

Modeling unidimensional categorization in monkeys

SIMON FARRELL

University of Bristol, Bristol, England

ROGER RATCLIFF

Ohio State University, Columbus, Ohio

and

ANIL CHERIAN and MARK SEGRAVES

Northwestern University, Evanston, Illinois

The categorization performance of monkeys on a unidimensional perceptual categorization task was examined with reference to decision bound and exemplar theories of categorization. Three rhesus monkeys were presented with stimuli varying along a single dimension, the displacement of a target light from a fixation point. Left or right saccade responses were probabilistically reinforced according to one of three functions, two of which were nonmonotonic at one end of the stimulus space. The monkeys all showed a monotonic increase in response probability as a function of target light displacement in this region, consistent with decision bound theory. Fits of a single-boundary model (GRT; Ashby & Gott, 1988) and two exemplar models—one using a probabilistic response function (GCM; Nosofsky, 1986), the other using a deterministic response function (DEM; Ashby & Maddox, 1993)—revealed overall support for the decision bound model. The results suggest that monkeys used a perceptual decision boundary to perform the task.

Categorization is thought to be fundamental to information processing (Ashby, Boynton, & Lee, 1994), as well as behaviorally adaptive (Anderson, 1991). It has therefore been given extensive empirical treatment in the human cognition literature (see, e.g., Erickson & Kruschke, 1998; Maddox & Ashby, 1993, 1996; Maddox & Bohil, 2001; Medin & Schaffer, 1978; Nosofsky, 1986, 1991), and several well-formulated models have been proposed to explain processes underlying categorization (Anderson, 1991; Ashby, 2000; Ashby & Townsend, 1986; Bussemeyer & Myung, 1992; Erickson & Kruschke, 1998; Kruschke, 1992; Nosofsky, 1986; Nosofsky & Palmeri, 1997; see also Ashby, 1992a).

A body of recent research has complemented these developments by focusing on categorization behavior in animals. Although animal categorization has not enjoyed the extensive theoretical development witnessed in human categorization (Shimp, Long, & Fremouw, 1996), some

recent research has taken advantage of the advanced state of theory in the human literature by focusing on examining animals' categorization under conditions previously examined in humans (Fagot, Kruschke, Dépy, & Vauclair, 1998; Herbranson, Fremouw, & Shimp, 1999; Medin & Dewey, 1984; Sigala, Gabbiani, & Logothetis, 2002; Sigala & Logothetis, 2002; Smith, Minda, & Washburn, 2004). In several cases, this work has been complemented by application of well-developed quantitative theories of human categorization to the behavior of animals (Fagot et al., 1998; Herbranson et al., 1999; Sigala et al., 2002).

Our article contributes to this theoretical development by applying three popular models of human categorization—a decision bound model (Ashby & Townsend, 1986) and two versions of exemplar theory (Ashby & Maddox, 1993; Nosofsky, 1986)—to the performance of rhesus monkeys on a probabilistic perceptual categorization task. We relate contemporary models of human categorization to research in animal categorization and examine the categorization behavior of monkeys using an experimental paradigm recently found to be useful for distinguishing between different accounts of human categorization (Rouder & Ratcliff, 2004). We then present simulations comparing the quantitative accounts of exemplar and decision bound models, and the results suggest that monkeys' performance can be explained with reference to a single boundary in decision bound theory. The experiment and simulations give further insight into the representations and decision processes underlying simple probabilistic perceptual categorization.

This article was supported by National Institute of Mental Health Grants R01-MH59893, R37-MH44640, and K05-MH01891 and National Eye Institute Grant R01-EY08212. We thank Angela Nitzke and the staff of Northwestern's Center for Experimental Medicine for animal care, and Northwestern's Instrument and Electronics Shops for machining and electronic hardware. We also thank Eric-Jan Wagenmakers and several anonymous reviewers for their comments. Correspondence relating to this article may be sent to S. Farrell, Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Clifton, Bristol BS8 1TN, England (e-mail: simon.farrell@bristol.ac.uk).

Comparing Models of Categorization

Three general principles have been posited to underlie categorization abilities in humans. These form the basis of prototype, exemplar, and decision bound models (see also Anderson, 1991; Gluck & Bower, 1988). *Prototype* theories (e.g., Reed, 1972) hold that when presented with members from categories, people form a representation of the central tendency for each category (i.e., a prototype). People are then thought to respond to a new stimulus based on its similarity to each category's prototype (Ashby & Maddox, 1993; Reed, 1972, 1978). In *exemplar* models, exposure to a category member results in the storage of a representation of that stimulus and its category (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986, 1992). Categorization of a new item occurs by matching the item to all of the stored exemplars in each category (Nosofsky, 1986). The probability of making the response for a particular category is generally determined by use of a normalized choice probability rule (Nosofsky, 1986), although several authors have noted that a deterministic rule may also be used in these models (Nosofsky, 1991; Ashby & Maddox, 1993). Finally, *decision bound* models, which are grounded in signal detection theory (Green & Swets, 1966), assume that the stimulus space is partitioned into separate regions by decision boundaries (Ashby, 2000; Ashby & Gott, 1988; Ashby & Townsend, 1986). New stimuli are categorized according to the region of stimulus space in which they fall. Variability in the encoding of stimuli and the placement of boundaries (Ashby & Maddox, 1993) results in miscategorization of stimuli (and hence probabilistic responding) as a function of their distance from the boundaries.

Although a number of studies have provided varying support for all three theories, and particularly for exemplar and decision bound theories (see, e.g., Ashby et al., 1994; Ashby & Maddox, 1993; Erickson & Kruschke, 1998; Maddox & Ashby, 1996; Maddox, Ashby, & Gottlob, 1998; Maddox & Bohil, 2001; Medin & Schaffer, 1978; Nosofsky,

1991; Nosofsky & Alfonso-Reese, 1999), it has been argued that exemplar and decision bound models can closely mimic each other's behavior, thus compromising attempts to quantitatively distinguish the models (Estes, 1992; Kalish & Kruschke, 1997; Rouder & Ratcliff, 2004). Recognizing the methodological difficulties involved in distinguishing decision bound and exemplar models, Rouder and Ratcliff (2004; see also Kalish & Kruschke, 1997) addressed this mimicking problem in a probabilistic, unidimensional categorization task by using irregular mappings of stimulus values along a single perceptual dimension to categorical responses (see also Espinoza-Varas & Watson, 1994; Ratcliff & Rouder, 1998; Ratcliff, Thapar, & McKoon, 2001). Rouder and Ratcliff used probabilistic feedback, in which the probability of a response being designated as correct was nonmonotonically related to the single stimulus dimension, thus providing a qualitative test of exemplar models, which predict that response probabilities should follow the reinforcement probabilities in functional form (Nosofsky & Palmeri, 1997). Consider, for example, the middle panel of Figure 1 (labeled *Dip*). The probability that a particular response is reinforced (in this case, a rightward saccade) is plotted in crosses as a function of the value of a stimulus along a single perceptual dimension (in this case, the separation between two lights). Notice that on the left of the graph, as the distance between the two lights decreases, the probability of reinforcement drops from .95 to .65. Exemplar theory (the predictions of which are illustrated by a dashed line) predicts that the probability of responding will also drop; although the response probabilities do not exactly match the reinforcement probabilities, over the stimulus values they do have a functional change similar to that of the reinforcement probabilities. Conversely, decision bound models predict that the response probability function will be monotonic as long as there is only one boundary; if boundaries are placed optimally, this should be the case if (in a two-choice task) the reinforcement probabilities for a particular response cross .5 only at one

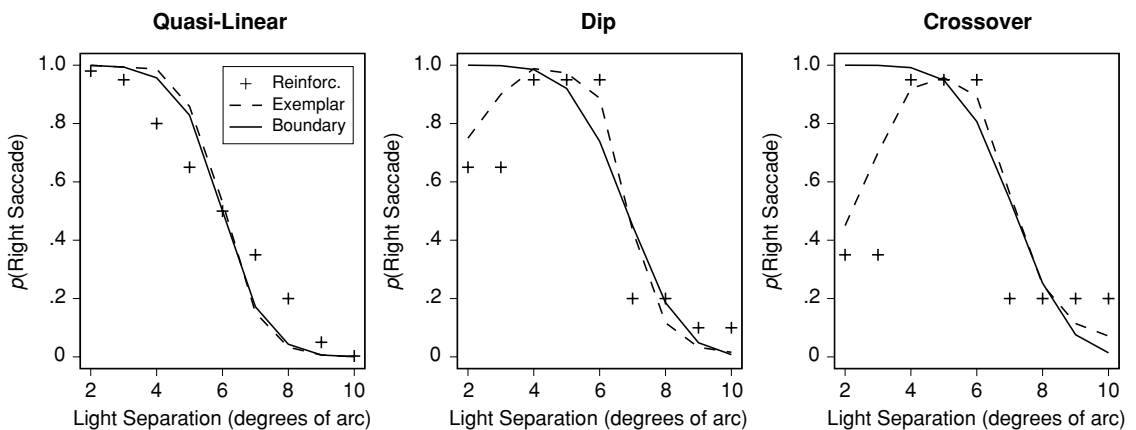


Figure 1. Qualitative predictions from exemplar theory and decision bound theory for three reinforcement probability functions (quasi-linear, dip, and crossover conditions). The crosses give the probability that a right saccade was reinforced at each light separation value. Qualitative predictions (the probability of a right saccade) from the exemplar and decision bound models are represented as dashed and solid lines, respectively.

point along the stimulus dimension (see, e.g., the left and middle panels of Figure 1). For the example in the middle panel of Figure 1, this prediction is illustrated by the solid line. Although the reinforcement probabilities drop as the stimulus values decrease, decision bound theory predicts a monotonic increase in the probability of responding as stimuli move away from the boundary (here assumed to be placed optimally between light separations 6 and 7).

