Journal of Experimental Psychology:
Learning, Memory, and Cognition
1988, Vol. 14, No. 1, 33–53

0278-7393/88/$00.75

Decision Rules in the Perception and Categorization of Multidimensional Stimuli

F. Gregory Ashby
University of California, Santa Barbara

Ralph E. Gott
Ohio State University

This article examines decision processes in the perception and categorization of stimuli constructed from one or more components. First, a general perceptual theory is used to formally characterize large classes of existing decision models according to the type of decision boundary they predict in a multidimensional perceptual space. A new experimental paradigm is developed that makes it possible to accurately estimate a subject's decision boundary in a categorization task. Three experiments using this paradigm are reported. Three conclusions stand out: (a) Subjects adopted deterministic decision rules, that is, for a given location in the perceptual space, most subjects always gave the same response; (b) subjects used decision rules that were nearly optimal; and (c) the only constraint on the type of decision bound that subjects used was the amount of cognitive capacity it required to implement. Subjects were not constrained to make independent decisions on each component or to attend to the distance to each prototype.

Whenever an organism voluntarily responds to a stimulus, some decision process must evaluate the available perceptual and cognitive information and select an appropriate response. Decision processes are therefore of fundamental importance in all areas of psychology. In fact, even if one's interest is at some other level, an understanding of decision processes is vital. For example, most experimental paradigms that study perceptual processes in human subjects require voluntary overt responding and thus involve decision as well as perceptual processing. These two separate processes are not usually directly observable, but both must be understood before a complete description of human behavior in perceptual tasks is possible.

In this article, we study decision processes in perception. Specifically, we focus on decision rules in the perception and categorization of stimuli constructed from one or more stimulus components. Many such rules have been described in the literature. One difficulty in testing between them is that they are described in the languages of different theoretical perspectives and so are difficult to compare. We begin by describing a very general perceptual theory that enables us to characterize virtually all possible decision rules in a way that makes them simple to compare. We then introduce an experimental procedure that allows direct and detailed observation of the decision process in a categorization task. When the decision process is observable, testing different decision models becomes relatively simple.

To illustrate the application of this procedure, three experiments will then be described. These experiments required subjects to categorize simple geometric figures constructed from lines of varying length. The decision rules adopted by subjects in these experiments turned out to be remarkably efficient. Perhaps the most surprising outcome, however, was that we obtained strong evidence of deterministic responding. For a given perceptual effect, subjects appeared never to guess. This result contradicts may alternative decision rules that have been described in the literature.

Overview of Models of the Decision Process

Depending on the way components of stimuli are thought to be processed before a response is made, perceptual theories or models can be categorized into two major classes (Shaw, 1982): (a) independent decisions models and (b) information integration models. Independent decisions models assume a two-stage decision process. In the first stage, a separate decision is made about the presence of each stimulus component. The second stage uses the results of these decisions to select a response. This type of model enjoys great popularity. For example, the idea of independent decisions forms the basis of high-threshold models (e.g., Blackwell, 1963) as well as feature-analytic models of letter recognition (e.g., Townsend & Ashby, 1982). In feature-analytic models, separate decisions are first made about the presence or absence of each particular visual feature. In the next stage, the list of perceived features is compared to a stored list of the features known to be contained in each letter. A response is selected on the basis of this matching process.

Information integration models assume a one-stage decision process. The total perceptual information is combined in some fashion and a response is selected by examining (i.e., processing) the integrated information. Typically, the integrity

Parts of this research were presented at the 1983 Meetings of the Midwestern Psychological Association, Chicago, Illinois, and at the Seventeenth Annual Mathematical Psychology Meetings at the University of Chicago. Some of this research was conducted at Ohio State University and some was conducted while F. Gregory Ashby was visiting Harvard University under National Science Foundation Grant 606 7517/2. He is grateful to W. K. Estes for the opportunity the year provided. Both authors would like to thank Jerry Busemeyer and Rob Nosofsky for helpful comments.

Correspondence concerning this article should be addressed to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, California 93106.

or vitality of the perceptual representation of each stimulus component is assumed to be continuous. Thus, for example, on some trials weak evidence might be obtained that component *X* was contained in the stimulus. Instead of making an immediate decision about the presence of component *X* (as in independent decisions models), integration models assume this information is collated with similar information about the other components and then a response is selected on the basis of the entire ensemble of perceptual information.

The assumption that subjects optimize performance in some fashion, for example, by maximizing response accuracy, generally implies information integration. As we will see, in certain special circumstances, independent decisions are optimal, but optimal responding across a variety of experimental paradigms can only be predicted by certain integration models. Thus, for example, ideal observer models in signal detection theory (e.g., Green & Swets, 1966) all assume information integration.

Another large class of integration models comes from multidimensional scaling (MDS; e.g., Torgerson, 1958). In a typical MDS task, subjects are presented with pairs of stimuli and are asked to respond with a numerical rating of their perceived similarity. These ratings are then used to produce a geometric representation in which each stimulus is identified with a point in a multidimensional perceptual space and with the property that similarity and psychological distance are inversely related. The decision rule implicit in MDS models assumes that subjects base their similarity judgements on the distance between the raw perceptual representations of two stimuli. The process of computing a distance metric is one way of integrating information from the various perceptual dimensions.

Still other examples of information integration can be found in the distributed memory models that are based upon ideas of matched filtering, cross correlation, autocorrelation, or convolution (e.g., Hinton & Anderson, 1981; Murdock, 1982). In these models the perceptual sample is compared with a list of stored representations of the alternative stimuli via a process such as cross correlation. The response is determined by the alternative that correlates most highly with the input. Distributed memory models frequently assume that the stimulus is sampled at different time points. In our language, the different samples can be thought of as stimulus components. This kind of representation seems especially appropriate in cases of dynamic stimulus presentation, in which the characteristics of the stimulus change over time. Information in the time varying perceptual repesentations is integrated via the correlation or convolution operation.

This brief survey reveals that many models of the decision process are very differently formulated and so are difficult to compare, both analytically and empirically. For example, how can we decide between the independent decisions of a feature-analytic model and the minimum distance rule of an MDS model? The approach we follow is to find a common language through which to interrelate the different models. The language we choose is that of a very general perceptual theory. Using this language and the experimental procedure described later makes deciding between competing decision models a fairly straightforward task.

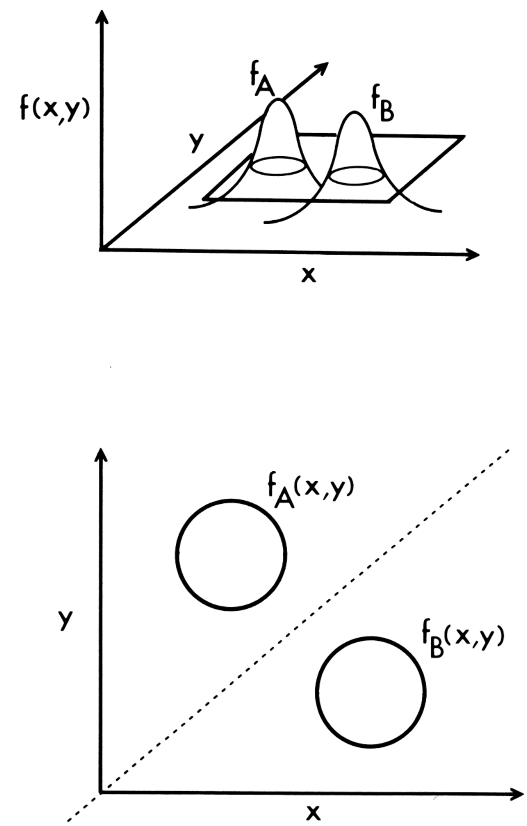

Figure 1. Top: The distributions of perceptual effects in an experiment with two stimuli, S_A and S_B, that are each composed of two stimulus components; bottom: the contours of equal probability associated with the distributions of Figure 1(top).

General Recognition Theory

The general recognition theory that motivates both our classification of decision models and the experimental procedure introduced later was originally described by Ashby and Townsend (1986). However, because our goals are different and because it may not yet be widely familiar, we briefly reintroduce the theory here.

In this article, we are specifically interested in the decision processes involved when perceptual information is available along several different dimensions, and so we shall focus on stimuli that consist of several separate components. Each of these stimulus components may vary along some physical dimension that we denote as *X*, *Y*, and so forth. On any particular trial the experimenter will present the subject with some stimulus containing one or more components and we will denote the perceptual effects of these components as *x*, *y*, and so forth.

From the perceptual effects of a particular stimulus, the subject's decision process must assign a response. We will denote this response function by r(*x*, *y*). Of course, the form of this function will depend on the type of experimental paradigm in use. For instance, in a yes–no detection task, r will assign either a yes or a no to every possible pair (*x*, *y*). It is the form of this response function that concerns us in this article.

To account for observed variability in subject performance, we assume that presentation of the same stimulus (X, Y) does not always produce the same perceptual effect (x, y), or in other words, that the perceptual effects x and y are random. Call their joint probability distribution (i.e., their density function) $f(x, y)$. Figure 1 shows an example in which the stimulus ensemble contains two stimuli, S_A and S_B. The top of Figure 1 shows the distribution of the perceptual effects when the two stimuli are presented. Note that presentation of either stimulus could generate a perceptual effect anywhere in the perceptual space. The plane cutting through the two joint density functions describes the equal probability contours of the two distributions. The bottom of Figure 1 is a view of this plane from above. Every perceptual effect associated with any point on either circle is equally like to occur (i.e., is associated with the same probability density).

Although the experimental technique described later is technically a categorization paradigm, its development was motivated by considering how the general recognition theory might account for a simple two-stimulus identification task in which both stimuli are composed of a pair of components that vary along the same two physical dimensions. In this paradigm, on each experimental trial the subject is shown either stimulus S_A or S_B. In most of the experiments reported in this article, stimuli S_A and S_B are each composed of a vertical and horizontal line segment joined at an upper left corner. The line segments in the two stimuli differ in length and it is on this basis that accurate responding is possible. Prototypical stimuli are shown in Figure 2.

Because these stimuli are always constructed from two stimulus components (a horizontal and a vertical line segment), let us assume that the perceptual space and also the contours of equal probability are two dimensional, as in Figure 3, with a dimension associated with each component. For convenience, suppose x is the dimension associated with the horizontal component, y is the dimension associated with the vertical component, and the perceptual distribution (i.e., probability density function) corresponding, say, to stimulus S_A is $f_A(x, y)$.

The presentation of any stimulus then, induces a perceptual effect (x, y). The subject's response function $r(x, y)$ uses this perceptual effect to select one of the two responses R_A and R_B. This effectively divides up the x, y space into two regions, one associated with each possible response alternative. The

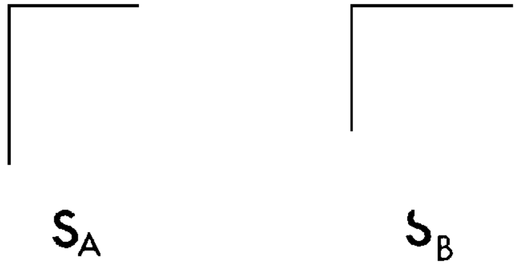

Figure 2. Prototypical stimuli from Experiments 1–3.

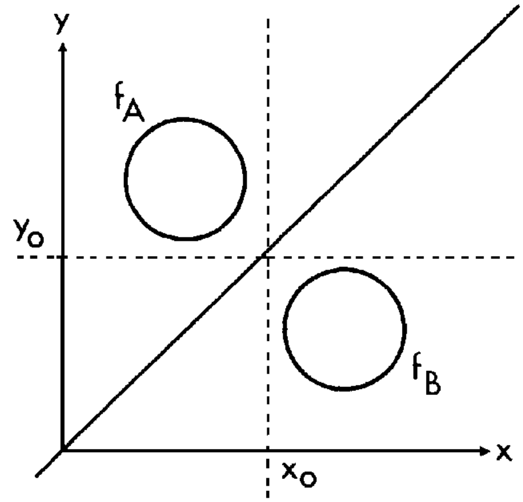

Figure 3. Contours of equal probability and two possible sets of decision bounds in a two-stimulus identification experiment.

response is determined by the region into which the perceptual effect (x, y) falls.

In Figure 3, the solid line marks one possible boundary between the two response regions. Any perceptual effect falling above the solid line elicits an R_A response and any falling below elicits an R_B. Note that a subject using this boundary will never guess, unless perhaps a perceptual sample happens to fall exactly on the decision boundary. This is not a necessary prediction of the theory. The dotted-line boundaries in Figure 3 postulate frequent guessing. A sample falling in the upper left region elicits an R_A response, a sample falling in the lower right region elicits an R_B response, but a sample falling in either other quadrant causes the subject to guess. In this model, the probability of responding say, R_A, depends only on which quadrant a sample happens to fall into and not on where a sample falls within a given quadrant.

