

# Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification

ROBERT M. NOSOFSKY

*Indiana University, Bloomington, Indiana*

and

THOMAS J. PALMERI

*Vanderbilt University, Nashville, Tennessee*

The authors compared the exemplar-based random-walk (EBRW) model of Nosofsky and Palmeri (1997) and the decision-bound model (DBM) of Ashby and Maddox (1994; Maddox & Ashby, 1996) on their ability to predict performance in Garner's (1974) speeded classification tasks. A key question was the extent to which the models could predict facilitation in the correlated task and interference in the filtering task, in situations involving integral-dimension stimuli. To obtain rigorous constraints for model evaluation, the goal was to fit the detailed structure of the response time (RT) distribution data associated with each individual stimulus in each task. Both models yielded reasonably good global quantitative fits to the RT distribution and accuracy data. However, the DBM failed to properly characterize the interference effects in the filtering task. Apparently, a fundamental limitation of the DBM is that it predicts that the fastest RTs in the filtering task should be faster than the fastest RTs in the control task, whereas the opposite pattern was observed in our data.

An important direction in the recent development of formal models of perceptual categorization has been to account for the time course of classification decision making. As current models grow increasingly sophisticated, many are able to account for the same fundamental patterns of results regarding rates of learning, overall accuracy, and patterns of generalization. By requiring competing models to simultaneously account for classification response time (RT), additional constraints are provided for choosing among them, and a deeper understanding of the processes underlying perceptual categorization may be obtained.

The purpose of the present research was to test two recently developed models on their ability to account for fundamental patterns of results involving multidimensional speeded perceptual classification. The models under investigation are the *exemplar-based random walk model* proposed by Nosofsky and Palmeri (1997; Palmeri, 1997) and the *decision-bound model* proposed by Ashby and Maddox (1991, 1994; Ashby, Boynton, & Lee, 1994; Maddox & Ashby, 1996). An infinite variety of speeded classification paradigms can be designed for comparing the models, but we hoped to initiate the investigation by collecting data that researchers might regard as fundamental. With this goal in mind, we used as an experimental paradigm the classic set of speeded categorization tasks used by Garner (1974) and his colleagues for distinguish-

ing between integral and separable dimensions. These tasks provide canonical information about how basic manipulations of stimulus structure influence multidimensional classification RT, so we judged them to serve well as a fundamental testing ground. Also, Ashby and Maddox (1994) and Maddox and Ashby (1996) have already begun extensive theoretical and empirical investigations into the ability of the decision-bound model to account for performance in these tasks, so this arena seemed like a particularly attractive one in which to conduct the model comparisons.

Beyond comparing the models on their ability to simply predict mean RTs in these tasks, we followed Maddox and Ashby (1996) by focusing on the ability of the models to predict the structure of the entire set of RT distributions. An important advantage of this research approach is that the RT distributions provide far more constraints for model fitting than do the means alone (e.g., Ratcliff & Murdock, 1976). Indeed, one of the major themes in the present research involves the demonstration that, although certain versions of the models yield excellent accounts of the mean RTs in these tasks, they are falsified in striking fashion when one considers the full structure of the RT distributions.

We organize our article by first briefly reviewing the exemplar-based random-walk model and the decision-bound model of speeded classification. Next, we review the speeded classification tasks under investigation and the main patterns of performance that have been observed in previous research. We then introduce some important issues concerning how the alternative models might fare in accounting for the main patterns of data in these tasks. Finally, in the main section of the article, we conduct rig-

---

This work was supported by Grant PHS R01 MH48494-05 from the National Institute of Mental Health to R.M.N. Correspondence should be addressed to R. M. Nosofsky, Department of Psychology, Indiana University, Bloomington, IN 47405 (e-mail: nosofsky@indiana.edu), or to T. J. Palmeri, Department of Psychology, Vanderbilt University, 301 Wilson Hall, Nashville, TN 37240 (e-mail: palmerit@ctrvax.vanderbilt.edu).

orous quantitative tests of the ability of the models to fit the actual RT distributions and accuracy rates observed for individual stimuli in the Garner speeded classification tasks.

### EXEMPLAR-BASED RANDOM-WALK MODEL

The exemplar-based random-walk (EBRW) model proposed by Nosofsky and Palmeri (1997) combines elements of Nosofsky's (1984, 1986) *generalized context model* of categorization and Logan's (1988, 1992) *instance-based model of automaticity*. According to the EBRW model, people represent categories by storing individual exemplars of categories in memory. When an item is presented to be classified, it acts as a retrieval cue for the stored exemplars. The exemplars race to be retrieved, with rates determined by how similar they are to the presented item. These retrieved exemplars then provide incremental information that enters into a random-walk process for making classification decisions (e.g., Bussemeyer, 1982; Link & Heath, 1975; Luce, 1986; Townsend & Ashby, 1983).

We start by reviewing the similarity assumptions in the model. In the EBRW model, exemplars are represented as points in a multidimensional psychological space, and similarity between exemplars is a decreasing function of their distance in the space. Selective attention processes systematically modify the structure of the space in which the exemplars are embedded. Specifically, the distance between exemplars  $i$  and  $j$ , ( $d_{ij}$ ), is given by

$$d_{ij} = \left[ \sum_{m=1}^M w_m |x_{im} - x_{jm}|^\gamma \right]^{1/\gamma}, \quad (1)$$

where  $x_{im}$  is the psychological value of exemplar  $i$  on dimension  $m$ , and  $M$  is the number of dimensions composing the objects. The value of  $\gamma$  in Equation 1 determines the distance metric of the space. Common values are  $\gamma = 1$ , which yields a city-block distance metric; and  $\gamma = 2$ , which yields a Euclidean metric (Garner, 1974; Shepard, 1964, 1991). The parameters  $w_m$  ( $0 \leq w_m \leq 1$ ,  $\sum w_m = 1$ ) form a set of "attention weights" given to each dimension  $m$ . Research suggests that observers are inclined to attend selectively to those dimensions that are relevant for performing a classification and to ignore dimensions that are irrelevant. However, the extent to which this selective attention process operates depends on the types of dimensions that compose the objects. Selective attention operates readily for stimuli composed of "separable" dimensions but is more difficult for stimuli composed of "integral" dimensions (e.g., Garner, 1974; McKinley & Nosofsky, 1996; Nosofsky, 1987; Nosofsky & Palmeri, 1996; Shepard, 1964; Shepard & Chang, 1963). Finally, the similarity between exemplars  $i$  and  $j$  is an exponential decay function of their distance in the psychological space,

$$s_{ij} = \exp(-c \cdot d_{ij}), \quad (2)$$

where  $c$  is an overall scaling parameter (Shepard, 1987).

When an item is presented to be classified, it causes all exemplars in memory to be "activated." Because of factors such as recency of presentation, exemplars may reside in memory with differing strengths. Let  $H_j$  denote the memory strength for exemplar  $j$ . Then, when item  $i$  is presented, the activation for exemplar  $j$  is given by

$$a_{ij} = H_j \cdot s_{ij}. \quad (3)$$

That is, exemplar activation is determined jointly by the exemplar's strength in memory and by its similarity to the test item.

All activated exemplars race to be retrieved from memory. The race process is stochastic. Specifically, the exemplars race exponentially with rates determined by their activation values. (For similar assumptions concerning exemplar race processes, see, for example, Bundesen, 1990, Logan, 1997, and Marley, 1992.) The probability density that exemplar  $j$  completes its race at time  $t$ , given test item  $i$ , is given by

$$f(t) = a_{ij} \exp(-a_{ij} \cdot t). \quad (4)$$

The first exemplar to complete its race is retrieved. Thus, those exemplars most likely to be retrieved are the ones with the greatest memory strengths and that are highly similar to the test item.

The retrieved exemplars drive a random-walk process, as illustrated in Figure 1. There is a random-walk counter with initial value zero. The observer establishes criteria representing the amount of evidence needed to execute either a Category A response (+A) or a Category B response (-B). If the retrieved exemplar belongs to Category A, then the counter is increased by unit value, whereas if the retrieved exemplar belongs to Category B, then the counter is decreased by unit value. If the counter reaches either the +A or the -B criterion, then the appropriate categorization response is made. Otherwise, a new race is initiated, a second exemplar is retrieved, and the process continues.

The time to take each step in the random walk is given by

$$\Delta T = \alpha + t_x, \quad (5)$$

where  $\alpha$  is a constant term associated with each step, and  $t_x$  is the time that it takes to retrieve the winning exemplar. Note that on each step of the random walk, the category label associated with the retrieved exemplar needs to be extracted, and the appropriate information then is accumulated on the counter. A psychological interpretation for the  $\alpha$  parameter is that it represents the time needed for this category-label extraction and accumulation process.

The random-walk process just described determines the time for classification decision-making to occur. Total RT is also affected by the duration of other stages of processing, such as encoding and response-execution. For simplicity, we assume that the overall duration of these other residual stages of processing is normally distributed with mean  $\mu_R$  and variance  $\sigma_R^2$ . Furthermore, we assume

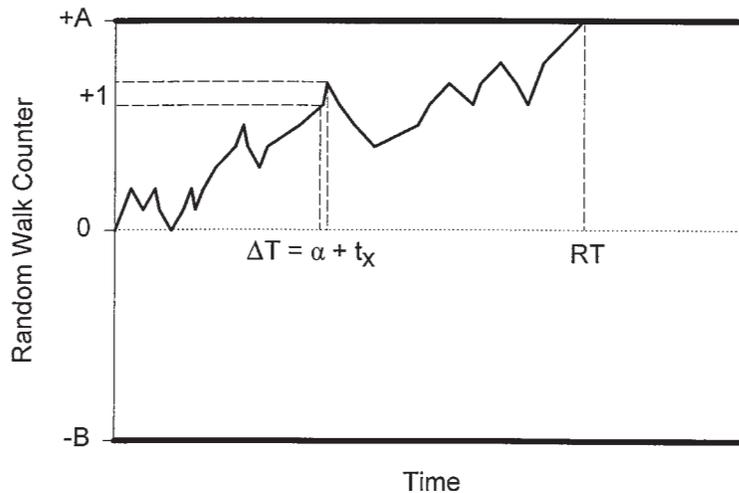


Figure 1. Schematic illustration of the random-walk process in the EBRW model.

that the distribution of residual times is the same for all stimuli and is independent of the duration of the decision-making stage.<sup>1</sup>

It is useful at this point to discuss some main conceptual predictions associated with the EBRW model. Note that, according to the model, classification RT is determined jointly by the number of steps required to complete the random walk and by the speed with which these individual steps are made. One major prediction is that classification RT should be fastest (and accuracy greatest) for objects that are similar to the exemplars of one category and dissimilar to the exemplars of the contrast category. Under such circumstances, each exemplar that is retrieved tends to belong to the same category, and the random walk marches consistently to the appropriate category criterion. By contrast, items that are similar to the exemplars of both categories result in slow RTs. The reason is that the random walk will tend to wander back and forth, sometimes retrieving exemplars from one category and other times retrieving exemplars from the other category.