By examining the behavior of human participants categorizing unidimensional stimuli under such reinforcement functions, Rouder and Ratcliff (2004) found overall support for decision bound theory; however, they also discovered that support for decision bound or exemplar models depended upon the discriminability of the stimuli. For example, in their Experiment 2, square size constituted the single stimulus dimension. When the increment in size between successive squares was small, results supported a decision bound model. However, when the step size was large, and stimuli were thus highly discriminable, the patterns of performance were more in line with exemplar models. Rouder and Ratcliff concluded that the data were consistent with a "hybrid" model in which both exemplar- and rule-based mechanisms could be used, and they suggested that rules are used for less discriminable stimuli because of the difficulty of accurately identifying and matching exemplars in such cases.

The paradigm employed by Rouder and Ratcliff (2004) closely parallels research in animal learning, in which the reinforcement of responses has often been probabilistic rather than deterministic, resulting in various qualitative patterns of behavior (e.g., Estes, Burke, Atkinson, & Frankmann, 1957; Graf, Bullock, & Bitterman, 1964; Meyer, 1960; Shimp, 1973; Wilson, Oscar, & Bitterman, 1964a, 1964b). In some cases, "maximizing" behavior has been observed, with an animal responding in a deterministic manner; thus, if the probability of reinforcement of a response given a stimulus is above chance (e.g., above .5 in a two-choice situation), that response will nearly always be given to the stimulus (see, e.g., Estes et al., 1957; Meyer, 1960; Treichler, Conner, & Ricciardi, 1970; Wilson et al., 1964b). This behavior has been named *maximizing* because it achieves the maximal reinforcement probability, and it is consistent with the processes of decision bound theory discussed above. In other cases, probability matching has been witnessed, with an animal's response probabilities approximating the reinforcement probabilities, so that the same stimulus does not always evoke the same response (e.g., Robbins & Warner, 1973; Wilson & Oscar, 1966; Wilson et al., 1964a). It has been noted that findings of probability matching or maximizing have implications for stochastic learning theory (see, e.g., Estes, 1950) and questions of optimality in animal discrimination.

Building on this animal learning research, investigations have focused on more complex categorization in animals (see Astley, Peissig, & Wasserman, 2001; Astley & Wasserman, 1992, 1999; Bhatt, Wasserman, Reynolds, & Knauss, 1988; Dépy, Fagot, & Vauclair, 1997; Fagot et al., 1998; Herrnstein, Loveland, & Cable, 1976; Medin & Dewey, 1984; Shimp et al., 1996; Smith et al., 2004;

Spinozzi, 1996). Notably, several of these animal categorization studies have explicitly focused on models of human categorization. Shimp et al. (1996) exposed pigeons to binomial samples representing the repeated tossing of one of two biased "coins." In their experiment, a green light flash represented "heads" and a red light flash indicated "tails," and pigeons were reinforced for correctly indicating from which binomial generator a sequence of lights had come. Shimp et al. found that the pigeons' response probabilities exceeded the binomial probabilities for small sample sizes (number of observations of the "coin") and undermatched them for large sample sizes. Furthermore, Shimp et al. found a nonlinear relationship between the likelihood that a sample was generated by a particular coin and the probability of responding in favor of that coin, indicating that the pigeons did not simply match probabilities. In a similar vein, Herbranson et al. (1999) tested pigeons on a multidimensional categorization task in which the members of each category, rectangles, were normally distributed along two stimulus dimensions, width and height. This experiment was a replication of tasks previously employed to examine decision bound accounts of human perceptual categorization (Ashby & Gott, 1988; Ashby & Maddox, 1992). Herbranson et al. found that the boundaries estimated from the pigeons' performance approximated the boundaries necessary to perform the task optimally, a result previously found with humans (Ashby & Gott, 1988; Ashby & Maddox, 1992; Maddox & Ashby, 1993). Surprisingly, this was the case even when the optimal boundary was nonlinear (see, e.g., Figure 10 of Herbranson et al., 1999). Although the experiment was not intended to test any particular model, the patterns of boundary use suggest that complex representations may underlie the categorization behavior of animals, representations that perhaps approach the complexity of those underlying humans' performance on such perceptual tasks (Herbranson et al., 1999).

Sigala et al. (2002) conducted a study that explicitly compared models of categorization. They compared seven computational models of categorization using data gathered from rhesus monkeys and humans on a "subordinate" categorization task, in which categories were distinguished by small changes in stimulus elements. Two classes of stimuli were used: line drawings of faces and of fish. Each class consisted of four dimensions along which stimuli could vary; for example, in the case of faces, the varying dimensions were eye height, eye separation, mouth height, and nose length (see also Nosofsky, 1991; Reed, 1972). Sigala et al. found that for both monkeys and humans, exemplar and decision bound models gave better accounts of the observed response probabilities than did a prototype model and probability-based models. The exemplar and decision bound models, however, could not be distinguished on the basis of their fit, reflecting the mimicking observed by Rouder and Ratcliff (2004) in human categorization.

Both the theoretical and empirical inquiries motivating these efforts point to the likelihood that common representations and processes underlie perceptual categorization in humans and animals and that general theories may

be applied to different species. In light of this research, we were interested in examining the performance of animals in a task that could dissociate exemplar and decision bound models. Accordingly, an experiment was run on monkeys using the two-alternative probabilistic unidimensional categorization paradigm of Rouder and Ratcliff (2004). The data obtained using this procedure allowed exemplar and decision bound models to be quantitatively evaluated in order to determine the processes underlying the monkeys' behavior.

The Present Experiment

The experiment was run as part of a larger project aimed at determining the neural correlates of choice probabilities and decision times in monkeys (see Ratcliff, Cheria, & Segraves, 2003). The subjects performed a simple visual categorization task in which the monkeys were to respond by making a saccade to the left or right depending on the displacement of a target light from a fixation point (see the Method section). There were nine values of displacement, ranging between 2° and 10° of visual arc in steps of 1° .

Three different reinforcement functions were employed, as shown in Figure 1. The three reinforcement functions (the "+" symbols in Figure 1) give the probability that a right saccade was reinforced for increasing light displacement (in degrees). Along with these functions, Figure 1 contains schematic depictions of the predictions of decision bound and exemplar theories.

The first reinforcement function, the *quasi-linear* condition shown in the first panel of Figure 1, was a linear function with negative slope that flattened out at the ends to allow probabilistic reinforcement at the extreme values. As shown in Figure 1, exemplar and decision bound models are expected to make indistinguishable predictions for this reinforcement structure.

The second reinforcement function, in the second panel of Figure 1, was named the *dip* condition. The dip referred to occurs at the left of the panel, where (heading toward the left) the probability of reinforcement drops from .95 at a light displacement of 4° to .65 at a light displacement of 3° . For our purposes, this region will be referred to as the *critical region*, since the predicted differences between the models are expected to be focused here. Figure 1 shows that decision bound models predict consistent responding in the critical region, whereas exemplar models predict that the response function will follow the reinforcement function in form.

The third reinforcement structure, shown in the right panel of Figure 1 and called the *crossover* condition, is similar to that of the dip condition in the middle panel, with the important difference that the reinforcement probabilities for a right saccade response drop to .35 (i.e., below .5) at displacements of 2° and 3° . The predictions of exemplar theory should be qualitatively similar to the predictions for the dip condition. Specifically, the probability of responding as light separation decreases in the critical region is expected to be nonmonotonic, because many exemplars in this region will be associated with reinforcement for a left saccade response. Decision bound models

may predict one of two patterns, in this case. If monkeys use a single boundary to perform the task, predictions should be very similar to those in the dip condition. That is, the response probability should continue to monotonically increase as stimulus separation decreases. However, if a second boundary is introduced to allow more accurate responding, then nonmonotonicity should occur, leading to predictions similar to those of exemplar theory.

METHOD

Subjects

The experiment was conducted on 3 female rhesus monkeys (*Macaca mulatta*) weighing 6–9 kg and identified here as MK03, MK05, and MK07. The procedures used in training, surgery, and experiments with the monkeys were all approved by Northwestern University's Animal Care and Use Committee. Each monkey received preoperative training followed by an aseptic surgery to implant a subconjunctival wire search coil and a stainless steel receptacle to allow the head to be held still during behavioral paradigms. As preparation for neural recordings conducted in parallel for another project, one or two stainless steel recording cylinders were fixed to the skull surface during this surgery. Surgical anesthesia was induced with the short-acting barbiturate Brevital (11 mg/kg), injected through an IV line, and maintained using halothane (1%) inhaled through an endotracheal tube. All of these methods have been described in detail elsewhere (Dias & Segraves, 1999; Segraves, 1992).

Apparatus and Procedure

We used a two-choice task in which the monkey decided to saccade left or right according to the displacement of a test stimulus light from a fixation light (Figure 2). To indicate their responses, the monkeys were trained to make a saccade to one of two response target lights, one in the left visual hemifield and the other in the right hemifield. The response target lights were kept on throughout the running of the task. The fixation light was always presented at the center of the display screen. The moveable test stimulus light

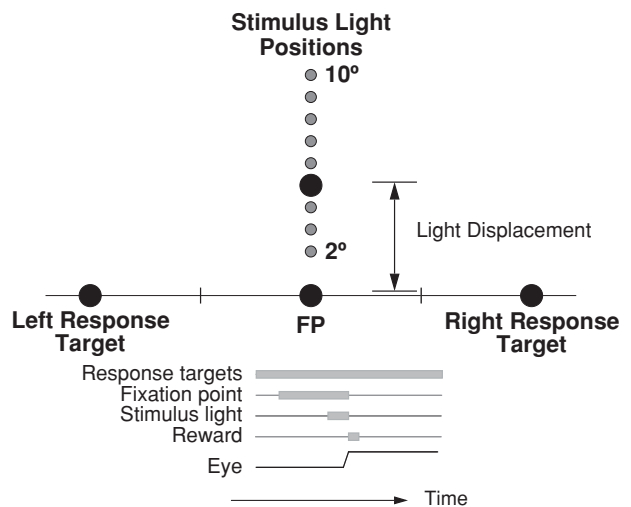


Figure 2. Schematic depiction of the experiment setup and procedure. The top part of the figure shows the experimental array; the horizontal array presents the response targets and fixation point (FP), and the vertical lights are the stimuli, which varied from 2° to 10° of separation from FP. The bottom part of the figure shows the temporal structure of the trial.

was positioned vertically above the fixation light by a mirror galvanometer (General Scanning, Inc.; www.gsig.com), and its distance from the fixation light varied randomly from trial to trial over a range of 2° – 10° , in 1° increments. The fixation point appeared at the beginning of each trial, and after the monkey achieved fixation, the fixation point remained illuminated for a randomly varied interval of 500–1,500 msec. At the end of this interval, the test stimulus light was turned on. The test stimulus light and fixation point then remained on for up to 5,000 msec while the monkey was given the opportunity to make a saccade to one of the response target lights.