A different way to incorporate guessing into the response process is to associate a different guessing distribution with each point in the space. For example, the probability of responding R_A might continuously decrease with the distance of the sample from the mean of the S_A perceptual distribution. Note that with this type of guessing mechanism, there is no true decision bound, although it still might be of interest to examine the $P(R_A) = .5$ contour.

The most familiar version of the general recognition theory ssumes that the perceptual distributions are multivariate normal. This special case, which Ashby and Townsend (1986) called the general Gaussian recognition model, is related to the (Case 1) model of Thurstone's law of categorical judgment (Thurstone, 1927; see also Torgerson, 1958; Hefner, 1958; Zinnes & MacKay, 1983), but it can also be viewed as a multidimensional generalization of signal detection theory (e.g., Green & Swets, 1966; Tanner, 1956; Wandell, 1982).

The contours of equal probability of a bivariate normal distribution are always ellipses or circles (as in Figure 3). The shape is determined by the variances and by the correlation or the covariance parameter. As the covariance moves away from zero the contour becomes more elliptical and begins to tilt (i.e., so that the major and minor axes of the ellipse no longer agree with the x and y coordinate axes). Similarly, the contours tend to become more elliptical as the difference in the magnitude of the two variances increases. These parameters are conveniently catalogued in a structure known as the covariance matrix, which is a symmetric matrix with variances on the major diagonal and covariances elsewhere.

The General Recognition Randomization Technique

Suppose that on every trial we are able to observe not only the stimulus and the subject's response but also the subject's perceptual representation of the stimulus (x, y). Over the course of many trials, the perceptual samples will be distributed throughout the x, y space. By noting the subject's response at each point in the space it would be very easy to test between say, the solid-line and the dotted-line decision rules of Figure 3. For example, if the subject is using the dotted-line boundaries, there should be two quadrants in the space (upper right and lower left) containing a random mixture of R_A and R_B responses.

Of course, in most experimental paradigms the perceptual representations are not observable. To reduce accuracy, large amounts of internal perceptual noise are usually induced either by limiting exposure duration through tachistoscopic presentation or by limting contrast by reducing stimulus intensity.

These techniques guarantee variability in the perceptual representations but they provide no way of measuring the perceptual noise they induce. With the stimuli of Figure 2, the general recognition theory assumes that internal noise causes the subject to somewhat misperceive the lengths of the horizontal and vertical segments of a particular stimulus, making identification difficult. In most psychophysical tasks, the degree of this misperception is unknown. In our paradigm, which we call the general recognition randomization technique, it is specified as accurately as possible by the experimenter.

To test between the solid-line and dotted-line decision rules of Figure 3, the general recognition randomization technique proceeds as follows. First a set of numerical parameter values that completely specify a pair of bivariate (normal) distributions is chosen so that the resulting contours of equal probability agree with those in Figure 3. On each experimental trial, a random sample (X_s, Y_s) is drawn from one of these two distributions. A stimulus is then constructed with horizontal length X_s and vertical length Y_s and is presented to the subject. The subject's task is to classify the stimulus sample as a member of stimulus class A or stimulus class B.

Note that with this technique the subject may never be shown the Figure 2 stimuli exactly. For example, on S_A trials the horizontal and vertical line segments of the stimulus actually shown to the subject may each be either longer or shorter than those shown in Figure 2. A stimulus like the one marked S_B in Figure 2 could therefore be a sample from either distribution, although it is more likely to have been drawn from $f_B(X, Y)$. In fact, in all of the experiments reported here, stimulus parameters were chosen so that an ideal observer could correctly classify only 80% of the stimulus samples. In addition, perceptual noise was minimized by making stimulus contrast high and by using response terminated displays.

On each trial a record is kept of the stimulus (S_A or S_B), the actual line lengths (X_s, Y_s), and the subject's response (R_A or R_B). The actual line lengths correspond to a point somewhere in the X, Y plane. The decision rule used by the subject can be determined by observing the pattern of their responses throughout the plane. Thus, if a line can be found in the plane such that the subject responds R_A to all stimulus samples falling above the line and R_B to all below it, then the solid line decision rule of Figure 3 is supported and the dotted-line rule is falsified.

Note that the task used in this paradigm is one of categorization rather than identification, because it uses a many-to-one stimulus to response mapping rather than a one-to-one map. The experimental results we report are therefore directly relevant to the categorization literature. Indeed, we believe this technique is a particularly good one for studying rules of categorization. Typically, categorization is studied in designs that use only a few stimuli (i.e., 5 to 15). With stimulus representations sparsely scattered in the perceptual space, many very different decision rules will make identical predictions. On the other hand, the present paradigm, with its potentially infinite number of stimuli, guarantees a perceptual space as densely packed with stimulus representations as the experimenter desires.

In addition to its potential contribution to the study of the categorization process, we feel that the results of the experiments reported in this article are also relevant to the study of decision processes in identification tasks. If the general recognition theory is correct, then on trials of an identification task in which stimuli are degraded through say, tachistoscopic exposure, the same stimulus will not always elicit the same percept. However, with two stimuli, the subject's decision process must select one of the two responses. Thus, just as in categorization, the mapping from the perceptual space to the response is many-to-one. We therefore, believe that the decision problems encountered in categorization and in identification are very similar. This argument is supported by Nosofsky (1984, 1986) who, after allowing for changes in selective attention, successfully accounted for subjects' categorization performance on the basis of their identification responses, thus suggesting that subjects use the same set of decision strategies in each task.

It is important to note that we are not claiming that the external noise added in the experiments reported here has the same statistical properties as perceptual noise induced through standard techniques (e.g., tachistoscopic stimulus presentation). Even so, we feel that if the statistical properties differ, the high degree of observability afforded by the general recognition randomization technique and the similarity of the decision problems facing the subject makes it a useful one for studying decision processes in identification.

We are also not claiming a necessarily close correspondence between the stimulus and the perceptual spaces. For example, perceived length and physical length may be nonlinearly related. However, no problems occur so long as the relation between the physical and perceptual dimensions is monotonic (as both Steven's law and Fechner's law predict). In this case, the optimal bound in the stimulus space is mapped onto the optimal bound in the perceptual space. Thus, in the absence of perceptual noise, optimal responding in the perceptual space will manifest itself as optimal responding in the stimulus space.

The general recognition randomization technique is related to several other experimental paradigms. One of these is the "contrived control" experiment described by Sperling (1984). The most widely known of the related paradigms, however, is the numerical decision task originally developed by Lee and Janke (1964, 1965; Hammerton, 1970; Healy & Kubovy, 1977; Kubovy & Healy, 1977; Kubovy, Rapoport, & Tversky, 1971; Ward, 1973; Weissmann, Hollingsworth, & Baird, 1975). In this task, two univariate normal distributions of numbers are specified. On each trial, a number is sampled from one of the distributions and presented to the subject. The subject's task is to decide which distribution the sample came from. Lee experimented with nonnumeric stimuli, including the position of a dot on a piece of paper (Lee, 1963; Lee & Janke, 1964, 1965; Lee & Zentall, 1966) and the grayness of a patch of paper (Lee & Janke, 1964; Lee & Zentall, 1966). With the exception of Lee (1963), all of these studies used univariate stimulus distributions and therefore they provide no information about the form of a subject's decision bound when categorizing or identifying multidimensional stimuli. Lee (1963) was also not interested in this issue. Instead, his study, along with each of the others, focused on whether subjects use a deterministic or a probabilistic decision rule. With univariate stimulus distributions, a deterministic rule involves a fixed criterion x_c. For example, a rule might read

$$\text{if } x > x_c \text{ then respond } R_A.$$

If the criterion x_c varies from trial to trial, we say there is criterial noise. In this case the decision rule is still deterministic, in the sense that, for fixed values of x and x_c, the subject either always responds R_A or always responds R_B. When we speak of a probabilistic decision rule, we will mean one for which there exist fixed values of x and x_c (where $x \neq x_c$ and, of course, when x_c is a relevant parameter), for which the probability of responding R_A is neither 0 nor 1.

In the numeric decision task, Kubovy and his colleagues (Kubovy & Healy, 1977; Kubovy et al., 1971) found that a fixed cutoff accounted for the data significantly better than a probabilistic decision rule (Lee's, 1963, micromatching model), but that it still mispredicted a small percentage of the responses (5.87% in Kubovy et al., 1971). One possibility is that these mispredictions were the result of criterial noise. Although distinguishing between deterministic and probabilistic responding is not our primary goal, the general recognition randomization technique does allow us to examine this issue.

Independent Decisions Models

Before using the general recognition randomization technique, we need to examine the nature of the decision rules predicted by the various models. To begin, consider again the dotted-line decision rule illustrated in Figure 3. Recall that in this case, the decision is to respond R_A to a sample falling in the upper left quadrant, R_B to a sample falling in the lower right quadrant, and otherwise to guess.

Because dimension x corresponds to the horizontal component, a large value of x is strong evidence that the horizontal component of the presented stimulus is long and a small value, that it is short. One interpretation of the dotted-line Figure 3 rule, therefore, is that the subject sets a criterion x_0 on x such that the perceptual effect (x, y) is recognized as containing a long horizontal component if and only if $x > x_0$. Note that this decision does not depend on the value of y, the magnitude of the perception associated with the vertical segment. Thus, the subject makes an independent decision about the length of the horizontal component. Similarly, a separate independent decision is made about the length of the vertical segment by setting a criterion y_0 on y such that the perceptual sample is recognized as containing a long vertical segment if $y > y_0$ and a short vertical segment if $y < y_0$. The results of these independent decisions can be combined to form the compound rule of Figure 3: (a) If the horizontal segment is short ($x < x_0$) and the vertical segment is long ($y > y_0$), respond R_A; (b) if the horizontal segment is long ($x > x_0$) and the vertical segment is short ($y < y_0$), respond R_B; and (c) otherwise, contradictory information has been collected and so guess.

Independent decisions models need not postulate a guessing mechanism. For example, one sort of response bias would be to respond R_A if the sample falls in the upper left quadrant and to respond R_B otherwise. A different kind of response bias occurs if the subject guesses when contradictory information is obtained, but in a way that favors one alternative over the other. Thus, not all independent decisions models are equivalent. Many different rules can exist for specifying how a response is selected after the initial decisions have been made on all the components (e.g., Townsend & Ashby, 1982; Shaw, 1982). Even so, the one thing all independent decisions models have in common is that they always predict decision bounds that are parallel to the coordinate axes.

Ashby and Townsend (1986) identified decisions rules of this type with decisional separability. The subject's decision on one component does not depend on the level or the perceived magnitude of the other. They distinguish this kind of separability from perceptual separability, which occurs if the perceptual effect of one component does not depend on the level of the other. Independent decisions is a natural decision strategy to use when stimuli are constructed from perceptually separable components, because in many, but not all such cases, it will maximize response accuracy (Ashby & Townsend, 1986).

Information Integration Models

Because all independent decisions models predict the same kind of decision bounds, they will be fairly easy to identify empirically. Integration models, however, can predict many different kinds of decision boundaries. The exact shape and placement will depend on the characteristics of the decision rule and the nature of the added noise.

Minimum Distance Classifiers

In the categorization task described above, a very simple decision rule states that a subject will categorize a sample into class A or B depending on which prototype is most similar. In the categorization literature, models postulating this classification scheme are known as prototype models (e.g., Homa, Sterling, & Trepel, 1981; Posner & Keele, 1968, 1970; Rosch, 1973; Rosch, Simpson, & Miller, 1976). Prototype models typically use an MDS definition of similarity. The prototype that is most similar is the one that is nearest.

In the general recognition theory, this decision rule is equivalent to categorizing a perceptual sample according to the response associated with the perceptual distribution with the nearest mean. The subject is assumed to compute the distance from a given perceptual sample to the mean of each perceptual distribution and to "recognize" the sample as the stimulus associated with the minimum of these distances.

In the two-stimulus categorization task described earlier, this decision rule translates into a linear decision boundary placed midway between the means of the two perceptual distributions and orthogonal to a line passing through both means. The solid line boundary of Figure 3 is an example of minimum distance classification. Any perceptual sample falling above the line will be closer to the mean of the S_A distribution and any sample below the line will be closer to the S_B mean.