A second prediction is that increased experience with specific exemplars should facilitate performance on these items. As experience with a specific exemplar increases, more tokens of that exemplar come to be stored in memory. Thus, a greater number of tokens of the exemplar race to be retrieved when an item is presented at test. A fundamental statistical prediction of parallel race-horse models is that the greater the number of tokens of an exemplar that race to be retrieved, the faster the *winning* retrieval time tends to be (e.g., Logan, 1988; Raab, 1962; Townsend & Ashby, 1983). Intuitively, the greater the number of exemplar tokens that participate in the race, the greater is the probability that at least one of the retrieval times will be particularly fast. These faster winning retrieval times result in faster individual steps in the random-walk process (Equation 5). As will be seen, these properties of the EBRW model are important for explaining the performance patterns in the Garner speeded classification tasks.

## DECISION-BOUND MODEL

Ashby's decision-bound model (DBM) combines concepts from the general recognition theory (GRT) of Ashby and Townsend (1986) with a descriptive model of RT known as the *RT-distance hypothesis* (Ashby et al., 1994). In describing the model, we limit consideration to situations in which the stimuli vary along two dimensions, which is true of the Garner speeded classification tasks under current investigation. According to GRT, each stimulus gives rise to a bivariate normal distribution in a perceptual space. This assumption of variability in the underlying perceptual representation associated with multiple presentations of a stimulus lies at the heart of GRT. Each bivariate normal distribution is described by a mean and variance along each dimension and by a correlation between dimensions. Furthermore, an observer establishes a decision bound to partition the perceptual space into response regions. Any time a stimulus gives rise to a percept that falls in Region A, a Category A response is made.

Classification RTs are predicted by the model in accord with Ashby et al.'s (1994, p. 11) RT-distance hypothesis: "RT decreases with the distance in psychological space from the stimulus representation to the decision bound that separates the exemplars of the contrasting categories." Although highly descriptive in nature, the RT-distance hypothesis has a long history of success in unidimensional domains (see Ashby et al., 1994, for a review), and recent extensions to multidimensional RTs also seem extremely promising (Ashby et al., 1994; Maddox & Ashby, 1996; Thomas, 1996).

To derive quantitative predictions from the DBM, specific assumptions need to be introduced regarding the form of the function that relates RT to distance from the boundary. Following Maddox and Ashby (1996) and Murdock (1985), we assume that decision time ( $T_D$ ) is related to distance-from-bound ( $D$ ) by means of an exponential decay function,

$$T_D = k \cdot \exp(-\beta \cdot D), \quad (6)$$

where the parameter  $\beta$  determines the rate at which RT decreases, and  $k$  gives the maximum decision time (when  $D = 0$ ).<sup>2</sup> Furthermore, to test the model, specific assumptions need to be introduced regarding the form of the decision bounds that are used for partitioning the perceptual space into response regions. We discuss this issue in some depth after introducing the Garner speeded classification tasks.

Finally, the RT-distance hypothesis is intended to describe the time course of the classification decision-making stage. We follow Maddox and Ashby (1996) by assuming that the overall duration of residual stages of processing is normally distributed with mean  $\mu_R$  and variance  $\sigma_R^2$ , independent of the stimulus that is presented. Note that this assumption is the same as we made for the EBRW model.

### GARNER'S (1974) SPEEDED CLASSIFICATION TASKS

The basic setup for Garner's (1974, chap. 6) speeded classification tasks is illustrated schematically in Figure 2. There are four stimuli varying along two dimensions, with two values per dimension, and the dimension values are roughly equally discriminable. In all tasks, the requirement is to classify each stimulus into its assigned category as rapidly as possible without making errors.

In the *control* task, on each trial, one of two possible stimuli that vary along one dimension is presented. An example is to classify Stimulus A into Category 1 and Stimulus B into Category 2. The key aspect of the control task is that a single dimension is relevant for classifying each object (Dimension 1 in the A vs. B example), and

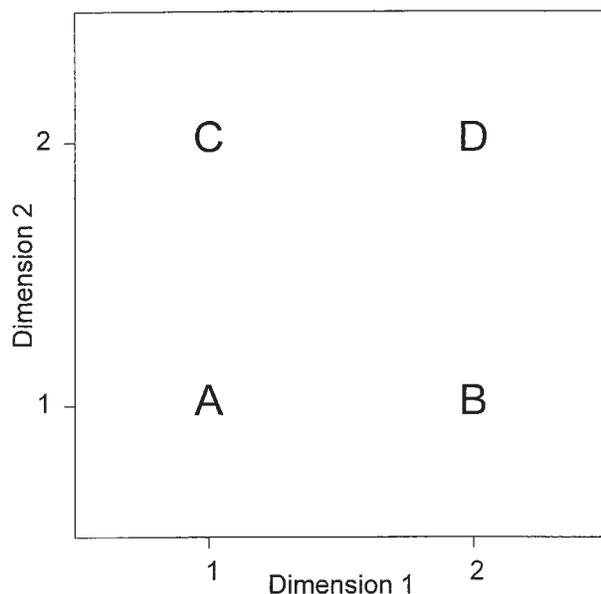


Figure 2. Schematic illustration of the stimulus configuration used in Garner's speeded classification tasks.

values on the irrelevant dimension are held constant. In the *filtering* task, on each trial, any one of the four stimuli in the complete set is presented. As is the case in the control task, a single dimension is relevant for classifying the objects, but now values along the irrelevant dimension vary. An example is to classify Stimuli A and C into Category 1 and Stimuli B and D into Category 2. In the *correlated* task, one of two stimuli is presented on each trial, but they differ along both dimensions. An example is to classify Stimulus A into Category 1 and Stimulus D into Category 2.

In situations involving moderately discriminable stimuli and relatively inexperienced participants, a well-known pattern of results emerges in these tasks (see Garner, 1974, 1976, for reviews). Consider first the case of highly *separable*-dimension stimuli. Such stimuli are ones in which the dimensions remain psychologically distinct when in combination, such as forms varying in their shape and color, or circles varying in their size and angle of orientation of a radial line. Here, to a first approximation, mean RTs are basically identical across the control, filtering, and correlated tasks.

A much different pattern emerges for stimuli varying along *integral* dimensions, however. Integral-dimension stimuli are ones in which the dimensions combine into relatively unanalyzable, integral wholes. Examples are colors varying in brightness and saturation, or tones varying in pitch and loudness. For integral-dimension stimuli, there is interference in the filtering task and facilitation in the correlated task. That is, for integral-dimension stimuli, mean RTs in the filtering task are markedly slower than in the corresponding control tasks, whereas mean RTs in the correlated tasks are markedly faster than in the control tasks.

We emphasize that the preceding summary is intended as an approximation to the detailed pattern of results obtained in these tasks. For example, in situations involving highly confusable stimuli and extended training, facilitation effects are sometimes observed in the correlated task even for separable-dimension stimuli (e.g., Garner & Felfoldy, 1970; Maddox & Ashby, 1996). Likewise, because the distinction between integral- versus separable-dimension stimuli exists on a continuum, it may not be possible to observe a pure case of either type of dimensional interaction. Thus, it would not be surprising to see some small amounts of interference observed in the filtering task for stimuli that are traditionally viewed as being composed of separable dimensions. The main point is that, relative to separable-dimension stimuli, integral-dimension stimuli show far more interference in the filtering task and far more facilitation in the correlated task.

The performance patterns observed in these speeded classification tasks are among the fundamental converging operations used by Garner (1974, 1976) for distinguishing among alternative dimensional interactions. Intuitively, optimal performance in the filtering task could be achieved if the observer attends to only the relevant dimension and ignores the irrelevant one. (A deeper the-

oretical basis for this intuition is provided in the section in which we consider the predictions of the formal models.) For highly separable-dimension stimuli, this goal of attending selectively to only the relevant dimension is easily accomplished. However, the selective attention process is far less efficient in the case of integral-dimension stimuli, so variation along the irrelevant dimension hinders performance in the filtering task when integral-dimension stimuli are used.

### APPLYING THE MODELS TO THE SPEEDED CLASSIFICATION TASKS

In this section, we discuss in a general conceptual fashion whether or not the EBRW model and the DBM can account for the main performance patterns observed in the control, filtering, and correlated tasks. The issues that are highlighted are then examined in more rigorous, quantitative fashion in the model-fitting section of the article.

#### Applying the EBRW Model

In discussing the predictions of the EBRW model, we assume for simplicity that the exemplars are positioned in multidimensional psychological space as illustrated in Figure 2. We emphasize, however, that, as in all past applications of the exemplar model, a complete account of performance requires assumptions about the locations of the exemplars in psychological space, which may not correspond precisely to the physical space defined by the experimenter (Nosofsky, 1992).

According to the EBRW model, when highly separable-dimension stimuli are used, participants attend selectively to the single dimension that they are instructed is relevant for solving the task. In the extreme situation in which all attention weight is placed on this single dimension (see Equation 1), the structures of the control, filtering, and correlated tasks are all identical to one another. Consider, for example, a situation in which the stimuli vary in shape (triangle vs. circle) and color (red vs. blue). Suppose that, in the control and filtering tasks, the relevant dimension is shape. Thus, in one of the control tasks, the observer classifies the red triangle into Category 1 and the red circle into Category 2. In the filtering task, the observer classifies the red and blue triangles into Category 1 and the red and blue circles into Category 2. From the perspective of the experimenter, there are two stimuli being classified in the control task and four stimuli being classified in the filtering task. However, in the extreme situation in which *all* attention is placed on only the shape dimension, then, from the perspective of the observer, there are only two stimuli in both tasks—namely, a triangle and a circle. The same holds true in the correlated task in the extreme situation in which all attention is placed on the single dimension of shape. Because the structures of the control, filtering, and correlated tasks are identical in this situation, it is evident that the EBRW model predicts identical performance in each task.<sup>3</sup>

Why, according to the EBRW model, is there facilitation in the correlated task when integral-dimension stim-

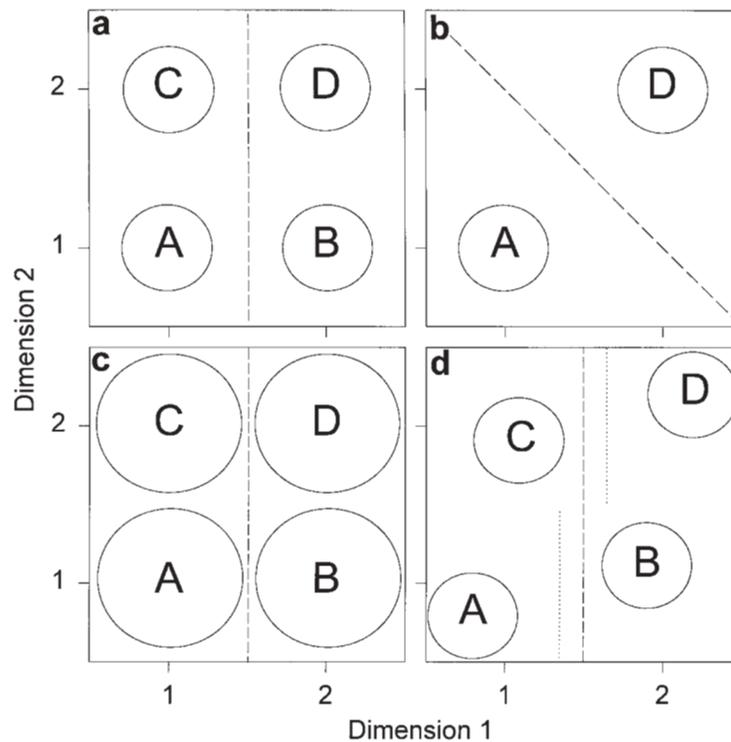
uli are used? Note that the stimuli in the correlated task (e.g., A and D) are less similar than are the stimuli in the control task (e.g., A and B; see Figure 2). According to the EBRW model, the greater discriminability of the stimuli in the correlated task causes less competition in the random-walk process. For example, when Stimulus A is presented, it will rarely cause exemplars of the highly dissimilar Stimulus D to be retrieved. Thus, the random-walk counter will move in consistent fashion to its appropriate response criterion. The counter wanders back and forth to a greater extent in the control task, because of the higher similarity of the exemplars from contrasting categories.