Feedback varied as a function of displacement according to the reinforcement functions in Figure 1. Since reinforcement was probabilistic, before each trial the experimental program quasi-randomly determined which response would be reinforced, according to the functions in Figure 1. If the monkeys made the selected response, they were given a sip of juice as a reward. In line with similar experiments (Fagot et al., 1998; Smith et al., 2004), and following procedures employed with human participants, a noncorrection procedure was used, meaning that if the selected response was not made, no opportunity was then given to make the selected response.

The monkeys first received training on the quasi-linear condition shown in the first panel of Figure 1. They then completed blocks of each of the conditions in Figure 1, with the order of blocks and the number of sessions per block different for each monkey. Generally, the condition for a monkey was changed when its behavior in a condition had stabilized, at which point the monkey would be switched to a new reinforcement condition.

The number of trials within each session varied according to the behavior of a monkey (i.e., whether it was making a large proportion of responses that were not unambiguously left or right saccades). On average, each daily session contained approximately 1,200–1,800 trials and lasted for 60–120 min.

RESULTS

Trials with saccades made to either response target light were scored as valid and included in the data used for this report, regardless of whether the saccade was rewarded or not. Trials on which the monkey failed to make a saccade, or made a saccade that did not terminate at either response target, were rejected. Responses were also ex-

cluded from analysis if a complete saccade was not made within 600 msec or if the saccade began within 160 msec after the onset of a test stimulus (these being defined as anticipatory responses). With the exclusion of these trials, the figures presented throughout give the probability of a definite right saccade (the alternative being a definite left saccade).

Session-to-session performance of the monkeys was quite variable, especially after a switch in reinforcement function (see also Treichler et al., 1970; Wilson et al., 1964a). Since we were not interested in learning curves, in order to obtain stable data for model fitting, a subset of the sessions was selected for analysis and model fitting. Three sessions were selected for each monkey in each condition according to two criteria. One criterion was that performance for the session should be near asymptote for the condition associated with that session; that is, the session should come close to the end of a block of sessions, after performance had stabilized. The second, related, criterion was that the performance in that session should be similar to performance in the other sessions selected for that monkey. Both of these conditions were satisfied by calculating response probabilities for each session and finding the three sessions whose mutual residuals were the lowest.¹

Figures 3–5 give the average right saccade probabilities for each condition; each figure gives the results under the three conditions for a particular monkey. The empirical results, as a function of light displacement, are depicted as unconnected open circles. Also shown are standard error bars, which were estimated from the observed response proportions, assuming a binomial distribution. Finally, the lines in each panel are the predictions of the models, which will be discussed after presentation of each of the models.

The results for the quasi-linear condition (first panel in Figures 3–5) are consistent with findings in the discrimination literature, where the relationship between reinforcement probability and response probability is typi-

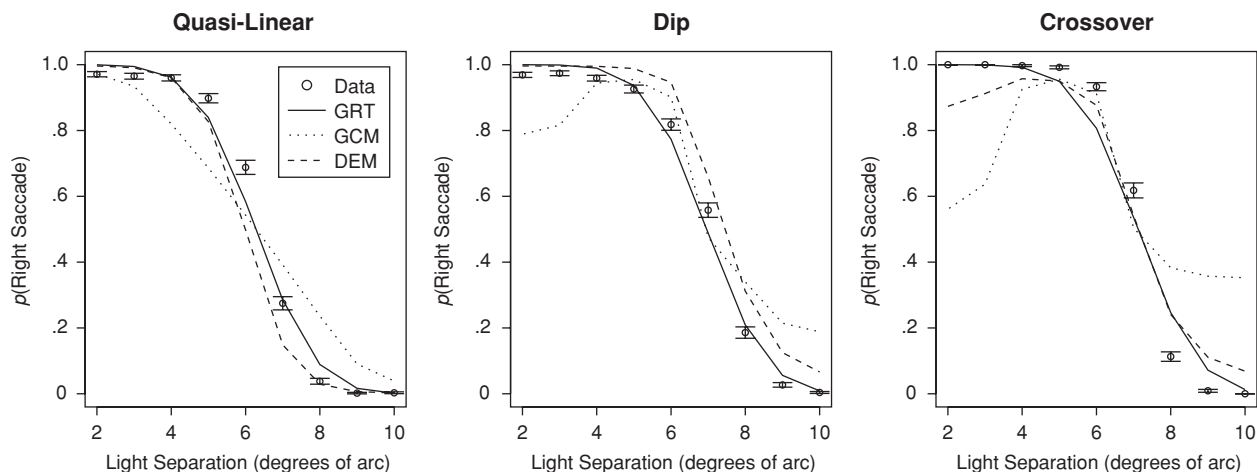


Figure 3. Proportion of right saccades at each light separation for Subject MK03, along with the maximum likelihood predictions of general recognition theory (GRT), the generalized context model (GCM), and the deterministic exemplar model (DEM). Data are represented by unconnected circles, and the predictions of GRT, GCM, and DEM are shown by solid, dotted, and dashed lines, respectively.

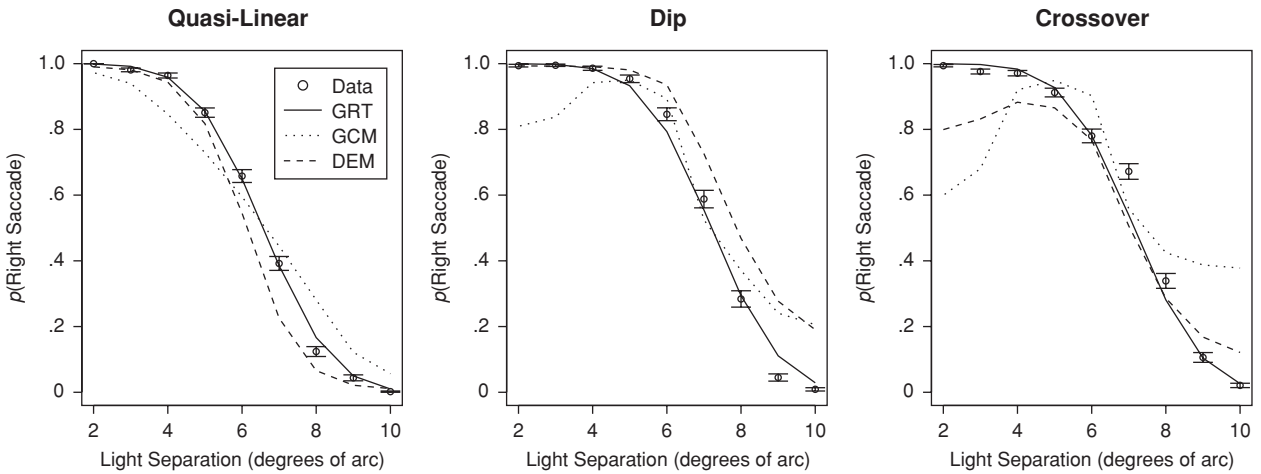


Figure 4. Proportion of right saccades at each light separation for Subject MK05, along with the maximum likelihood predictions of GRT, GCM, and DEM.

cally an ogival function (see, e.g., Shimp, 1973; recall that reinforcement probability was related to light separation by an approximately linear function, flattening at the ends, in the quasi-linear condition). Although this condition could not discriminate between exemplar and decision bound models, the results do confirm that the monkeys could perform the task, and they replicated findings from previous discrimination learning experiments of non-deterministic responding (Ratcliff et al., 2003; Shimp, 1973; Shimp et al., 1996; Wilson et al., 1964a).

The results in the dip condition (second panel in Figures 3–5) are much the same as those from the quasi-linear condition. In particular, whereas the reinforcement probabilities bend down from near 1.0 to .65 with decreasing light displacement in the critical region, the response probabilities show no such pattern and remain at ceiling for the low values of light separation. Although no dip in response probabilities in the critical region is predicted

by decision bound models, this result is inconsistent with the qualitative predictions of exemplar theory discussed earlier. The only noticeable departure from the results of the quasi-linear condition is that the response function appears to have shifted to the right by around 1° of light displacement. In fact, right saccade probabilities for light displacement of 7° were above .5, even though the objective reinforcement probability for this stimulus value was .2. Note, however, that the monkeys were sensitive to the change in reinforcement probabilities; this is reflected in an overall shift of the response function to the right.