Notice that with a minimum distance rule, the amount and nature of the perceptual variability associated with each stimulus does not affect the placement of the boundary. Only the means of the distributions are taken into account. In fact, once the means are fixed, both the slope and intercept of the decision boundary (i.e., the line of equidistance) are determined.

The General Linear Classifier

Another type of integration model, more general than the minimum distance model, is the general linear classifer (e.g., Medin & Schwanenflugel, 1981; Morrison, 1976; Nilsson, 1965; Townsend & Landon, 1983). In this model, the decision bounds are always linear but unlike the minimum distance classifier, no constraints are imposed on their slopes and intercepts.

This increased generality allows the general linear classifer to accommodate response bias. Perhaps the most straightforward way of accomplishing this is simply by changing the intercept. For example, in Figure 3 a bias toward response R_A can be introduced by decreasing the intercept of the solid line decision boundary.

Of all the linear bounds, the one that classifies the stimuli more accurately than any other linear rule is known as the optimal linear bound. This bound will often significantly outperform minimum distance classification; an example is shown in Figure 4. A subject using the solid line as a decision boundary will more accurately identify the stimuli S_A and S_B than a subject using a minimum distance rule. To see this note that a sample at (x_0, y_0) is closer to the S_A prototype but is more likely to be a sample from the S_B distribution. The optimal linear bound is not always more accurate than minimum distance classification. For example, no decision rule predicts better performance with the Figure 3 perceptual distributions than the minimum distance model. In the next section, within the context of the general Gaussian recognition model, we will discuss the exact conditions under which linear classifiers can outperform minimum distance models.

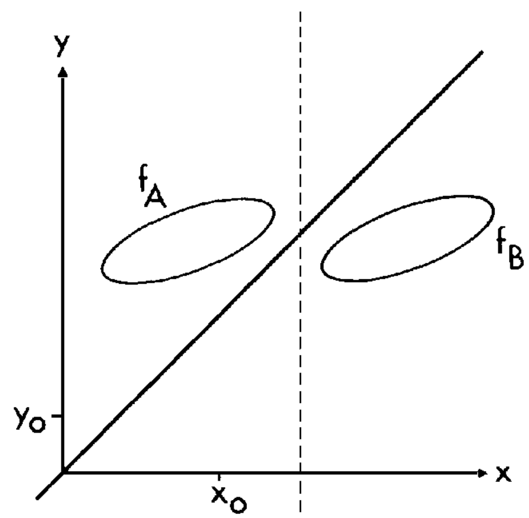

Figure 4. Contours of equal probability and two sets of decision bounds. (The solid line boundary is predicted by the optimal linear classifier and the dotted line is predicted by the minimum distance classifier.)

Recently, Medin and Schwanenflugel (1981) studied the ability of human subjects to learn general linear classification with stimuli consisting of four binary-valued components. They presented sets of linearly separable stimuli that turned out to be no easier for subjects to categorize than certain stimuli that were not linearly separable. They concluded that "linear separability is not a major constraint on human classification performance" (p. 355). We will argue that in many cases, linear classification is a very natural decision strategy and that when faced with an unfamiliar classification task, subjects often initially apply linear decision rules. However, if the resources to apply the optimal linear rule exceed available capacities, as was perhaps the case in Medin and Schwanenflugel's study, subjects may use simpler, less efficient rules (which may be linear), or if linear classification produces unsatisfactory performance, subjects apparently can turn to more complex strategies.

Optimal Decision Rules

Although the optimal linear classifier is guaranteed to be at least as accurate as the minimum distance classifier, only in fairly special circumstances will it be as accurate as the best possible decision rule. An example in which the optimal rule is not linear is given in Figure 5.

With two stimuli, S_A and S_B, and a perceptual sample at (x_s, y_s), an ideal observer (i.e., that maximizes accuracy)

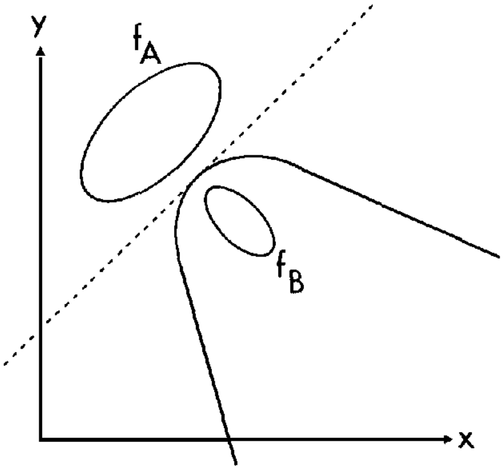

Figure 5. Contours of equal probability for which the optimal decision bound is nonlinear.

computes the likelihood ratio

$$\ell(x_s, y_s) = \frac{f_A(x_s, y_s)}{f_B(x_s, y_s)}$$

and compares it with a criterion value β (e.g., Fukunaga, 1972; Green & Swets, 1966; Noreen, 1981; Townsend & Landon, 1983). If the likelihood ratio is greater than β, response R_A is given and otherwise the subject responds R_B. In the unbiased case, $\beta = 1$. When the stimulus presentation probabilities are equal, a criterion of $\beta = 1$ also yields the most accurate responding.

As is well known, this kind of optimal responding model forms the basis of signal detection theory (or more specifically, of ideal observer theory, e.g., Green & Swets, 1966). However, in signal detection theory, interest has focused almost exclusively on unidimensional perceptual distributions. A noteworthy exception is found in an early article by Tanner (1956), who assumed that the sensory effects of noise and each of two auditory signals embedded in noise have bivariate normal distributions, all with covariance matrices equal to the identity (i.e., all variances equal to one and all correlations equal to zero). However, under these assumptions, the ideal response function is always a minimum distance classifier and so, with regard to understanding optimal decision rules in the recognition of multidimensional stimuli, Tanner's work is not particularly relevant.

If no distributional assumptions are made, not much can be said about the shape of the optimal decision boundary. In general, any shape is possible. However, if normality is assumed the picture becomes much simpler. Consider an identification task with two stimuli, S_A and S_B. Let $\boldsymbol{\mu}_i$ be a vector containing the coordinates of the mean of the S_i perceptual distribution. With the stimuli shown in Figure 3, for example, the first entry in the $\boldsymbol{\mu}_A$ vector would be the mean perceived length of the horizontal component of the S_A prototype and the second entry would be the mean perceived length of the vertical component. Let Σ_i be the covariance matrix associated with the S_i perceptual distribution and let

$$\mathbf{x}_s = \begin{bmatrix} x_s \\ y_s \end{bmatrix}$$

be the vector that contains the coordinates of the perceptual sample. Now define

$$\begin{aligned} &\mathrm{h}(x_s, y_s) \\ &\quad = \mathrm{h}(\mathbf{x}_s) = -\ln[\ell(\mathbf{x}_s)] = \tfrac{1}{2}(\mathbf{x}_s - \boldsymbol{\mu}_A)^T \textstyle\sum_A^{-1} (\mathbf{x}_s - \boldsymbol{\mu}_A) \qquad (1) \\ &\qquad - \tfrac{1}{2}(\mathbf{x}_s - \boldsymbol{\mu}_B)^T \textstyle\sum_B^{-1} (\mathbf{x}_s - \boldsymbol{\mu}_B) + \tfrac{1}{2} \ln\left(\dfrac{|\Sigma_A|}{|\Sigma_B|}\right), \end{aligned}$$

where T denotes transpose and $|\Sigma_i|$ is the determinant of Σ_i. Under these conditions, the optimal decision rule in the general Gaussian recognition model is to respond R_A if $\mathrm{h}(x_s, y_s) < 0$ and otherwise to respond R_B (e.g., Fukunaga, 1972). The decision boundary is the curve satisfying $\mathrm{h}(x, y) = 0$ or equivalently $\ell(x, y) = 1$.

The important thing about Equation 1 is that it always defines a decision bound that is quadratic or linear in x and y. In general, the function is quadratic when the covariance matrices associated with the two perceptual distributions are not equal. In this case, the contours of equal probability will not have the same shape.

When the covariance matrices are equal, so that $\Sigma_A = \Sigma_B = \Sigma$, the decision bound becomes a linear function satisfying

$$\begin{aligned} \mathrm{h}(x_s, y_s) &= \mathrm{h}(\mathbf{x}_s) \\ &= (\boldsymbol{\mu}_B - \boldsymbol{\mu}_A)^T \textstyle\sum^{-1} \mathbf{x}_s \\ &\quad + \tfrac{1}{2}(\boldsymbol{\mu}_A{}^T \textstyle\sum^{-1} \boldsymbol{\mu}_A - \boldsymbol{\mu}_B{}^T \textstyle\sum^{-1} \boldsymbol{\mu}_B) \qquad (2) \\ &= 0. \end{aligned}$$

In this case the contours of equal probability have exactly the same shape and are just centered at different points. For example, the contours in Figures 3 and 4 satisfy this constraint and so the optimal decision bound is linear in each case. In Figure 5 the contours are not simple shifts, suggesting that the associated covariance matrices are not equal, and that the optimal boundary is not linear (i.e., it is quadratic, as shown).

Finally, in the special case in which the covariance matrices both equal the identity or the same scalar multiple of it, the contours of equal probability are all circles of equal diameter and the optimal decision bound reduces to

$$\begin{aligned} \mathrm{h}(x_s, y_s) &= \mathrm{h}(\mathbf{x}_s) \qquad (3) \\ &= (\boldsymbol{\mu}_B - \boldsymbol{\mu}_A)^T \mathbf{x}_s + \tfrac{1}{2}(\boldsymbol{\mu}_A{}^T \boldsymbol{\mu}_A - \boldsymbol{\mu}_B{}^T \boldsymbol{\mu}_B). \end{aligned}$$

This is a linear bound that bisects and is perpendicular to the chord joining the S_A and S_B perceptual means, or in other words, a minimum distance classifier. Minimum distance classification is therefore optimal if the covariance matrices are all equal and are all scalar multiples of the identity.

It turns out that the decision rules of many distributed memory models are also equivalent to minimum distance classification. An alternative interpretation of the decision rule in Equation 3 is that the cross correlation between the perceptual representation and the mean stored representations of S_A and S_B are computed and then the difference is compared to a preset criterion.[1] A response is selected on the basis of this comparison. Note that, like minimum distance models, correlation classifiers therefore use only the mean stored representations and ignore the associated variances and covariances. Because of this, these models are optimal if very strong assumptions are made about these parameters (namely, that all covariance matrices are equal and are some scalar multiple of the identity).

Model Mimicry

In translating independent decision rules and the different types of integration rules into decision boundaries, it becomes evident that different models sometimes make identical predictions. For example, in Figure 3, minimum distance classification, the optimal linear classifier, and the ideal observer model all predict the same decision boundary. Thus, with this stimulus configuration, evidence that subjects are using the decision rule of Figure 3 allows us to rule out independent decisions, but it does not allow us to test between the many types of information integration. This identifiability problem is very common in all types of modeling (e.g., Greeno & Steiner, 1964; Townsend & Ashby, 1983), but with the general recognition randomization technique, it can largely be circumvented. For example, with the stimulus configurations of Figure 5, the optimal decision boundary is nonlinear, so independent decisions models, optimal linear models, and ideal observer models all make different predictions.

Some instances will also exist in which integration models and independent decisions models make identical predictions. For example, consider a complete identification experiment in which the stimuli are constructed by factorially combining two components, A and B, each of which is presented at two different experimental levels. Under these conditions, there will be four stimuli, A_1B_1, A_1B_2, A_2B_1, and A_2B_2. On each trial, one of these is randomly selected and then presented to the subject, whose task is to identify it uniquely. One possible model of a subject's identification performance in this task is illustrated in Figure 6. In this case, all models discussed in this article predict the solid line decision bounds. Thus, although a complete identification experiment with this stimulus ensemble might be advantageous for testing certain hypotheses about the perceptual processes (e.g., Ashby & Townsend, 1986), it is not a good choice if our interest is in decision processes.

In addition to the problem of exact equivalence, sometimes the models may be mathematically identifiable but still make predictions that are so similar they are empirically difficult to differentiate. An example arises with a popular experimental paradigm that uses the same stimuli, A_1B_1, A_1B_2, A_2B_1, and A_2B_2 as the complete identification task but that uses different response instructions. In this paradigm, the subject is instructed to respond "yes" if either component appears at Level 2. Thus, the subject's task is to respond "yes" to stimuli A_1B_2, A_2B_1, or A_2B_2 and to respond "no" only on A_1B_1 trials. Frequently, the levels are defined so that Level 1 is component absence and Level 2 is component presence. In this case, the subject is told to respond "yes" if the stimulus contains either component.

