

# On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization

W. TODD MADDOX  
*University of Texas, Austin, Texas*

Averaging across observers is common in psychological research. Often, averaging reduces the measurement error and, thus, does not affect the inference drawn about the behavior of *individuals*. However, in other situations, averaging alters the structure of the data qualitatively, leading to an incorrect inference about the behavior of individuals. In this research, the influence of averaging across observers on the fits of decision bound models (Ashby, 1992a) and generalized context models (GCM; Nosofsky, 1986) was investigated through Monte Carlo simulation of a variety of categorization conditions, perceptual representations, and individual difference assumptions and in an experiment. The results suggest that (1) averaging has little effect when the GCM is the correct model, (2) averaging often improves the fit of the GCM and worsens the fit of the decision bound model when the decision bound model is the correct model, (3) the GCM is quite flexible and, under many conditions, can mimic the predictions of the decision bound model, whereas the decision bound model is generally unable to mimic the predictions of the GCM, (4) the validity of the decision bound model's perceptual representation assumption can have a large effect on the inference drawn about the form of the decision bound, and (5) the experiment supported the claim that averaging improves the fit of the GCM. These results underscore the importance of performing single-observer analysis if one is interested in understanding the categorization performance of individuals.

The ability to categorize quickly and accurately is fundamental to survival. Everyday, we make hundreds of categorization judgments. Several detailed theories and quantitative models have been proposed to account for the perceptual and cognitive processes involved in categorization, the goal being to understand the categorization performance of *individual* behaving organisms. Despite the emphasis on individuals, it is common to analyze aggregate data—that is, data averaged across observers.

The use of aggregate data to inform us about the nature of individual behavior is common in psychology. The effects of averaging on psychological inquiry have been studied in some domains (Ashby, Maddox, & Lee, 1994; Estes, 1956; Massaro & Cohen, 1993; Siegler, 1987), but, in general, it is assumed that averaging does not affect the underlying structure of the data. Specifically, the implicit assumption is that the data represent a measure of

some psychological construct plus measurement error, which may be reduced by averaging. Under these conditions, averaging will affect only the measurement error, and inferences about individual behavior can be drawn meaningfully from the aggregate data.

However, another possibility exists. Averaging can alter the structure of the data in such a way that the correct model of individual behavior provides a poor account of the aggregate data and an incorrect model of individual behavior provides a good account of the aggregate data. Under these conditions, one might infer (erroneously) that the incorrect model provides a meaningful description of individual behavior. For example, if several observers show all-or-none learning but on different learning trials, each individual learning profile would be well described by a step function, and an all-or-none learning model would provide the best description of individual behavior. The aggregate profile, on the other hand, would reveal a gradual increase in learning. The aggregate data would be best described by an incremental learning model and would be poorly described by an all-or-none model, even though each individual showed all-or-none learning. In this case, the analysis of aggregate data leads to an incorrect inference regarding individual behavior.

The main focus of the present research was to examine the effects of averaging on two classes of categorization models: decision bound models (Ashby, 1992a; Ashby & Maddox, 1993; Maddox, 1995; Maddox & Ashby, 1993, 1996) and generalized context models (GCMs; McKinley & Nosofsky, 1995, 1996; Nosofsky, 1986, 1989, 1992).

---

Portions of the research were presented at the 28th Annual Mathematical Psychology Meeting. This research was supported in part by National Science Foundation Grant SBR-9514331. Special thanks go to F. Gregory Ashby and David Brainard for tutoring on color system transformations and help in generating the perceptually integral distributions. Thanks go to Sergei Bogdanov, Jason Huntington, Erin Jennings, and Patrick Simmons for help in empirical data collection. Helpful comments and suggestions on an earlier version of this article were provided by Greg Ashby, John Kihlstrom, Lester Krueger, Štephan Lewandowsky, Robin Thomas, and an anonymous reviewer. Correspondence concerning this article should be addressed to W. T. Maddox, Department of Psychology, Mezes Hall 330, University of Texas, Austin, Texas 78712 (e-mail: maddox@psy.utexas.edu).

These two model classes were chosen for three reasons. First, these are two of the most successful models of categorization (see, e.g., Maddox & Ashby, 1993; McKinley & Nosofsky, 1995, 1996). Second, previous research suggests that averaging may have a large effect on the structure of data generated from these two models (Ashby, Lee, & Balakrishnan, 1992; Ashby et al., 1994; Maddox & Ashby, 1993). Finally, the implications of this work are extensive, because variants of the GCM have been applied in perceptual and cognitive psychology (Lamberts, 1995; McKinley & Nosofsky, 1995, 1996; Nosofsky, 1986, 1987, 1989, 1991, 1992), social psychology (E. R. Smith & Zarate, 1992), developmental psychology (L. B. Smith, 1989), and several other areas of psychology. Because most of these studies used aggregate data, the present results are extremely relevant.

Two factors have been identified that exert a large influence on categorization performance. First, the form of the categorization rule can have a large effect on performance. In particular, a distinction is often made between categorization rules that require (decisional) selective attention<sup>1</sup> and those that require attention to all stimulus dimensions (Ashby & Lee, 1991; Ashby & Maddox, 1994; McKinley & Nosofsky, 1996; Nosofsky, 1986). Second, the nature of the perceptual representation can have a large effect on performance. In particular, a distinction is made between stimulus dimensions that are perceptually separable and those that are perceptually integral (see, e.g., Ashby & Maddox, 1994; Ashby & Townsend, 1986; Garner, 1974; Maddox, 1992; Shepard, 1964). In short, perceptually separable dimensions are characterized by fairly simple perceptual representations, whereas perceptually integral dimensions yield complex perceptual representations (see details below). Interestingly, some research suggests an interaction between the nature of the perceptual representation and the ability to attend selectively. For example, Nosofsky (1987; see, also, Garner, 1974; McKinley & Nosofsky, 1996) suggests that selective attention is more difficult with integral-dimension than with separable-dimension stimuli.

Monte Carlo simulations were used in the present study to investigate the effects of averaging on fits of the decision bound model and the GCM when perceptual separability or perceptual integrality was satisfied and when selective attention or attention to all dimensions was required. Two sets of Monte Carlo simulations were performed. In the first set, data were generated from a linear decision bound model, and both the GCM and the decision bound models were applied to the individual and aggregate observer data. In the second set, data were generated from the GCM, and both the GCM and the decision bound models were applied to the individual and aggregate observer data. An experiment was also conducted to test the empirical validity of the simulation results.

The next (second) section briefly outlines the basic assumptions of the GCM and decision bound models and outlines the separability–integrality distinction within each framework. The third section reviews the previous research devoted to the issue of averaging across ob-

servers. The fourth and fifth sections are devoted to the simulations. The sixth section describes an empirical application, and the final section summarizes the findings and discusses the implications for future research.

## MODEL ASSUMPTIONS

The models are outlined only briefly here. Detailed discussions can be found in numerous articles (e.g., Ashby, 1992a; Ashby & Maddox, 1993, 1994; Maddox, 1992; Maddox & Ashby, 1993, 1996, 1998; McKinley & Nosofsky, 1995, 1996; Nosofsky, 1986, 1987, 1998).

### Generalized Context Model

The GCM assumes that each exemplar is represented by a point in some multidimensional scaling (MDS) psychological space. In other words,

$$\mathbf{x}_{pi} = \mathbf{x}_i, \quad (1)$$

where  $\mathbf{x}_i$  represents the location in the MDS psychological space assigned to Stimulus  $i$  and  $\mathbf{x}_{pi}$  is the perceptual effect on any given trial. In other words, repeated presentations of the same stimulus yield the same perceptual effect on each trial. Thus, the perceptual effect across trials can be represented by a single point,  $\mathbf{x}_i$ . In a simple two-category problem, when required to categorize a target exemplar, the probability of responding Category A is given by the summed similarity of the target exemplar to all Category A exemplars stored in memory divided by the summed similarity to all relevant category exemplars stored in memory. In other words, the probability that Stimulus  $i$  is classified as a member of Category A,  $P(A|i)$ , is given by

$$P(A|i) = \frac{\beta \left( \sum_{j \in C_A} \eta_{ij} \right)^\gamma}{\beta \left( \sum_{j \in C_A} \eta_{ij} \right)^\gamma + (1-\beta) \left( \sum_{j \in C_B} \eta_{ij} \right)^\gamma}, \quad (2)$$

where  $j \in C_j$  represents all stored exemplars of Category  $J$ ,  $\eta_{ij}$  is the similarity of Stimulus  $i$  to exemplar  $j$ ,  $\beta$  (which ranges from 0 to 1) is the bias toward response A, and  $\gamma = 1$  (the  $\gamma$  parameter will be discussed shortly). Similarity is assumed to be a monotonically decreasing function of interpoint distance where

$$\eta_{ij} = e^{-d_{ij}^\theta}. \quad (3)$$

When  $\theta = 1$ , the exponential decay similarity function results, and when  $\theta = 2$ , the Gaussian similarity function results (Shepard, 1987). Assuming two perceptual dimensions,  $x_1$  and  $x_2$ , the interpoint distance is given by

$$d_{ij} = c \left[ w |x_{1i} - x_{1j}|^r + (1-w) |x_{2i} - x_{2j}|^r \right]^{\frac{1}{r}}, \quad (4)$$

where  $w$  (which ranges from 0 to 1) is a parameter that represents the amount of selective attention to Dimension 1,  $c$  is a scaling constant, and  $r$  determines the distance

metric.<sup>2</sup> When  $r = 1$ , the city-block metric results, and when  $r = 2$ , the Euclidean metric results.

Ashby and Maddox (1993) proposed the deterministic exemplar model (DEM) that uses the Equation 2 response rule but leaves  $\gamma$  as a free parameter to be estimated from the data. The  $\gamma$  parameter allows the model to predict responding that is more deterministic (i.e., less based on probability matching) than that predicted by the GCM. Notice that the GCM is a special case of the DEM in which  $\gamma = 1$ . The GCM predicts probability matching because the GCM response probabilities depend exclusively on the relative strength of evidence for the two categories. In the DEM, when  $\gamma > 1$ , overmatching is predicted, and responding is more like that predicted from an all-or-none model. In other words, more extreme response probabilities (near 0 or 1) are predicted. When  $\gamma < 1$ , undermatching is predicted, and responding is less deterministic than that predicted by the GCM. In other words, response probabilities near .5 are predicted.

Within the framework of the GCM (and the DEM), the integrality or separability of the stimulus dimensions is determined by the nature of the distance metric in Equation 4. Specifically, stimulus dimensions are defined as separable when a city-block distance metric (i.e.,  $r = 1$ ) is used and as integral when a Euclidean distance metric (i.e.,  $r = 2$ ) is used (Maddox, 1992).

### Decision Bound Model

Decision bound models assume that, because of stimulus noise, sensory noise, and perceptual noise, there is trial-by-trial variability in the perceptual representation of all stimuli (Ashby & Lee, 1993). Thus, over trials, a stimulus is represented by a multivariate probability distribution of perceptual effects. Let the vector  $\mathbf{x}_i$  represent the observer's mean perceptual effect for Stimulus  $i$ . Because of these noise sources, the observer's percept of Stimulus  $i$ , on any trial, is represented by

$$\mathbf{x}_{pi} = \mathbf{x}_i + \mathbf{e}_{pi}, \quad (5)$$

where  $\mathbf{e}_{pi}$  is generally assumed to be multivariate normally distributed with mean vector 0 and covariance matrix  $\Sigma_{pi}$ . In other words, repeated presentations of the same stimuli yield, as a rule, different perceptual effects. Thus, over trials, the stimulus is represented by a distribution of perceptual effects. The experienced observer divides up the perceptual space into separate regions and assigns a category response to each region. The partitions between the separate category response regions are called *decision bounds*. On each trial, the observer determines into which region the perceptual effect falls and emits the associated response. Notice that the information accessed on each trial to be used to generate a categorization response is very different from that postulated in the GCM. In the GCM, a categorization response is generated by accessing all exemplar information in memory on each trial. In the decision bound model, trial-by-trial exemplar information is important in the assignment of responses to regions, but exemplar information stored in memory does not need to be accessed in order for a categorization re-

sponse to be generated. The experienced observer uses the following decision rule:

$$\begin{aligned} &\text{If } h(\mathbf{x}_{pi}) + e_c < 0; \text{ then respond "A";} \\ &\text{otherwise, respond "B",} \end{aligned} \quad (6)$$

where  $h$  determines the shape and location of the decision bound, and  $e_c$  (normally distributed with mean 0 and variance  $\sigma_c^2$ ) represents the effects of criterial noise. Criterial noise implies that the memory for the location of the decision bound can vary from trial to trial. For simplicity, no criterial noise was assumed (i.e.,  $e_c = 0$ , and  $\sigma_c^2 = 0$ ). Depending on the nature of the function  $h$ , the decision bound can take on one of many forms. In this article, we restrict attention to either a linear or a quadratic decision bound. A case in which the decision bound is linear is depicted in Figures 1b and 1d. A case in which the decision bound is quadratic is depicted in Figures 1c and 1e.