Why is there interference in the filtering task? One reason, according to the EBRW model, is that, for any given item that is presented, there are twice as many exemplars in memory that are identical to that item in the control task as there are in the filtering task (e.g., in the control task, Stimulus A is presented on half the trials, whereas in the filtering task, it is presented on one fourth of the trials). A test item is most likely to retrieve exemplars to which it is identical. As explained previously, the greater the number of these exemplars in memory, the faster the winning retrieval times tend to be; thus, the random walk finishes more quickly.

This line of reasoning about the basis for interference in the filtering task assumes equal memory strengths for all exemplars. A more plausible assumption is that memories for previous exemplars get weaker each time a subsequent test item is presented. Introducing this assumption, however, yields the same interference predictions. In the control task, an exemplar that is identical to the test item will be presented on half of the immediately preceding trials; however, in the filtering task, an identical exemplar is presented on only one fourth of the immediately preceding trials. Because the activation rates with which exemplars race are influenced by their memory strengths (Equation 3), the retrieval of identical exemplars is more rapid in the control task than in the filtering task.

#### Applying the DBM

A simplified GRT representation for the four stimuli in the Garner tasks is shown in Figure 3a. As illustrated in the figure, each stimulus gives rise to a bivariate normal distribution. The distributions are represented schematically by “contours of equal likelihood”—namely, sets of points that are equally likely to be produced by the distribution (Ashby & Gott, 1988). The center of each contour gives the mean of the bivariate normal distribution. The expanse of the contour along each dimension represents the variability of the distribution along that dimension. For simplicity in this discussion, we assume equal variances along both dimensions, zero correlation between dimensions, and that the variances are identical for all stimuli within a given task. Also, stimuli with identical physical values on a given dimension are assumed to have the same mean on that dimension in the perceptual representation. Although some of the simplifying assump-



**Figure 3.** Panel a: Simplified general recognition theory (GRT) representation of the stimulus configuration in the Garner tasks when separable-dimension stimuli are used. Panel b: GRT representation for the correlated task when a diagonal decision boundary is used. Panel c: GRT representation with increased variances due to uncertainty effects. Panel d: GRT representation illustrating a mean-shift integrality.

tions could be relaxed, this GRT representation is the type expected for stimuli varying along highly separable dimensions (see Ashby & Townsend, 1986, for a more detailed discussion).

In a control or filtering task in which Dimension 1 is relevant, we assume that the decision bound would be as illustrated in Figure 3a—namely, a linear boundary orthogonal to Dimension 1. Ashby and Maddox (1994, pp. 443–444) suggested that, regardless of the types of stimulus dimensions, it was plausible that a naive observer (such as the participants in our study) would use this type of boundary. If naive observers also use this orthogonal linear boundary in the correlated task, then the DBM accounts in straightforward fashion for the canonical pattern of results observed for separable-dimension stimuli. In particular, the distribution of distances of percepts from the decision bound is identical across the control, filtering, and correlated tasks; therefore, the predicted mean RTs would be identical.

Why, according to the DBM, is there facilitation in the correlated task when integral-dimension stimuli are used? A likely reason is that, in such situations, participants adopt a more optimal linear boundary—namely, a diagonal boundary that takes advantage of both dimensions of variation in the correlated task. Such a boundary is il-

lustrated in Figure 3b. As can be seen, percepts tend to be further from this boundary in the correlated task than from the orthogonal linear boundary used in the control task; therefore, facilitation is observed. Indeed, Maddox and Ashby (1996) provided evidence that highly experienced observers learn to use such a diagonal linear boundary even when rapidly classifying separable-dimension stimuli.

Although Ashby and Maddox (1994) did not propose an explicit model, they discussed two main ideas regarding why interference might be observed in the filtering task when integral-dimension stimuli are used. First, because the filtering task has four stimuli, whereas the control task has only two, stimulus uncertainty effects may play a role (Ashby & Maddox, 1994, p. 427). Ashby and Maddox (1994, p. 452) proposed to model stimulus uncertainty effects in terms of increased variances of the perceptual distributions associated with each stimulus (cf. Durlach & Braida, 1969; Luce, Green, & Weber, 1976; Marley & Cook, 1984). As illustrated in Figure 3c, with this increase in variance, more percepts would lie close to the decision bound in the filtering task than in the control task. Although this increase in variance would also result in an increased number of percepts being located *farther* from the bound in the filtering task than in the con-

control task (compare Figures 3a and 3c), the nonlinear function relating RT to distance (Equation 6) can result in a slower overall mean RT when averaging over all percepts.

Taken by itself, however, this increased variance assumption makes a strong prediction: Simply put, the very fastest RTs in the filtering task should be faster than the very fastest RTs in the control task. This strong prediction is one of the reasons why it is critical, when evaluating the models, to require them to fit the entire set of RT distributions, instead of simply the means.

A second idea is that interference in the filtering task may result from what Ashby and Maddox (1994, pp. 432, 444) term a *mean-shift integrality*. This idea is illustrated in Figure 3d. According to the hypothesis, although the stimuli vary orthogonally on Dimensions 1 and 2 in the physical space, they occupy the vertices of a parallelogram in the underlying psychological space. The decision boundaries that would presumably be used in the control tasks are illustrated by the dotted lines, and the decision boundary in the filtering task is illustrated by the dashed line. Given appropriate parameter settings, this model can predict interference in the filtering task (see Ashby & Maddox, 1994, pp. 448–449). The intuition is that Stimuli B and C, which lie very close to the boundary in the filtering task, result in very long RTs. Although Stimuli A and D are far from the bound, the nonlinear function relating RT to distance can again result in a slower overall mean RT in the filtering task than in the control tasks after averaging over all stimuli.

The hypothesis of mean-shift integrality draws support from previous work concerned with the classification of integral-dimension stimuli. For example, Melara and Marks (1990) conducted Garner tasks in which the stimuli were tones that varied in pitch (low, high) and loudness (soft, loud). In the correlated tasks, they found that the A–D discrimination (low–soft vs. high–loud) was performed more rapidly than the B–C discrimination (high–soft vs. low–loud)—a result that they attributed to an effect of *dimensional congruity*. The mean-shift integrality illustrated in Figure 3d is consistent with such a pattern of results. Likewise, Kingston and Macmillan (1995) conducted Garner tasks in which the stimuli were vowel sounds varying along dimensions of nasalization and first-formant frequency. Although the dependent measure in their study was accuracy, not RT, Kingston and Macmillan observed strong interactions between these perceptual dimensions analogous to the effects obtained by Melara and Marks (1990). They modeled these perceptual interactions in terms of a mean-shift integrality representation, such as the one shown in Figure 3d.

We do not question the critical importance of mean-shift integrality in providing a full account of speeded perceptual classification. Nevertheless, in combination with the RT–distance hypothesis, the mean-shift integrality hypothesis makes a strong prediction in the present context: As can be seen from Figure 3d, one of the diagonal pairs in the filtering task, either Pair A–D or Pair B–C, should have faster mean RTs than the pairs tested in the control tasks, A–B and C–D (which pair depends on the

precise shape of the parallelogram). If there is some combination of mean-shift integrality *and* increased variance due to uncertainty effects, the DBM is not forced into this prediction. However, even this more complicated hypothesis predicts that the fastest RTs for one of the diagonal pairs in the filtering task should be faster than those in the control tasks. Thus, it is again critical to examine the RT distributions for individual stimuli across the tasks to evaluate the merits of these hypotheses.

## EXPERIMENT

The data that are analyzed and modeled in this article were obtained in an experiment reported previously by Nosofsky and Palmeri (1997). In this previous study, we demonstrated that the EBRW model was consistent with the global pattern of results obtained across the control, correlated, and filtering tasks. However, only the mean RTs and accuracies across tasks were reported and analyzed. By contrast, in the present study, we have the goal of modeling the complete structure of the RT distributions and accuracies observed for each of the individual stimuli across the tasks; therefore, the aims are far more ambitious.

To obtain a data set suitable for quantitative fitting, Nosofsky and Palmeri (1997) conducted a battery of the Garner speeded classification tasks under the umbrella of a single experiment. Whereas Maddox and Ashby (1996) had tested a few individual participants for extended periods of time in speeded classification tasks involving separable-dimension stimuli, we decided to test multiple participants in a single session in tasks involving integral-dimension stimuli. The main reason we decided to focus on integral-dimension stimuli is that the patterns of RT results appear to provide a greater challenge to the models. For example, there is very little variability in performance across the control and filtering tasks when separable-dimension stimuli are used. In general, it seems less challenging for a model to fit identical patterns of performance across these tasks than to fit large variations in performance. Indeed, in the previous section, we already pointed out reasons to believe that the DBM might have difficulty accounting for large performance variations across the control and filtering tasks.

The issue of testing individual participants over extended periods of time, as opposed to testing multiple participants in a single experimental session, is a complex matter. Both research approaches have their advantages and disadvantages. A disadvantage of testing multiple participants and modeling group data is that, if the parameters for individual participants vary substantially, then the group data are likely to obscure patterns at the individual participant level. Unfortunately, however, the fundamental patterns of RT results systematized by Garner in his classic work were obtained in testing conditions with relatively inexperienced participants. Testing a single participant over an extended period of time may well produce changes in psychological processes that are not representative of the novice observer.

For example, our theoretical analyses involving the DBM assume that the observer adopts an orthogonal linear boundary in the control and filtering tasks (see Figure 3d). As argued by Ashby and Maddox (1994), this assumption seems highly plausible in the case of relatively inexperienced observers, because the single linear boundary is presumably easy to implement and the participants are given explicit instructions to classify on the basis of a single dimension. But, with extensive training, these participants may learn to adopt much more complex decision boundaries (e.g., curvilinear boundaries that come closer to optimizing performance than do linear bounds). Because an infinite variety of decision bounds are possible, one's ability to test the DBM depends on conducting experimental conditions that yield plausible constraints on the form of the bound. Therefore, we felt it was important to initiate the investigation by testing relatively inexperienced participants where the orthogonal linear-bound hypothesis was reasonable.

Likewise, from the perspective of the EBRW model, it may be that, with extensive training, an observer can become highly efficient at selectively attending even to integral-dimension stimuli. Thus, the fundamental interference effect might be greatly diminished, thereby removing the challenge posed to the competing models. Although studies that examine the effects of extensive training on performance in the Garner tasks are of a great deal of interest, we chose to first investigate performance under more standard testing conditions.

Because any single observer contributes limited data in the course of a single session, in this article, we adopted the procedure known as *vincentizing* for creating group RT distributions (Ratcliff, 1979). Briefly, in this procedure, each observer's RT data are divided into quantiles. The mean RTs in each quantile are then averaged across observers to create the group distribution. Ratcliff illustrated with various examples that, if the individual RT distributions have the same basic "shape," then the group distribution created by the method of vincentizing also has this shape. Thus, if behavior at the individual participant level is fairly homogeneous, then the group RT distribution should be representative of the patterns of performance observed at the individual participant level.