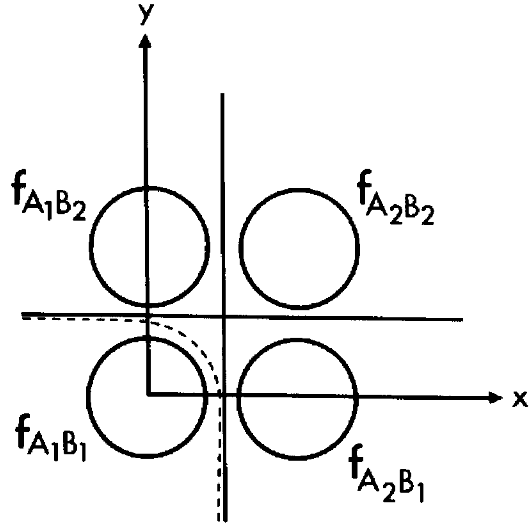

Figure 6. Contours of equal probability for a task with stimuli A_1B_1, A_1B_2, A_2B_1, A_2B_2. (The solid lines are the optimal bounds for a complete identification experiment and the dotted line is the optimal bound in a task where the subject is instructed to respond "no" to stimulus A_1B_1 and otherwise to respond "yes.")

In this design, we expect subjects making independent decisions on the two components to use the decision bounds of Figure 6 and to assign the response "yes" to all quadrants other than the one associated with stimulus A_1B_1, to which they assign a "no" response. This rule is not optimal and so independent decisions models and models postulating optimal integration of perceptual information are mathematically identifiable. The problem is that the optimal rule, the dotted-line bound in Figure 6, is very similar to the independent

---

[1] To see that Equation 3 describes a correlation classifier, note that it can be rewritten as

$$\text{if } \boldsymbol{\mu}_B{}^T\mathbf{x}_s - \boldsymbol{\mu}_A{}^T\mathbf{x}_s \lesseqgtr \tfrac{1}{2}(\boldsymbol{\mu}_B{}^T\boldsymbol{\mu}_B - \boldsymbol{\mu}_A{}^T\boldsymbol{\mu}_A) \text{ then respond } \begin{cases} R_A \\ R_B \end{cases}.$$

The product $\boldsymbol{\mu}_i{}^T\mathbf{x}_s$ is the correlation between $\boldsymbol{\mu}_i$ and $\mathbf{x}_s$. Suppose the vector $\mathbf{x}_s$ consists of perceptual representations sampled at different points in time and that the dimensions of the perceptual space correspond to the different times of sampling so that, for example, the jth component of the vector represents the mean perceptual value of stimulus S_A at the jth time point (i.e., at time t_j). In this case the correlation can be rewritten as

$$\boldsymbol{\mu}_i{}^T\mathbf{x}_s = \sum_{j=1}^{n} \mu_i(t_j)x_s(t_j),$$

where a total of n samples are taken and $\mu_i(t_j)$ is the jth component of $\boldsymbol{\mu}_i$ (and similarly for $x_s[t_j]$). In the case of continuous sampling, the correlation becomes an integral, that is,

$$\sum_{j=1}^{n} \mu_i(t_j)x_s(t_j) \rightarrow \int_0^T \mu_i(t)x_s(t)\,dt,$$

where T is the duration of sampling. This is the standard form of a cross correlation that is used in many distributed memory models.

decisions rule. Running an experiment powerful enough to discriminate between the two would be an ambitious undertaking.

We are now ready to use the general recognition randomization technique to study human pattern classification. In particular we are interested in determining which type of classifier—independent decisions, minimum distance, optimal linear, or ideal observer—best describes human performance. The three experiments described in this article are very similar and so the general methods used in each will be described first and additional details given as each experiment is discussed.

General Method

Subjects

All subjects were volunteers who were paid a base rate plus a bonus of 5¢ for every percentage point correct above 70%. The subjects all had either 20/20 vision or vision corrected to 20/20. Three subjects served in each experiment except for Experiment 2, in which 5 subjects participated. No subject was used for more than one experiment. Participants were either volunteers from the Harvard University community who were paid a base rate of $3 for the 1-hr experimental session or were work-study students in the Ohio State University Psychology Department. The base rate for these subjects was $4 per hour.

Stimuli

The stimuli, as illustrated in Figure 2, were simple two-line figures formed by a vertical and a horizontal line orthogonally joined at an upper left corner. The figures were computer generated and displayed on a Hewlett-Packard CRT in a dimly lit room. In every experiment, the stimuli were one of two types (S_A or S_B). Each type was associated with a specific bivariate logistic distribution. The logistic was chosen because it is very similar to the normal distribution but has a simple closed-form expression for its cumulative distribution function.

The process of generating a stimulus on each trial proceeded as follows. First, stimulus type (S_A or S_B) was determined by randomly sampling from a uniform distribution. Each type was equally likely to be selected. Then, a random sample (X_s, Y_s) was drawn from the bivariate distribution associated with the stimulus type selected for that trial. A figure was then generated in which the length of the horizontal segment was X_s and the length of the vertical segment was Y_s.

For each experiment, the parameters of the bivariate distributions were changed. However, they were always selected so that an ideal observer could correctly classify the samples with probability .80.

Procedure

On every trial a stimulus was presented that subjects categorized as an S_A or S_B figure by pressing the appropriate button. Subjects were told that an "expert" would be correct about 80% of the time and they were told they would receive a bonus (as described earlier) for every percentage point correct above 70%. Accuracy was stressed much more than speed. The stimulus display was terminated by the subject's response or in case the subject had not yet responded, after 5 s. Feedback showing the correct response appeared on the screen immediately after a response was given. There was a 3-s pause between trials.

Table 1
Stimulus Parameter Values: Experiment 1

	S_A	S_B
Horizontal *M*	400	500
Vertical *M*	500	400
Horizontal *SD*	84	84
Vertical *SD*	84	84
Horizontal–vertical covariance	0	0

Note. The units here are arbitrary screen units in which there are about 270 units per degree of visual angle.

The first 100 trials were part of a practice session that allowed the subject to become familiar with the S_A and S_B distributions. A pause separated the practice session from an experimental session that consisted of 300 trials; during the pause the subject was allowed to ask questions about the procedure. The practice and experimental sessions both began with a prototype learning block in which the prototypes (i.e., the means of each distribution) were displayed alternately with category labels until each had been shown five times. In the experimental session the prototype learning block was followed by four experimental blocks, each with 75 trials. There was a 30-s break between blocks to allow subjects to rest.

Experiment 1

Introduction

The purpose of Experiment 1 was to determine whether an independent decisions model or an integration model best describes a subject's behavior in the kind of simple pattern classification task described in the General Method section. In particular, we were interested in testing between the dotted-line and the solid-line decision rules of Figure 3, so we chose stimulus distributions corresponding to the ones illustrated in Figure 3. The prototypes therefore resembled the figures shown in Figure 2 with prototype S_A having a long vertical component and a short horizontal component and prototype S_B having the opposite configuration. Within each distribution, the variances on each dimension were equal and the covariance term was essentially zero.[2] The exact parameter values are given in Table 1.

All of the major integration models discussed earlier predict the solid line decision bound $y = x$ illustrated in Figure 3 with these stimulus distributions. Note that this rule is equivalent to one that compares the horizontal and vertical lengths and responds R_A if the vertical segment is longer and R_B if the horizontal segment is longer (or to one that compares the ratio of the perceived vertical length to the perceived horizontal length and responds R_A whenever the ratio exceeds 1.0).

On the other hand, as we mentioned earlier, all independent decisions models predict that the response space will consist of the four quadrants demarcated by the dotted lines of Figure 3. In discriminating between the two classes of models, the lower left and the upper right quadrants are critical. If these contain a random mixture of R_A and R_B responses (in not

[2] The algorithm we used to generate random stimuli required a nonzero covariance term. Therefore, in experiments in which we desired uncorrelated variates the covariance was set to 0.1.

necessarily equal numbers) then independent decisions is indicated whereas information integration is supported if the quadrants are bisected by a boundary that perfectly partitions the R_A and R_B responses.

As previously mentioned, a subject using the optimal rule $y = x$ will, in the long run, be correct with probability .80. On the other hand, a subject making independent decisions could be correct with probability .72. Thus, the overall predicted accuracy difference between the two rules is not large and so the two strategies might be difficult to discriminate empirically on the basis of the dependent variable percentage correct. Using the general recognition randomization technique, however, makes them easy to differentiate.

An important attendant concern is whether the stimulus components in Figure 2 are separable or integral. If the components are integral we would not expect subjects to use independent decisions even if they normally do so when stimuli are constructed from separable components.

On the one hand, intuition suggests that the components of Figure 2 might be separable because it seems easy to attend separately to one while ignoring the other (Shepard, 1964). On the other hand, several researchers have argued that it is possible to attend selectively to integral components in unspeeded conditions (e.g., Foard & Nelson, 1984; Garner, 1974; Lockhead, 1972). In addition, if redundant line segments are added to the stimuli of Figure 2, then rectangles can be created and several studies have reported evidence that the components of rectangles are integral (e.g., Felfoldy, 1974). First, however, none of the present experiments involve speed stress and second, if the components of Figure 2 are integral, they are certainly not integral in the same way that the perceptual components of say, hue and brightness are integral. In the case of line segments, one can easily focus attention on one segment and ignore the other, but it is very difficult or impossible to focus attention on, say, hue and ignore brightness. Fortunately, strong evidence exists that subjects can separately attend to the components in Figure 2.

Townsend, Hu, and Ashby (1980, 1981) reported the results of a complete identification experiment with four stimuli created from all possible combinations of a vertical and horizontal line segment. If we let V represent the vertical component and H the horizontal, then the four stimuli were VH (both present), V (only the vertical component present), H, and Ø (neither component present). Stimulus VH was in the same configuration as the Figure 2 prototypes. Stimuli were tachistoscopically presented to each of 4 subjects who were told to identify the presented stimulus on each trial (as VH, V, H, or Ø). Townsend et al. (1981) very successfully accounted for the conditional response probabilities (i.e., the confusion matrices) from this experiment by using a feature-analytic model that assumes independent decisions. Ashby and Perrin (1988) obtained equally successful accounts with a version of the general recognition theory that assumed independent decisions. In fact, they showed that this model accounted for the data as well as the biased choice model (Luce, 1963; Shepard, 1957), which for 20 years has been the most successful model at accounting for the confusions made in identification experiments.

Further evidence that it is possible to attend separately to these components comes from Ashby and Townsend (1986), who checked this same Townsend et al. (1981) data for a number of conditions that they developed specifically to test for separability. In each case, separability was strongly supported. These findings do not imply that in the present experiment subjects will adopt an independent decisions strategy but only that they could, if so motivated. The stimulus conditions are different from those of Townsend et al., and there is no reason to expect subjects to use identical decision rules under differing stimulus conditions. The important point, however, is that these stimulus components are not integral in the way that hue and brightness are integral because, under the appropriate conditions, subjects were able to attend separately to the two components.

The present series of studies would be of interest, however, even with stimuli created from integral components. We might expect subjects to use an integration rule in such a case, but as we have seen, there are many different ways of integrating information. The fact that certain stimulus components are associated with interference in a filtering task (e.g., Garner, 1974) might tell us they are integral, but it does not tell us whether subjects confronted by such stimuli will use minimum distance, general linear, or ideal observer classification. The general recognition randomization technique permits enough observability to investigate decision processes at this more microscopic level.

Results and Discussion

Figure 7 shows the stimulus space and the responses during the 300 trials of the experimental session for Subject 1. The other subjects displayed a very similar pattern of responses. An x in Figure 7 indicates an R_A response to a sample with that particular horizontal and vertical line length. A filled circle indicates an R_B response, and the squares indicate that more than one sample with those coordinates was shown and that the subject responded inconsistently[3] (R_A one time and R_B another). A casual examination of Figure 3 reveals that responses do not appear random in any of the critical quadrants. Instead, it appears that Subject 1 used a decision rule strikingly similar to the solid-line boundary of Figure 3. On first examination, it appears that subjects integrated the information from the two dimensions before initiating any decision processes.

The optimal bound, $y = x$, is also shown in Figure 7. Note that it gives a very good account of the data, in the sense that very few R_A responses fall below it and very few R_B responses fall above it, and those that do tend to miss by only a small amount. In fact, all 3 subjects responded at a near optimal level. This conclusion is corroborated by their response accuracy, given in Table 2, which was 88%, 82%, and 79%, respectively. It may be recalled that the stimulus parameters were selected so that, in the long run, an ideal observer would

[3] In order to prepare these figures, some collapsing of the response space was necessary. Thus, each symbol in Figure 7 represents a square 10 screen units long on each side.