Within the framework of the decision bound model, the separability or integrality of the stimuli is determined by the relationship among the perceptual distributions for each stimulus (see, e.g., Ashby, 1988; Ashby & Lee, 1991; Ashby & Maddox, 1991, 1994; Ashby & Townsend, 1986; Maddox, 1992; Maddox & Ashby, 1996). Perceptual separability is satisfied in Figures 1d and 1e. Specifically, perceptual separability holds along Dimension 1 if the perceptual effects along Dimension 1 are unaffected by the level along Dimension 2—in other words, if the marginal perceptual distributions along Dimension 1 are identical across levels of Dimension 2 (see Ashby & Townsend, 1986, for details). Perceptual integrality denotes a violation of perceptual separability and holds in Figures 1b and 1c.

Before proceeding, it is worth discussing briefly the important, but often misunderstood, difference between the operational definitions of integrality and separability assumed by the GCM (see, e.g., Garner, 1974) and perceptual separability and perceptual integrality as defined in decision bound theory (Ashby & Townsend, 1986). The integrality and separability of a set of stimuli, as operationally defined, is based on the results of a series of experimental tests. For example, if the city-block distance metric ( $r = 1$ ) provides a better account of a set of similarity ratings than the Euclidean distance metric ( $r = 2$ ), the stimuli are defined as separable. If the Euclidean distance metric provides a better account of a set of similarity ratings than the city-block distance metric, the stimuli are defined as integral. Perceptual separability and perceptual integrality, on the other hand, are theoretically motivated, rather than operationally defined (see Ashby & Maddox, 1994; Ashby & Townsend, 1986, for details).

If a pair of stimulus dimensions have been found to be separable, on the basis of a series of operational tests, in the GCM and DEM a city-block distance metric (Equation 4) and an exponential decay similarity function (Equation 3) generally will be assumed (Nosofsky, 1984). If a pair of stimulus dimensions have been found to be integral, on the basis of a series of operational tests, in the GCM and DEM a Euclidean distance metric and a

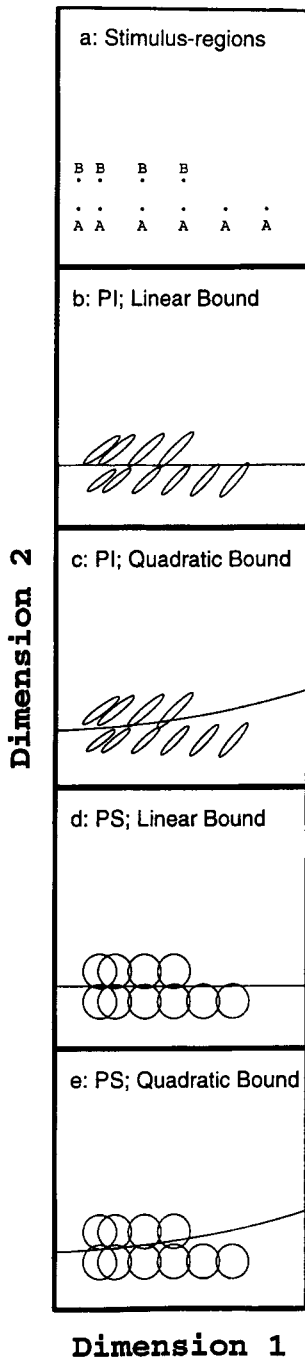


Figure 1. (a) A schematic representation of a typical categorization problem. The label "A" denotes stimuli assigned to Category A and the label "B" denotes stimuli assigned to Category B. Hypothetical contours of equal likelihood for this categorization problem when (b) a complex perceptual integrality and a linear decision bound are assumed, (c) a complex perceptual integrality and a quadratic decision bound are assumed, (d) a stimulus-invariant,  $\sigma^2\mathbf{I}$  perceptually separable representation and a linear decision bound are assumed, and (e) a stimulus-invariant,  $\sigma^2\mathbf{I}$  perceptually separable representation and a quadratic decision bound are assumed. PS, perceptual separability; PI, perceptual integrality.

Gaussian similarity function generally will be assumed. If the marginal perceptual distributions along a stimulus dimension are equal, perceptual separability is satisfied. This requires that the means and variances be equal. If the means or variances are not equal, perceptual separability is violated.

Often, when applying the decision bound model to categorization data, a simple perceptually separable representation is assumed. Specifically, it is often assumed that the perceptual variance is equal across dimensions and stimuli and that the covariance terms are zero. This is termed a *stimulus-invariant,  $\sigma^2\mathbf{I}$  perceptually separable representation* because the perceptual distribution is identical for all stimuli (i.e., stimulus invariant), and the perceptual covariance matrix is a scalar multiple of the identity matrix (i.e.,  $\sum p_i = \sum p = \sigma^2\mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix). Under these conditions, and assuming two perceptual dimensions, the contours of equal likelihood that describe the perceptual distributions are always circles with a constant diameter, and perceptual separability is satisfied. Consider, for example, the categorization problem depicted in Figure 1a. In this problem, there are 10 two-dimensional stimuli, 6 assigned to Category A, and 4 to Category B. In Figures 1d and 1e, a stimulus-invariant,  $\sigma^2\mathbf{I}$  perceptually separable representation is assumed. Figures 1b and 1c depict a much more complex perceptual representation in which stimulus invariance is violated (i.e., the contours of equal likelihood differ across stimuli) and the perceptual variances differ for the two dimensions. In this case, perceptual integrality holds.

Many (but not all) applications of the decision bound model assumed a stimulus-invariant,  $\sigma^2\mathbf{I}$  perceptually separable representation. (For applications in which this simple representation was not assumed, see Ashby & Lee, 1991; Maddox & Ashby, 1996.) This simple representation has (at least) two important properties. First, without additional evidence to suggest a more complex perceptual representation, this is the most parsimonious initial assumption. Second, and perhaps more important, this very simple perceptual representation produces a similarity function that is equivalent to the measure used by the GCM and DEM when equal amounts of attention are allocated to each stimulus dimension—that is, when  $w = 0.5$  from Equation 4 (Ashby & Perrin, 1988; see, also, Ashby & Maddox, 1993). It is important to note that this assumption may be incorrect in some cases. In particular, this representation may be inappropriate when the stimulus dimensions are characterized by complex perceptual integralities, exposure duration is short, or response time is emphasized over accuracy. Thus, it is important to understand the effects of complex perceptual integralities on the ability of the models to account for categorization data.

## EFFECTS OF AVERAGING

Recently, Ashby et al. (1994) examined the influence of averaging on fits of MDS models to similarity ratings (see, e.g., Kruskal, 1964a, 1964b; Shepard, 1962a, 1962b;

Torgerson, 1958) and on fits of the similarity choice model (SCM) to identification data (Luce, 1963; Shepard, 1957). Similarity ratings were generated from a model that was incompatible with the distance-based MDS model (see, e.g., Shepard, 1957). Specifically, similarity ratings were generated from a model for which the value of  $r$  from Equation 4 was less than 1. Distance-based models are valid only when  $r \geq 1$ . When the value of  $r \geq 1$ , the triangle inequality holds. (The triangle inequality states that the distance between two stimuli,  $i$  and  $j$ , plus the distance between the stimuli  $j$  and  $k$  must be greater than or equal to the distance between stimuli  $i$  and  $k$ , for any three stimuli,  $i$ ,  $j$ , and  $k$ .) When  $r < 1$ , the triangle inequality is violated. As a result, the MDS model provided a poor account of the individual observer data (generated from a model incompatible with the MDS model). However, when the similarity ratings were averaged, the model that generated the individual observer data provided a poor account of the averaged data, and the distance-based MDS model provided an excellent account of the averaged data. Ashby et al. (1994) concluded that this result was due to the fact that the averaging operation artificially increased the symmetry in the data (i.e., increased the likelihood that the similarity between the stimuli  $i$  and  $j$  was equal, or nearly so, to the similarity between the stimuli  $j$  and  $i$ ), thus eliminating violations of the triangle inequality. In the individual observer data, there were 33 triangle inequality violations on average (out of a possible 84 in each data set). In the averaged data, on the other hand, there were no violations of the triangle inequality. Similar results were obtained for the identification analyses. Specifically, poor fits of the SCM were obtained for individual observer data, but excellent fits of the SCM were obtained for the averaged data (see, also, Ashby et al., 1992).

Although clearly relevant to the study of similarity and identification, these results also have important implications for categorization research, in particular applications of the GCM, for two reasons. First, the GCM assumes that category exemplars are represented perceptually as points in some MDS psychological space, where similarity is inversely related to interpoint distance. Second, the GCM utilizes the same response rule as does the SCM. In fact, the GCM is a special case of the SCM in which similarity is inversely related to distance in an MDS psychological space. It is likely, then, that the GCM will also benefit from averaging.

A study by Maddox and Ashby (1993) provides some support for this claim. Maddox and Ashby (1993) applied the GCM, and the linear and quadratic decision bound models (both of which assumed the stimulus-invariant,  $\sigma^2I$  perceptually separable representation, described above) to two data sets collected by Nosofsky (1986, 1989). Both studies (Nosofsky, 1986, 1989) used the same stimulus dimensions and approximately equivalent categorization conditions. The major difference between the two studies was that the Nosofsky (1986) observers were highly experienced, and the data were analyzed separately for each observer, whereas the Nosofsky (1989) observers were relatively inexperienced, and

the analyses were based on data collapsed across observers. Decision bound models generally provided a large improvement in fit over the GCM when applied to the single, highly experienced observer data, but the models provided nearly equivalent fits when applied to the averaged, novice observer data. These data suggest that the GCM might benefit from averaging. We turn now to a set of Monte Carlo simulations that test this hypothesis and attempt to outline the conditions under which averaging alters the structure of the data in this manner.

## DECISION BOUND MODEL SIMULATIONS

This section summarizes the simulation approach taken when the decision bound model was assumed to be correct. A series of Monte Carlo simulations were conducted to examine the effects of averaging across observers on the fits of the GCM, DEM, the linear decision bound model, and the quadratic decision bound model. Each simulated experiment utilized 10 hypothetical observers who behaved in accordance with decision bound theory. The "true" model (i.e., the model that accurately described the individual observer performance) assumed each observer compared the perceptual effect on each trial with a fixed linear decision bound that partitioned the perceptual space into a Category A region and a Category B region. Thus, each hypothetical observer used a single fixed linear decision bound on every trial. However, each observer used a different linear decision bound (i.e., the slope and intercept of the linear decision bound was different for each hypothetical observer). Two hundred *percepts* were sampled randomly from each perceptual distribution. Across simulations, assumptions about the categorization rule (selective attention or attention to both dimensions), perceptual representation (perceptual separability or perceptual integrality), and individual differences in the observers' decision bound were varied systematically. A summary of the conditions investigated is outlined in Table 1. The first column denotes the categorization condition, and the second and third columns denote the decision bound assumptions.

### Categorization Rules and Perceptual Representations

This report focused on two qualitatively different categorization rules. The (decisional) *selective attention* (SA) rule required that the observer place no importance, during the categorization decision process, on the dimensional value along one (irrelevant) stimulus dimension and set a criterion along the other (relevant) stimulus dimension. This is referred to as decisional selective attention because, although the observer may be aware (perceptually) of the value along the irrelevant dimension, it should not enter into the categorization decision. The *equal attention* (EA) rule required that the observer place approximately equal importance, during the categorization decision process, on each stimulus dimension.