Surprisingly, this monotonicity in the response function in the critical region was also apparent in the results for the crossover condition (see last panels in Figures 3–5). Even when the probability of being rewarded for a right saccade dropped below .5 in the critical region, the monkeys’ right saccade probabilities remained at ceiling. This finding was unexpected for both the exemplar and deci-

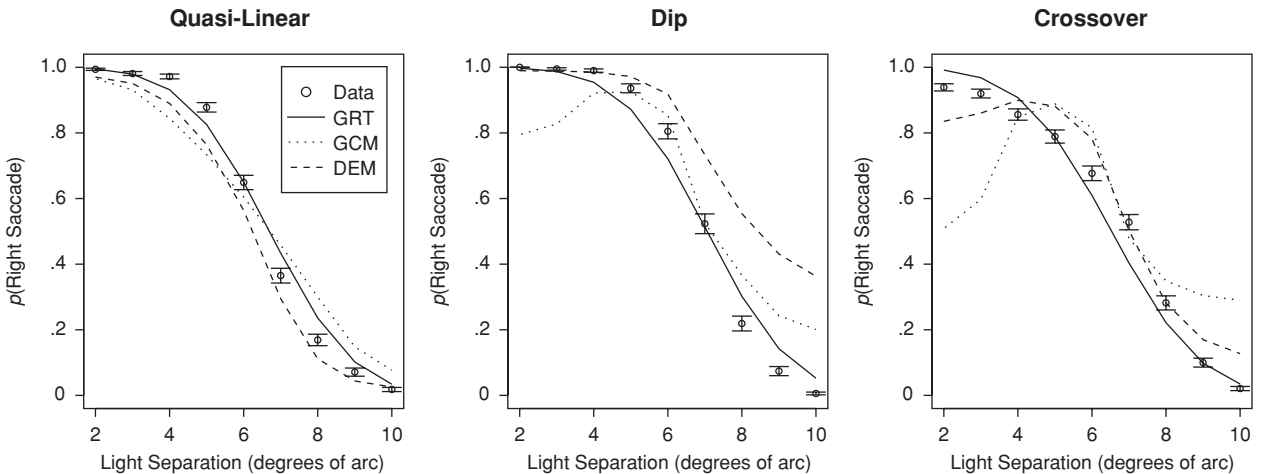


Figure 5. Proportion of right saccades at each light separation for Subject MK07, along with the maximum likelihood predictions of GRT, GCM, and DEM.

sion bound theories. Exemplar theory predicts that performance in this condition should be nonmonotonic and drop to some extent with decreasing light displacement. Decision bound theory, based on optimal boundary placement (Herbranson et al., 1999), predicts that a second boundary will be placed somewhere between light separations 3 and 4, so in that model response probabilities would also be expected to decrease with decreasing light separation. However, these results are consistent with the use of a single boundary placed as in the quasi-linear and dip conditions.

One possible explanation for the monotonicity in the critical region in the crossover condition is that the monkeys had overlearned the quasi-linear and dip conditions, and that this prevented the monkey from learning a qualitatively different task. To check this possibility, the monkeys were also trained under a fourth, *probability reversal* condition (see, e.g., Treichler et al., 1970), in which the quasi-linear reinforcement function shown in Figure 1 was flipped along the abscissa around a light displacement of 6°. Notably, the monkeys learned this reverse rule quite quickly: Performance in this condition generally stabilized in three or four sessions. The observation that the monkeys were able to switch to the new reversed rule in two or three sessions mitigates against perseverative responding as an explanation for the monotonic responding in the critical region in the crossover condition.

Finally, there is little evidence of any substantial individual differences in behavior among the 3 monkeys. Although there were slight deviations in quantitative results (for example, the response functions of MK05 and MK07 appear slightly flatter than those for MK03), the general pattern of results was very consistent across the 3 monkeys.

In general, these results appear to run contrary to the qualitative predictions of exemplar theories and provide support for single-boundary versions of decision bound theory. The results in the crossover condition imply that if monkeys were using boundaries to perform the task, only a single boundary was used even in cases in which the use of two boundaries was required for optimal performance.

Model Fits

Three models—namely general recognition theory (GRT; see, e.g., Ashby & Maddox, 1993; Ashby & Townsend, 1986); the exemplar-based generalized context model (GCM) of Nosofsky (1986, 1991); and a deterministic version of GCM, the deterministic exemplar model (DEM; Ashby & Maddox, 1993)—were fit to the experimental data in order to substantiate our conclusions about the superiority of decision bound models. A strength of using quantitative models is that models can be fit (using an objective minimization algorithm) to ensure that verbally derived predictions are consistent with the actual best-fitting performance of the models. We introduce the models in turn, and then evaluate their fits.

GRT. GRT (Ashby & Gott, 1988; Ashby & Maddox, 1993; Ashby & Townsend, 1986), a decision bound model, assumes that a stimulus is represented as a point in a

multidimensional stimulus space. The space can be partitioned by boundaries so that regions of space correspond to particular categories or responses. A stimulus is categorized by determining in which region it lies. Although GRT uses nonlinear boundaries if these are optimal for a category structure (Ashby & Maddox, 1992), in the case of the unidimensional categorization task a boundary is a single point on the stimulus dimension. GRT assumes that stimulus processing is contaminated by perceptual noise; although the same physical stimulus might be presented several times during an experiment, it will result in a different perceptual state each time because of noise in the perceptual system.

GRT was applied as it was by Rouder and Ratcliff (2004). A stimulus on a particular trial was assumed to be sampled from a normal distribution with mean equal to the true value of light displacement and standard deviation σ representing the amount of perceptual noise. If the sampled value was higher than the boundary value, a left saccade was recorded; otherwise, a right saccade was identified. The probability of categorizing a light separation l_i as Category B (a right saccade) was given by

$$p(B) = \int_{-\infty}^{\delta} f(x, l_i, \sigma^2) dx, \quad (1)$$

where δ is the position of the boundary, f is the normal probability density function, x is the stimulus space (i.e., the dimension of light separation), and σ^2 is the variance of perceptual noise, which was assumed to be constant across stimuli and conditions (see Rouder & Ratcliff, 2004).

One important factor that may have affected performance in these choice tasks is systematic bias not attributable to the stimuli. For example, in choice situations, animals may often develop response or position habits, preferring one spatial response over the other (see, e.g., Nash, 1970). Animals might also show bias due to the overall reinforcement probability for each response independent of the presented stimulus, or due to sensitivity to feedback in previous trials (see below). To enable GRT to naturally capture these possible sources of systematic responding, the boundary δ was allowed to be a free parameter.

GCM. GCM (Nosofsky, 1986, 1991), an exemplar model, also represents the stimulus as a point in multidimensional space (which is usually derived from the physical stimulus space using multidimensional scaling [MDS]; Maddox & Ashby, 1993; Nosofsky, 1986, 1991). The model makes three key assumptions. First, learning involves the storage of exemplars in memory, where each exemplar is associated with a category or response. Second, categorization of a new stimulus is based on its similarity to all exemplars in memory. Similarity is computed using an exponential gradient (Shepard, 1987): s_{ij} , the similarity between the stimulus l_i and exemplar y_j , is given by

$$s_{ij} = \exp(-\lambda d_{ij}), \lambda > 0, \quad (2)$$

where d_{ij} is the geometric distance between exemplar y_j and the stimulus. In the unidimensional paradigm used here, $d_{ij} = |l_i - y_j|$, the absolute difference between the

stimulus l_i and the exemplar (see also Rouder & Ratcliff, 2004). The third assumption is that response probabilities are given by a choice rule based on the sums of the similarities of exemplars in each category. The sum of similarities for a left saccade is given by

$$a = \sum_{j \in A} s_j, \quad (3)$$

and for a right saccade

$$b = \sum_{j \in B} s_j, \quad (4)$$

where the sum is over exemplars s_j that belong to either category. The summed similarities a and b are transformed to response probabilities using the Luce–Shepard choice rule (Luce, 1963; Shepard, 1957). By this rule, the probability of Response B is given by

$$p(B) = \frac{b}{a+b}. \quad (5)$$

As mentioned in the discussion of GRT, the monkeys might be expected to show some bias in responses from condition to condition. To enable GCM to capture these effects, a bias parameter, β , was incorporated into GCM, in keeping with Nosofsky (1986). The probability of a right saccade, $p(B)$, was given by

$$p(B) = \frac{\beta b}{(1-\beta)a + \beta b}, \quad 0 \leq \beta \leq 1. \quad (6)$$

Examination of Equations 2–6 suggests how GCM might accommodate the data. The parameter λ in Equation 2 scales the gradient of similarity. If this parameter is large, similarity will fall off quickly with increasing distance d_{ij} , and if it is small, the similarity gradient will be roughly flat. This component will be especially important for determining performance in the critical region in the present experiments. If the parameter is set to a high value, stimuli will only bear strong similarity to their own exemplars (keeping in mind that the dimension is discrete, so each stimulus will correspond to one of the nine exemplars), and responding will tend to track the reinforcement function (Figure 1). If this parameter has a low value, then the exemplars in the critical region will also be similar to the adjacent exemplars from the other category (light separation values 4 and 5), increasing the probability of a right saccade for light separations of 2 and 3.

DEM. The final model examined was a deterministic exemplar model (DEM; Ashby & Maddox, 1993; Maddox & Ashby, 1993). Nosofsky (1991) used a deterministic decision process with GCM in relating classification and recognition, whereby a response was made depending on whether the difference in summed similarities exceeded a criterion. Ashby and Maddox (1993) pointed out that this model did not formally relate to the standard GCM (e.g., Nosofsky, 1986), which uses the similarity choice rule (Equation 5), and so they introduced a generalization of the decision rule in GCM that is identical to a deterministic model. Specifically, they showed that using the decision rule

$$\text{Respond } B \text{ if } \log(b) - \log(a) > \delta$$

(where a and b are from Equations 3 and 4 and δ is a cutoff value) is equivalent to using a probabilistic rule:

$$p(B) = \frac{\beta b^\gamma}{(1-\beta)a^\gamma + \beta b^\gamma}. \quad (7)$$

The parameter γ in Equation 7 scales the consistency of responding in the model. If $\gamma > 1$, the model will tend to respond consistently (i.e., always give the same response to a stimulus), but as γ approaches 1 the model will approach probability matching—that is, it will display variability in responding. As γ moves from 1 to 0, the model's predictions will approach chance (.5 for the two-choice task). Nosofsky and Palmeri (1997) recently showed that an exemplar-based random-walk model could respond deterministically; in fact, the decision rule resulting from Nosofsky and Palmeri's model is formally identical to Equation 7. We used DEM to model consistency of responding in exemplar theory because of its simplicity of application and because the model has performed well with this paradigm in previous applications (Rouder & Ratcliff, 2004).