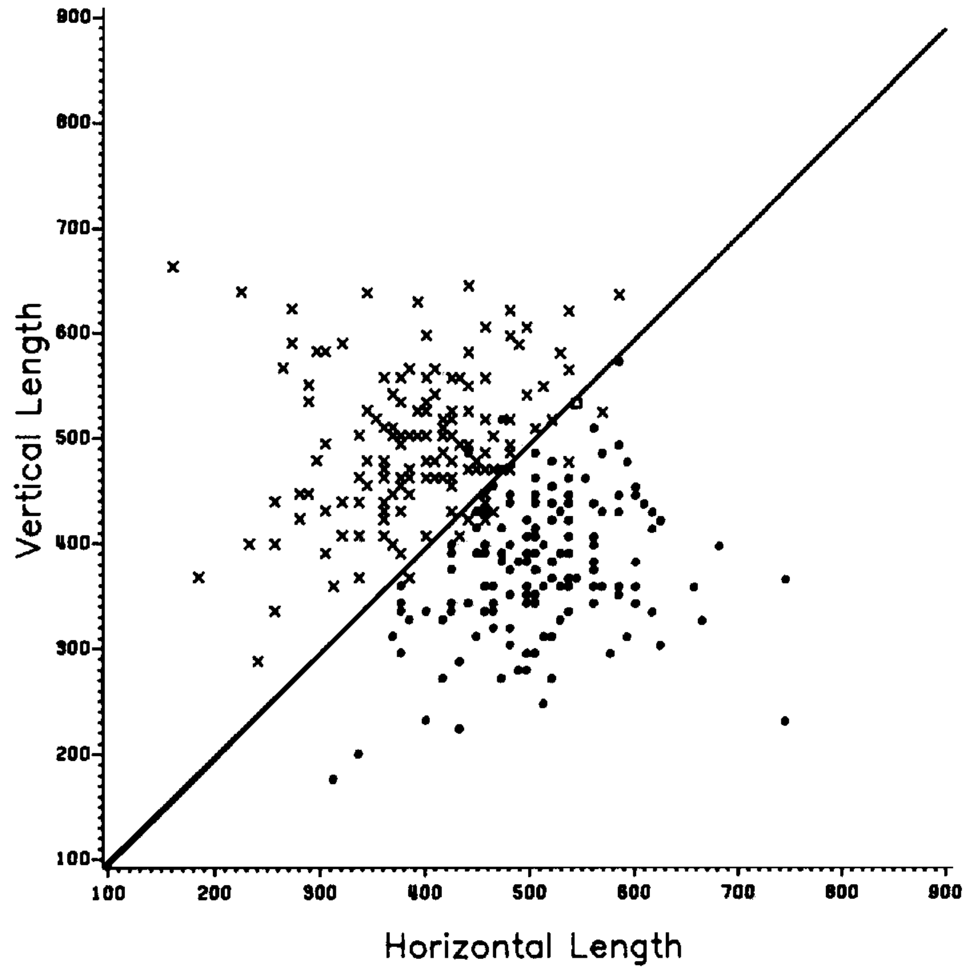

Figure 7. Response data for Subject 1 in Experiment 1. (An x indicates an R_A response, a filled circle an R_B response, and a square indicates response inconsistency.)

correctly classify the stimuli on 80% of the trials. Subjects 1 and 2 exceeded this level because random sampling from the two stimulus distributions happened to generate 300 samples somewhat easier to classify than expected (i.e., the sample variances were smaller than the population variances). As Table 2 indicates, if subjects had perfectly used the optimal rule, $y = x$, they would have correctly classified 89%, 83%, and 82% of the stimulus samples, respectively. Thus, all three subjects responded within 3% of the optimal level.

If, on the other hand, the subjects had used the best possible independent decisions rule, they would have correctly classified 74%, 80%, and 75% of the samples, respectively. All 3 subjects therefore responded more accurately than the best independent decisions device.

To clarify the distinction even further, a statistical test can be conducted. Consider a subject using an independent decisions rule. The response of such a subject to any sample falling in the lower left or upper right quadrant will not depend on which side of the bound $y = x$ the sample happens to fall. Let P equal the probability that the subject responds optimally in these two quadrants (i.e., R_A to any sample falling above the $y = x$ bound and R_B to any sample falling below): then under the null hypothesis that subjects used an independent decisions rule, $P = .5$.

The parameter P is easily estimated from each subject's response data. For the three subjects of Experiment 1, estimates of P were .89, .90, and .90, respectively. In all three cases the null hypothesis that $P = .5$ is rejected in favor of the alternative that $P > .5$ with $\alpha = .001$. The data overwhelmingly support information integration over independent decisions.

To test even further the degree to which subjects used the optimal rule, a computer search was implemented to determine the linear bound that best accounts for the subject's responses. A prediction error occurs whenever an R_A response falls below the bound or an R_B response falls above, and we defined the magnitude of a prediction error as the (Euclidean) distance from an incorrectly predicted response to the decision bound (i.e., as the distance from the coordinates in the horizontal–vertical stimulus space of an incorrectly predicted

Table 2
Results of Experiment 1

Subject	% of stimuli accounted for by optimal bound	% of stimuli accounted for by independent decisions rule	% correct	Best fitting linear bound	% of responses accounted for by best fitting bound	% of responses accounted for by optimal bound
1	89	74	88	$y = x - 12$	95	96
2	83	80	82	$y = x - 12$	96	96
3	82	75	79	$y = .95x + 23$	96	97

response to the decision bound). The linear bounds that minimized the sum of the magnitudes of all prediction errors are given in Table 2.

First note that as expected, the best fitting linear bounds are very close to the optimal $y = x$. Second, note that the bounds for 2 of the 3 subjects are parallel to the optimal bound but have negative intercepts. This result is nicely predicted by the horizontal–vertical illusion. When the vertical and horizontal segments of a stimulus are of equal length the vertical line is perceived to be longer and so the subject responds R_A.

Table 2 also lists the percentage of responses correctly predicted by the best fitting linear bound. Note that even though the bounds were not selected to maximize this value, in all three cases at least 95% of all responses are correctly accounted for by a linear rule. Our guess is that the few responses not accounted for are due to perceptual noise. Thus, in addition to the horizontal–vertical illusion, it appears that subjects sometimes slightly misperceive true line length so that, for example, a stimulus actually falling above the subject's decision bound appears to fall below it and so elicits an R_B response. Although some perceptual noise is evident, its overall contribution is very small. Thus, at least under these stimulus conditions, the decision process can be directly observed.

Another striking aspect of the data in Figure 7 is the apparently deterministic nature of the response function $r(x, y)$. With the exception of samples falling near the line $y = x$, the probability of responding, say R_A was essentially either 0 or 1 (depending on whether the sample was above or below the bound). Subjects did not guess but instead consistently used a single simple decision rule. These results therefore can not be predicted by any model that depends heavily on a guessing strategy or that postulates competing response tendencies that are associated with every point in the perceptual space.

Of course, specific models that postulate competing response tendencies may have enough flexibility in their parameters to approximate the deterministic responding of Figure 7. In particular, we have in mind the context model proposed by Medin and Schaffer (1978) and elaborated by Nosofsky (1984, 1986), which predicts that

$$P(R_A \mid \text{sample } \mathbf{x}_s) = \frac{\sum_i s(\mathbf{x}_{a_i}, \mathbf{x}_s)}{\sum_i s(\mathbf{x}_{a_i}, \mathbf{x}_s) + \sum_i s(\mathbf{x}_{b_i}, \mathbf{x}_s)}, \quad (4)$$

where $s(\mathbf{x}_{a_i}, \mathbf{x}_s)$ is the perceived similarity of sample $\mathbf{x}_s$ to $\mathbf{x}_{a_i}$, the ith exemplar of category A. In the case of separable stimulus components,

$$s(\mathbf{x}_{a_i}, \mathbf{x}_s) = \exp(-c \textstyle\sum_j |x_{a_{ij}} - x_{s_j}|),$$

where $x_{a_{ij}}$ is the perceived value of exemplar $\mathbf{x}_{a_i}$ on dimension j. This model has one free parameter, c. It best approximates the deterministic responding of Figure 7 when c is large. R. M. Nosofsky (personal communication, March 1985) has shown that when c is infinite and with an infinite number of exemplars in each category, Equation 4 can be reduced to

$$P(R_A \mid \text{sample } \mathbf{x}_s) = \frac{f_A(\mathbf{x}_s)}{f_A(\mathbf{x}_s) + f_B(\mathbf{x}_s)},$$

where $f_i(\mathbf{x}_s)$ is the height (i.e., the likelihood) of the probability density function of the exemplars of category i at the sample point $\mathbf{x}_s$. With the stimulus parameters in Experiment 1, this function monotonically decreases as $\mathbf{x}_s$ moves from the upper left to the lower right of Figure 7, but much more gradually than the data indicate. For example, in the case of a sample falling exactly at the S_A mean ($\mathbf{x}_s$ = [400, 500]), the data indicate that subjects always respond R_A, but the best the context model can predict[4] is that response R_A is given with probability .80. Similarly, subjects virtually never responded R_A to a sample falling at the S_B mean but the context model predicts they should at least 20% of the time. Clearly, the context model can not account for these data.

If subjects use deterministic decision rules, why do probabilistic models such as the context model successfully account for such a wide variety of other data? One possibility is because the presence of perceptual noise can obscure deterministic responding (for another possibility see Martin & Caramazza, 1980). For example, consider an x in Figure 7 that falls below the optimal bound. This point corresponds to a stimulus sample in which the horizontal component is longer than the vertical but to which the subject responds R_A. There are two possibilities. First, the subject could have correctly perceived that the horizontal component was longer and responded R_A anyway. This would indicate a nondeterministic decision rule. The second possibility is that the subject incorrectly perceived the vertical component to be longer and then used the follow-

[4] As Table 2 indicates, the samples that each subject received were slightly easier to categorize than the population as a whole. Thus, for the subjects in this experiment, the upper limit allowed by the context model on the probability of responding R_A to a sample exactly at the S_A mean will be slightly greater than .8. Even so, for each subject the observed proportion exceeds this upper limit.

ing deterministic decision rule:

if $y > x$ respond R_A, otherwise respond R_B.

Given the experimental paradigm used here, there is no way to distinguish between these possibilities. Criterial noise (e.g., Gravetter & Lockhead, 1973; Nosofsky, 1983; Wickelgren, 1968) will cause the same identifiability problems. Whether a certain response function is deterministic will only become evident after these sources of internal variability are eliminated.

The results of Experiment 1 strongly indicate that subjects can very easily and accurately integrate the stimulus components of Figure 2. To what extent does this conclusion generalize to other types of stimulus components? Although the components of Figure 2 have been demonstrated to be separable (in the sense of Ashby & Townsend, 1986), they still possess a special property not normally associated with separability. The relevant perceptual dimension associated with each component is perceived length, thus making it very easy to compare and hence also to integrate information from the two components. In fact, the results of Experiment 1 indicate that it is so easy to integrate information from these components that the possibility exists that subjects treated the vertical–horizontal difference, $y - x$, as a single psychological dimension. This would make the Experiment 1 categorization a one-dimensional task, in which case independent decisions and integration models make the same prediction. Although our subjects did have an easier time learning decision rules with a slope of 1.0 than any other decision rules, they were able to perform sufficiently well in tasks that required rules with slopes different from 1.0 to make us doubt that vertical–horizontal difference is the primary psychological dimension. For example, in a task for which the optimal bound was $y = 900 - x$, 2 subjects performed nearly optimally (77% and 78% correct, respectively) and 2 subjects had some difficulty (69% correct for each) but the best fitting linear bounds had negative slopes in each case.

Unlike the components of Figure 2, many separable stimulus components are associated with different perceptual dimensions, which makes information integration a more difficult task. For example, consider stimuli constructed from a semicircle of varying radius with a line projecting from its center to the semicircle's edge and varying in angle of rotation. Stimuli of this type were originally investigated by Shepard (1964) and have a long research history. Like the perceptual components associated with the Figure 2 stimuli, the components of these semicircles, perceived radius and perceived angle of rotation, are also separable, but in this case are much more difficult to compare. Use of the decision rule employed by subjects in Experiment 1 requires comparing the perceived radius with the perceived angle of rotation and deciding which is greater. Can subjects learn such a decision rule?