Because the experimental method has been described in depth in the article by Nosofsky and Palmeri (1997), we only briefly review it here.

## Method

**Subjects.** The subjects were 26 graduate and undergraduate students associated with the Indiana University Psychology Department.

**Stimuli and Apparatus.** The stimuli were tones varying in pitch (frequency) and loudness (intensity). Previous work indicates that such stimuli have integral-dimension characteristics (Grau & Kemer Nelson, 1988; Melara & Marks, 1990). The stimulus set was constructed by combining orthogonally frequency values of 900 and 950 Hz, with intensity values of 60 and 70 dB. Thus, the physical values corresponding to Stimuli A–D in Figure 2 were A = (900, 60), B = (950, 60), C = (900, 70), and D = (950, 70). Previous work reported by Melara and Marks (1990) and Melara and Mounds (1994) indicated that the pitch and loudness differences used are roughly equally discriminable.

**Procedure.** The experiment was organized into 13 conditions as shown in Table 1. However, only the first 8 conditions listed in Table 1 are analyzed and modeled in the present article. The stimuli that were eligible for presentation in each condition, as well as their division into categories, are reported in Column 2 of the table. The ordering of conditions was balanced according to two Latin squares. Each condition consisted of 96 trials, with each eligible stimulus presented with equal frequency. Ordering of stimulus presentations was randomized for each individual participant and condition.

## Results

The experimental conditions of interest in the present article are the first 8 listed in Table 1. There are two control–pitch conditions (A vs. B, and C vs. D), two control–loudness conditions (A vs. C, and B vs. D), two correlated conditions (A vs. D, and B vs. C), a filter–pitch condition (A,C vs. B,D), and a filter–loudness condition (A,B vs. C,D). Note that there is a total of 20 condition  $\times$  stimulus combinations involved across these tasks (i.e., two stimuli for each of the four control and two correlated tasks, and four stimuli for each of the two filtering tasks). Our theoretical goal was to simultaneously model the RT distributions and accuracies associated with all 20 combinations. The first 24 trials of each condition were considered practice. Only the final 72 trials of each condition were included in the analyses and the modeling.

**Preliminary data transformation.** We took two preliminary steps before creating the group vincentized distributions. First, the RT data for each individual participant in each individual condition were examined. RTs greater than 3 standard deviations above the mean for each individual data set were eliminated, as were RTs less than 100 msec. This procedure led to the removal of less than 0.1% of the total observations.

Second, to create a more nearly homogeneous data set appropriate for vincentizing, we attempted to remove the influence of task-order position on the RT data. Presumably, as participants gain increased experience in these speeded classification tasks, overall performance improves, such that tasks tested early in the sequence will tend to

**Table 1**  
Experimental Design and Mean Correct Response Times  
(in Milliseconds) and Error Rates in Each Condition

Task	Stimuli	MRT	<i>P</i> (E)
Control–Pitch 1	A vs. B	397.8	.037
Control–Pitch 2	C vs. D	387.0	.027
Control–Loudness 1	A vs. C	421.5	.036
Control–Loudness 2	B vs. D	415.7	.032
Correlated 1	A vs. D	367.5	.016
Correlated 2	B vs. C	371.0	.023
Filter–Pitch	A,C vs. B,D	439.1	.043
Filter–Loudness	A,B vs. C,D	447.2	.047
Stretch–Pitch	A,G vs. B,H	—	—
Stretch–Loudness	A,E vs. C,F	—	—
Condensation	A,D vs. B,C	—	—
Focus (25%)	A vs. B,C,D	—	—
Focus (50%)	A vs. B,C,D	—	—

Note—MRT = mean response time; *P*(E) = proportion of errors. Stimuli G and H in the stretch–pitch condition had frequency values of 1000 Hz. Stimuli E and F in the stretch–loudness condition had intensity values of 80 dB.

have slower and more variable RTs than will tasks tested late. Figure 4 plots the means and standard deviations of all RTs in the eight tasks as a function of position of testing. To estimate the influence of task position on the RTs, we fitted to each plot a power function that minimized the sum of squared deviations between predicted and observed data points. We used power functions because of their flexibility and previous suggestions in the literature that they provide good descriptions of the decreases in RT with practice (e.g., Logan, 1992; Newell & Rosenbloom, 1981). These fitted power functions are illustrated along with the observed data in Figure 4. For the mean RTs, the best-fitting power function was  $y = 217.8 + 238.7 \cdot x^{-0.12}$ ; for the standard deviations, the best-fitting function was  $y = 131.8 + 57.7 \cdot x^{-2.11}$ .

We then used a two-step standardization procedure to remove the estimated effect of task position on the RTs. (Nobel, 1996, used a similar transformation procedure to remove effects of extraneous variables in creating RT distributions in a series of memory experiments.) Let  $P_M(i)$  denote the mean RT predicted by the power function for position  $i$ , and let  $G_M$  denote the grand mean of these mean RTs across the 13 positions of testing. In the first step, all RT data in position  $i$  ( $X_i$ ) were standardized by using the transformation

$$X'_i = X_i + [G_M - P_M(i)]. \quad (7a)$$

This transformation removes the estimated effect of task position on the mean RTs. Likewise, let  $P_S(i)$  denote the standard deviation of RTs predicted by the power function for position  $i$ , and let  $G_S$  denote the grand mean of these standard deviations. In the second step, all of the

mean-standardized data in position  $i$ ,  $X'_i$ , were further standardized by using the transformation

$$X''_i = [G_S/P_S(i)] \cdot [X'_i - G_M] + G_M. \quad (7b)$$

This transformation removes the estimated effect of task position on the standard deviations of the RTs, while leaving the means unchanged. To the extent that the power-function estimates are exact, the net effect of these transformations is to produce RT distributions with the same mean and standard deviation across task position. We believe that this standardization procedure is sensible given our goal of creating group RT distributions suitable for quantitative testing of the models. We emphasize, however, that when we model the group RT distributions obtained from the raw data without the standardization procedure, the same patterns of results are obtained, and none of our conclusions are changed.

**Mean RTs and accuracies across tasks.** Before presenting the analysis of the individual stimulus RT distributions, we briefly characterize the global patterns of results observed across the tasks. The mean correct RTs and accuracies for each of the eight tasks are presented in Table 1. The main results replicate the classic patterns for integral-dimension stimuli. Mean RT was significantly slower in the filtering tasks than in the control tasks, indicating interference [pitch,  $t(25) = 3.86, p < .001$ ; loudness,  $t(25) = 2.81, p < .01$ ]. Mean RT in the correlated tasks was significantly faster than in either type of control task, indicating facilitation [pitch,  $t(25) = 6.29, p < .001$ ; loudness,  $t(25) = 5.63, p < .001$ ]. Mean RTs for discriminating pitch were somewhat faster than for discriminating loudness [control,  $t(25) = 2.90, p < .01$ ; filtering,

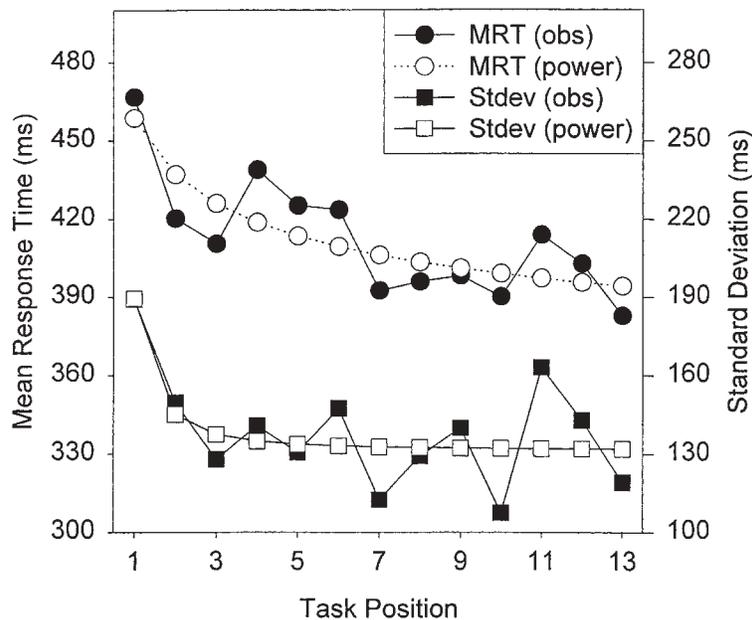


Figure 4. Mean RTs (MRTs) and standard deviations of RTs (Stdevs) plotted as a function of task position. The power functions that best fit these curves are also illustrated.

$t(25) = 0.38, p > .05]$ . Thus, in the ensuing modeling analyses, we expect the psychological representation for the stimuli to be slightly elongated along the pitch dimension. In addition, the probability of errors across the eight tasks mirrors the pattern for the mean RTs; therefore, the main qualitative findings do not appear to be attributable to a speed–accuracy tradeoff.

These main patterns of results appear to be reasonably homogeneous at the individual-participant level. For example, all 26 participants had faster mean RTs in the correlated tasks than in the control tasks, and 23 of the 26 participants had slower mean RTs in the filtering tasks than in the control tasks. In subsequent analyses, we report an important result that mean RTs and error rates were lower for the A–D stimulus pair than for the B–C stimulus pair in the filtering and correlated tasks (see MRT column of Table 2). Even at this more fine-grained level of analysis, the patterns are reasonably homogeneous at the individual-participant level. Of the 26 observers, 21 had faster mean RTs on the A–D pair than on the B–C pair, averaged across tasks. The key question now is the extent to which the alternative models can account quantitatively for the detailed performance patterns that underlie these main findings.

**Vincentized RT distributions.** We created group RT distributions for each individual stimulus in each task by the method of vincentizing (see Ratcliff, 1979, pp. 448–449, for a detailed illustration of the procedure). Specifically, taking each stimulus in each task separately, we divided each participant’s RT data into deciles (i.e., bins containing 10% of the RT observations). Only correct responses were included in forming these deciles. The mean RT in each decile was computed. The group RT distribution for each stimulus in each task was then constructed

by averaging these decile means across the 26 participants. The 20 resulting group RT distributions are reported in Table 2. In addition, the table presents the overall mean RTs and error probabilities for each of the individual stimuli across the tasks. These data form the bedrock for most of the ensuing modeling and statistical analyses reported in this article.

Conceptually, each individual bin mean in each vincentized distribution summarizes an interval containing 10% of the RT observations. We adopt the following conventions in modeling and displaying these distributions. First, the cutoff for each interval of RTs is assumed to be located midway between the means of adjacent intervals. For example, in Distribution 1 in the control–pitch task (CP-A), the means for the first three deciles are 295.5, 324.4, and 343.8. Thus, the lower cutoff for Interval 2 is 310.0, and the upper cutoff is 334.1. For purposes of displaying the data, the width (W) of each interval is then given by the distance between the lower and upper cutoffs.<sup>4</sup> Because each interval contains 10% of the observations, the height (H) of each interval is then defined as  $H = 0.1/W$ .