Two different SA and EA categorization rules were utilized. For the perceptually integral simulations, the

**Table 1**  
**Decision Bound Assumptions for the Monte Carlo**  
**Simulations of the Linear Decision Bound Model**

Condition	Slope and Intercept Assumptions	
	(Approximately) Optimal Decision Bounds	Nonoptimal Decision Bounds
Perceptual Integrality Application		
Selective attention	SA; intercept variability	No SA; slope and intercept variability
Equal attention	Not applicable	Slope = 1.25, intercept = -24; slope and intercept variability
Perceptual Separability Application		
Selective attention	SA; intercept variability	No SA; slope and intercept variability
Equal attention	Slope = 1, intercept = 0; slope and intercept variability	Slope < or > 1, intercept < or > 0; slope and intercept variability

Note—The slope and intercept values specified are population mean decision bound parameters. In the perceptual integrality application the slope and intercept standard deviations were .4 and 4.0, respectively. In the perceptual separability application, the slope and intercept standard deviations were .2 and .4, respectively. SA, (decisional) selective attention.

stimulus–response mappings for the SA and EA categorization rules were similar to those from a study by McKinley and Nosofsky (1996; Experiment 1). The stimulus-to-response mappings are depicted in Figures 2a and 2b, where the “A” denotes stimuli assigned to Category A, the “B” denotes stimuli assigned to Category B, and the “T” denotes unassigned transfer items. The simulations were modeled after experiments in which observers participated in a training phase followed by a transfer phase. During the training phase, only the Category A and Category B exemplars were presented. During the transfer phase, these original training items were supplemented with a set of unassigned transfer items. These categorization problems were chosen not because they are special in any way, but rather because, in the original experiment, the stimuli were iso-hue Munsell color chips that varied in saturation and brightness, which are highly integral dimensions. As we will see shortly, knowledge of the Munsell color system and other color systems will be used to estimate perceptually integral representations for these stimuli. For the perceptually separable simulations, the stimulus–response mappings for the SA and EA categorization rules were similar to those from a study by Nosofsky (1989). The stimulus-to-response mappings are depicted in Figures 2c and 2d. These categorization problems were chosen because, in the original experiment, the stimuli were a semicircle with an embedded radial line, which are thought to be separable (however, see Ashby & Lee, 1991; Ashby & Maddox, 1990).

**Perceptual integrality application.** This application focused on two category structures used by McKinley and Nosofsky (1996). The stimuli were iso-hue Munsell color chips that varied in saturation and brightness, which are thought to be highly integral (Garner, 1974). To simulate data from a perceptually integral representation, one must define the nature of this representation. Specifically, the parameters of each perceptual distribution (i.e., the means, variances, and covariance) must be defined. One approach would be to investigate arbitrary representations (such as a mean shift or variance shift in-

tegrality; see Ashby & Maddox, 1994; Maddox, 1992) that satisfy perceptual integrality, but this seems less than satisfactory. Rather it would be advantageous to use knowledge of the Munsell color system to guide the choice of perceptual representations. The Munsell color system was developed in an attempt to identify a set of physical dimensions that correspond closely to possible psychological dimensions of color. The Munsell color system is a discrete (physical) system, whereas the possible psychological dimensions of color—hue, saturation, and brightness—are continuous valued. Of course, other psychological representations have been offered (e.g., Wandell, 1995). Unfortunately, the Munsell color system is discrete, and, in order to specify a perceptual distribution, a continuous-valued space is necessary. Fortunately, several transformations have been developed that allow one to transform from the discrete-valued Munsell space to the continuous-valued CIE color space (Brainard, 1995; Wandell, 1995; Wyszecki & Stiles, 1982). The transformations and all other details of the simulations are outlined in detail in Appendix A.<sup>3</sup> The resulting perceptual representations are depicted in Figures 3a and 3b. Notice that the stimulus coordinates for the EA condition (Figure 2b) show a generally negative trend, whereas the means of the perceptual distributions for this condition (Figure 3b) show a generally positive trend. This occurs for two reasons. First, the transformation from the Munsell space to the color system in Figure 3 is nonlinear. Second, as part of this transformation, a rotation was performed to bring the Dimension 2 perceptual distribution means in the SA condition (Figure 3a) in line with the Dimension 2 stimulus coordinates in Figure 2a. This rotation was also applied to the perceptual distributions in Figure 3b. The nonlinearity of the color transformation along with the rotation led to the different trends in Figures 2b and 3b. (The details are outlined in Appendix A.)

**Perceptual separability application.** This application focused on two categorization conditions used by Nosofsky (1989). The stimuli varied in the size of a semicircle and the orientation of an embedded radial line,

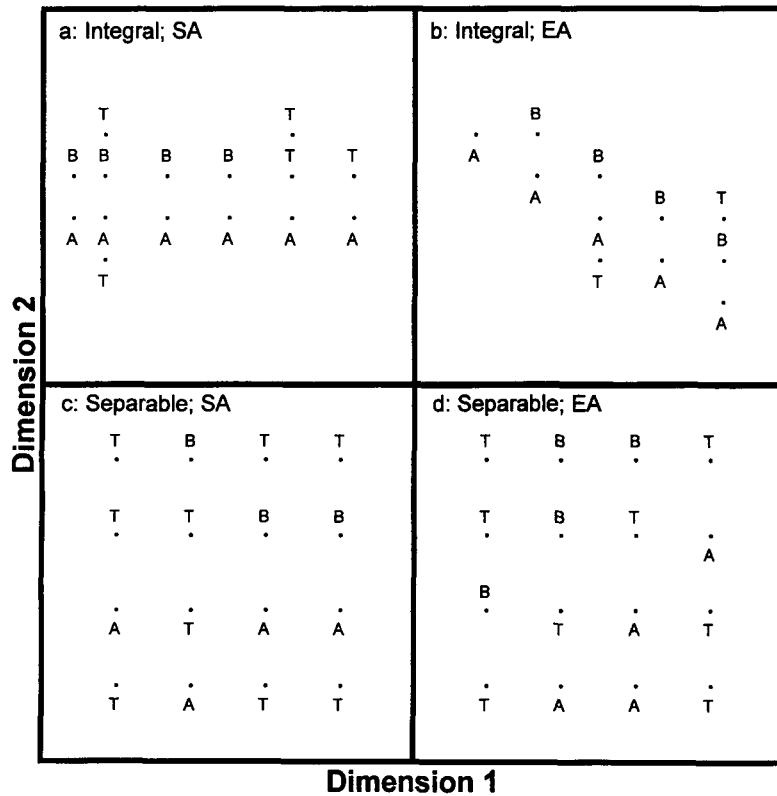


Figure 2. Schematic illustrating the categorization problems for the (a) selective attention and (b) equal attention conditions of the perceptual integrality application for the (c) selective attention and (d) equal attention conditions of the perceptual separability application. The label “A” denotes stimuli assigned to Category A, the label “B” denotes stimuli assigned to Category B, and the label “T” denotes unassigned transfer stimuli. SA, selective attention; EA, equal attention.

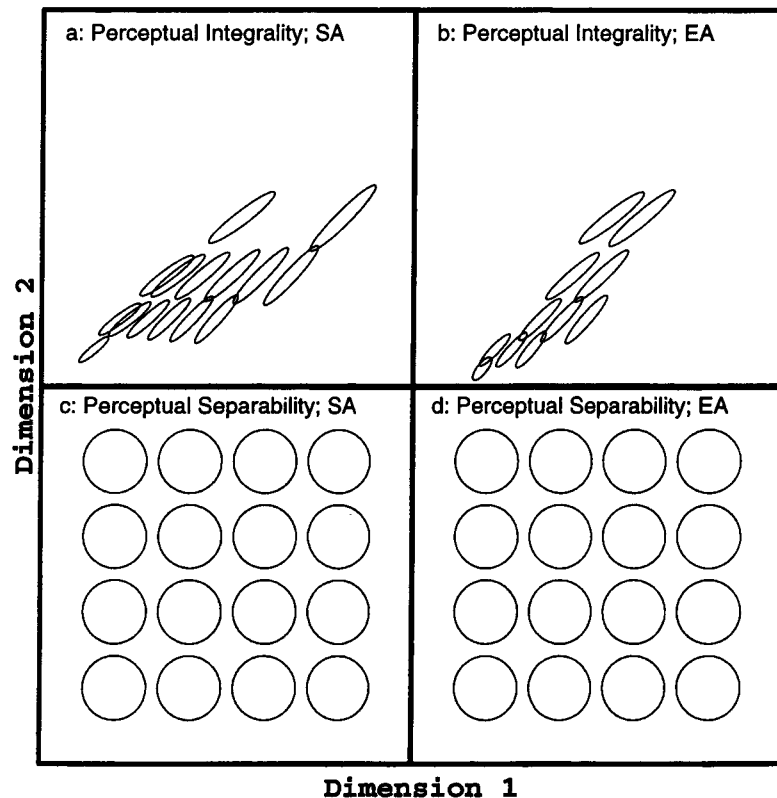
which are thought to be separable. To simulate data in this application, the simple stimulus-invariant,  $\sigma^2I$  perceptually separable representation (discussed earlier) was assumed. The resulting perceptual representations are depicted in Figures 3c and 3d. One other stimulus-invariant, perceptually separable representation was examined. In this representation, the variance along Dimension 1 was larger than the variance along Dimension 2. Nosofsky (1989) found discriminability differences across dimensions, so this seemed to be a reasonable perceptual representation to investigate. The simulation results were nearly identical to those for the stimulus-invariant,  $\sigma^2I$  perceptually separable representation, so this representation will not be discussed further.

**Decision Bounds**

The next step in the simulation approach was to specify the slope and intercept of the linear decision bound used by each hypothetical observer. For each categorization condition and perceptual representation, two types of decision bound assumptions were investigated. The first assumed that each observer used a linear decision bound that was approximately optimal. The second assumed nonoptimal decision bounds. Although some observers were assumed to use nonoptimal decision bounds,

all the observers with low accuracy rates were excluded. Specifically, in the perceptual integrality application, all the observers with accuracy rates lower than 85% were excluded. In the perceptual separability application, all the SA observers whose accuracy was lower than 80% and all the EA observers whose accuracy was lower than 70% were excluded. These values are similar to those used in the literature (e.g., McKinley & Nosofsky, 1996; Nosofsky, 1989). The second column of Table 1 denotes the assumptions made when the decision bound was approximately optimal, and the third column denotes the assumptions made when the decision bound was nonoptimal. In the EA condition of the perceptual integrality application, no linear decision bound was approximately optimal. In other words, nearly optimal performance (i.e., performance that maximized accuracy) required the use of a nonlinear decision bound. Therefore, only nonoptimal (linear) decision bounds were investigated. An illustration of the decision bound assumptions for the perceptual integrality and perceptual separability applications are depicted in Figures 4 and 5, respectively.

To summarize, when optimality was assumed, each observer attended selectively in the SA condition and attended approximately equally to each dimension in the EA condition. When optimality was violated, it was gen-



**Figure 3.** Contours of equal likelihood for the (a) selective attention and (b) equal attention conditions when perceptual integrity holds and for the (c) selective attention and (d) equal attention conditions when perceptual separability holds. SA, selective attention; EA, equal attention.

erally assumed that half of the observers used one type of bound and half used another type of bound. The idea that averaged categorization data contain a mixture of data collected from individuals using different decision strategies forms the basis for some recent categorization models, such as Ashby, Alfonso-Reese, Turken, and Waldron's (1998) competition between verbal and implicit systems (COVIS) model and Nosofsky, Palmeri, and McKinley's (1994) rule-plus-exceptions (RULEX) model.