To answer this question, we replicated Experiment 1 in every detail except that instead of the line segments of Figure 2 we used these semicircles.[5] Three subjects were used. Two of these were naive, but the third subject, R.G., was the second author of this article. A sophisticated subject was used to learn whether prior knowledge of the optimal decision rule would improve performance. As it turned out, all subjects performed at about the same level. It appeared that prior knowledge of the optimal decision rule provided little advantage.

The only other difference from Experiment 1 was that we tested subjects for more than 1 hr. We reasoned that if subjects could learn the optimal rule, it would take them longer to do so than with the line-segment stimuli of Figure 3. Each subject therefore participated in 1-hr experimental sessions (as described earlier in the General Method section) on successive days until reaching a criterion of 78% correct during the 300 trial experimental block of a particular session or until their accuracy over sessions asymptoted.

All 3 subjects surpassed the 78% accuracy criterion. Subject 1 achieved 83% correct on Day 4; Subject 2 and Subject R.G. achieved 79% correct on Day 2. The accuracy of each subject monotonically increased across sessions. If subjects had used the independent decisions rule of Figure 3, their accuracy would have been 82%, 77%, and 74% for Subjects 1, 2, and R.G., respectively. Thus, all 3 subjects responded more accurately than was predicted by the independent decisions strategy of Figure 3.

The proportion of times that each subject responded according to the optimal rule $y = x$ to a stimulus sample falling in either the upper right or the lower left quadrant was .73, .60, and .71 for Subjects 1, 2, and R.G., respectively. Recall that the independent decisions rule predicts these proportions to be .5. Although these values are substantially smaller than the corresponding estimates obtained from the Experiment 1 data, they are still large enough to allow us to confidently reject the null hypothesis of an independent decisions strategy (i.e., that $P = .5$) with $\alpha = .001$ for Subjects 1 and R.G. and $\alpha = .02$ for Subject 2. As in Experiment 1, the best fitting linear bound was very close to the optimal bound and although fewer responses were accounted for by the best fitting linear bound than was the case in Experiment 1, the bound still accounted for a substantial percentage of each subject's responses (88%, 91%, and 94% for Subjects 1, 2, and R.G., respectively).

A trial-by-trial examination of each subject's responses corroborates these findings. Plots of each subject's responses look very much like the plots of the responses of the Experiment 1 subjects, except that they show more variability about the decision bound. Thus, as in Experiment 1, subjects clearly were integrating information from the two separable components. We believe their responses show more variability than was observed in Experiment 1 only because it is inherently more difficult (but as we have shown, not impossible) to integrate perceived radius and perceived angle of rotation than it is to integrate perceived length and perceived length.

[5] Mean radius for the S_A distribution was 500 units and for S_B it was 400 units. Mean angle of rotation was $\pi/5$ radians for S_A and $\pi/4$ radians for S_B. Standard deviation on the radius dimension was 84 units for both distributions and on the angle dimension was $21\pi/500$ radians for both distributions. To determine the optimal bound, it is necessary to equate screen units with angle units. Because they are both ratio scales, there is no reason to expect one of our screen units to psychologically equal one radian. We assumed $\pi/2$ radians (a quarter of a circle) was psychologically equal to 1000 screen units (about a quarter of the screen width).

In this experiment and in Experiment 1, subjects integrated perceptual information, even though the stimulus components were perceptually separable. In both of these experiments, the stimulus conditions were arranged so that subjects could earn more money (i.e., through larger bonuses for accuracy) by properly integrating stimulus information. It may be that subjects naturally use an independent decisions strategy but will integrate information if given sufficient incentive. The conditions of Experiment 2 were therefore designed so that an independent decisions rule was optimal (i.e., maximized accuracy and payoffs). This task should be especially easy if subjects naturally use independent decisions strategies.

Experiment 2

Introduction

The parameter values for the stimulus distributions used in this experiment are given in Table 3. Note that the horizontal means and variances are identical in the two distributions and that both covariances are zero. Thus, all information differentiating the two is contained in the length of the vertical component. The optimal decision bound, $y = 440$, therefore ignores the horizontal component and sets a criterion on the length of the vertical component. Response R_A is given if the vertical length exceeds this criterion and response R_B is given if it does not. If the subject makes independent decisions about the components, this task should be fairly easy because the optimal rule depends only on the outcome of the decision about whether the vertical component is long.

Results and Discussion

Because of the importance of this condition, we tested 5 naive subjects instead of 3. To verify the efficacy of the stimulus generation procedure the percentage of samples correctly classified by the optimal bound $y = 440$ was computed for each subject. The results are given in Table 4. Note that the percentages range from 81% to 88%. In contrast, the bound $y = x$ accounts for between 70% and 80% of the stimulus samples. For each subject, the accuracy of the optimal bound exceeds the accuracy of the $y = x$ bound by at least 7 percentage points. Thus, it appears that the statistical properties of the stimulus samples generated in Experiment 2 were satisfactory.

Table 3
Stimulus Parameter Values: Experiment 2

	S_A	S_B
Horizontal M	500	500
Vertical M	511	369
Horizontal SD	84	84
Vertical SD	84	84
Horizontal–vertical covariance	0	0

Note. The units here are arbitrary screen units in which there are about 270 units per degree of visual angle.

Table 4 also contains the equations of the best fitting linear bound and summarizes other relevant results. Figures 8 and 9 show the responses of 2 of the subjects over the course of the experimental sessions along with the optimal decision bound, $y = 440$.

Note first that, compared with Experiment 1, the optimal rule poorly predicts the response data. Subject 2 is the only subject whose best fitting linear bound has a slope of zero. Thus, only this subject successfully ignored the length of the horizontal component and so was able to invoke an independent decisions rule. The data of Subject 2 indicates however, that it is possible to attend separately to these components. Notice however, that although Subject 2 appears to have used a rule that is virtually optimal, this subject displays more variability in the use of this rule and so is relatively less accurate than any of the subjects of Experiment 1. This can be seen in Figure 9 as well as in Table 4, which shows that the best fitting linear bound accounts for 5% fewer responses than any of the best fitting bounds of Experiment 1.

Why should the rule $y = 440$ be harder to learn than the rule $y = x$? One possibility is that the former rule requires an internal referent, whereas the latter rule does not. For example, in Experiment 2, an ideal observer might try to store in memory the representation of a vertical segment exactly 440 screen units long. When a stimulus is presented, its vertical component can be compared to this referent and a response made depending on which is longer. On the other hand, the rule $y = x$ requires no internal referent. When a stimulus is presented, the vertical component can be compared to the horizontal component and a response made depending on which is longer. The extra variability displayed by Subject 2 in Experiment 2 suggests that such an internal referent is inherently noisy.

This factor nicely accounts for the performance decrement of Subject 2, but it cannot account for the relatively poor

Table 4
Results of Experiment 2

Subject	% of stimuli accounted for by optimal bound	% of stimuli accounted for by the bound $y = x$	% correct	Best fitting linear bound	% of responses accounted for by best fitting bound	% of responses accounted for by optimal bound
1	84	73	82	$y = .66x + 104$	91	84
2	82	71	77	$y = 436$	90	91
3	81	70	75	$y = .92x - 4$	94	84
4	81	74	75	$y = .31x + 291.8$	87	86
5	88	80	79	$y = .90x + 9.3$	78	80

Note. The best fitting bound minimizes the sum of the distances from all incorrectly predicted responses to the decision bound.

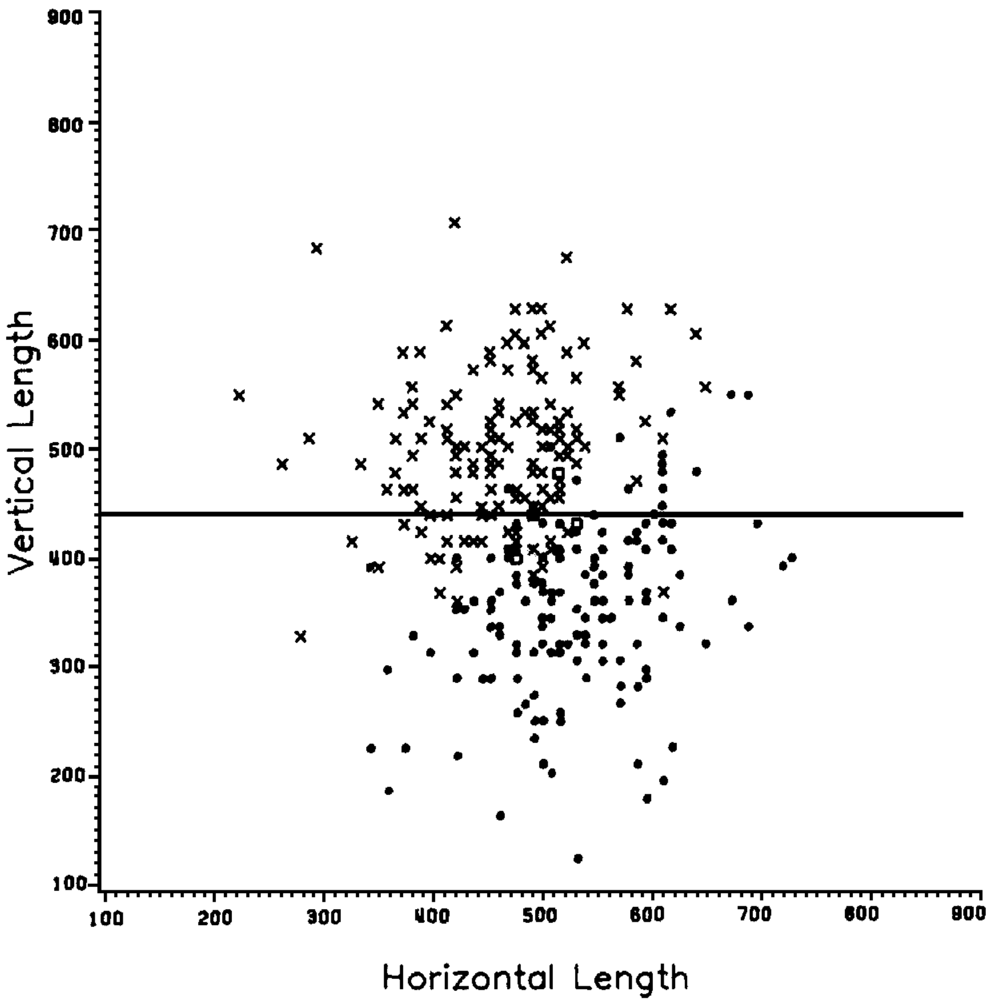

Figure 8. Response data for Subject 1 in Experiment 2. (An x indicates an R_A response, a filled circle an R_B response, and a square indicates response inconsistency.)

performance of the other subjects. Note that the linear bound that best describes the response strategy of Subjects 1, 3, 4, and 5 has a relatively large positive slope. In the case of Subjects 1 and 3, this bound accounts for a substantially greater percentage of responses than the optimal rule. Instead of ignoring the horizontal component, these subjects considered its length when choosing a response. For Subjects 4 and 5, the best fitting linear bound accounts for about the same percentage of responses as the optimal bound. These subjects apparently either continually changed strategies or made extensive use of guessing. In either case, they did not learn the optimal rule as successfully as did the subjects of Experiment 1.

It seems likely that the desire of our subjects to integrate information is due, in part, to the ease with which these components are integrated. We suspect that a replication of this experiment that used Shepard (1964) circles as stimuli would find greater evidence of independent decisions. Clearly, more research is needed in this area.

The data from each of these experiments support the hypothesis that the natural tendency of subjects is to integrate the information from the stimulus components of Figure 2 before making any decisions. This is not to say that all optimal rules that require information integration are easier to learn than an independent decisions rule, even when the integration rules are of the minimum distance variety. We ran several more conditions in which the optimal rule was minimum distance classification that also required integration of information.

In one of these, stimulus distributions were chosen[6] so that a correlation existed between the lengths of the horizontal and vertical stimulus components and so that the optimal bound, $y = x + 200$, had a large intercept. The 3 subjects in

[6] Mean horizontal length was 400 units for the S_A distribution and 500 units for S_B. Mean vertical length was 700 units for S_A and 600 units for S_B. The standard deviation on each dimension was 71.5 units and the horizontal–vertical correlation was −.39.