To gain insight into the overall structure of the RT distribution data, we computed global distributions corresponding to the control, correlated, and filtering tasks. Each global distribution was computed by averaging over the vincentized values corresponding to tasks of a given type. For example, the global control distribution was computed by averaging over Distributions 1–8 in Table 2. The structures of the resulting global distributions are illustrated in Figure 5a. As expected, the figure illustrates that RTs are fastest in the correlated tasks, intermediate in the control tasks, and slowest in the filtering tasks. The distributions are not simply shifted along the RT axis, how-

**Table 2**  
**Vincentized RT Distributions Associated with Each Individual Stimulus in Each Task,**  
**Together with the Mean RTs (in Milliseconds) and Proportions of Errors**

Task	Decile										MRT	P(E)
	1	2	3	4	5	6	7	8	9	10		
CP-A	295.5	324.4	343.8	361.9	378.8	395.9	414.9	439.3	478.3	568.5	400.0	.029
CP-B	293.1	323.7	342.4	359.4	374.9	391.9	410.8	432.5	470.9	555.3	395.6	.044
CP-C	288.6	315.8	336.0	352.8	368.5	384.6	401.0	420.3	461.6	542.2	386.6	.020
CP-D	289.8	317.1	335.5	352.0	367.6	385.6	403.1	423.9	461.5	536.9	387.3	.035
CL-A	309.0	341.6	361.3	379.8	397.3	413.7	438.7	466.1	521.4	624.4	425.0	.047
CL-C	304.7	329.9	347.4	366.2	386.0	407.7	431.7	464.6	517.1	628.9	417.9	.026
CL-B	314.0	343.3	363.7	382.6	400.4	418.0	441.2	472.3	510.1	614.0	425.4	.038
CL-D	299.0	326.1	347.7	365.5	382.8	403.7	422.0	447.9	485.7	580.7	405.9	.026
Cr-A	280.7	311.1	327.8	342.2	358.1	371.2	386.5	402.6	432.3	506.8	371.3	.020
Cr-D	264.2	294.5	315.8	331.4	346.0	358.9	374.0	395.4	437.7	517.8	363.6	.012
Cr-B	281.0	307.4	324.9	338.5	353.2	369.0	388.1	409.3	443.8	529.8	374.6	.030
Cr-C	272.8	301.3	320.3	338.5	353.1	367.8	384.8	402.4	430.5	504.2	367.3	.015
FP-A	317.3	341.9	361.6	378.4	392.2	413.8	445.5	483.8	541.0	662.9	433.7	.026
FP-B	324.3	353.5	375.9	394.9	419.7	447.2	475.7	506.0	553.2	670.2	452.4	.077
FP-C	313.5	347.5	373.9	395.4	414.0	437.5	462.4	495.9	550.3	681.4	447.7	.057
FP-D	310.1	338.3	357.9	373.3	391.3	411.3	436.9	467.6	524.6	622.4	422.4	.011
FL-A	321.9	350.8	369.1	387.7	406.4	428.9	454.1	487.0	544.0	651.9	440.6	.039
FL-B	329.1	362.1	380.9	396.4	417.9	439.0	464.9	499.5	560.5	659.9	450.8	.053
FL-C	326.5	352.6	374.6	395.6	413.8	444.0	478.4	520.2	577.4	682.9	457.3	.072
FL-D	314.0	338.6	357.8	377.9	405.0	426.5	450.6	500.1	560.0	665.1	440.0	.022

Note—Letters to the left of each hyphen denote the type of task, and the letter to the right denotes the individual stimulus. CP = control–pitch; CL = control–loudness; Cr = correlated; FP = filter–pitch; FL = filter–loudness. MRT denotes mean response time, and P(E) denotes probability of error.

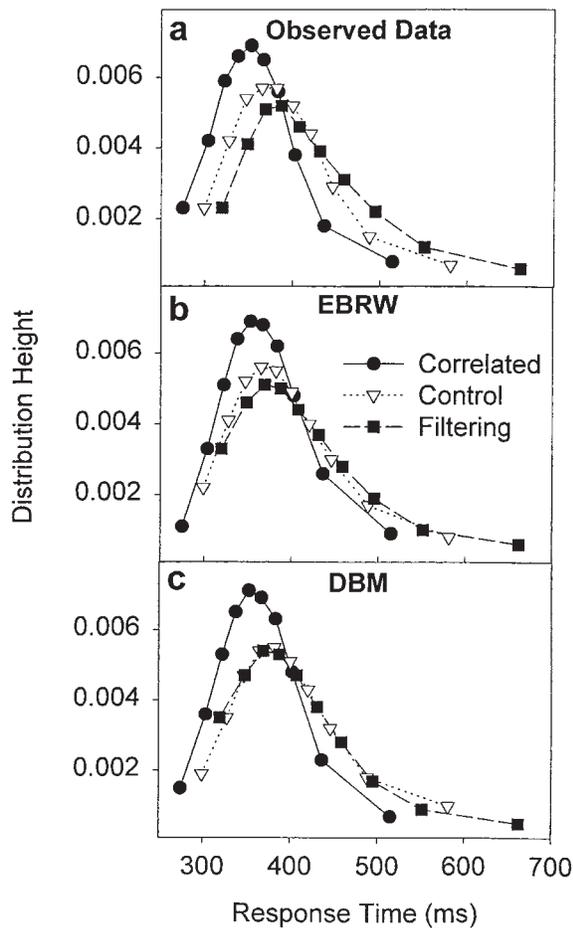


Figure 5. Panel a: Global RT distributions observed for the control, correlated, and filtering tasks. Panel b: Global distributions predicted by the EBRW model. Panel c: Global distributions predicted by the DBM.

ever. Rather, they have different shapes. Although all three distributions are positively skewed, the distribution for the correlated task rises rapidly to a high peak and is relatively narrow, whereas the filtering distribution is much flatter and has a long right tail. The control distribution is intermediate between these extremes. The figure also indicates that the fastest filtering RTs are slower than the fastest control RTs, suggesting that some of the strong predictions that we discussed previously for the DBM may not be confirmed. However, more rigorous examination of this issue requires modeling the individual RT distributions corresponding to each stimulus rather than the global distributions. We turn, therefore, to the modeling section of the article.

## THEORETICAL ANALYSIS

### EBRW Free Parameters

The version of the EBRW model fitted to the RT data had the following free parameters. First, to calculate dis-

tances and similarities, each stimulus was represented by two coordinate parameters—namely, a value along Dimension 1 and a value along Dimension 2. Because distances between stimuli are invariant under horizontal and vertical translations of the space, the coordinates for Stimulus A could be held fixed at (0,0) without loss of generality. Thus, there were six free coordinate parameters. Note also that because the stimulus coordinates were free parameters, the value of the scaling factor ( $c$ ) in Equation 2 could be held fixed at 1.0 without loss of generality. We also estimated a single attention-weight parameter ( $w_{att}$ ) used in the distance function (Equation 1). In tasks in which Dimension 1 was relevant, we set  $w_1 = w_{att}$  (with  $w_2 = 1 - w_{att}$ ), and vice versa for tasks in which Dimension 2 was relevant. In the correlated tasks, in which both dimensions were relevant, we assumed  $w_1 = w_2 = .50$ , although this latter assumption has little influence on the overall fit of the model. A final free parameter involved in the calculation of similarities was the value of  $\gamma$  in the distance function. Because the dimensions of pitch and loudness seem to lie intermediate along the integrality–separability continuum, we expected the best-fitting value of  $\gamma$  to lie intermediate between the canonical values of 1 and 2 (see Tversky & Gati, 1982, pp. 135–136, for previous evidence along these lines in experiments involving color stimuli).

For simplicity, the memory strengths for all stimuli were set at 1.0. More psychologically realistic versions of the EBRW model would be obtained by introducing assumptions about memory decay into the modeling; however, to reduce the number of free parameters, we decided to first pursue this simpler version.

The parameters involved in the random-walk process included the following. First, we defined an integer-valued criterion parameter ( $crit$ ) representing the amount of evidence needed for executing a response. Because of the symmetry of the category structures, we assumed unbiased responding, so that  $A = B = crit$  (see Figure 1). Another free parameter was the constant ( $\alpha$ ) in the step-time function (Equation 5). Also, because the RT predictions of the EBRW model are in arbitrary units, a scaling constant ( $k$ ) was needed for transforming these units into milliseconds. Finally, we needed to estimate the mean ( $\mu_R$ ) and standard deviation ( $\sigma_R$ ) of the duration of the residual processing stages.

In summary, the version of the EBRW tested here has 13 free parameters: 6 free coordinate parameters (the values of  $x_{im}$  in Equation 1), the value of  $\gamma$  in Equation 1, the attention weight given to relevant dimensions ( $w_{att}$ ) in Equation 1, the value of  $\alpha$  in the step-time function, the criterion parameter  $crit$ , the scaling constant  $k$ , and the residual parameters  $\mu_R$  and  $\sigma_R$ .

### DBM Free Parameters

Like the EBRW model, the DBM used 6 free coordinate parameters representing the means of each bivariate normal distribution [with the mean for Stimulus A held fixed at (0,0)]. We also defined free parameters repre-

senting the standard deviations along Dimensions 1 and 2 ( $s_1$  and  $s_2$ , respectively), and the correlation between Dimensions 1 and 2 ( $\rho$ ). All stimuli were assumed to have common values of the standard deviation and correlation parameters, however. To account for potential stimulus uncertainty effects in the filtering conditions, the standard deviations along both dimensions were multiplied by a free parameter,  $\kappa$ , in these conditions. (When  $\kappa = 1$ , there are no uncertainty effects, and the standard deviations are identical across conditions.) Other free parameters included the distance-from-bound rate parameter ( $\beta$ ) in Equation 6, the scaling constant ( $k$ ) for transforming the predictions of decision time into milliseconds (Equation 6), and the residual-stage parameters  $\mu_R$  and  $\sigma_R$ . This model makes use of 13 free parameters. (Because the stimulus means and standard deviations are free parameters, the  $\beta$  parameter can be held fixed at 1.0 without loss of generality.)

As explained previously, we assumed that the participants adopted orthogonal linear boundaries in the control and filtering conditions. Because of the symmetric category structures, we assumed an unbiased placement of the decision bound (analogous to our assumption of unbiased response criteria in the EBRW model). Thus, the bound was located midway between the means of the stimuli from contrasting categories in each condition. (Of course, in the conditions in which Dimension 1 was relevant, the bound was orthogonal to Dimension 1, so only the Dimension 1 means were relevant to the calculation of the bound; whereas, in the conditions in which Dimension 2 was relevant, only the Dimension 2 means were used.) In the correlated conditions, we assumed that the participants adopted a diagonal linear bound, as illustrated previously in Figure 3b. For simplicity, and to reduce the number of free parameters, we assumed specifically that the participants used a “minimum-distance bound” to classify the stimuli (Ashby & Gott, 1988). Such a boundary is a straight line located midway along, and perpendicular to, the chord connecting the means of the stimuli from the contrasting categories. In a subsequent analysis, we consider an extended version of the model that allows more flexible linear boundaries but find that it leads to minimal improvements in overall fit.

In sum, note that the baseline versions of the EBRW model and the DBM tested here used the same number of free parameters. We aimed for an equal number of parameters in the hope of achieving fair comparisons between the models.