Details of Fitting

The models were fit to the data using maximum likelihood estimation; details of the model fitting appear in Appendix A. One assumption made in fitting the models to the data was that the psychological stimulus scale was identical to the physical stimulus scale. That is, light displacement, in degrees, was directly used as x in Equation 1 and d in Equation 2. As noted above, GCM typically takes MDS-transformed stimuli as input. MDS is a procedure that maps physical stimuli onto dimensions more closely reflecting the hidden structure underlying the data (Young & Hamer, 1987). However, an MDS solution was not possible in this case, because similarity ratings could not easily be obtained from the monkeys. Nevertheless, it is worth noting that Sigala et al. (2002), who used multidimensionally varying stimuli, found that an exemplar model based on physical dimensions was as successful in accounting for their results as was the MDS-based GCM.

Another assumption made in the model fitting related to sequential dependencies in the data. Previous work has shown that examination of sequential effects in choice tasks may reveal data patterns that can discriminate between quantitative models better than examination of average unconditionalized performance can (see, e.g., Ratcliff, Van Zandt, & McKoon, 1999; Stewart, Brown, & Chater, 2002). Findings of sequential dependencies in discrimination (Overall & Brown, 1963; Robbins & Warner, 1973; Steiner, 1970; Wilson et al., 1964b) suggest that such dependencies might exist in the performance of the monkeys here, and that the models, particularly GCM and DEM, might be distinguished on the basis of these dependencies. Analysis of sequential dependencies did indeed reveal that the monkeys' responses on a trial depended in part on events in the preceding trial (see Appendix B, which presents the results of these analyses). Here, we considered crossing the response on trial $n-1$ with reinforcement on trial $n-1$. The response on the previous trial could

have been a left or right saccade, and reinforcement might or might not have been presented, leading to a 2×2 factorial upon which performance on trial n could be conditionalized. Sequential dependencies were modeled by allowing the bias term in each model (δ in GRT and β in GCM and DEM) to vary. Thus, each model incorporated 12 bias parameters (2 possible previous responses \times 2 previous reinforcement outcomes \times 3 experimental conditions).

Simulation Results

For GRT, the maximum likelihood parameter estimates for the global parameter σ were 1.277, 1.466, and 1.804 for Subjects MK03, MK05, and MK07, respectively. With GCM, values of 1.657, 1.409, and 1.147 were estimated for λ . For DEM, λ took on values of 0.162, 0.094, and 0.054, and γ was estimated at 10.72, 14.68, and 22.94 for the respective monkeys. Bias parameters are presented in Tables B1–B3 and discussed in Appendix B. Here, it is sufficient to make two observations. First, it was found that the bias parameters tended to be above their neutral values (i.e., $\beta > .5$ for GCM and DEM, and $\delta > 6$ for GRT). This indicates that there was an overall response bias toward right responding, regardless of events on the previous trial or the stimulus presented on the current trial. This is not surprising, since the global reinforcement probability for right saccades (i.e., regardless of the stimulus) was above .5 for the dip and crossover conditions. Second, the bias parameters, though variable, reflected the sequential dependencies observed in the monkeys' behavior. For example, in the crossover condition, a tendency was observed for MK07 to respond using the same response that was made in the previous trial; this is reflected in the bias parameter values for all of the models for MK07 in this condition.

The predictions of the models from these parameter values, averaging across sequential dependencies, are shown in Figures 3–5. The quantitative predictions of the models conform with the qualitative predictions illustrated in Figure 1. GCM gave a relatively poor fit of the data. Notably, GCM was unable to produce the monotonicity observed in the lower light separations. It predicted an extensive dip in response probabilities for the dip and crossover conditions and also tended to predict overly shallow response functions for the quasi-linear condition. DEM produced the same qualitative trends as GCM, but gave noticeably better fits of the data. This model did show roughly monotonic responding at lower light separations in the dip condition, but at the expense of an overestimation of response probabilities for higher light separations. Also, DEM predicted nonmonotonicity in responding at the lower light separations for the crossover condition. In contrast, GRT produced sigmoid-shaped response functions for all monkeys and conditions, and thus captured the monotonicity apparent in the monkeys' saccade probabilities at lower light separations.

Visual inspection of the model fitting results provides convincing support for GRT. This support is confirmed by examination of the goodness-of-fit values for the 3 monkeys under the four sequential dependencies. Table 1 pre-

sents log likelihood estimates for all of the models, which were converted to Akaike information criterion (AIC; Akaike, 1973, 1974) scores; such scores trade off quantitative fit against the number of parameters to select the best and simplest model. To measure the relative goodness of fit of each model, the rightmost column of Table 1 gives the difference in AIC (Δ AIC) between each model and the best-fitting model for a particular condition (a value of 0 indicates that a model was the best-fitting model for that condition). Although AIC scores can be open to misinterpretation (see Wagenmakers & Farrell, 2004, for a discussion), these AIC differences are large enough to be considered nonarbitrary (see, e.g., Burnham & Anderson, 2002, p. 70). In all cases, the GRT model gave a superior account of the data over both GCM and DEM. Thus, the modeling results provide unequivocal support for the GRT model.

DISCUSSION

Model fitting revealed that overall, the GRT model of perceptual categorization outperformed both exemplar models examined. However, before accepting these results as convincing evidence for GRT, several potential objections regarding the quantitative fits need to be addressed. One is that when fitting the exemplar models, each experimental condition was treated as separate and independent. That is, it was assumed that exposure to each new condition resulted in the accumulation of a new set of exemplars in GCM and DEM, so that exemplars were not retained between conditions. This is not a major consideration, however, since we were concerned with steady-state behavior in each condition and generally focused on behavior at the end of a block of sessions for a particular condition.

Nonetheless, it might still be argued that if carryover of exemplars were taken into account, the exemplar models, particularly GCM, would give a better fit in the

Table 1
Goodness-of-Fit Table for General Recognition Theory (GRT), the Generalized Context Model (GCM), and the Deterministic Exemplar Model (DEM)

	$\ln L$	AIC	Δ AIC
GRT ($n = 13$)			
MK03	-2,095.51	4,217.02	0.00
MK05	-3,187.37	6,400.74	0.00
MK07	-3,067.69	6,161.38	0.00
GCM ($n = 13$)			
MK03	-4,034.05	8,094.10	3,877.08
MK05	-4,159.63	8,345.26	1,944.52
MK07	-3,694.21	7,414.42	1,253.04
DEM ($n = 14$)			
MK03	-2,878.19	5,784.38	1,567.36
MK05	-3,342.64	6,713.28	312.54
MK07	-3,096.37	6,220.74	59.36

Note— n for each model indicates the total number of free parameters. The last column gives the difference in Akaike information criterion (AIC) between each model and the best-fitting model. $\ln L$, log maximum likelihood.

crossover condition, because the exemplars in the critical region from the earlier conditions would add to the similarity value for the right saccade response (b in Equations 6 and 7) and increase right saccade responding in the critical region in these conditions. However, examination of performance in the later sessions of the experiment argues against this possibility. As mentioned earlier, later sessions also involved a fourth training rule in which the monkeys were trained on a reversed quasi-linear condition. The monkeys' ability to learn this new function argues against overlearning of the quasi-linear and dip functions as an explanation for the monotonic performance in the crossover condition. More telling for the possibility of exemplar carryover is the performance of the monkeys on sessions involving the crossover reinforcement function that immediately followed sessions with the reversed rule. In these cases, the exemplars from the reverse condition would be expected to be strong because of their recent presentation, so exemplars from the reverse condition should give a boost to the summed similarity for a left saccade in Equations 6 and 7. However, even under this sequence of trials, the monkeys still showed no lowering of response probabilities in the critical region. Together, these results suggest that carryover effects due to trial ordering cannot be the sole mechanism for the behavior we witnessed in the monkeys.

A final possibility is that exemplar models with learned attentional shifting, such as ALCOVE (Kruschke, 1992), might give a better account of the data here, given previous investigations of attentional shifting in animal discrimination and categorization (Fagot et al., 1998). The attentional learning in ALCOVE, however, is irrelevant here; since stimuli varied along only a single dimension, shifting of attention between dimensions could have no role. However, Kalish and Kruschke (2000) recently explored an extension of ALCOVE, called CORNER, in which attention was shifted between stimulus values along a particular dimension, as well as between dimensions. CORNER offers the possibility of handling the present results, since it includes learning to shift attention to or from exemplars surrounding the critical region in order to increase maximizing in all conditions. However, exploratory simulations revealed that this model also made predictions that were not consistent with the data from the monkeys; in fact, CORNER predicted sharpening of the response function at points along the stimulus dimension where the probabilities rose or dropped. For example, in the dip condition, where the reinforcement probabilities dropped from near 1.0 to .65, the CORNER model predicted that probabilities would drop from .9 to .4, which is clearly not consistent with the performance of the monkeys.

We now turn to the implications of the modeling results for the theoretical issues we have been addressing. One obvious conclusion is that given the superior fits from GRT, the monkeys appear to have partitioned the stimulus space into equivalence classes and responded to the stimuli on the basis of their position with respect to the boundary. This is not a trivial implication, since it indicates that verbal formulation of rules is not necessary for acquisition

of perceptual categories to take place (cf. Ashby, Alfonso-Reese, Turken, & Waldron, 1998). As Herbranson et al. (1999) pointed out, the finding that even pigeons can use nonlinear boundaries in a multidimensional categorization task supports the idea that the nature of the rules in theories such as GRT is basic and perceptual. However, there do appear to be limits to such learning. Smith et al. (2004), in a recent replication of Shepard, Hovland, and Jenkins's (1961) well-known category learning experiment, observed that rhesus monkeys could not learn an exclusive-or (xor) category membership rule. Under such conditions, monkeys' performance appeared to be consistent with cue-conditioning and stimulus-generalization theories, but humans appeared to make use of explicit verbal rules.