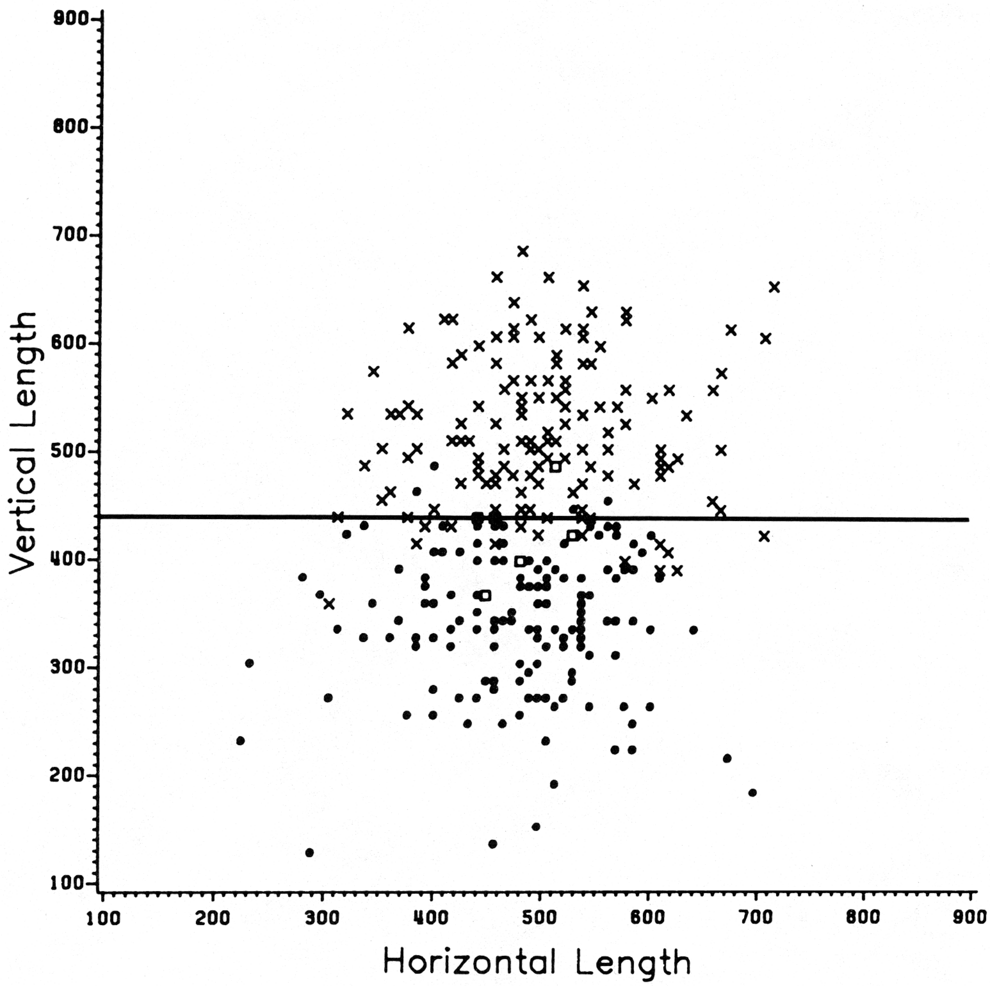

Figure 9. Response data for Subject 2 in Experiment 2. (An x indicates an R_A response, a filled circle an R_B response, and a square indicates response inconsistency.)

this experiment all responded nearly optimally. Their accuracy was 82%, 81%, and 81%, respectively and the best fitting linear bounds were $y = x + 180$, $y = x + 200$, and $y = x + 208$, respectively. Furthermore, these bounds accounted for more than 90% of each subject's responses (92%, 93%, and 92%, respectively) and thus, as in Experiment 1, subjects rarely guessed. With these stimulus distributions, the best independent decisions strategy places bounds at $x = 450$ and $y = 650$. As in Experiment 1, the upper right and lower left quadrants determined by these bounds are critical when testing independent decisions against optimal integration. The proportion of times that each subject responded according to the optimal rule $y = x + 200$ to a stimulus sample falling in either of these quadrants was .80, .83, and .78, respectively. Each of these values are large enough to reject the null hypothesis of independent decisions (i.e., that $P = .5$) with $\alpha = .001$. Therefore, in spite of the large intercept, as in Experiment 1, subjects accurately integrated the stimulus information.

On the other hand, with the more complex rule, $y = 2x - 460$, subjects adopted nonoptimal strategies. The best fitting linear bounds for all three subjects in the latter experiment had slopes that were closer to 1.0 than to 2.0 (i.e., .8, .63, and 1.1). Two of these subjects very consistently applied these nonoptimal rules. In both cases, the best fitting linear bound accounted for 97% of the responses. The third subject displayed more variability and appeared to be guessing on many trials (82% of the responses were accounted for by the best fitting linear bound).

In all these experiments, the optimal rule, minimum distance classification, required knowledge only of the distribution means. Recall that during the prototype learning session subjects were presented with and encouraged to learn the distribution means. They therefore had available enough information to achieve optimal performance before beginning the first practice block. We suspect that with sufficient practice, subjects would have gradually shifted toward the optimal rule but it definitely appears that, although subjects can learn minimum distance classification, they first try even simpler rules (e.g., is $y - x$ positive or negative?). Only when their performance with very simple decision rules becomes noticeably inadequate (according to whatever criterion is relevant) do they exert the effort required to learn more complex rules.

Experiment 3

Introduction

Experiments 1 and 2 provide evidence that subjects spontaneously combine or integrate information when classifying the Figure 2 stimuli. We now try to determine the nature of this integration process. Recall that the three types of integration models discussed in the introduction were (a) the minimum distance model, (b) the optimal linear classifier, and (c) the ideal observer model. With the stimulus distributions used in the first two experiments all three of these models make the same predictions and so neither experiment can tell us which of the three best describes human classification. Therefore, in Experiment 3, stimulus parameters were chosen so that the minimum distance model predicts a different boundary from the optimal linear and the ideal observer models.

In our earlier discussion of optimal linear and minimum distance classification we presented the distributions associated with Figure 4 as an example in which the two models make differential predictions. We thus chose to use the Figure 4 contours of equal probability in Experiment 3. The exact parameter values are given in Table 5. Note that, although the means on the vertical dimension are identical, both stimulus classes described in Figure 4 are characterized by a large positive correlation between the lengths of the vertical and horizontal segments (hence the tilt in the major axes). Because of this correlation, the vertical length is informative and so should not be ignored.

Both the optimal linear and the ideal observer models are sensitive to stimulus component correlation and so they predict the solid-line decision boundary shown in Figure 4. Minimum distance models, on the other hand, are sensitive only to stimulus means (i.e., the prototypes) and so they predict the dotted-line decision boundary of Figure 4.

Note that the minimum distance rule of Figure 4 is similar to the one that subjects had difficulty learning in Experiment 2. In both cases, the minimum distance model predicts that subjects will respond as if ignoring one of the two components. In Experiment 2, subjects had some difficulty ignoring the horizontal component. In addition, we found evidence that some subjects in that experiment initially adopted a rule similar to the optimal linear rule of Figure 4 (especially Subject 3). One interpretation of Experiment 2 then, is that minimum distance classification is not a natural constraint on human pattern classification. Even when it was optimal, subjects had difficulty applying the minimum distance criterion.

What then can we learn from Experiment 3? If subjects have a predilection to minimum distance classification but for some reason have difficulty applying it in the Experiment 2 case then we should expect the same pattern of results in the two experiments. On the other hand, if subjects are sensitive to correlation between stimulus components, as ideal observer or optimal linear classification requires, then the decision bound adopted by subjects in Experiment 3 should be significantly closer to the line $y = x$ than in Experiment 2.

Table 5
Stimulus Parameter Values: Experiment 3

	S_A	S_B
Horizontal M	450	550
Vertical M	500	500
Horizontal SD	122	122
Vertical SD	107	107
Horizontal–vertical covariance	11,475	11,475

Note. The units here are arbitrary screen units in which there are about 270 units per degree of visual angle.

Results and Discussion

To verify that the stimulus samples had the desired statistical properties, the percentage of samples that a subject who perfectly used the optimal rule $y = x$ could correctly classify was computed for each subject and compared with the percentage that could be classified by the minimum distance rule $x = 500$. The results are given in Table 6. Note that optimal responding will increase accuracy at least 20 percentage points relative to minimum distance classification. Subjects capable of detecting the horizontal–vertical correlation could therefore substantially increase their response accuracy. Table 6 also contains the equation of the best fitting linear bounds and summarizes other results. Figure 10 shows the responses of 1 of the 3 subjects along with the optimal decision bound $y = x$.

Note first that, for all 3 subjects, the best fitting linear bound is very close to the optimal bound, $y = x$. In fact, for Subjects 2 and 3 it is virtually identical (except that both subjects exhibit a small effect of the vertical–horizontal illusion). In addition, although none of the subjects applied the optimal rule with as much consistency as the subjects in Experiment 1, the optimal bound does account for a large percentage of their responses. On the other hand, in each case, the minimum distance bound accounts for more than 20% fewer responses. By comparing the slopes of the best fitting linear bound for each subject in Experiment 3 (.89, 1.0, and 1.0) with the slopes from each subject of Experiment 2 (.66, 0.0, .92, .31, and .90), it is also clear that the decision rule used by subjects of Experiment 3 was closer to the rule $y = x$ than was the decision rule used by subjects of Experiment 2. It therefore seems clear that all 3 subjects of Experiment 3 were sensitive to the horizontal–vertical correlation that, in this case, makes minimum distance classification a relatively poor response strategy.[7]

[7] Another possibility is that subjects initially adopted the rule $y = x$ and then were reinforced by feedback. Although this possibility cannot be ruled out, it seems unlikely because subjects were shown the prototypes and their labels at the beginning of both the practice and experimental sessions.

Table 6
Results of Experiment 3

Subject	% of stimuli accounted for by optimal bound	% of stimuli accounted for by bound $x = 500$	% Correct	Best fitting linear bound	% of responses accounted for by best fitting bound	% of responses accounted for by optimal bound	% of responses accounted for by bound $x = 500$
1	88	66	85	$y = .89x + 37$	95	92	70
2	86	65	84	$y = x - 2$	91	89	67
3	88	66	86	$y = x - 5$	94	93	70

Note. The best fitting bound minimizes the sum of the distances from all incorrectly predicted responses to the decision bound.

An experiment reported by Ashby and Perrin (1988) that used the same experimental paradigm introduced here provides additional evidence against minimum distance classification. In Ashby and Perrin's experiment, stimuli were horizontal lines that differed only in length and so the S_A and S_B distributions were unidimensional. Stimulus parameters were selected so that the two distributions had equal means but different variances. Thus, any model that assumes subjects classify by computing the distance to each prototype must predict chance performance under these conditions. Subjects in this experiment responded at levels significantly better than chance. In fact, their performance was close to optimal (80% correct), thus falsifying any minimum distance model.

Of all the decision rules described in this article, the two most difficult to distinguish are the ideal observer and optimal linear models. We saw earlier that if the perceptual distributions are multivariate normal, the models make divergent predictions only when the two distributions have different covariance matrices.