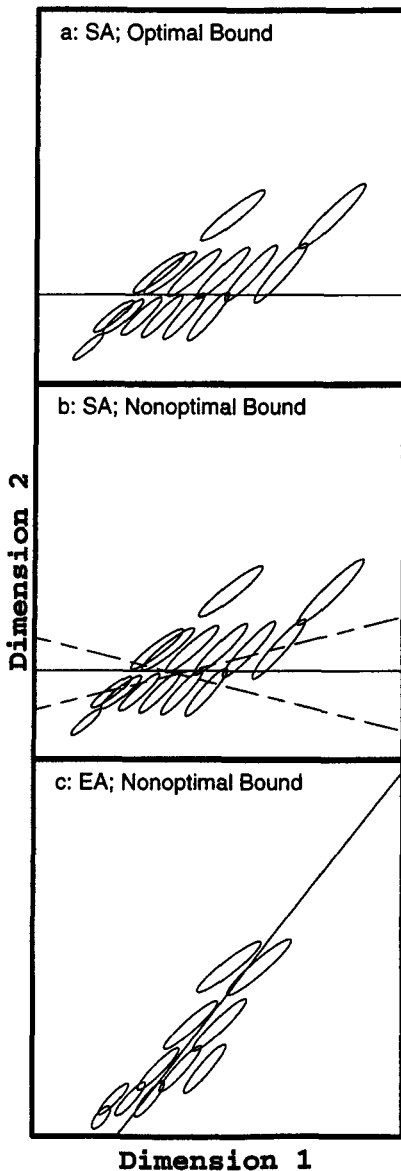
Because the goal of this research was to examine the influence of averaging across observers, it was crucial that reasonable individual differences exist among observers. Although individual differences can be specified at one of many levels, the strategy chosen here was to assume that each observer used a slightly different decision bound. The decision bound parameters specified in Table 1 and in Appendix A (see also Figures 4 and 5) might be better termed the *population mean decision bound parameters*. In other words, the decision-bound slope and intercept values outlined in Table 1 and in Appendix A represent the mean (or average) slope and intercept from the population distribution of hypothetical observers. The decision-bound slope and intercept for each observer is determined by taking a random sample from the population slope distribution and a random sample from the population intercept distribution. Both distributions are assumed

to be univariate normal, with the mean and variance outlined in Appendix A. Notice that, although the exact linear decision bound (i.e., slope and intercept) differs across observers, each individual observer uses a fixed linear decision bound and thus yields responding that is highly deterministic (i.e., all-or-none responding).

#### Models Tested

Several models were applied to the data from each observer and to the aggregate data. Three decision bound models were applied to the data. One model was the "true" decision bound model (denoted DBMTRUE). This model assumed the correct perceptual representation and a linear decision bound. The second model was a linear decision bound model that assumed a stimulus-invariant,  $\sigma^2 I$  perceptually separable representation (denoted GLCSI for general linear classifier with stimulus invariance). For the perceptual separability application, the DBMTRUE and GLCSI models are equivalent, because they make identical perceptual representation and decision bound assumptions. Thus, when applied to the data, the two models will yield identical fit values (see Figure 7). For the perceptual integrity application, the DBMTRUE and GLCSI models make different perceptual representation assumptions but identical decision bound assumptions. A comparison of these two models





**Figure 4.** Decision bound assumptions for the perceptual integrality application when (a) selective attention and optimal bounds, (b) selective attention and nonoptimal bounds, and (c) equal attention and nonoptimal bounds are assumed. SA, selective attention; EA, equal attention.

in the perceptual integrality application will allow us to explore the effects of incorrect perceptual representation assumptions on the ability of the decision bound model to account for the data. The third model was a quadratic decision bound model that assumed a stimulus-invariant,  $\sigma^2 I$  perceptually separable representation (denoted GQCSI for general quadratic classifier with stimulus invariance). For the perceptual separability application, this model makes the correct perceptual representation assumptions but assumes a quadratic decision bound. For the perceptual integrality application, this model makes

an incorrect perceptual representation assumption but allows for a quadratic decision bound. A comparison of this model with the DBMTRUE and GLCSI models in the perceptual integrality application will allow us to explore the ability of the extra decision bound parameters to accommodate for the incorrect perceptual representation assumptions.

In order to apply the GCM and DEM to categorization data, one must select a distance metric and similarity function (see Equations 4 and 3, respectively). Within the MDS literature, a city-block distance metric defines separable-dimension stimuli, and a Euclidean metric defines integral-dimension stimuli. In the original formulation of the GCM, it was assumed that the city-block metric should be paired with the exponential decay similarity function and that the Euclidean metric should be paired with the Gaussian similarity function (see Nosofsky, 1984, for details). Recently, these assumptions have been relaxed, and the most common applications of the GCM assume a Euclidean distance metric and an exponential decay similarity function. All three versions of the GCM and DEM were applied to the data. In all cases, the Euclidean/Gaussian version provided the best account of the data. Thus, we restrict attention to this version of the models. The models were applied to the data from each *individual* observer, using a maximum likelihood estimation procedure (see Ashby, 1992b; Maddox & Ashby, 1993; Wickens, 1982, for details). The individual observer data were then averaged, and the models were applied to the aggregate data, again using maximum likelihood. Specifically, the model parameters were estimated by minimizing the negative of the natural log of the likelihood function (i.e.,  $-\ln L$ ). Because the decision bound models and GCM are not *nested* (i.e., one model cannot be derived from the other by setting some model parameters to constants), it could be argued that Akaike's information criterion (AIC; see Ashby, 1992b, for details), which penalizes a model for each free parameter, would be more appropriate. In the present application, however, the goal was to have models that yield identical trial-by-trial predictions have identical fit values. This approach simplified the qualitative model comparisons. For example, if the GQCSI and GLCSI models yielded exactly the same trial-by-trial predictions, the  $-\ln L$  values would be identical. However, based on AIC, the GQCSI fit would be worse simply because of the additional parameters. Even so, all the general trends to be discussed shortly would hold for both  $-\ln L$  and AIC.

## Results

The results for the perceptual integrality and perceptual separability applications are depicted in Figures 6 and 7, respectively. In each case, the smaller the value along the ordinate, the better the fit of the model. The five bars on the left of each figure depict the averaged fits to the individual observer data (denoted *so ave*), and the five bars on the right depict the fits to the averaged data (denoted *ave obs*).

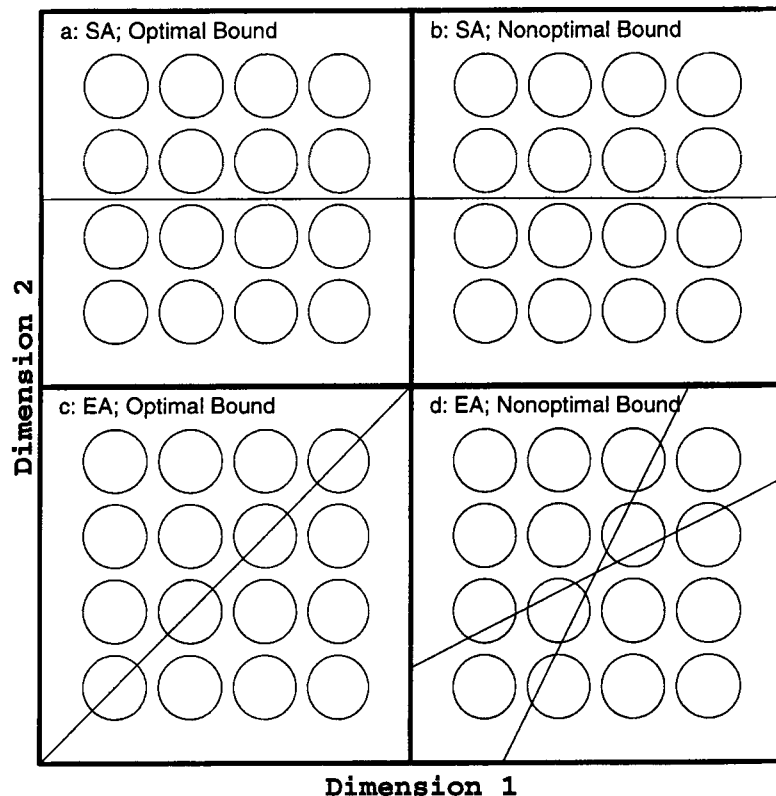


Figure 5. Decision bound assumptions for the perceptual separability application when (a) selective attention and optimal bounds, (b) selective attention and nonoptimal bounds, (c) equal attention and optimal bounds, and (d) equal attention and nonoptimal bounds are assumed. SA, selective attention; EA, equal attention.

**DBM versus the GCM and DEM.** Two results stand out when comparing the decision bound models with the GCM and DEM. First, in general, when each observer used a decision bound that was approximately optimal (Figures 6a, 7a, 7c), the fits of the averaged data were very similar to those of the individual observer data. If, in addition, perceptual separability was satisfied (Figures 7a, 7c), the fits of all the models were nearly equal. These results were expected, because previous research suggests that, under certain conditions, the equivocality contour (i.e., the set of coordinates for which the summed similarity to Category A is equal to the summed similarity to Category B) predicted by the GCM (and DEM) is very similar to the optimal decision bound (Ashby & Alfonso-Reese, 1995; Ashby & Maddox, 1993; Nosofsky, 1990). The GCM (and DEM) parameters allow the model enough flexibility to account for small deviations from the optimal bound and, thus, can mimic the performance of observers who behave approximately optimally. The fact that averaging had little effect on the model fits also seems reasonable. Because each observer used a similar decision bound, the averaging operation affected only the measurement error and, thus, led to little or no change in the fit of the models to the averaged data.

Second, when each observer's decision bound was nonoptimal (Figures 6b, 6c, 7b, 7d), the averaging opera-

tion yielded a large *improvement* in fit for the GCM (and usually the DEM) and a *decrement* in fit for the decision bound models, in particular, the DBMTRUE. This effect was magnified when perceptual integrality was satisfied. In other words, although the DBMTRUE was clearly superior to the GCM (and DEM) at the level of the individual observer data, once averaged, the superiority was strikingly reduced. Clearly, the averaging operation led to a qualitative change in the nature of the data. However, the data were not altered in a haphazard fashion; rather, the data changed in such a way that the ability of the GCM and DEM to describe the data was improved and the ability of the true model to describe the data worsened.

**GCM versus DEM.** A comparison of the GCM and DEM fits yielded the following results. First, the fits of the GCM and DEM to the average data were more similar than the GCM and DEM fits to the individual observer data. The DEM advantage for the single observer data was expected, because the decision bound model yields responding that is more deterministic (i.e., less based on probability matching) than that predicted by the GCM. The DEM can completely or partially account for this level of response determinism by increasing the value of the  $\gamma$  parameter. However, when highly deterministic individual observer data are averaged, the aggregate data appear much less deterministic (recall the example of

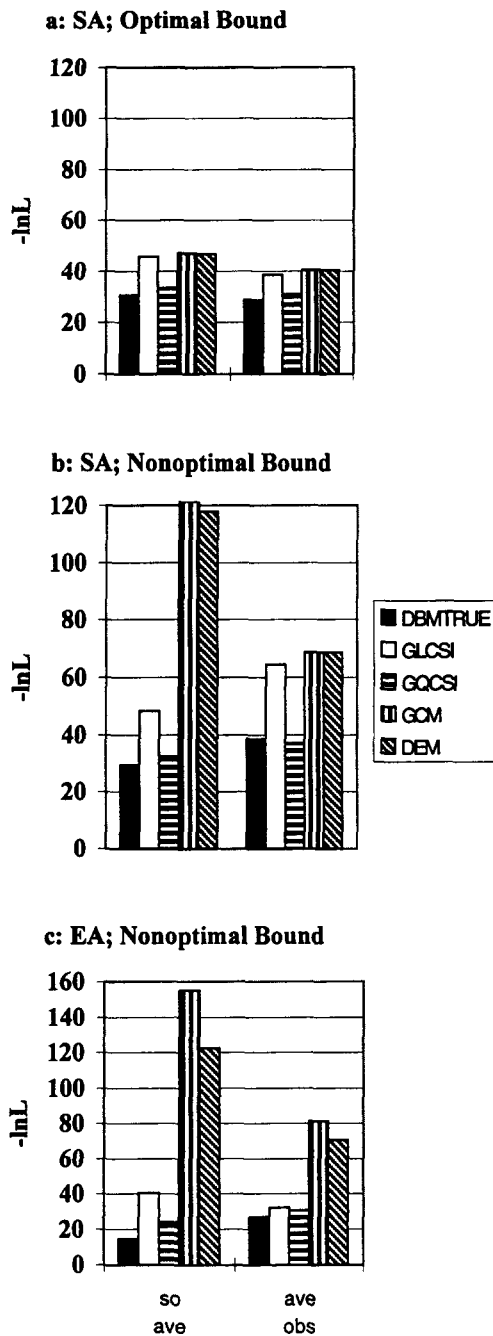


Figure 6. Goodness-of-fit ( $-\ln L$ ) values for the Monte Carlo simulations in which a linear decision bound was assumed and perceptual integrality held (see text for details). In each panel, the five bars on the left depict the average of the single-observer data fits (denoted *so ave*). The five bars on the right depict the fits to the average observer (denoted *ave obs*). SA, selective attention; DBMTRUE, decision bound model with the true perceptual representation and a linear decision bound; GLCSI, linear decision bound model with a stimulus-invariant,  $\sigma^2 I$  perceptually separable representation; GQCSI, quadratic decision bound model with a stimulus-invariant,  $\sigma^2 I$  perceptually separable representation; GCM, generalized context model with a Euclidean distance metric and a Gaussian similarity function; DEM, deterministic exemplar model with a Euclidean distance metric and a Gaussian similarity function.

all-or-none learning in the introduction). Since the GCM predicts probability matching, it is better able to account for the aggregate than for the individual observer data.