Another outcome relates to the impressive fits of DEM. It is not immediately apparent how the addition of the single parameter γ should lead to extensively better fits for DEM over GCM. Inspection of the maximum likelihood parameter estimates for DEM reveals that the estimated value of γ was quite large for all 3 monkeys, with a maximum of 22.94 for MK07. This indicates that the processing in the model approximated deterministic rule-based use (see Ashby & Maddox, 1993), which led to consistent responding in the model. That is, whenever the summed similarity for a right saccade was higher than that for a left saccade, the model showed a strong tendency to respond with a right saccade. However, this does not explain why DEM did not respond in the wrong direction in the critical region for the crossover condition: The reinforcement probabilities here were below .5, so left saccade summed similarities should have been greater than those of the right saccade, and consistency in responding should have led to right saccade probabilities being near 0. An explanation lies in the fitted values of λ for DEM; these were small relative to those for GCM, meaning that the similarity gradients for DEM were very flat in comparison with the gradients for GCM. Because of this, in DEM there was some similarity between a stimulus and even distant exemplars. Thus, for the crossover condition, the model could actually get a greater summed similarity for a right saccade in the critical region because of the contribution of exemplars of light separations 4, 5, and 6, which provided support for a right saccade. Note, however, that this improved fit in the critical region came at a cost for the model. In particular, it led the model to overpredict right saccade probabilities at the highest light separation values, since those stimuli overlapped highly with the right saccade exemplars at light separations 4–6.

In summary, GRT gave superior fits for all 3 monkeys. The large disparity in fit between GCM and DEM also suggests that a large contributor to performance of the models was the extent to which they incorporated consistency in responding.

GENERAL DISCUSSION

The performance of 3 monkeys was examined on a one-dimensional categorization task in which popular theories

of categorization predicted either deterministic or more variable responding. The results of the experiment were consistent with the qualitative predictions from decision bound theories of categorization and did not display the variability in responding that was qualitatively predicted by exemplar theory. Quantitative application of a decision bound theory, GRT (Ashby, 2000; Ashby & Gott, 1988), and of two exemplar theories, GCM (Nosofsky, 1986) and DEM (Ashby & Maddox, 1993; see also Nosofsky, 1991; Nosofsky & Palmeri, 1997), gave overall support to GRT.

One interesting result is that the boundary-based solution found by the monkeys cannot achieve maximal accuracy. There are two likely reasons for this outcome. A methodological reason relates to the fact that the monkeys received some training on sessions for which a single boundary was sufficient before being trained on the crossover condition. Although we argued that there was little retention of exemplars between conditions, the monkeys may have developed some qualitative rigidity in their approach to the task by continuing to use only a single boundary. It is possible that the costs of learning an additional boundary did not warrant the small increase in probability of reinforcements that would have resulted. From the perspective of optimality, an additional factor not considered here was the payoff matrix for the task. In this task, a sip of juice was the reward for a "correct" response (i.e., a response probabilistically determined to be correct on that trial). It might be that making the reward greater, or making the consequences of an incorrect response aversive, could induce use of a second rule in the monkeys.

Another possibility is that this result reflects basic limits in the capacity of the monkeys to learn the task. However, there is reason to doubt this explanation. Herbranson et al. (1999) showed that pigeons were able to learn nonlinear boundaries in a multidimensional categorization task, closely approximating the rule use found in humans (Ashby & Lee, 1992; Ashby & Maddox, 1992). On this basis, we would expect higher order animals such as monkeys to be able to learn more than one of the simple rules in order to perform optimally. Given the exploratory nature of the present experiment, this question is left as an open one.

Relation to Previous Animal Studies

As mentioned previously, several studies have considered application of theories of human categorization to animals. One conclusion from these previous studies was that monkeys and baboons appear not to rely on prototypes when performing typical categorization tasks (Dépy et al., 1997; Sigala et al., 2002). Although we did not examine prototype models here, such models would be expected to perform more poorly than GCM. Both exemplar and decision bound theories have found support in previous animal studies (Dépy et al., 1997; Medin & Dewey, 1984; Shimp et al., 1996; Sigala et al., 2002); in particular, Sigala et al. (2002), who directly compared the fits of exemplar and decision bound models to monkeys' categorization of two-dimensional stimuli, found that categorization was performed both on the basis of similarity of stimuli to

exemplars and with respect to distance of the stimuli from a putative objective boundary. Sigala et al. examined in detail the performance of only 1 monkey, and their data indicated the use of both exemplars and boundaries within the monkey. Although it might be that monkeys can use both exemplars and decision boundaries, the mimicking observed between exemplar and decision bound theories prevents firm conclusions from being drawn.

Relation to Human Studies

Rouder and Ratcliff (2004) applied an experimental paradigm very similar to the one employed here to examining the performance of humans on a light/dark discrimination task with probabilistic feedback, and they found evidence supporting both exemplar and decision bound models. Although in general GRT gave a better fit across their experiments, Rouder and Ratcliff found that DEM appeared to be superior in cases in which the stimuli were not confusable. On this basis, Rouder and Ratcliff suggested that humans might alternate in their use of exemplars or decision bounds to perform a categorization task, depending on the discriminability of stimuli (and thus the usefulness of individual exemplars); such a possibility has been considered in other models of human categorization (Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994). Although application of a hybrid exemplar–decision bound model might be possible with the present results, such models have not received a great deal of theoretical development and suffer from problems of parameter identifiability (Bamber & van Santen, 2000); given the mimicking problem that motivated the use of our paradigm, fitting a hybrid model would not have allowed us to accurately identify the contribution of individual components in such a model. We therefore chose to restrict our attention to simple, interpretable models. The validity of our modeling results is highlighted by the consistency of their superior fit of decision bound theory with the modeling results of Rouder and Ratcliff.

Theoretical Implications

The main conclusion from the experiment and the modeling results is that monkeys partition perceptual space in order to perform perceptual categorization tasks of the type discussed here. However, the large disparity in fit between GCM and DEM is also highly suggestive. In particular, it implies consistency in responding, regardless of the underlying representations employed; the deterministic responding in GRT was likely a major contributor to the relative success of that model. Other researchers have drawn similar conclusions when quantitatively comparing this set of models with respect to human categorization performance (Maddox & Ashby, 1993).

Conclusions

The experiment and model fitting presented in this article demonstrate that quantitative modeling of perceptual categorization by animals can complement research in human categorization. This research also offers extensions to a domain wider than categorization probabilities. Along

with research on response times in categorization (Ashby, 2000; Lamberts, 1998; Nosofsky & Palmeri, 1997; Ratcliff et al., 2003) and investigations relating categorization performance and neural activity (Ashby & Ell, 2001; Ratcliff et al., 2003; Sigala & Logothetis, 2002), the present study suggests the potential for theoretically meaningful integration and extension of research on the processes underlying categorization.

REFERENCES

- AKAIKE, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267-281). Budapest: Akademiai Kiado.
- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.
- ANDERSON, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, **98**, 409-429.
- ASHBY, F. G. (Ed.) (1992a). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.
- ASHBY, F. G. (1992b). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, **44**, 310-329.
- 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., BOYNTON, G., & LEE, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics*, **55**, 11-27.
- ASHBY, F. G., & ELL, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, **5**, 204-210.
- ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.
- ASHBY, F. G., & LEE, W. W. (1992). On the relationship among identification, similarity, and categorization: Reply to Nosofsky and Smith (1992). *Journal of Experimental Psychology: General*, **121**, 385-393.
- ASHBY, F. G., & MADDOX, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 50-71.
- ASHBY, F. G., & MADDOX, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.
- ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.
- ASTLEY, S. L., PEISSIG, J. J., & WASSERMAN, E. A. (2001). Superordinate categorization via learned stimulus equivalence: Quantity of reinforcement, hedonic value, and the nature of the mediator. *Journal of Experimental Psychology: Animal Behavior Processes*, **27**, 252-268.
- ASTLEY, S. L., & WASSERMAN, E. A. (1992). Categorical discrimination and generalization in pigeons: All negative stimuli are not created equal. *Journal of Experimental Psychology: Animal Behavior Processes*, **18**, 193-207.
- ASTLEY, S. L., & WASSERMAN, E. A. (1999). Superordinate category formation in pigeons: Association with a common delay or probability of food reinforcement makes perceptually dissimilar stimuli functionally equivalent. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 415-432.
- BAMBER, D., & VAN SANTEN, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, **44**, 20-40.
- BHATT, R. S., WASSERMAN, E. A., REYNOLDS, W. F., JR., & KNAUSS, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 219-234.
- BURNHAM, K. P., & ANDERSON, D. R. (2002). *Model selection and multimodel interference: A practical information-theoretic approach* (2nd ed.). New York: Springer.
- BUSEMEYER, J. R., & MYUNG, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General*, **121**, 177-194.
- DÉPY, D., FAGOT, J., & VAUCLAIR, J. (1997). Categorisation of three-dimensional stimuli by humans and baboons: Search for prototype effects. *Behavioural Processes*, **39**, 299-306.
- DIAS, E. C., & SEGRAVES, M. A. (1999). Muscimol-induced inactivation of monkey frontal eye field: Effects on visually and memory-guided saccades. *Journal of Neurophysiology*, **81**, 2191-2214.
- ELIASON, S. R. (1993). *Maximum likelihood estimation: Logic and practice*. Newbury Park, CA: Sage.
- ERICKSON, M. A., & KRUSCHKE, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, **127**, 107-140.
- ESPINOZA-VARAS, B., & WATSON, C. S. (1994). Effects of decision criterion on response latencies of binary decisions. *Perception & Psychophysics*, **55**, 190-203.
- ESTES, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, **57**, 94-107.
- ESTES, W. K. (1992). Models of categorization and category learning. In D. L. Medin, R. Taraban, & G. Nakamura (Eds.), *Psychology of learning and motivation: Special volume on categorization* (pp. 15-56). San Diego: Academic Press.
- ESTES, W. K., BURKE, C. J., ATKINSON, R. C., & FRANKMANN, J. P. (1957). Probabilistic discrimination learning. *Journal of Experimental Psychology*, **54**, 233-239.
- FAGOT, J., KRUSCHKE, J. K., DÉPY, D., & VAUCLAIR, J. (1998). Associative learning in baboons (*Papio papio*) and humans (*Homo sapiens*): Species differences in learned attention to visual features. *Animal Cognition*, **1**, 123-133.
- GLUCK, M. A., & BOWER, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 227-247.
- GRAF, V., 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. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- HERBRANSON, W. T., FREMOUW, T., & SHIMP, C. P. (1999). The randomization procedure in the study of categorization of multidimensional stimuli by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 113-135.
- HERRNSTEIN, R. J., LOVELAND, D. H., & CABLE, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **2**, 285-302.
- KALISH, M. L., & KRUSCHKE, J. K. (1997). Decision boundaries in one-dimensional categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1362-1377.
- KALISH, M. L., & KRUSCHKE, J. K. (2000). The role of attention shifts in the categorization of continuous dimensioned stimuli. *Psychological Research*, **64**, 105-116.
- KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.
- KULLBACK, S., & LEIBLER, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.
- LAMBERTS, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 695-711.
- LUCE, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103-189). New York: Wiley.
- 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. (1996). Perceptual separability, decision separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 795-817.
- MADDOX, W. T., ASHBY, F. G., & GOTTLÖB, L. R. (1998). Response time