### Model-Fitting Method

The models were fitted to the data by searching for the free parameters that minimized the sum of squared deviations (SSD) between predicted and observed data values. The total SSD was a weighted sum of two component SSDs, one corresponding to the RT data and the other corresponding to the accuracy data. The RT SSD was computed as follows. Given the parameters in the model, we calculated the proportion  $p_i$  of observations predicted

to lie in each RT bin for each stimulus in each task. Given the nature of our vincentizing procedure, the observed proportion of RTs in each bin is equal to .10. Thus, the overall RT SSD [SSD(RT)] is given by

$$\text{SSD(RT)} = \sum (p_i - .10)^2,$$

where the sum is over all 10 bins in each of the 20 individual stimulus distributions (a total of 200 data points, with 180 degrees of freedom). The accuracy SSD [SSD(A)] was defined as the sum of squared deviations between predicted and observed proportions of errors in each of the eight tasks, averaged over stimuli. The total SSD [SSD(T)] was then defined as  $\text{SSD(T)} = \text{SSD(RT)} + 25 \cdot \text{SSD(A)}$ . Any method for combining SSD(RT) and SSD(A) into an overall fit is arbitrary—we weighted the accuracy SSD by 25 because SSD(RT) was composed of 200 squared deviations, whereas SSD(A) was composed of only 8 squared deviations.<sup>5</sup>

The parameter-search program that was used incorporates a heuristic search procedure that alternates between making trial-and-error adjustments of individual parameters and pattern adjustments in the entire set of parameters. A variety of initial configurations were used in an effort to avoid local minima. Although minimizing SSD is not a maximum-likelihood procedure, note that because all observed data cells in the vincentized RT distributions are equal to .10, they have the same error variance for the SSD (RT). Thus, minimizing SSD seemed to be a reasonable criterion of fit.

Nosofsky and Palmeri (1997) derived analytic predictions from the EBRW model for the mean classification RTs and accuracies given any stimulus at a particular stage of learning. Because we have been unable to derive such analytic predictions for the RT distributions, however, we fit these data by means of computer simulation. The simulations were based on the construction of 10,000 random stimulus sequences satisfying the constraints of our experimental design. Each sequence consisted of 96 stimulus presentations, with all eligible stimuli occurring with equal frequency. For each stimulus, the random-walk process and residual times were simulated, and the resulting category choice and RT were recorded. We then computed the proportion of RTs occurring in each vincentized bin, using the cutoff procedure described previously (and where the lowest and highest bins were unbounded). As was the case in computing the observed data, only the final 72 trials of each sequence were used to generate the predictions, and only those RTs associated with correct categorization choices were included. Although analytic prediction methods are available for the DBM, for purposes of direct comparability with the EBRW model, as well as certain individual-participant analyses that we describe later, we used the same simulation methods for that model. Because the RT and accuracy predictions were virtually identical when alternative initial random seeds were used across runs of the models, we are confident that the number of simulations is sufficient for our model-fitting goals.

**Model-Fitting Results**

The predicted probabilities of RTs in each bin, as well as the predicted and observed overall accuracies in each task, are given in Tables 3 and 4 for the EBRW model and the DBM, respectively. (Recall that the predicted bin probabilities should be compared with observed values of .10 to assess the fits of each model.) The best-fitting parameters for each model are reported in Table 5. Overall, the EBRW model yields  $SSD(T) = .095$ , with component fits of  $SSD(RT) = .080$  and  $SSD(A) = .00059$ . The DBM yields  $SSD(T) = .109$ , with  $SSD(RT) = .099$  and  $SSD(A) = .00036$ . The overall quantitative fit of the EBRW model is slightly better with respect to predicting the RT distributions. Both models yield good fits to the error probabilities, the fits of the DBM being somewhat better than those of the EBRW model.

Figure 6 plots, for each model separately, the predicted RT distributions against the observed RT distributions for each individual stimulus in each task. Considering that each model uses only 13 free parameters to predict the 200 data points (plus the accuracy data), the overall fits to these detailed RT-distribution data strike us as fairly impressive. Nevertheless, each model also suffers important and systematic limitations, as we soon demonstrate.

To gain insight into the overall strengths and weaknesses of each model, we computed the global control, filtering, and correlated distributions that each model predicted, and we compared these with the observed global distributions (Figure 5a). Figures 5b and 5c plot the predicted distributions in relation to one another, as was done for the observed global distributions in Figure 5a. Both models appear to capture the basic shape of the RT distributions quite well. However, Figure 5c illustrates that the DBM fails to separate the main body of the global RT distributions corresponding to the control and filtering

tasks. Thus, when the DBM is required to fit the detailed structure of the RT distributions, it apparently fails to capture the fundamental interference effect observed for these integral-dimension stimuli. The EBRW model fares much better in accounting for this key qualitative result, although it somewhat underestimates the magnitude of the interference (see Figure 5b). Both the EBRW model and the DBM make the correct qualitative prediction of facilitation in the correlated task and also predict nicely the peaked shape associated with the correlated distribution. However, a weakness for the EBRW model is that it underestimates the degree of facilitation in the correlated task.

To gain another view of the overall strengths and weaknesses of the models, Figures 7 and 8 plot together on the same graphs the predicted and observed global RT distributions. The EBRW model (Figure 7) predicts with extremely good accuracy the control distribution. It also yields fairly good fits to the filtering distribution, although it overpredicts the proportion of very fast RTs (i.e., the probability of RTs in the first vincentized bin). The EBRW model's clearest shortcoming is with respect to the correlated task, however, where the predicted distribution is shifted systematically to the right of the observed distribution. It is interesting to note that the EBRW model also underpredicts the error rate in the correlated tasks (see Table 3a). This pattern suggests the possibility of a speed-accuracy tradeoff at work, a point to which we return later in this article.

As can be seen in Figure 8, the DBM predicts RTs in the control task that are too slow, and predicts RTs in the filtering task that are too fast. This combination of mispredictions probably represents a compromise of the best-fitting parameters, because the model predicts virtually identical global distributions across the control and fil-

**Table 3**  
**EBRW Model Predictions of Vincentized Bin Probabilities and Overall Accuracies in Each Task**

Task	Decile										$P(E)$	
	1	2	3	4	5	6	7	8	9	10	Pre.	Obs.
CP-A	0.080	0.094	0.097	0.100	0.094	0.092	0.094	0.101	0.120	0.128	.030	.037
CP-B	0.075	0.092	0.091	0.092	0.092	0.094	0.090	0.103	0.126	0.143		
CP-C	0.065	0.085	0.094	0.096	0.097	0.092	0.089	0.116	0.134	0.130	.020	.027
CP-D	0.069	0.084	0.090	0.095	0.102	0.100	0.093	0.109	0.124	0.133		
CL-A	0.129	0.122	0.104	0.097	0.085	0.089	0.089	0.096	0.097	0.093	.037	.036
CL-C	0.100	0.087	0.093	0.106	0.111	0.104	0.101	0.103	0.102	0.093		
CL-B	0.144	0.122	0.109	0.100	0.086	0.084	0.087	0.079	0.090	0.098	.033	.032
CL-D	0.086	0.096	0.104	0.098	0.104	0.094	0.087	0.093	0.113	0.125		
Cr-A	0.076	0.104	0.098	0.110	0.109	0.100	0.097	0.107	0.125	0.074	.003	.016
Cr-D	0.034	0.074	0.093	0.100	0.101	0.105	0.126	0.162	0.141	0.066		
Cr-B	0.056	0.075	0.079	0.086	0.100	0.114	0.118	0.123	0.143	0.105	.010	.023
Cr-C	0.040	0.068	0.088	0.099	0.097	0.104	0.105	0.113	0.150	0.137		
FP-A	0.137	0.102	0.097	0.083	0.089	0.114	0.110	0.098	0.090	0.079	.045	.043
FP-B	0.152	0.117	0.103	0.103	0.102	0.089	0.073	0.069	0.087	0.104		
FP-C	0.130	0.137	0.124	0.098	0.090	0.084	0.078	0.082	0.091	0.086		
FP-D	0.121	0.110	0.096	0.093	0.099	0.102	0.096	0.100	0.092	0.091		
FL-A	0.155	0.110	0.095	0.093	0.094	0.088	0.085	0.090	0.092	0.098	.059	.047
FL-B	0.180	0.123	0.083	0.086	0.085	0.078	0.080	0.088	0.087	0.109		
FL-C	0.148	0.105	0.104	0.090	0.098	0.104	0.091	0.080	0.078	0.103		
FL-D	0.114	0.092	0.098	0.120	0.110	0.087	0.106	0.102	0.080	0.090		

Note—CP = control-pitch; CL = control-loudness; Cr = correlated; FP = filter-pitch; FL = filter-loudness;  $P(E)$  = probability of error; Pre. = predicted; Obs. = observed.

**Table 4**  
**DBM Predictions of Vincentized Bin Probabilities and Overall Accuracies in Each Task**

Task	Decile										P(E)	
	1	2	3	4	5	6	7	8	9	10	Pre.	Obs.
CP-A	0.084	0.090	0.094	0.098	0.095	0.094	0.094	0.100	0.113	0.139	.040	.037
CP-B	0.079	0.089	0.088	0.090	0.092	0.096	0.092	0.103	0.120	0.152		
CP-C	0.071	0.082	0.090	0.093	0.095	0.093	0.090	0.116	0.127	0.143	.031	.027
CP-D	0.074	0.081	0.086	0.092	0.100	0.101	0.093	0.107	0.117	0.148		
CL-A	0.089	0.095	0.089	0.092	0.086	0.098	0.101	0.114	0.118	0.119	.039	.036
CL-C	0.068	0.063	0.075	0.094	0.107	0.110	0.115	0.123	0.127	0.118		
CL-B	0.107	0.099	0.098	0.099	0.091	0.093	0.100	0.090	0.106	0.117	.032	.032
CL-D	0.063	0.074	0.085	0.089	0.102	0.099	0.097	0.109	0.130	0.151		
Cr-A	0.092	0.108	0.098	0.109	0.108	0.100	0.096	0.106	0.115	0.068	.003	.016
Cr-D	0.047	0.080	0.096	0.099	0.102	0.104	0.126	0.161	0.126	0.060		
Cr-B	0.068	0.079	0.080	0.087	0.102	0.117	0.119	0.123	0.127	0.098	.011	.023
Cr-C	0.051	0.073	0.090	0.099	0.099	0.105	0.106	0.111	0.140	0.126		
FP-A	0.174	0.116	0.109	0.091	0.093	0.113	0.097	0.076	0.069	0.061	.043	.043
FP-B	0.175	0.128	0.110	0.108	0.106	0.085	0.064	0.060	0.076	0.087		
FP-C	0.157	0.149	0.135	0.101	0.090	0.081	0.070	0.068	0.077	0.072		
FP-D	0.163	0.129	0.108	0.103	0.106	0.099	0.085	0.078	0.065	0.064		
FL-A	0.134	0.104	0.096	0.097	0.102	0.099	0.091	0.094	0.094	0.091	.042	.047
FL-B	0.162	0.121	0.087	0.093	0.097	0.088	0.088	0.092	0.085	0.089		
FL-C	0.128	0.098	0.104	0.097	0.109	0.117	0.099	0.080	0.083	0.085		
FL-D	0.111	0.088	0.099	0.127	0.123	0.096	0.112	0.098	0.077	0.070		

Note—CP = control–pitch; CL = control–loudness; Cr = correlated; FP = filter–pitch; FL = filter–loudness; P(E) = probability of error; Pre. = predicted; Obs. = observed.

tering tasks (Figure 5c). Finally, as was the case for the EBRW model, the DBM predicts RTs in the correlated task that are too slow.