Second, the fits of the GCM and DEM to the individual observer data were more similar in the SA than in the EA condition. The GCM provided a poor account of the individual observer data in both cases, and the DEM provided a poor account of the data in the SA condition. Interestingly, however, when perceptual separability was satisfied in the EA condition, the DEM fit was nearly identical to that of the DBMTRUE. In order to account for the decision bound used in the SA condition, the GCM and DEM must attend more heavily to Dimension 2 than to Dimension 1. This is accomplished by setting the  $w$  parameter to a value near 0 ( $w$  represents the attention placed on Dimension 1). This stretches the psychological space along Dimension 2 and shrinks the psychological space along Dimension 1. Stretching the psychological space increases the Dimension 2 interpoint distances and, thus, makes the predictions from the models more extreme. In other words, the model predictions become more deterministic. Consequently, although the  $w$  parameter is generally assumed to represent the effects of selective attention only, it has a secondary effect of increasing response determinism. The  $c$  parameter can increase or decrease response determinism by expanding or contracting the psychological space uniformly. Interestingly, this leads to a large amount of parameter interdependence. Consequently, when  $c$  is large, the ability of  $w$  to affect the equivocality contour is reduced. Similarly, when  $w$  is extreme (near 0 or 1), the effect of  $c$  is reduced. This can have a strong effect on the ability of the GCM to mimic individual observer data generated from the decision bound model. For example, when a nearly optimal decision bound was assumed, only the intercept was free to vary across observers, and the decision-bound slope was set equal to zero. Under these conditions, the best fitting value of  $w$  was near zero, and the  $c$  parameter was adjusted to approximate the level of response determinism. The addition of the  $\gamma$  parameter in the DEM had little effect, because the values of  $c$  and  $w$  were such that responding was already highly deterministic. Thus, the DEM and GCM provided nearly equivalent fits to the data and were able to predict the data as well as the DBMTRUE.

However, when responding was nonoptimal (Figures 6b, 7b), both models performed poorly. When responding was nonoptimal, the individual observer decision-bound slopes were allowed to differ slightly from zero. A close examination of the GCM and DEM fits suggested the following. When the decision-bound slope was slightly negative, both models were able to mimic the decision bound and high level of determinism by setting  $w$  near zero and adjusting  $c$  (and  $\gamma$  for the DEM) to approximate the response determinism. In these cases, the GCM and DEM fit the data nearly as well as the DBMTRUE. However, when the decision-bound slope was slightly positive, both models performed poorly, because no set of parameter values was able to mimic the decision bound.

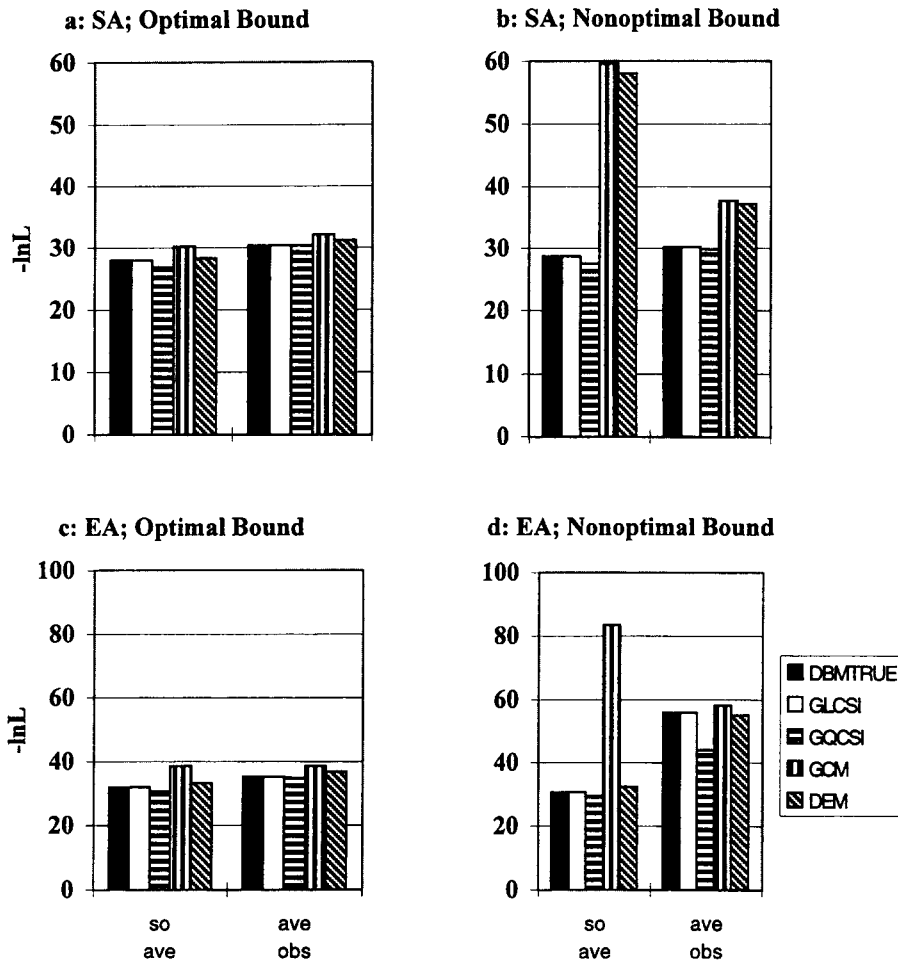


Figure 7. Goodness-of-fit ( $-\ln L$ ) values for the Monte Carlo simulations in which a linear decision bound was assumed and perceptual separability held (see text for details). In each panel, the five bars on the left depict the average of the single-observer data fits (denoted *so ave*). The five bars on the right depict the fits to the average observer (denoted *ave obs*). SA, selective attention; EA, equal attention; DBMTRUE, decision bound model with the true perceptual representation and a linear decision bound; GLCSI, linear decision bound model with a stimulus-invariant,  $\sigma^2 I$  perceptually separable representation; GQCSI, quadratic decision bound model with a stimulus-invariant,  $\sigma^2 I$  perceptually separable representation; GCM, generalized context model with a Euclidean distance metric and a Gaussian similarity function; DEM, deterministic exemplar model with a Euclidean distance metric and a Gaussian similarity function.

Recall that the GCM and DEM yield an equivocality contour that is similar in spirit to a decision bound, where the equivocality contour is the line or curve through the psychological space at which the summed similarity to the two categories is equal. If the model parameters can be adjusted so that the equivocality contour can mimic the decision bound, the model may be able to account for the data. If no set of parameter values exists that will allow the equivocality contour to mimic the decision bound, the models will perform poorly. When no selective attention exists (i.e., when  $w = .5$ ), the equivocality contour in the two SA conditions (see Figure 2c for the perceptually separable application and Figure 3a for the perceptually integral application) is linear with a slope of zero for moderate values along Dimension 1, bends up for small

values along Dimension 1, and bends down for large values along Dimension 1. When all attention is placed on Dimension 2 (i.e., when  $w = 0$ ), the equivocality contour in the two SA conditions is linear with a slope of zero. As the attention weight shifts from .5 to 0, the equivocality contour becomes more linear but will have a generally negative slope to it. Thus, when data are generated from a linear decision bound with a negative slope, the GCM and DEM are able to adjust the  $w$  (and other) parameters to approximately mimic the true linear decision bound. On the other hand, when the true decision bound slope was positive, no set of GCM and DEM parameter values exist that can mimic this decision bound. The best the GCM and DEM could do was to set  $w$  equal to zero, which yielded a zero slope. The  $c$  and  $\gamma$  parameters could be adjusted to

approximate the response determinism, but the poor account of the decision bound led both the GCM and DEM to provide poor accounts of the data.

In the perceptually separable, EA condition, when the decision bounds were nearly optimal (Figure 7c), each individual observer's slope was near 1, with variation across observers. Both the GCM and the DEM were able to mimic the appropriate slope by adjusting  $w$  to a value near .5 and using the  $c$  (and  $\gamma$  for the DEM) to approximate the response determinism. In these cases, the GCM and DEM fit the data nearly as well as the DBMTRUE. When the decision bounds were nonoptimal, some individuals used bounds with slopes around .5, and others used bounds with slopes around 2 (see Appendix A for details). For the GCM to account for these decision bounds, the  $w$  parameter needed to be different from .5 (less than .5 when the slopes were near .5, and greater than .5 when the slopes were near 2). In addition, the high level of response determinism required large values of  $c$ . But, as stated earlier, large values of  $c$  diminish the power of the  $w$  parameter. In short, the two parameters were in competition, and poor fits to the data resulted. The DEM, on the other hand, was able to account for the response determinism by adjusting the  $\gamma$  parameter. This allowed the value of  $c$  to stay small, which increased the power of the  $w$  parameter to account for the decision bound. In support of this claim, the average value of  $c$  (across observers) was 3.390 and 0.525 for the GCM and DEM when applied to the perceptual separability, EA condition in which nonoptimal bounds were assumed. In addition, the average value of  $\gamma$  from the DEM was quite large (15.29), suggesting that the  $\gamma$  parameter was able to account for the high level of response determinism.

In the perceptually integral, EA condition (Figure 6c), both the GCM and the DEM provided poor accounts of the data. In short, the equivocality contour was nonlinear because of the perceptual integrality inherent in these stimuli. Thus, no set of parameter values could mimic the true decision bound predictions. Although the addition of the  $\gamma$  parameter for the DEM did provide some improvement in fit over the GCM, the model was unable to mimic the performance of the DBMTRUE.

**GLC versus GQC.** Recall that two of the decision bound models applied to the data assumed a stimulus-invariant,  $\sigma^2I$  perceptually separable representation. One assumed a linear decision bound (the GLCSI), and one assumed a quadratic decision bound (GQCSI). The third decision bound model was the model that generated the data. This model assumed the true perceptual representation and a linear decision bound. In the perceptual separability application, the true perceptual representation was a stimulus-invariant,  $\sigma^2I$  perceptually separable representation, and, not surprisingly, all three models provided essentially equivalent fits of the individual observer data (Figure 7). In the perceptual integrality application, on the other hand, the "true" perceptual representation was not a stimulus-invariant,  $\sigma^2I$  perceptually separable

representation, and thus the representation assumptions of the GLCSI and GQCSI models were incorrect. Under these conditions, the GQCSI model enjoyed a consistent advantage over the GLCSI model (Figure 6). It appears that when the model makes an incorrect assumption about the perceptual representation, the extra *decision bound* parameters of the GQCSI model allowed it to account for some of the error that resulted from the incorrect *perceptual representation* assumption. This finding underscores the importance of making the correct perceptual representation assumptions. Despite the fact that individual observers used a linear decision bound, when the incorrect perceptual representation assumptions were made, it appeared as if the observer was using a quadratic decision bound.