- distributions in multidimensional perceptual categorization. *Perception & Psychophysics*, **60**, 620-637.
- MADDOX, W. T., & BOHIL, C. J. (2001). Feedback effects on cost-benefit learning in perceptual categorization. *Memory & Cognition*, **29**, 598-615.
- MEDIN, D. L., & DEWEY, G. I. (1984). Learning of ill-defined categories by monkeys. *Canadian Journal of Psychology*, **38**, 285-303.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.
- MEYER, D. R. (1960). The effects of differential probabilities of reinforcement on discrimination learning by monkeys. *Journal of Comparative & Physiological Psychology*, **53**, 173-175.
- NASH, A. (1970). Response probabilities for individual stimuli in learning tasks. *Psychonomic Science*, **19**, 5-6.
- NELDER, J. A., & MEAD, R. (1965). A simplex method for function minimization. *Computer Journal*, **7**, 308-313.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOSOFSKY, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 3-27.
- NOSOFSKY, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363-393). Hillsdale, NJ: Erlbaum.
- NOSOFSKY, R. M., & ALFONSO-REESE, L. A. (1999). Effects of similarity and practice on speeded classification response times and accuracies: Further tests of an exemplar-retrieval model. *Memory & Cognition*, **27**, 78-93.
- NOSOFSKY, R. M., & PALMERI, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.
- NOSOFSKY, R. M., PALMERI, T. J., & MCKINLEY, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.
- OVERALL, J. E., & BROWN, W. L. (1963). Responses of rhesus monkeys to probabilistic sequential dependencies. *Journal of General Psychology*, **69**, 69-74.
- RATCLIFF, R., CHERIAN, A., & SEGRAVES, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of Neurophysiology*, **90**, 1392-1407.
- RATCLIFF, R., & ROUDER, J. F. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347-356.
- RATCLIFF, R., THAPAR, A., & MCKOON, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology & Aging*, **16**, 323-341.
- RATCLIFF, R., VAN ZANDT, T., & MCKOON, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, **106**, 261-300.
- REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.
- REED, S. K. (1978). Category vs. item learning: Implications for categorization models. *Memory & Cognition*, **6**, 612-621.
- ROBBINS, D., & WARNER, P. L. (1973). Individual organism probability matching with rats in a two-choice task. *Bulletin of the Psychonomic Society*, **2**, 405-407.
- ROUDER, J. N., & RATCLIFF, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, **133**, 63-82.
- SEGRAVES, M. A. (1992). Activity of monkey frontal eye field neurons projecting to oculomotor regions of the pons. *Journal of Neurophysiology*, **68**, 1967-1985.
- SHEPARD, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, **22**, 325-345.
- SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.
- SHEPARD, R. N., HOVLAND, C. I., & JENKINS, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, **75**(13, Whole No. 517).
- SHIMP, C. P. (1973). Probabilistic discrimination learning in the pigeon. *Journal of Experimental Psychology*, **97**, 292-304.
- SHIMP, C. P., LONG, K. A., & FREMOUW, T. (1996). Intuitive statistical inference: Categorization of binomial samples depends on sampling context. *Animal Learning & Behavior*, **24**, 82-91.
- SIGALA, N., GABBIANI, F., & LOGOTHETIS, N. K. (2002). Visual categorization and object representation in monkeys and humans. *Journal of Cognitive Neuroscience*, **14**, 187-198.
- SIGALA, N., & LOGOTHETIS, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, **415**, 318-320.
- SMITH, J. D., MINDA, J. P., & WASHBURN, D. A. (2004). Category learning in rhesus monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General*, **133**, 398-414.
- SPINOZZI, G. (1996). Categorization in monkeys and chimpanzees. *Behavioural Brain Research*, **74**, 17-24.
- STEINER, J. (1970). Observing responses and uncertainty reduction: II. The effect of varying the probability of reinforcement. *Quarterly Journal of Experimental Psychology*, **22**, 592-599.
- STEWART, N., BROWN, G. D. A., & CHATER, N. (2002). Sequence effects in the categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 3-11.
- TREICHLER, F. R., CONNER, J. B., & RICCIARDI, A. M. (1970). Probability learning and reversal by experimentally naive rhesus and squirrel monkeys. *Learning & Motivation*, **1**, 411-419.
- WAGENMAKERS, E.-J., & FARRELL, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, **11**, 192-196.
- WICKENS, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.
- WILSON, W. A., JR., & OSCAR, M. (1966). Probability learning in monkeys with lateral frontal lesions. *Journal of Comparative & Physiological Psychology*, **62**, 462-464.
- WILSON, W. A., JR., OSCAR, M., & BITTERMAN, M. E. (1964a). Probability learning in the monkey. *Quarterly Journal of Experimental Psychology*, **16**, 163-165.
- WILSON, W. A., JR., OSCAR, M., & BITTERMAN, M. E. (1964b). Visual probability-learning in the monkey. *Psychonomic Science*, **1**, 71-72.
- YOUNG, F. W., & HAMER, R. M. (1987). *Multidimensional scaling: History, theory, and applications*. Hillsdale, NJ: Erlbaum.

NOTE

1. This selection procedure was employed, rather than simply examining terminal performance, because the monkeys were variable in their behavior from session to session. Even close to the end of a condition block, in some sessions monkeys still showed close-to-chance performance or made large numbers of ambiguous responses. Our selection procedure produced data representative of terminal performance with controlled variability.

APPENDIX A
Details of Model Fitting

Models were fit to the data using maximum likelihood estimation. Maximum likelihood estimation is a powerful method for parameter estimation (Maddox & Ashby, 1993) because it produces estimators that are consistent (meaning that the probability of the difference between the estimator and the real parameter value being above some small value is 0 in the limit of sample size). Estimators that are asymptotically efficient (i.e., that have the smallest variance) will also be found if they exist (Eliason, 1993; Wickens, 1982). The likelihood for a series of responses given a model for a particular condition c is

$$L_c(r_1, r_2, \dots, r_n) = \prod_{i=1}^n p(A|I_i)^{1-r_i} p(B|I_i)^{r_i}, \tag{A1}$$

where I_i denotes the stimulus presented on trial i and r_i denotes the response made on trial i ; $r_i = 0$ if Response A was made and $r_i = 1$ if Response B was made (see, e.g., Ashby, 1992b; Maddox & Ashby, 1993). The total likelihood across the conditions is given by multiplying together the likelihoods for each condition. For the simulations, Equation A1 was adapted to give the log likelihood for condition c as

$$\ln L_c = \sum_{i=1}^k \exp[F(A|I_i)] \ln[1 - p(A|I_i)] + \exp[F(B|I_i)] \ln p(B|I_i), \tag{A2}$$

where $F(A|I_i)$ and $F(B|I_i)$ are the respective empirical frequencies of left or right saccades given in response to stimulus I_i across all three conditions (quasi-linear, dip, and crossover).^{A1}

An advantage of using maximum likelihood is that it is possible to compare models on goodness of fit while correcting for differences in number of free parameters used to fit the data. Akaike (1973, 1974) related maximum likelihood to a measure of information, the Kullback–Leibler distance (Kullback & Leibler, 1951), which measures the information lost when one model is used to approximate another; in this case, the model being approximated is the unknown process underlying the monkeys’ data. Akaike (1973) showed that an unbiased estimate of maximum likelihood should take into account the number of parameters between models. This corrected measure, known as the Akaike information criterion, is given by

$$AIC_i = -2 \ln L_i + 2K_i, \tag{A3}$$

where K_i is the number of parameters in the model and $\ln L_i$ is the log maximum likelihood obtained by summing the log likelihoods.

The simplex minimization algorithm of Nelder and Mead (1965) was used for each model to find the parameter values that maximized the likelihood of obtaining the data from that model. To ensure that a global maximum was found, the simplex routine was run 100 times per model, each run using a different vector of starting parameters, and the highest maximum likelihood across the runs was taken as the best solution.