Although we made no attempt to predict the mean RTs in each task when fitting the models, some insights into the models’ performance can be gained by considering the means. Figure 9 plots, for each model separately, the observed mean RTs in each of the eight tasks against the predicted mean RTs. To achieve clearer comparisons, the results for each filtering task are broken down into two means. The first mean corresponds to Stimulus Pair A–D, whereas the second mean corresponds to Stimulus Pair B–C. On this criterion of predicting the mean RTs, the EBRW model performs far better than the DBM. The correlation between predicted and observed mean RTs is  $r = .968$  for the EBRW model, but only  $r = .809$  for the DBM. Whereas the EBRW model makes only small errors in predicting the ordering of mean RTs across tasks, the DBM suffers the critical qualitative shortcoming of predicting virtually identical overall mean RTs in the control–pitch and filter–pitch tasks and in the control–loudness and filter–loudness tasks. Indeed, it predicts a slightly faster mean RT for Stimulus Pair A–D in the filter–pitch task than for either Pair A–B or Pair C–D in the control–pitch tasks. Likewise, it predicts a slightly faster mean RT for Pair A–D in the filter–loudness task than for either Pair A–C or Pair B–D in the control–loudness tasks. The observed data, however, are decidedly in the opposite direction.

Beyond the main interference and facilitation effects that we have already discussed in this article, Figure 9 reveals the following more fine-grained qualitative effects: (1) mean RTs in the control–pitch tasks were faster than in the control–loudness tasks; (2) mean RTs for corre-

sponding pairs of stimuli were faster in the filter–pitch tasks than in the filter–loudness tasks; (3) within the control–pitch tasks, Pair C–D was discriminated more rapidly than Pair A–B; (4) within the control–loudness tasks, Pair B–D was discriminated more rapidly than Pair A–C; (5) in the correlated tasks, Pair A–D was dis-

**Table 5**  
**Best-Fitting Parameters for the Baseline and Extended Versions of the Exemplar-Based Retrieval (EBRW) Model and the Decision-Bound Model (DBM)**

	EBRW		DBM	
	Baseline	Extended	Baseline	Extended
$x_{B1}$	1.123	1.097	1.173	1.147
$x_{B2}$	0.021	–0.018	0.018	0.041
$x_{C1}$	0.134	0.097	0.059	0.061
$x_{C2}$	1.031	1.016	0.981	1.006
$x_{D1}$	1.378	1.290	1.309	1.284
$x_{D2}$	1.039	1.036	1.045	1.048
$w_{att}$	0.595	0.608	—	—
$\gamma$	1.384	1.634	—	—
$\alpha$	0.199	0.199	—	—
$\beta$	—	—	3.824	4.348
$crit$	6	6	—	—
$s_1$	—	—	0.336	0.324
$s_2$	—	—	0.278	0.278
$\rho$	—	—	–0.049	0.006
$\kappa$	—	—	1.039	1.048
$\mu_R$	277.193	281.261	332.184	332.161
$\sigma_R$	37.365	37.375	43.742	44.377
$k$	36.785	37.142	512.065	512.573
$p$	—	0.325	—	—
$l_{crit}$	—	3	—	—
$\beta^*$	—	—	—	3.333

Note—The value of  $\beta$  for the DBM can be set at 1.0 without loss of generality by rescaling the  $x_{im}$ ,  $s_1$ ,  $s_2$ , and  $\kappa$  parameters, so the baseline DBM uses 13 free parameters. The values of  $x_{A1}$  and  $x_{A2}$  are held fixed at zero because the origin of the space is arbitrary. \*A separate value of  $\beta$  was estimated for the filtering conditions.

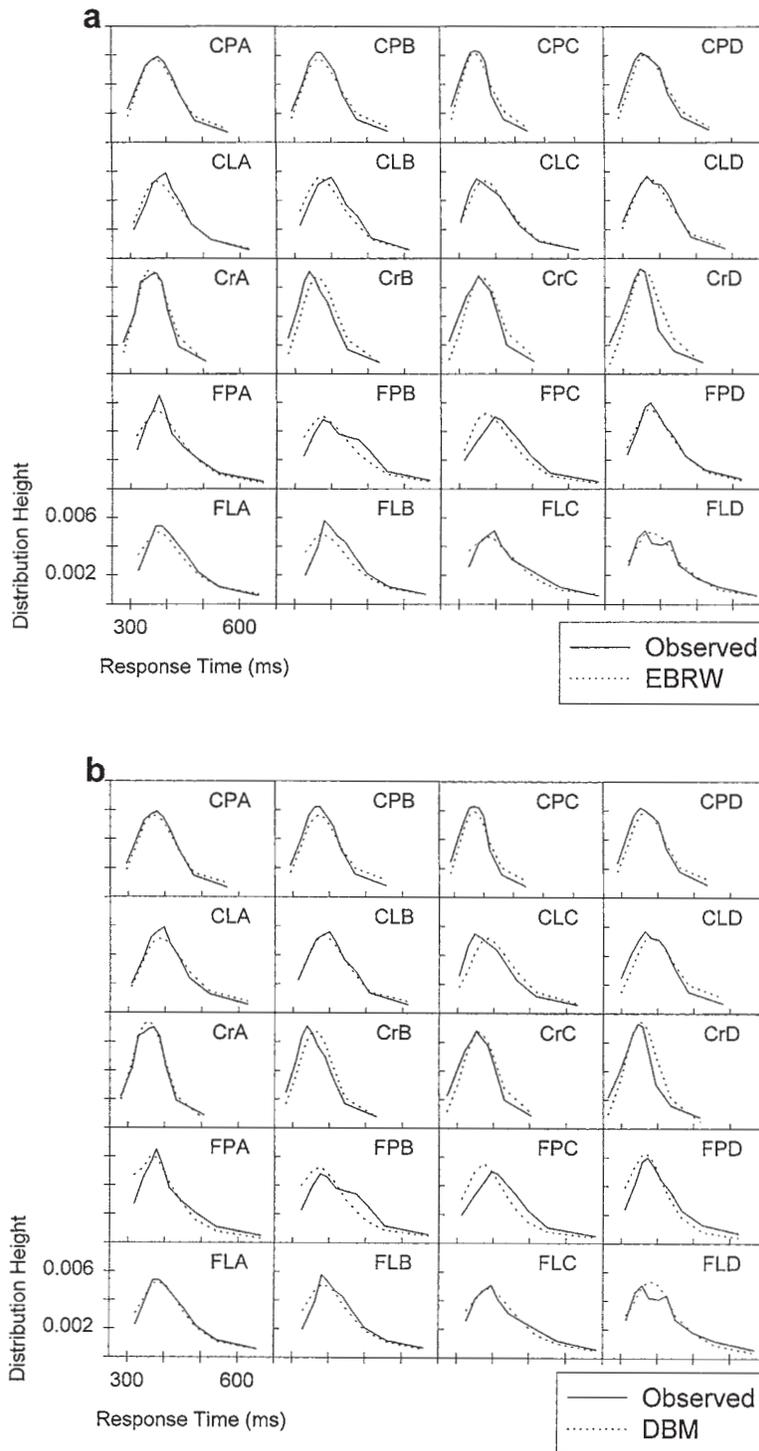


Figure 6. Individual RT distributions for each individual stimulus in each task. Panel a: Predictions for the EBRW model. Panel b: Predictions for the DBM. The first two letters in each plot denote the type of task, and the third letter denotes the individual stimulus. CP = control-pitch; CL = control-loudness; Cr = correlated; FP = filter-pitch; FL = filter-loudness.

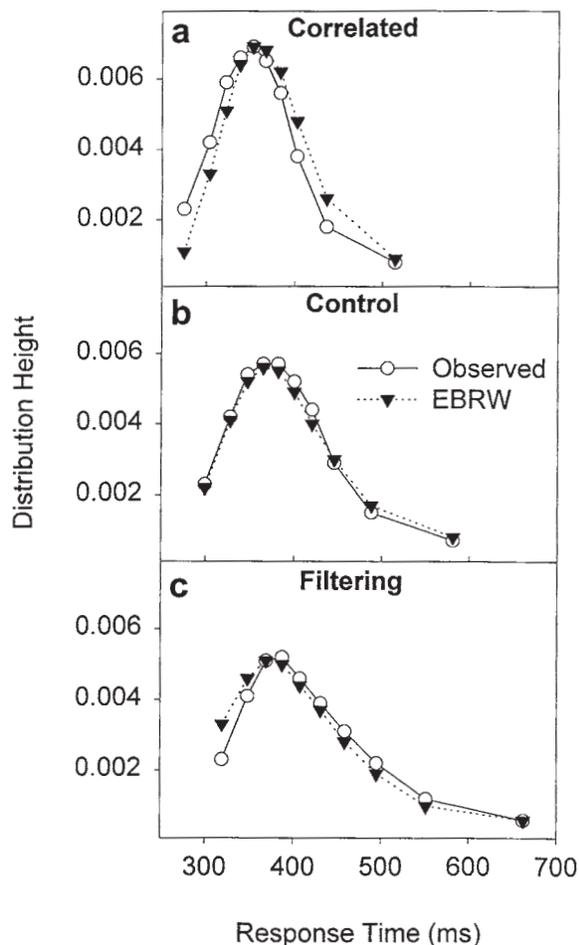


Figure 7. Observed global distributions for each type of task plotted with predicted global distributions from the EBRW model.

criminated more rapidly than Pair B–C; (6) in both the filter–pitch task and the filter–loudness task, Pair A–D was discriminated more rapidly than Pair B–C. As can be seen in Figure 9a, the EBRW model accounts correctly for all of these qualitative effects. It does so by positioning the four stimuli in a roughly trapezoidal configuration that is slightly stretched along the pitch dimension, as illustrated in Figure 10a (see stimulus-coordinate parameter estimates in Table 5). The DBM produces a similar configuration, as illustrated in Figure 10b. Thus, the hypothesis of a mean-shift integrality (Ashby & Maddox, 1994; Kingston & Macmillan, 1995), which we discussed earlier in this article, is strongly supported by these modeling analyses. (It is important to note that the mean-shift integrality configuration also allowed both models to make the correct qualitative prediction that the error rate on the A–D pair was lower than on the B–C pair in both filtering tasks.) However, despite the existence of the mean-shift integrality, the DBM fails to account for the

key qualitative finding of a uniform interference effect in the filtering conditions relative to the control conditions.