This is an important result, because it suggests that the incorrect perceptual representation assumption might lead one to an incorrect inference about the form of the decision bound. Recently, McKinley and Nosofsky (1996) outlined categorization conditions for which they argued a linear decision bound was predicted. They found that the GQCSI model provided a superior account of the data and used these findings to argue against the validity of decision bound theory. Their stimuli were iso-hue Munsell color chips that varied in saturation and brightness and, thus, are highly integral. However, they applied the GQCSI and GLCSI models that assume perceptual separability. Because it is likely that the perceptual representation assumption made by McKinley and Nosofsky (1996) was incorrect, it is possible that they drew an incorrect inference about the form of the decision bound.

### Brief Summary

The most critical findings from these simulations can be summarized as follows. First, when individual differences are small and observers are approximately optimal, all the models make similar predictions, and averaging has little effect. Second, when reasonable individual differences exist, averaging can have a large effect on the ability of the models to account for the data. In particular, averaging tends to improve the fits of the GCM and DEM and to worsen the fits of the decision bound model, even though the DBMTRUE perfectly described the behavior of individual observers. Third, the parameters of the GCM and DEM can be highly interdependent. Under some conditions, this interdependence does not adversely affect the ability of the models to mimic the data (e.g., when nearly optimal decision bounds are assumed), whereas, under other conditions, the interdependence severely constrains the models' predictions (e.g., when nonoptimal bounds are assumed). Even so, the potential adverse affects of parameter interdependence are reduced when the data are averaged. Finally, incorrect perceptual representation assumptions made by decision bound models can lead to an incorrect inference about the nature of the observer's true decision bound.

## GENERALIZED CONTEXT MODEL SIMULATIONS

This section summarizes the simulation approach taken when the GCM was assumed to be correct. The details are presented in Appendix B. The “true” model (i.e., the model that accurately described the individual observer performance) assumed every item was represented by a point in some MDS psychological space (Equation 1). When presented with an item to be categorized, the observer computed the similarity between the target exemplar and all exemplars in memory and used the Equation 2 response rule with  $\gamma = 1$  (see Equations 2–4). Several distance metric and similarity function pairings were examined. Across simulations, assumptions about the categorization rule (selective attention or attention to both dimensions), perceptual representation (separable or integral), and individual differences in the observers’ value of  $w$  and  $\beta$  were varied systematically.

As in the previous set of simulations, SA and EA conditions were examined. In the previous set of simulations, different SA and EA stimulus-to-response mappings were used for the perceptually separable and perceptually integral applications. This approach was taken because, within the framework of decision bound theory, the perceptual representations (i.e., the perceptual means, variances, and covariances) are strongly influenced by integrality or separability of the stimulus dimensions. Within the framework of the GCM, the integrality–separability distinction affects the nature of the distance metric and does not affect the stimulus coordinates (i.e., Equation 1). Thus, for simplicity, the SA and EA conditions depicted in Figures 2c and 2d were utilized.

### Integrality–Separability Distinction

Following the MDS literature, the Euclidean distance metric defined integral-dimension stimuli, and the city-block distance metric defined separable-dimension stimuli. As originally proposed, the GCM paired the Euclidean metric with the Gaussian similarity function and the city-block metric with the exponential decay similarity function. Simulations of both of these versions of the GCM were conducted. More recently, a version of the GCM that paired the Euclidean metric with the exponential decay similarity function has been examined (e.g., McKinley & Nosofsky, 1995). This version of the GCM was also simulated.

### Individual Differences and Categorization Rule

The next step in the simulation approach was to specify the parameters used by each hypothetical GCM observer. As in the previous set of simulations, optimal and nonoptimal parameter settings were investigated for each categorization condition and perceptual representation. The details are outlined in Appendix B. To summarize, when optimality was assumed, each observer attended selectively in the SA condition (i.e.,  $w = 0$ ) and attended approximately equally in the EA condition (i.e.,  $w$  near .5). When optimality was violated, it was generally as-

sumed that half of the observers used one set of parameter values and half used another set of parameter values. As in the previous simulations, to investigate the influence of averaging on the fits of the models, it was crucial that reasonable individual differences exist among observers. The three parameters of the GCM are the  $c$ ,  $w$ , and  $\beta$  parameters. Individual differences were incorporated by allowing the attention weight parameter,  $w$ , and the category bias parameter,  $\beta$ , to vary across observers. [The  $c$  parameter was held fixed ( $c = 1.5$ ) for all simulations in order to yield accuracy rates in accordance with empirically observed values.] The  $w$  and  $\beta$  parameters specified in Appendix B might be better termed the *population mean parameters*. In other words, the GCM parameter values in Appendix B represent the mean (or average) parameter values from the population distribution of hypothetical observers. The values of  $w$  and  $\beta$  for each observer were determined by taking a random sample from the population  $w$  distribution and a random sample from the population  $\beta$  distribution. Both distributions are assumed to be univariate normal, with the mean and variance outlined in Appendix B. Notice that, although the exact values of  $w$  and  $b$  differed across observers, each observer used a single value of  $w$  and  $\beta$ .

### Models

The GLCSI model, the GQCSI model, and the three versions of the GCM (i.e., city-block/exponential, Euclidean/exponential, Euclidean/Gaussian) were applied to the data from each observer and to the aggregate data, using a maximum likelihood procedure. The DEM was excluded from these analyses for two reasons. First, because the GCM can account perfectly for the data from each individual observer (inasmuch as it was used to generate the data), the fit of the DEM and GCM would be identical, and the additional  $\gamma$  parameter of the DEM would estimate to one (as is implied in the GCM). Second, although it is possible that the DEM might provide a slightly better fit to the averaged data than the GCM, this finding would not be very informative.

### Results

The distinction between optimal and nonoptimal parameter settings had little effect on the model fits, so the data were collapsed across optimal and nonoptimal parameter settings. The results for the three versions of the GCM are depicted in Figure 8. Figures 8a–c present the results from the SA condition, and Figures 8d–f present the results from the EA condition. The three columns depict the three distance metric/similarity function pairings. The results can be summarized as follows. First, averaging had little effect on the fits of any of the models. This finding was expected, because Ashby et al. (1994) and Ashby et al. (1992) had shown that the MDS model and the SCM yield single-observer and average data that have similar structures. In other words, averaging does not alter the structure of the data. In light of this fact, it is not surprising that the fits of the single observer and aggregate data are nearly identical. Second, the decision

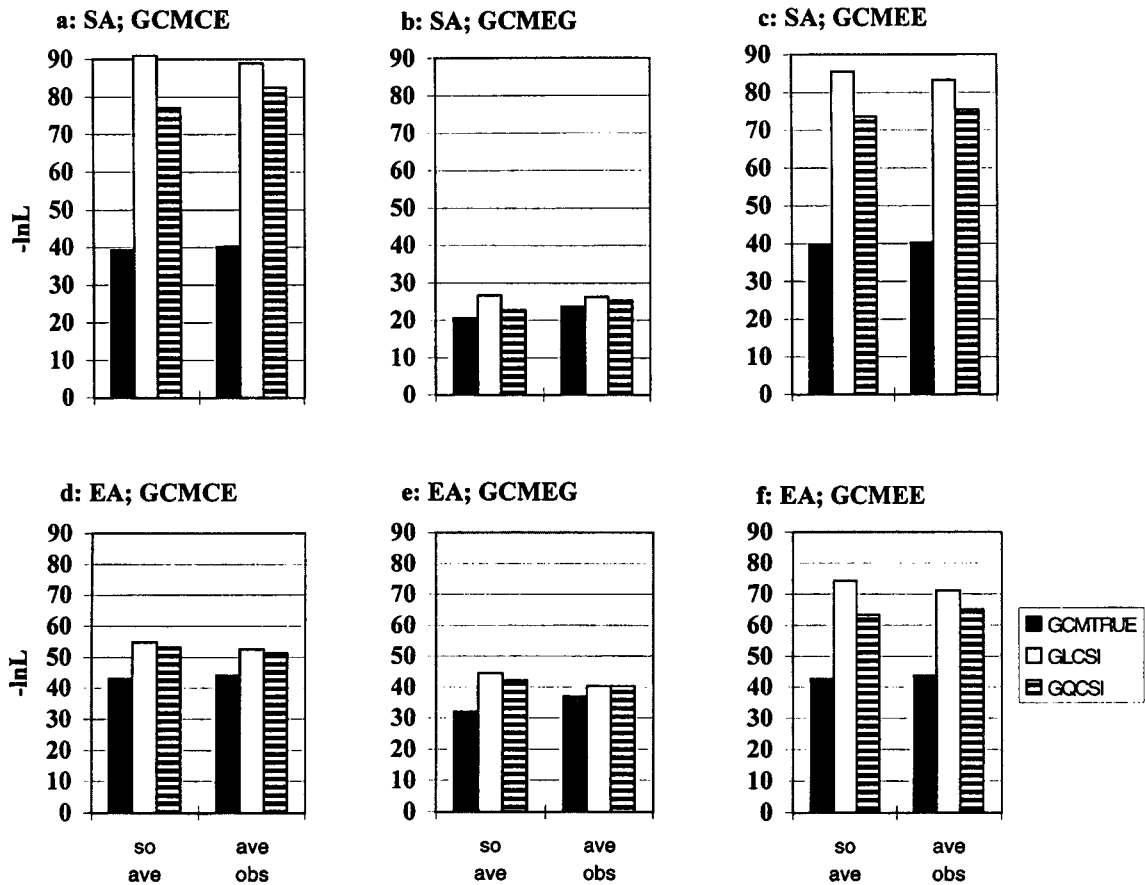


Figure 8. Goodness-of-fit ( $-\ln L$ ) values for the Monte Carlo simulations in which the generalized context model (GCM) was assumed to be correct (see text for details). In each panel, the three bars on the left depict the average of the single-observer data fits (denoted *so ave*). The three bars on the right depict the fits to the average observer (denoted *ave obs*). SA, selective attention; EA, equal attention; GCMCE, the GCM that assumed a city-block distance metric and an exponential decay similarity function; GCMEG, the GCM that assumed a Euclidean distance metric and a Gaussian similarity function; GCMEE, the GCM that assumed a Euclidean distance metric and an exponential decay similarity function; GCMTRUE, GCM with the correct distance metric and similarity function assumptions; GLCSI, linear decision bound model with a stimulus-invariant,  $\sigma^2 I$  perceptually separable representation; GQCSI, quadratic decision bound model with a stimulus-invariant,  $\sigma^2 I$  perceptually separable representation.

bound models and GCMTRUE yielded similar fits when the Euclidean/Gaussian model was assumed, but the decision bound model yielded much worse fits for the city-block/exponential and Euclidean/exponential versions of the GCM. Most likely, this is due to the fact that, under certain conditions, equivalencies can be derived between the decision bound models and the Euclidean/Gaussian version of the GCM (see Ashby & Maddox, 1993, for details). Finally, although not depicted in Figure 8, when the city-block/exponential version was correct, the correct version and the Euclidean/exponential version provided nearly equivalent fits. Analogously, when the Euclidean/exponential version was correct, the correct version and the city-block/exponential version provided nearly equivalent fits. However, when the Euclidean/Gaussian version was correct, only the correct version of the GCM provided a good account of the data. The two versions that assumed an exponential decay

similarity function provided poor descriptions of the data. Apparently, the correct choice of the similarity function is more important than the correct choice of the distance metric.

In summary, averaging had little effect on the fits of the GCM or the decision bound models. In addition, the decision bound models were always outperformed by the GCMTRUE when applied to the single-observer and aggregate data.

## EMPIRICAL APPLICATION

In this section, an empirical test of the averaging hypothesis is offered. In short, 9 observers classified lines that varied in length and orientation into one of two categories. The stimulus-response mappings were such that attention to both length and orientation was necessary to solve the problem (see Figure 9).

