of compound configuration on stimulus selection in the pigeon. *Journal of Experimental Psychology: Animal Behavior Processes, 3*, 229–239.

Lockhead, G. R. (1979). Holistic versus analytic process models: A reply. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 746–755.

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

Maddox, W. T., & Ashby, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and Nosofsky (1996). *Journal of Experimental Psychology: Human Perception and Performance, 24*, 301–321.

Maddox, T. W., Filoteo, J. V., Delis, D. C., & Salmon, D. P. (1996). Visual selective attention deficits in patients with Parkinson's disease: A quantitative model based approach. *Neuropsychology, 10*, 197–218.

McKinley, S. C., & Nosofsky, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Performance and Perception, 22*, 294–317.

Medin, D. L., & Dewey, G. I. (1984). Learning of ill-defined categories by monkeys. *Canadian Journal of Psychology, 38*, 285–303.

Nevin, J. A. (1981). Psychophysics and reinforcement schedules. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analyses of behavior: Vol. 1. Discriminative properties of reinforcement schedules* (pp. 3–27). Cambridge, MA: Ballinger.

Rachlin, H. C., Battalio, R., Kagel, J., & Green, L. (1981). Maximization theory in behavioral psychology. *Behavioral and Brain Sciences, 4*, 371–417.

Riley, D. A., & Roitblat, H. L. (1978). Selective attention and related cognitive processes in pigeons. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 249–276). Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 673–605.

Shelly, M. W., & Bryan, G. L. (Eds.). (1964). *Human judgments and optimality.* New York: Wiley.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1*, 54–87.

Shimp, C. P. (1966). Probabilistically reinforced choice behavior in pigeons. *Journal of the Experimental Analysis of Behavior, 9*, 443–455.

Shimp, C. P. (1969). Optimal behavior in free-operant experiments. *Psychological Review, 76*, 97–112.

Shimp, C. P. (1973). Probabilistic discrimination learning in the pigeon. *Journal of Experimental Psychology, 97*, 292–304.

Shimp, C. P., & Friedrich, F. J. (1993). Behavioral and computational models of spatial attention. *Journal of Experimental Psychology: Animal Behavior Processes, 19*, 26–37.

Shimp, C. P., Long, K. A., & Fremouw, T. (1996). Intuitive statistical inference: Effects of sample size on categorizations of binomial samples. *Animal Learning and Behavior, 24*, 82–91.

Silberberg, A., & Ziriax, J. M. (1985). Molecular maximizing characterizes choice on Vaughan's (1981) procedure. *Journal of the Experimental Analysis of Behavior, 43*, 83–96.

Staddon, J. E. R. (1992). Rationality, melioration, and law-of-effect models for choice. *Psychological Science, 3*, 136–141.

Staddon, J. E. R., & Ettinger, R. H. (1989). *Learning: An introduction to the principles of adaptive behavior.* San Diego, CA: Harcourt Brace Jovanovich.

Stebbins, W. C., & Berkely, M. A. (1990). *Comparative perception: Complex signals.* New York: Wiley.

Stephens, D. W., & Krebs, J. J. (1986). *Foraging theory.* Princeton, NJ: Princeton University Press.

Tinbergen, L. (1960). The natural control of insects in pinewoods: I. Factors influencing the intensity of predation by songbirds. *Archives Neerlandaises de Zoologie, 13*, 265–343.

Vaughan, W., Jr. (1987). Reply to Silberberg and Ziriax. *Journal of the Experimental Analysis of Behavior, 48*, 333–340.

Vaughan, W., Jr., & Green, S. L. (1984). Pigeon visual memory capacity. *Journal of Experimental Psychology: Animal Behavior Processes, 10*, 256–271.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Wasserman, E. A., Hugart, J. A., & Kirkpatrick-Steger, K. (1995). Pigeons show same-different conceptualization after training with complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 21*, 248–252.

Wittgenstein, L. (1953). *Philosophical investigations.* New York: Macmillan.

Wixted, J. T. (1993). A signal detection analysis of memory for nonoccurrence in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 19*, 400–411.

Wright, A. A. (1993). When is a stimulus a pattern? In T. R. Zentall (Ed.), *Animal cognition: A tribute to Donald A. Riley* (pp. 35–41). Hillsdale, NJ: Erlbaum.