NOTE

A1. In cases where the models predicted that response proportions were at zero or unity, these proportions were arbitrarily set to 10^{-8} and $(1 - 10^{-8})$, respectively, to allow the log transform in Equation A2.

APPENDIX B
Sequential Dependencies in Data and Models

The empirical results of the sequential effect analyses for all monkeys are presented in Figures B1–B3. Each panel shows four right saccade probability functions, each curve being the response function conditionalized on some combination of (1) the response on the previous trial and (2) whether or not reinforcement was received on the previous trial. It is evident from visual inspection that the monkeys do not all display the same sequential dependencies. The four response functions are mostly indistinguishable in MK03 (Figure B1), indicating that this monkey’s responses were not systematically affected by the response or reinforcement on the previous trial, but were only dependent on the stimuli on the current trial. MK05 (Figure B2) displays clearer sequential dependencies: Reinforcement of the previous response appears to decrease the probability of a right saccade on the following trial in the quasi-linear condition. MK05 also shows some evidence of sequential dependencies in the dip and crossover conditions, but these effects are not consistent across all light separations. For MK07 (Figure B3), there is clear evidence of sequential dependencies in the crossover condition. For this condition, MK07 appears to perseverate responses, so that a right saccade was more likely to be followed by another right saccade. For the other two experimental conditions, however, this subject displays no systematic effects across stimulus level.

APPENDIX B (Continued)

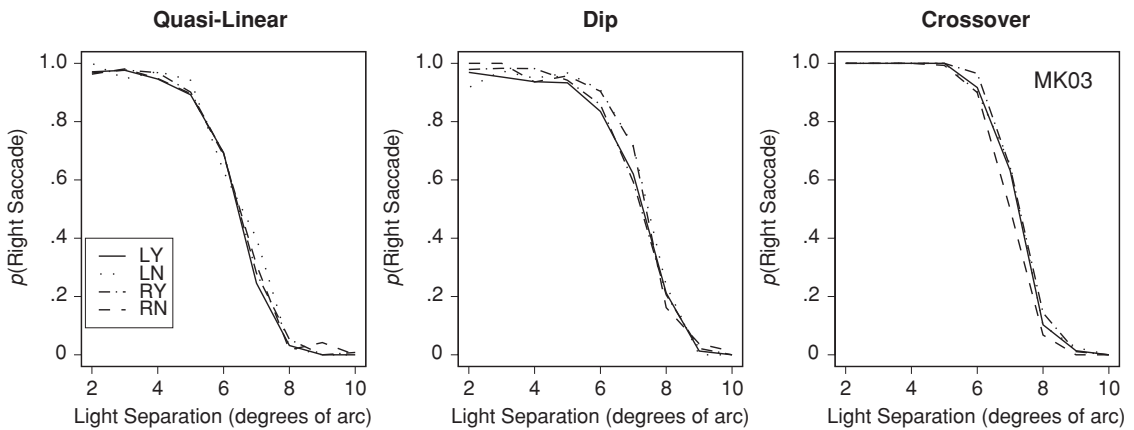


Figure B1. Proportion of right saccades for Subject MK03, conditionalized on the response and reinforcement for the previous trial. The first letter in the figure legend indicates the previous response (L, left saccade; R, right saccade), and the second letter indicates whether the previous response was reinforced (Y [yes], response was reinforced; N [no], response was not reinforced).

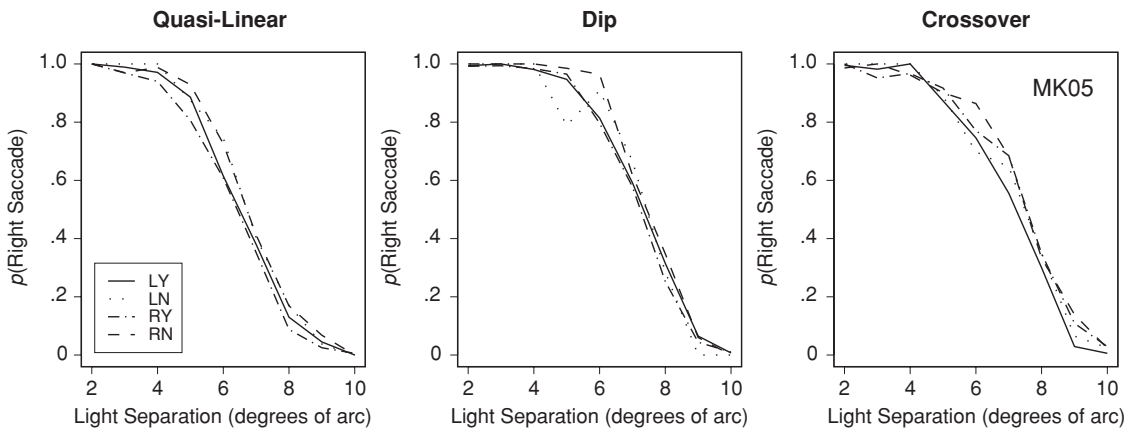


Figure B2. Proportion of right saccades for Subject MK05, conditionalized on the response and reinforcement for the previous trial. For an explanation of the figure legend, see the caption to Figure B1.

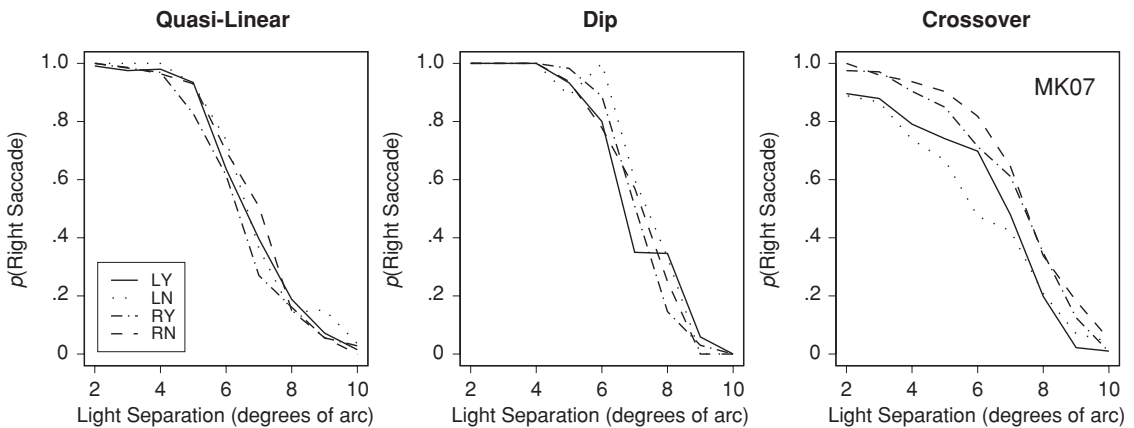


Figure B3. Proportion of right saccades for Subject MK07, conditionalized on the response and reinforcement for the previous trial. For an explanation of the figure legend, see the caption to Figure B1.

APPENDIX B (Continued)

Tables B1, B2, and B3 present the estimated bias parameters for MK03, MK05, and MK07, respectively. A higher bias parameter corresponds to a bias to respond with a right saccade. Overall, the patterns of bias parameters confirm the sequential dependencies identified visually from Figures B1–B3. In the quasi-linear condition, the bias parameters for GCM and GRT show no systematic pattern (see the first column of each table), but the parameters for DEM suggest a switching strategy in which the monkey alternated responses, since a right response was more likely if the previous response had been a left saccade. The results from the three models for the dip condition reveal some interesting patterns. For GRT, it appears that right saccade probabilities were increased following reinforcement of a right saccade or nonreinforcement of a left saccade. For GCM, there were no real differences in bias parameters between conditionalizations. However, for DEM there were clear effects of conditionalization on bias, since reinforcement of any response appears to have decreased right saccade probabilities on the following trial. Finally, for the crossover condition, all three models suggest that the monkeys (particularly MK05 and MK07) perseverated responding, so that they repeated responses regardless of whether or not they were reinforced.

Table B1
Bias Parameter Values for Each Model for Fits to Sequential Dependencies in Subject MK03

Reinforced?	Quasi-Linear		Dip		Crossover	
	Left	Right	Left	Right	Left	Right
GRT						
Yes	6.11	6.35	6.77	7.04	7.12	7.30
No	6.38	6.25	7.07	6.97	7.20	6.84
GCM						
Yes	0.506	0.599	0.674	0.662	0.684	0.720
No	0.538	0.527	0.738	0.588	0.664	0.664
DEM						
Yes	0.626	0.655	0.379	0.444	0.741	0.772
No	0.548	0.162	0.904	0.699	0.487	0.832

Table B2
Bias Parameter Values for Each Model for Fits to Sequential Dependencies in Subject MK05

Reinforced?	Quasi-Linear		Dip		Crossover	
	Left	Right	Left	Right	Left	Right
GRT						
Yes	6.50	6.29	7.14	6.29	6.91	7.25
No	6.74	6.75	6.74	7.42	7.17	7.25
GCM						
Yes	0.584	0.585	0.688	0.668	0.701	0.713
No	0.596	0.614	0.705	0.656	0.674	0.730
DEM						
Yes	0.695	0.588	0.324	0.304	0.764	0.799
No	0.695	0.203	0.855	0.798	0.319	0.880

Table B3
Bias Parameter Values for Each Model for Fits to Sequential Dependencies in Subject MK07

Reinforced?	Quasi-Linear		Dip		Crossover	
	Left	Right	Left	Right	Left	Right
GRT						
Yes	6.62	6.49	6.95	7.21	5.97	7.01
No	6.93	6.75	7.24	6.84	5.87	7.28
GCM						
Yes	0.606	0.602	0.648	0.646	0.580	0.645
No	0.626	0.582	0.716	0.589	0.443	0.718
DEM						
Yes	0.719	0.625	0.123	0.134	0.752	0.836
No	0.818	0.094	0.849	0.814	0.036	0.861