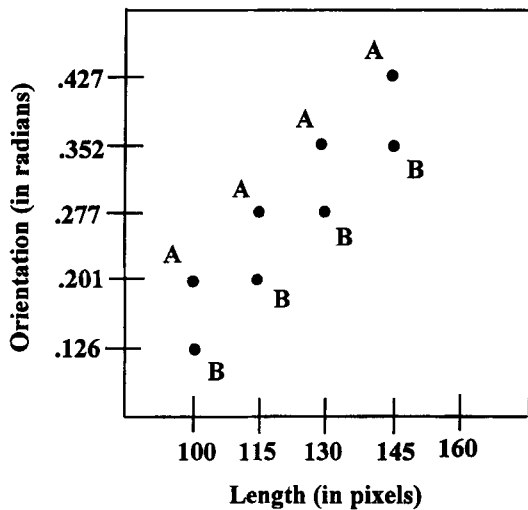


Figure 9. A schematic illustration of the stimulus-response mappings for the empirical application. The label "A" denotes a stimulus assigned to Category A, and the label "B" denotes a stimulus assigned to Category B. The stimuli were lines of varying length (in pixels) and orientation (in radians).

## Method

**Subjects.** Nine observers participated for pay (\$6 per session) in the experiment. All the observers were volunteers from the Arizona State University community. All observers claimed to have 20/20 vision or vision corrected to 20/20.

**Stimuli and stimulus-response mappings.** The stimulus ensemble consisted of eight lines of varying length and orientation. Four stimuli were assigned to Category A, and four to Category B. Figure 9 depicts a schematic of the categorization problem and stimuli. The stimuli were computer generated and displayed on a 14-in. SVGA monitor with  $1,024 \times 768$  resolution. The stimuli were white on a black background and subtended approximately  $1^\circ$  of visual angle.

**Procedure.** The experiment was conducted in a dimly lit sound-proof chamber. The observer was seated approximately 1 m from the computer monitor, with no head or chin restraints. All the observers completed two sessions of 800 trials each. The first session was considered practice and was excluded from subsequent analyses. Each session consisted of four blocks of 200 trials, for a total of 800 trials. Within each 200-trial block, each of the eight stimuli were presented 25 times. The presentation sequence was randomized within each 200-trial block. The observers were instructed to maximize accuracy and not to worry about speed of responding. A typical trial proceeded as follows: A stimulus was presented and was pattern masked after 800 msec; the mask remained on the screen until the observer responded by pressing a button marked "A" or "B" corrective feedback was provided for 1,000 msec, followed by a 500-msec intertrial interval.

## Results and Theoretical Analysis

The GCM and DEM (both assuming a city-block metric and an exponential decay similarity function) and the GLCSI and GQCSI models were applied to the data from each observer and to the aggregate data, using a maximum likelihood procedure. The city-block metric and exponential decay similarity function were assumed because these

stimulus dimensions have been found to be separable, on the basis of a series of operational tests (Garner, 1974). Because the stimuli were highly confusable and were presented briefly, it is likely that the stimulus-invariant,  $\sigma^2I$  perceptually separable representation assumption of the decision bound models is incorrect. Even so, without additional information to suggest the appropriate perceptual representation assumptions, this is the most parsimonious initial assumption. The goodness-of-fit values for the (averaged) single-observer data and the aggregate data for each model are displayed in Table 2.

The results can be summarized as follows. First, and most important, the fits of the models were more similar for the averaged than for the single-observer data. In particular, whereas the GQCSI model provided the best fit to the single-observer data, the DEM, GQCSI model, and GLCSI model all provided roughly equivalent fits to the aggregate data. Thus, on the basis of the aggregate data, it would be difficult to determine which approach (decision bound, or exemplar-similarity) provided the best description of the data. Second, the fits of all the models were better to the aggregate data than to the single-observer data, but the largest improvement was obtained for the GCM. This result is most likely due to the fact that the averaged data were less deterministic (i.e., more like probability matching) than the individual observer data. Third, the GQCSI model was superior to the GLCSI model for the individual observer data but not for the aggregate data. The simulations offer a possible explanation for this finding. The simulation results suggest that the GQCSI model can provide a superior fit over the GLCSI model when the true perceptual representation violates the stimulus-invariant,  $\sigma^2I$  perceptually separable representation. These conditions also were found to lead to improvements in the fit of both models to aggregate data, a result obtained in this experiment. Because the stimuli were highly confusable and were presented briefly, it is likely that the stimulus-invariant,  $\sigma^2I$  perceptually separable representation assumption was violated.

This experiment was not meant to provide a definitive test of the empirical validity of the averaging hypothesis. Clearly, a number of factors must be studied empirically to test this hypothesis rigorously. Even so, the major prediction, that the fits of the GCM (and DEM) would improve when applied to averaged data, was supported. Although the decision bound model fits also improved, the improvement in fit was much smaller than that of the GCM. Most important, as suggested by Mad-

Table 2  
Goodness-of-Fit ( $-\ln L$ ) Values From the Empirical Application

Model	Single Observer Average	Average Observer
GCM	53.57	35.60
DEM	29.16	21.85
GLCSI	29.78	21.62
GQCSI	26.88	21.62

Note—Based on data from 9 observers.



dox and Ashby (1993) the fits of all the models were more similar when applied to the aggregate than to the single-observer data.

### SUMMARY AND CONCLUSIONS

It is common in psychology to draw inferences about individual human behavior from analyses of aggregate data. Often, this is a valid approach, because the averaging operation reduces measurement error and does not affect the qualitative structure of the data. This appears to be the case when the GCM provides the true description of individual observer categorization performance. Under these conditions, averaging has little effect on the fits of the models, and the GCM is clearly superior to the decision bound models when applied to the individual observer and aggregate data. However, this is not always the case. Building on the findings of Ashby et al. (1994; see also Ashby et al., 1992), this report suggests that inferences drawn from averaged categorization data should be interpreted with caution if the decision bound model provides the true description of individual observer categorization performance. When the decision bound model is superior at the individual observer level, it is often the case that averaging leads to improved fits of the GCM (and DEM) and to worse fits of the decision bound models. If the focus is on the averaged data, an incorrect inference regarding individual performance might result. If the goal is to understand the behavior of individuals, caution must be taken before averaged data are interpreted.

It is important at this point to make a distinction between intraobserver variability and interobserver variability. One might argue that this article addresses only interobserver variability and ignores intraobserver variability. On the contrary, one of the strengths of theories such as decision bound or signal detection theory is that they acknowledge, and make fundamental, the importance of intraobserver variability. Within the framework of decision bound theory, intraobserver variability is specified at the level of the perceptual effect (i.e., the existence of perceptual noise is recognized) and at the level of the decision (i.e., the existence of criterial noise or memory for the decision bound is recognized). These sources of variability are captured by associated parameters in the model and will not lead to incorrect psychological inference. On the other hand, regardless of the existence or absence of intraobserver variability, there is still the potential to draw an incorrect inference about individual observer performance from analyses of aggregate data.

The results also make it clear that the GCM and, especially, the DEM are extremely flexible and are often able to mimic the predictions from decision bound models. In many cases, the parameters can be adjusted to mimic the form of the decision bound and the level of response determinism. The decision bound models, on the other hand, are less able to mimic the performance of the GCM,

especially when the GCM assumes an exponential decay similarity function. The results also suggest that an adequate test of decision bound theory requires the correct assumption about both the perceptual representation and the decision bound. When the perceptual representation violates perceptual separability and the observer uses a linear decision bound, the quadratic decision bound model will be superior to the linear decision bound model when both assume that perceptual separability is satisfied. In other words, the incorrect perceptual representation assumption of the models will lead to an incorrect inference regarding the shape of the observer's decision bound.

Obviously, the results of this article and the recent development of models that acknowledge the fundamental importance of individual differences in categorization (e.g., COVIS and RULEX) have important implications for modeling categorization performance and for the appropriateness of aggregate data analysis, but the implications go beyond the realm of categorization research. The GCM and DEM are grounded in an MDS approach to similarity and the probabilistic response rule of the SCM (although the DEM can mimic a wide range of responding from highly probabilistic to highly deterministic). Ashby, Maddox, and Lee (1994) showed both of these major components to be strongly affected by averaging. Because variants of the GCM, and thus the MDS and SCM assumptions, have been applied in several areas of psychology, one must give serious thought to the reported success of the GCM in these areas, inasmuch as nearly all of this work is based on aggregate data. Certainly, in some of this work, the averaging operation affects only measurement error, but in other realms this may not be the case.

### REFERENCES

- ASHBY, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics*, *44*, 195-204.
- ASHBY, F. G. (1992a). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449-483). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G. (1992b). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G., & ALFONSO-REESE, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*, 216-233.
- ASHBY, F. G., ALFONSO-REESE, L. A., TURKEN, A., & WALDRON, E. (1998). Competition between verbal and implicit rules of category learning. *Psychological Review*, *105*, 442-481.
- ASHBY, F. G., & LEE, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, *120*, 150-172.
- ASHBY, F. G., & LEE, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Foundations of perceptual theory* (pp. 369-397). New York: Elsevier, North-Holland.
- ASHBY, F. G., LEE, W. W., & BALAKRISHNAN, J. D. (1992). Comparing the biased choice model and multidimensional decision bound models of identification. *Mathematical Social Sciences*, *23*, 175-197.
- ASHBY, F. G., & MADDUX, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, *16*, 598-612.

- ASHBY, F. G., & MADDOX, W. T. (1991). A response time theory of perceptual independence. In J. P. Doignon & J. C. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 389-413). New York: Springer-Verlag.
- ASHBY, F. G., & MADDOX, W. T. (1993). Relations between exemplar, prototype, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.
- ASHBY, F. G., & MADDOX, W. T. (1994). A response time theory of perceptual separability and perceptual integrality in speeded classification. *Journal of Mathematical Psychology*, **38**, 423-466.
- ASHBY, F. G., MADDOX, W. T., & LEE, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, **5**, 144-150.
- 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.
- BRAINARD, D. H. (1995). Colorimetry. In M. Bass (Ed.), *Handbook of optics: Vol. 11. Fundamentals, techniques, and design* (pp. 26.1-26.54). New York: McGraw-Hill.
- ESTES, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, **53**, 134-140.
- GARNER, W. R. (1974). *The processing of information and structure*. New York: Wiley.
- KRUSKAL, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27.
- KRUSKAL, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, **29**, 115-129.
- LAMBERTS, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, **124**, 161-180.
- LUCE, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103-189). New York: Wiley.
- MACADAM, D. L. (1942). Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America*, **32**, 247-274.
- MADDOX, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147-180). Hillsdale, NJ: Erlbaum.
- MADDOX, W. T. (1995). Baserate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 288-301.
- MADDOX, W. T., & ASHBY, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.
- MADDOX, W. T., & ASHBY, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 795-817.
- MADDOX, W. T., & ASHBY, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and Nosofsky (1996). *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 301-321.
- MASSARO, D. W., & COHEN, M. M. (1993). The paradigm and the fuzzy logical model of perception are alive and well. *Journal of Experimental Psychology: General*, **122**, 115-124.
- MCKINLEY, S. C., & NOSOFSKY, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 128-148.
- MCKINLEY, S. C., & NOSOFSKY, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 294-317.
- NOSOFSKY, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 104-114.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOSOFSKY, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 87-109.
- NOSOFSKY, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics*, **45**, 279-290.
- NOSOFSKY, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology*, **34**, 393-418.
- NOSOFSKY, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 3-27.
- NOSOFSKY, R. M. (1992). Exemplar-based approach to relating categorization, identification and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363-393). Hillsdale, NJ: Erlbaum.
- NOSOFSKY, R. M. (1998). Selective attention and the formation of linear decision boundaries: Reply to Maddox and Ashby (1998). *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 322-339.
- NOSOFSKY, R. M., PALMERI, T. J., & MCKINLEY, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.
- 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. (1962a). The analysis of proximities: Multidimensional scaling with unknown distance function I. *Psychometrika*, **27**, 125-140.
- SHEPARD, R. N. (1962b). The analysis of proximities: Multidimensional scaling with unknown distance function II. *Psychometrika*, **27**, 219-246.
- SHEPARD, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, **1**, 54-87.
- SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.
- SIEGLER, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, **116**, 250-264.
- SMITH, E. R., & ZARATE, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, **99**, 3-21.
- SMITH, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, **96**, 125-144.
- TORGERSON, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-353.
- WANDELL, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer.
- WICKENS, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.
- WYSZECKI, G., & STILES, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae* (2nd ed). New York: Wiley.