Pilot studies indicate that quadratic bounds are difficult for subjects to learn. In a task where the ideal quadratic classifier predicted 80% accuracy and the most accurate linear classifier

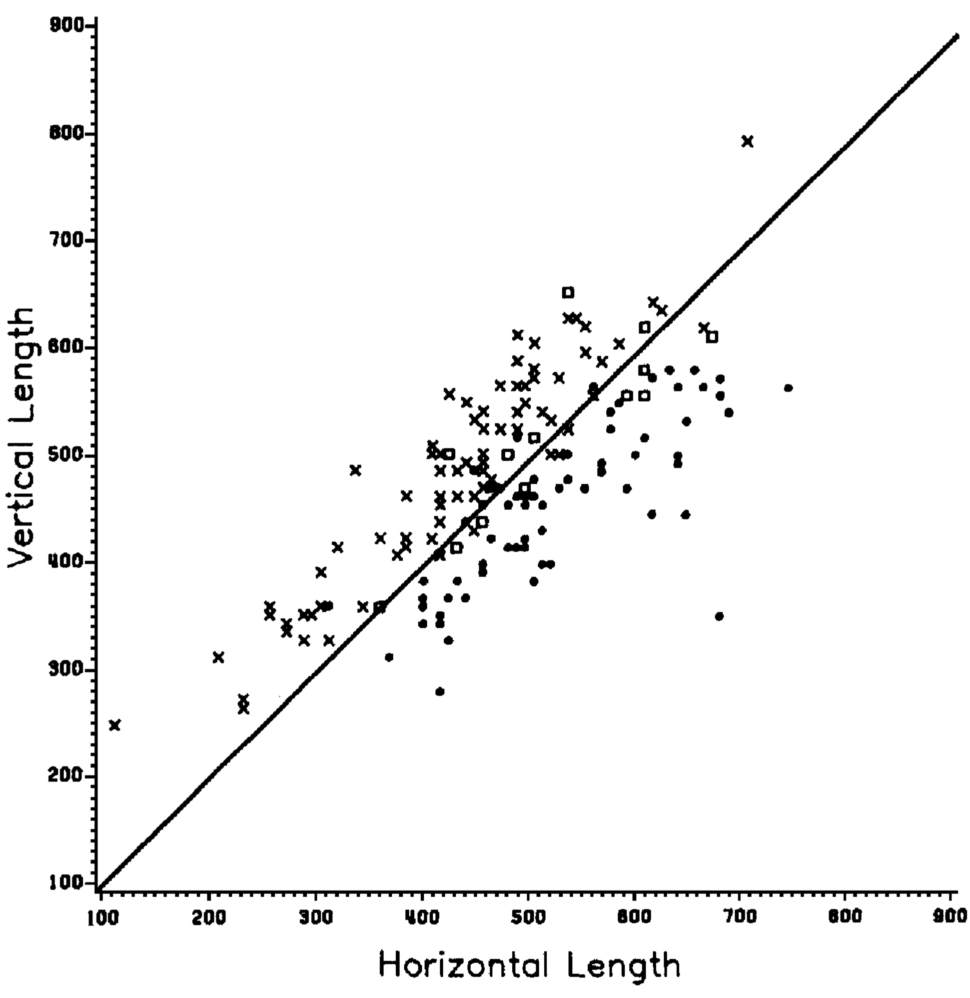

Figure 10. Response data for Subject 2 in Experiment 3. (An x indicates an R_A response, a filled circle an R_B response, and a square indicates response inconsistency.)

predicted 73% (about the same difference that existed between the integration and independent decisions models in Experiment 1), 3 subjects failed to adopt the quadratic bound, even after 4 days' experience. Instead they consistently responded according to the most accurate linear rule.

Does this mean humans are constrained by linear classification? There are several reasons we believe this not to be the case. First, in other pilot studies with different stimulus distributions, we found that subjects could adopt a nonlinear decision rule. For example, when the contours of equal probability were circles of different diameter (i.e., each covariance matrix was a different scalar multiple of the identity), subjects learned to place a closed boundary around the prototype of the less variable stimulus class. If a sample fell within this closed region, the response associated with the less variable class was made whereas a sample falling outside this region evoked the opposite response.

Medin and Schwanenflugel (1981) also reported evidence that humans are not constrained by linear classification. As we mentioned earlier, they presented subjects with sets of stimuli consisting of four binary valued components, some of which were possible to perfectly categorize with a linear decision rule. Medin and Schwanenflugel found that, for the stimulus sets they chose, the linearly separable stimuli were no easier to categorize than stimulus sets that were not linearly separable. The degree to which human categorization depends on nonlinear decision rules demands further study.

Summary and Conclusions

The goal of this article has been to theoretically characterize and to empirically investigate decision rules in the perception and categorization of stimuli composed of several separate components. Toward this end the general recognition theory was described, which assumes only that a single presentation of a stimulus induces a random perceptual effect $\mathbf{x}_s$ in a multidimensional space and that a response function $r(\mathbf{x}_s)$ somehow then selects a response.

This article makes several contributions. First, by using the general recognition theory, we showed how it is possible to formally characterize competing decision models in a way that makes them easy to discriminate. This is done by specifying the type of decision bound they predict in the multidimensional perceptual space. For example, we saw that independent decisions models always predict decision bounds that are parallel to the coordinate axes. This translation makes it possible to directly compare models expressed in very different languages. As an example, this approach indicates that the decision rules implicit in prototype-based MDS models are identical to the decision rules of distributed memory models that use correlation and convolution operations.

Second, to empirically investigate the response function $r(\mathbf{x}_s)$, the general recognition randomization technique was developed. In this paradigm the unobservable perceptual variability postulated by the general recognition theory is replaced by external stimulus variability controlled by the experimenter. The nature of the response function $r(\mathbf{x}_s)$ is easily determined by noting, for a given sample $\mathbf{x}_s$, which response was selected.

Finally, three experiments that used this technique were reported and three conclusions stand out. First, we found strong evidence that subjects used deterministic decision rules. In this sense our results corroborate findings obtained with the numerical decision task (e.g., Kubovy et al., 1971). In Experiment 1, a deterministic rule accounted for at least 95% of each subject's responses. The few unpredicted responses were to stimuli lying very near the decision bound. We believe the random responding here was due to the small amount of unobservable perceptual noise our technique did not control. The deterministic responding of subjects in Experiment 1 cannot be predicted by many popular models of human pattern classification (e.g., the context model of Medin & Schaffer, 1978). More response variability was displayed in the other experiments but we believe this was either because they required an internal referent, which induced criterial noise (as in Experiment 2) or because they required subjects to learn about second-order parameters of the stimulus distributions in addition to learning about the stimulus prototypes (as in Experiment 3).

This finding does not imply that subjects never use probabilistic response functions. In fact, the extensive literature on probability learning (e.g., Estes, 1976) clearly indicates that under certain circumstances, subjects do use probabilistic rules. However, we feel that the result suggests that in many perceptual tasks, in particular, in classification and identification, subjects may often be employing a much more deterministic response function than has been commonly believed. We believe that deterministic responding has been difficult to detect because it is almost always accompanied by perceptual and criterial noise. The combination of deterministic responding and internal noise is, at present, empirically indistinguishable from probabilistic responding. Experiment 1 was successful because it minimized the internal noise. Of course, more research is needed to test the generalizability of this hypothesis.

The second important finding was that subjects used decision rules that were very nearly optimal. This was especially true in Experiments 1 and 3. An unresolved issue concerns the role of decision bounds with a slope of 1.0 in this conclusion. Subjects had an easier time learning decision rules with a slope of 1.0 and it may be that when they are consistently required to learn decision bounds with different slopes they may resort to nonoptimal strategies.

The third striking conclusion to be drawn from these data is that there appear to be few natural constraints on human pattern classification. Humans are not constrained to make independent decisions and in fact, in Experiment 2 they had difficulty doing so even when such a strategy was to their advantage. They are not constrained by minimum distance classification and, although further research is required, we believe that they are not constrained by linear classification.

A plausible model suggested by the experiments reported here is that subjects first choose very simple deterministic decision rules that are easy to implement and place a minimal burden on available cognitive capacities. An example is the linear rule,

$$\text{if } y - x > 0 \text{ then respond } R_A, \text{ otherwise respond } R_B,$$

which requires no internal referent. If, by whatever the relevant criterion, the subject decides that this rule produces satisfactory performance, then it will be retained even if nonoptimal. On the other hand, if the simple rule leads to inadequate performance, the subject will expend a little more effort and adopt a more accurate rule.

The conclusions drawn from these experiments must be somewhat restricted because of the nature of the Figure 2 stimuli and because of the specific stimulus distributions that were chosen. We did replicate the near-optimal responding found in Experiment 1 with semicircles of varying radius with a spoke whose angle of rotation varied. This result strengthens the conclusion that subjects are often able to integrate information from separable stimulus components. More experimentation is needed with other stimulus components and with a wider array of stimulus distributions in order to test the generalizability of our other findings. We believe however, that the general recognition randomization technique, together with our characterization of decision models according to the type of decision boundary they predict, provide a powerful framework within which to conduct this experimentation.

References

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review, 95,* 124–150.

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

Blackwell, H. R. (1963). Neural theories of simple visual discriminations. *Journal of the Optical Society of America, 53,* 129–160.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review, 83,* 37–64.

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

Foard, C. F., & Nelson, D. G. K. (1984). Holistic and analytic modes of processing: The multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General, 113,* 94–111.

Fukunaga, K. (1972). *Introduction to statistical pattern recognition.* New York: Academic Press.

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

Gravetter, F., & Lockhead, G. R. (1973). Criterial range as a frame of reference for stimulus judgment. *Psychological Review, 80,* 203–216.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Greeno, J. G., & Steiner, T. E. (1964). Markovian processes with identifiable states: General considerations and applications to all-or-none learning. *Psychometrika, 29,* 309–333.

Hammerton, M. (1970). An investigation into changes in decision criteria and other details of a decision-making task. *Psychonomic Science, 21,* 203–204.

Healy, A. F., & Kubovy, M. A. (1977). A comparison of recognition memory to numerical decision: How prior probabilities affect cutoff location. *Memory & Cognition, 5,* 3–9.

Hefner, R. A. (1958). *Extensions of the law of comparative judgment to discriminable and multidimensional stimuli.* Unpublished doctoral dissertation, University of Michigan.

Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory.* Hillsdale, NJ: Erlbaum Associates.

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

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

Kubovy, M., Rapoport, A., & Tversky, A. (1971). Deterministic vs. probabilistic strategies in detection. *Perception & Psychophysics, 9,* 427–429.

Lee, W. (1963). Choosing among confusably distributed stimuli with specific likelihood ratios. *Perceptual and Motor Skills, 16,* 445–467.

Lee, W., & Janke, M. (1964). Categorizing externally distributed stimulus samples for three continua. *Journal of Experimental Psychology, 68,* 376–382.

Lee, W., & Janke, M. (1965). Categorizing externally distributed stimulus samples for unequal molar probabilities. *Psychological Reports, 17,* 79–90.

Lee, W., & Zentall, T. R. (1966). Factorial effects in the categorization of externally distributed stimulus samples. *Perception & Psychophysics, 1,* 120–124.

Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review, 79,* 410–419.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbood of mathematical psychology* (Vol. 1, pp. 103 - 189). New York: Wiley.

Martin, R. C., & Caramazza, A. (1980). Classification in well-defined and ill-defined categories: Evidence for common processing strategies. *Journal of Experimental Psychology: General, 109,* 320–353.

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

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 335–368.

Morrison, D. F. (1976). *Multivariate statistical methods* (2nd ed.). New York: McGraw-Hill.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Nilsson, N. J. (1965). *Learning machines.* New York: McGraw-Hill.

Noreen, D. L. (1981). Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology* (pp. 237–279). Providence, RI: American Mathematical Society.

Nosofsky, R. M. (1983). Information integration and the identification of stimulus noise and criterial noise in absolute identification. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 299–309.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114.

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

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

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic Press.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 491–502.

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

chology, 14, 353–409.

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

Sperling, G. (1984). A unified theory of attention and signal detection. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 103–181). New York: Academic Press.

Tanner, W. P. (1956). Theory of recognition. *The Journal of the Acoustical Society of America, 28,* 882–888.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–286.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 834–864.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes.* New York: Cambridge University Press.

Townsend, J. T., Hu, G. G., & Ashby, F. G. (1980). A test of visual feature sampling independence with orthogonal straight lines. *Bulletin of the Psychonomic Society, 15,* 163–166.

Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). Perceptual sampling of orthogonal straight line features. *Psychological Research, 43,* 259–275.

Townsend, J. T., & Landon, D. E. (1983). Mathematical models of recognition and confusion in psychology. *Mathematical Social Sciences, 4,* 25–71.

Wandell, B. A. (1982). Measurement of small color differences. *Psychological Review, 89,* 281–302.

Ward, L. M. (1973). Use of Markov-encoded sequential information in numerical signal detection. *Perception & Psychophysics, 14,* 337–342.

Weissmann, S. M. Hollingsworth, S. R., & Baird, J. C. (1975). Psychophysical study of numbers: III. Methodological applications. *Psychological Research, 38,* 97–115.

Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology, 5,* 102–122.

Zinnes, J. L., & MacKay, D. B. (1983). Probabilistic multidimensional scaling: Complete and incomplete data. *Psychometrika, 48,* 27–48.

Received December 16, 1985
Revision received February 9, 1987
Accepted January 26, 1987 ■

Gallup Appointed Editor of the *Journal of Comparative Psychology*, 1989–1994

The Publications and Communications Board of the American Psychological Association announces the appointment of Gordon G. Gallup, Jr., State University of New York at Albany, as editor of the *Journal of Comparative Psychology* for a 6-year term beginning in 1989. As of January 1, 1988, manuscripts should be directed to

Gordon G. Gallup, Jr.
Department of Psychology
State University of New York at Albany
Albany, New York 12222

Manuscript submission patterns for the *Journal of Comparative Psychology* make the precise date of completion of the 1988 volume uncertain. The current editor, Jerry Hirsch, will receive and consider manuscripts until December 31, 1987. Should the 1988 volume be completed before that date, manuscripts will be redirected to Gallup for consideration in the 1989 volume.