## NOTES

1. It is important to make a distinction between perceptual and decisional selective attention (see, e.g., Maddox & Ashby, 1998). Although the details are beyond the scope of this article, suffice it to say that perceptual selective attention is relevant to situations in which the perceptual representation changes as a function of task demand. Decisional selective attention, on the other hand, holds when the categorization rule is such that only one stimulus dimension is relevant. This article focuses on decisional selective attention, but hereafter we will simply use the traditional term *selective attention*.

2. It is generally assumed that the  $w$  parameter represents *decisional* selective attention (see, e.g., Nosofsky, 1987).

3. While searching for a suitable transformation from the discrete Munsell color space to a continuous valued color space, it became apparent that several possible transformations existed. Although the covariance matrices that would result from various transformations would clearly differ, as long as the resulting representation was characterized by complex variance and dependence relations, the results of the simulations should be similar. In fact, several arbitrary perceptual integral representations, such as variance shift integralities (see Ashby & Maddox, 1994; Maddox, 1992), were investigated before the approach outlined in this report was discovered. In general, the results were similar to those reported here.

**APPENDIX A**  
**General Procedure for Monte Carlo**  
**Simulations of the Linear Decision Bound Model**

---

**Basic Assumptions**

1. Each simulation consisted of 10 hypothetical observers.
2. The perceptual representation (i.e., means and covariance matrix for each stimulus) was constant across observers.
3. Each stimulus was presented 200 times. In other words, for each hypothetical observer, 200 random samples were taken from each perceptual distribution.
4. No criterial noise existed (i.e.,  $e_c = 0$ , and  $\sigma_c = 0$ ; see Equation 6).

**Specifics of Perceptual Integrality Application****Perceptual Representation Assumptions**

This application used category structures similar to the orthogonal (termed SA here) and diagonal(a) (termed EA here) category structures used by McKinley and Nosofsky (1996). The stimuli were iso-hue Munsell color chips that varied in saturation and brightness. The goal was to use current vision research to transform from the discrete Munsell color space to a color space that was continuous and, thus, would allow a rigorous description of the contours of equal likelihood for each stimulus. The steps were as follows:

1. Transform the three-dimensional (3-D) Munsell coordinates to the 3-D CIE LUV uniform color space. LUV is a uniform color space because equal distances in the LUV space are approximately perceptually equal. The transformation from Munsell to LUV was accomplished relative to CIE Illuminant C and is a nonlinear transformation (Brainard, 1995; Wyszecki & Stiles, 1982). The result was a set of LUV coordinates, one for each Munsell stimulus.
2. Generate an iso-discrimination surface for each stimulus in the CIE LUV space. Because equal distances are approximately perceptually equal in LUV, the iso-discrimination surface was a sphere.
3. Transform each iso-discrimination sphere (in the CIE LUV space) to an iso-discrimination ellipsoid in the 3-D CIE XYZ space. Because the ellipsoidal nature of the XYZ iso-discrimination curves was first discovered by Macadam (1942), these are generally referred to as Macadam ellipsoids. In order to perform the transformations, assumptions about the viewing conditions were necessary. Two viewing conditions were examined. The first assumed daylight viewing conditions; the second assumed incandescent light, which has more yellow and less blue than daylight. The resulting contours of equal likelihood were very similar. The reported simulations assumed daylight illuminance.
4. Because the contours of equal likelihood for a multivariate normal distribution are always elliptical, the next step was to estimate the parameters of a trivariate normal distribution for each Macadam ellipse. The result was a set of trivariate normal distributions, one for each Munsell stimulus.
5. The mean vectors for each trivariate normal distribution were approximately coplanar. In other words, when the Munsell stimulus coordinates were transformed to the CIE XYZ space, they all fell on a common plane in XYZ. This result was expected because the stimuli were iso-hue Munsell color chips, differing only in saturation and brightness. Although hue is not directly related to X, Y, or Z in isolation, each hue is represented by a plane through XYZ space.
6. In light of this fact and to simplify the problem, the trivariate normal distributions were projected onto this iso-hue plane, and a simple rotation was applied so that the resulting perceptual representation was as similar to the Munsell values as possible (see Figures 3a and 3b).
7. The axes in this space are denoted by the labels Dimension 1 and Dimension 2. These arbitrary labels were used to make clear that these dimensions differ from the X, Y, and Z dimensions of the CIE color space. Even so, notice that the Dimension 2 values in this space are very similar to the brightness values in the Munsell space. This obtains for two reasons. First, brightness in the Munsell system is analogous to Y in the CIE XYZ space. In other words, Munsell stimuli of a fixed brightness are represented in the CIE XYZ space by a fixed Y value. Second, the projection and rotation, outlined above, preserved the relationship among the Y values in XYZ space. Thus, it is reasonable to interpret the Dimension 2 values in the Figure 3a space as directly related to the brightness values in the Munsell system.
8. Two additional caveats need mention. First, in the EA condition, transfer items 3–5 from McKinley and Nosofsky (1996) were excluded because the GCM and DEM yielded undefined predictions for these items (i.e., the summed similarity to Category A and Category B was zero). Second, in order to achieve the appropriate accuracy rates, the entries of the perceptual covariance matrices were multiplied by a factor of four.

### APPENDIX A (Continued)

#### Decision Bound Assumptions

1. Selective Attention Condition
  1. Approximately Optimal Decision Bound
    1. Simulation Set 1: mean slope = 0, mean intercept = 20.25.
    2. Simulation Set 2: mean slope = 0; Observers 1–5, mean intercept = 18.25; Observers 6–10, mean intercept = 22.25.
    3. For both sets of simulations, the slope was fixed at 0 for each observer. The intercept for each observer was obtained by taking a random sample from a normal distribution, with the mean defined above and  $\sigma_{\text{intercept}} = 4.0$ .
  2. Nonoptimal Decision Bound
    1. Simulation Set 1: same as Simulation Set 1 above.
    2. Simulation Set 2: same as Simulation Set 2 above.
    3. Simulation Set 3: Observers 1–5, mean slope = .25, mean intercept = 11.25; Observers 6–10, mean slope =  $-.25$ , mean intercept = 27.75.
    4. Simulation Set 4: Observers 1–5, mean slope = .25, mean intercept = 11.75; Observers 6–10, mean slope =  $-.25$ , mean intercept = 18.25.
    5. For all four sets of simulations,  $\sigma_{\text{slope}} = .4$ , and  $\sigma_{\text{intercept}} = 4.0$ .
2. Equal Attention Condition
  1. Nonoptimal Decision Bound
    1. No simulations were conducted that used approximately optimal decision bounds, because the perceptual representation is complex and no linear decision bound yields optimal performance levels.
    2. Mean slope = 1.25, mean intercept =  $-24$ .
    3. As before, the slope and intercept for each observer was obtained by taking a random sample from a normal distribution, with the mean defined above and  $\sigma_{\text{slope}} = .4$ , and  $\sigma_{\text{intercept}} = 4.0$ .

#### Specifics of Perceptual Separability Application

##### Perceptual Representation Assumptions

1. The perceptual means were set to the relevant stimulus level. For example, the stimulus at Level 1 along both stimulus dimensions had a mean perceptual effect of (1,1). The stimulus at Level 1 along Dimension 1 and Level 2 along Dimension 2 had a mean perceptual effect of (1,2), and so on.
2. A stimulus-invariant,  $\sigma^2 I$  perceptually separable representation in which  $\sigma_1 = \sigma_2 = .5$  was assumed.

#### Decision Bound Assumptions

1. Selective Attention Condition
  1. Approximately Optimal Decision Bound
    1. Simulation Set 1: mean slope = 0, mean intercept = 2.5.
    2. Simulation Set 2: mean slope = 0; Observers 1–5, mean intercept = 2; Observers 6–10, mean intercept = 3.
    3. For both sets of simulations and all observers, the slope remained fixed at zero. The intercept for each observer was obtained by taking a random sample from a normal distribution with a mean as defined above, and  $\sigma_{\text{intercept}} = .4$ , where  $\sigma_{\text{intercept}}$  represents the standard deviation for the intercept.
  2. Nonoptimal Decision Bound
    1. All procedures were identical, except that the slope was not held fixed at zero; rather, the slope for each observer was obtained by taking a random sample from a normal distribution, with a mean of zero and  $\sigma_{\text{slope}} = .2$ .
2. Equal Attention Condition
  1. Approximately Optimal Decision Bound
    1. Mean slope = 1, mean intercept = 0.
    2. The slope (and intercept) for each observer was obtained by taking a random sample from a normal distribution. For the slope, the mean was 1 and  $\sigma_{\text{slope}} = .2$ ; for the intercept, the mean was 0 and  $\sigma_{\text{intercept}} = .4$ .

---

**APPENDIX A (Continued)**


---

2. Nonoptimal Decision Bound
    1. Simulation Set 1: Observers 1–5, mean slope = 0.5, mean intercept = 1.25; Observers 6–10, mean slope = 2, mean intercept = –2.5.
    2. Simulation Set 2: Observers 1–5, mean slope = 0.5, mean intercept = 1.5; Observers 6–10, mean slope = 2, mean intercept = –3.
    3. Simulation Set 3: Observers 1–5, mean slope = 0.5, mean intercept = 1.75; Observers 6–10, mean slope = 2, mean intercept = –2.5.
    4. For all simulations, the slope (and intercept) for each observer was obtained by taking a random sample from a normal distribution, with  $\sigma_{\text{slope}} = .2$ , and  $\sigma_{\text{intercept}} = .4$ .
- 

**APPENDIX B**  
**General Procedure for Monte Carlo Simulations**  
**of the Generalized Context Model**

---

**Basic Assumptions**

1. Each simulation consisted of 10 hypothetical observers.
2. The MDS psychological space (i.e., stimulus coordinates) was constant across observers.
3. The scaling constant  $c$  was set to 1.5 to yield the appropriate accuracy rates.

**Perceptual Representation Assumptions**

1. The perceptual means were set to the relevant stimulus level. For example, the stimulus at Level 1 along both stimulus dimensions had a mean perceptual effect of (1,1). The stimulus at Level 1 along Dimension 1 and Level 2 along Dimension 2 had a mean perceptual effect of (1,2), and so on.
2. *Perceptual Integrity*: In line with the MDS literature and previous applications of the GCM, two distance metric and similarity function pairings were examined. One assumed a Euclidean distance metric and a Gaussian similarity function. The second assumed a Euclidean distance metric and an exponential decay similarity function.
3. *Perceptual Separability*: In line with the MDS literature and some applications of the GCM, a city-block metric and an exponential decay similarity function were assumed.

**Categorization Rules and Individual Differences**

1. Selective Attention Condition
    1. Approximately Optimal Parameters
      1. Simulation Set 1:  $w = 0$ ,  $\beta = .5$ .
      2. Simulation Set 2:  $w = 0$ ; Observers 1–5,  $\beta = .3$ ; Observers 6–10,  $\beta = .7$ .
      3. For both sets of simulations,  $w$  was fixed at 0 for each observer. The value of  $\beta$  for each observer was obtained by taking a random sample from a normal distribution, with the mean defined above and  $\sigma_{\beta} = .2$ .
    2. Nonoptimal Parameters
      1. Simulation Set 1: same as Simulation Set 1 above.
      2. Simulation Set 2: same as Simulation Set 2 above.
      3. For both sets of simulations,  $\sigma_w = .2$ , and  $\sigma_{\beta} = .2$ .
  2. Equal Attention Condition
    1. Approximately Optimal Parameters
      1.  $w = .5$ ,  $\beta = .5$ .
      2. The value of  $w$  (and  $\beta$ ) for each observer was obtained by taking a random sample from a normal distribution, with the mean defined above and  $\sigma_w = .2$ , and  $\sigma_{\beta} = .2$ .
    2. Nonoptimal Parameters
      1. Observers 1–5,  $w = .3$ ,  $\beta = .5$ ; Observers 6–10,  $w = .7$ ,  $\beta = .5$ .
      2. The value of  $w$  (and  $\beta$ ) for each observer was obtained by taking a random sample from a normal distribution, with the mean defined above and  $\sigma_w = .2$ , and  $\sigma_{\beta} = .2$ .
-