Another qualitative distinction between the models is revealed by examining the very fastest RTs in each type of task. Note that each simulation of a model across the eight tasks can be viewed as a simulation for a single observer. For the 10,000 individual-observer simulations of each model, we conducted the following analysis. On each simulation, we computed the median of the 10 fastest RTs in each control–pitch task and then averaged these medians. We also computed, on each simulation, the median of the 10 fastest RTs in the filter–pitch task. An analogous procedure was used for the control–loudness and filter–loudness tasks. For ease of description, let us refer to the median of these fast RTs as the MFRT. For the EBRW model, the control–pitch MFRT was faster than the filter–pitch MFRT on 65% of the simulations, and the control–loudness MFRT was faster than the filter–loudness MFRT on 65% of the simulations. By contrast, for the DBM, the control–pitch MFRT was faster than the

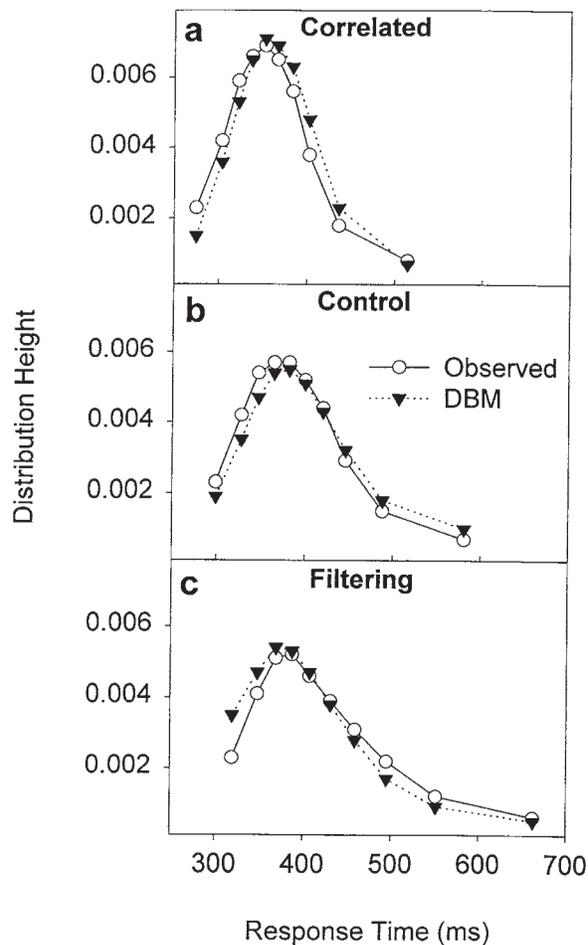


Figure 8. Observed global distributions for each type of task plotted with predicted global distributions from the DBM.

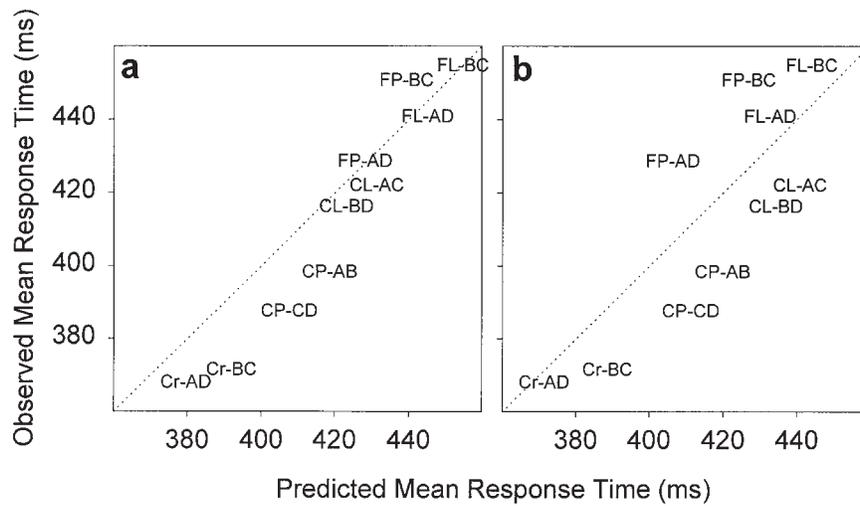


Figure 9. Observed mean RTs in each individual task plotted against predicted mean RTs. Letters to the left of each hyphen denote the type of task, and letters to the right denote the specific stimulus pair involved. CP = control–pitch; CL = control–loudness; Cr = correlated; FP = filter–pitch; FL = filter–loudness. Panel a: Predictions for the EBRW model. Panel b: Predictions for the DBM.

filter–pitch MFRT on only 48% of the simulations, and the control–loudness MFRT was faster than the filter–loudness MFRT on only 48% of the simulations. Thus, the EBRW model predicts an advantage for the control MFRTs, whereas the DBM predicts either no difference or a slight advantage for the filtering MFRTs. In our analysis of the actual observed data, we found that 65% of the participants had faster control–pitch MFRTs than filter–pitch MFRTs, and that 73% of the participants had faster control–loudness MFRTs than filter–loudness MFRTs. Overall, the observed MFRTs in the control tasks ( $M = 308.4$  msec) were significantly faster than the MFRTs in the filtering tasks ( $M = 326.3$  msec) [ $t(25) = 6.46, p < .001$ ]. Thus, with regard to the fastest RTs in the

control and filtering tasks, the qualitative prediction of the EBRW model is clearly supported over that of the DBM.

The shortcomings of the DBM can be conceptualized as follows. To account for filtering interference in the mean RTs, this model posits the existence of increased variance due to uncertainty effects (Figure 3c), mean-shift integrality (Figure 3d), or some combination of these two factors. The best-fitting parameters for the DBM seem to be trying to make use of both sources of interference. There is a mean-shift integrality (Figure 10b), and the standard-deviation multiplier in the filtering tasks is greater than one ( $\kappa = 1.04$ ; see Table 5). If the magnitude of these parameters were increased, the DBM would be more successful at predicting the overall interference effects in the

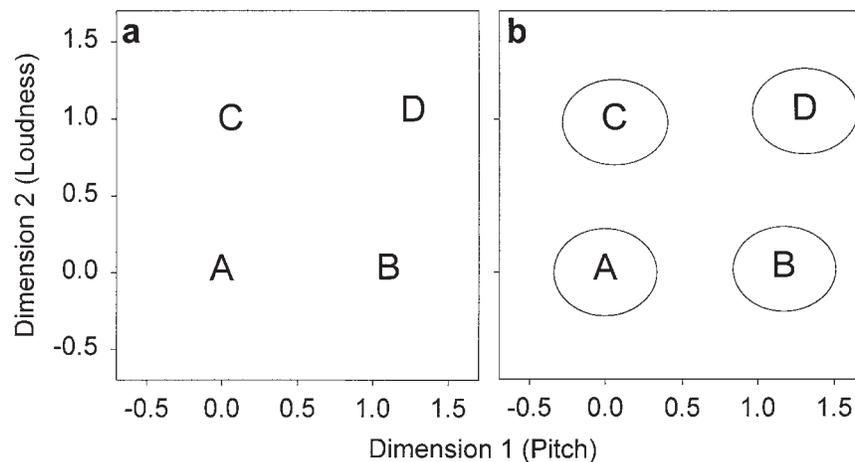


Figure 10. Panel a: Stimulus coordinates derived by fitting the EBRW model to the RT distribution data. Panel b: GRT representation derived by fitting the DBM to the RT distribution data.

mean RTs. Unfortunately, however, the DBM is caught in a bind. Increasing the magnitude of these parameters forces the DBM to simultaneously predict both a far greater proportion of very slow RTs and a far greater proportion of very fast RTs. Indeed, with the current parameter estimates, the DBM already makes the incorrect qualitative prediction of slightly faster MFRTs in the filtering tasks than in the control tasks. In sum, the constraints yielded by requiring the DBM to fit the actual RT distribution data for the individual stimuli pose an interesting challenge to the model.

### Examination of Best-Fitting Parameters

In this section, we briefly consider the best-fitting parameters from the EBRW model and the DBM (see Table 5). Interestingly, similar estimates were obtained of those parameters that play the same conceptual role across the two models, suggesting that the models may have much in common. For example, as discussed previously, in both models, the estimated stimulus coordinates gave rise to the roughly trapezoidal multidimensional scaling (MDS) configuration associated with mean-shift integrality (Figures 10a and 10b). Also, the estimated mean and standard deviation of residual times ( $\mu_R$  and  $\sigma_R$ ) were similar across the models.

In the distance function used by the EBRW model (Equation 1), the best-fitting value of the power exponent was  $\gamma = 1.38$ , which is consistent with our intuition that the dimensions of pitch and loudness lie intermediate along the integrality–separability continuum. Also, the estimated attention weight,  $w_{att} = .595$ , is consistent with our expectation that observers attempted to attend selectively to the relevant dimension in each task, but that the integrality of the dimensions prevented perfect selective attention. Finally, as noted previously for the DBM, the value of  $\kappa > 1$  is consistent with expectations emerging from that framework that the increased number of stimuli in the filtering task gives rise to uncertainty effects.

### Extended Versions of the Models

In this section, we briefly consider extended versions of the EBRW model and the DBM. Although we have not stated so explicitly, note that, in our tests of the baseline models in the previous section, we strove for parameter invariance across conditions.<sup>6</sup> The key issue was the extent to which the EBRW model and the DBM could predict *a priori* the fundamental interference and facilitation effects observed for these integral-dimension stimuli. A natural question, however, is whether improved fits can be obtained by allowing psychologically plausible variations in parameter settings across conditions.

The main shortcoming of the EBRW model is that it underestimated the degree of facilitation in RTs observed in the correlated tasks. At the same time, however, it underestimated the observed error rate in these tasks. A salient possibility, therefore, is that a speed–accuracy

tradeoff may have been operating. One way that the EBRW model can model speed–accuracy tradeoffs is in terms of shifts of the response–criterion parameter, *crit*. Setting *crit* at lower values, for example, leads the system to respond after retrieving fewer exemplars, which tends to speed RTs but decrease accuracy.

There are numerous ways to formalize this idea of varying criterion settings. As an illustrative example, we fitted a mixture model that assumed that, with probability  $p$ , an observer used a criterion setting in the correlated task that was identical to the one used in the control and filtering tasks and, with probability  $1-p$ , used a lower setting of the criterion parameter (*lcrit*). This model extends the baseline version by adding the free parameters  $p$  and *lcrit*. The fits of the extended model to the RT distribution and accuracy data were excellent. Although the baseline EBRW model was already doing fairly well at fitting the data, the extended model reduced SSD(RT) from .080 to .050 and reduced SSD(A) from .00059 to .00036 (the best-fitting parameters are reported in Table 5). The impressive performance of the model is summarized in Figure 11, where we plot the predicted global distributions against the observed global distributions in the control, correlated, and filtering tasks. The model still slightly overpredicts the percentage of filtering RTs in the first bin, but, otherwise, there is little room for improvement.

Recall that, in our tests of the baseline DBM, we assumed that the participants adopted orthogonal linear boundaries in the control and filtering conditions and adopted diagonal minimum distance boundaries in the correlated conditions. Also, we assumed an unbiased placement of these bounds. There is precedence for the assumption of orthogonal linear boundaries in the control and filtering tasks, particularly in situations in which inexperienced observers are instructed to classify on the basis of values on a single dimension (Ashby & Maddox, 1994). Furthermore, our assumption of unbiased placement of the bounds was based on the symmetric structure of all the categorization tasks. Nevertheless, to further investigate the issue, we assumed in an extended version of the model that the participants adopted linear decision bounds in all conditions, but we allowed the parameters of these linear boundaries to be completely free. Because there were eight conditions, and each linear boundary is described by two free parameters (a slope and intercept), this extended model used 16 additional free parameters, relative to the baseline version. Even with this large increase in the number of free parameters, however, the improvement in fit yielded by this extended model was minimal: SSD(RT) was reduced from .099 to .095, and SSD(A) from .00036 to .00032. Furthermore, the model still failed to predict the interference in RTs observed in the filtering tasks. Analyses were also conducted that allowed for variability in the criterion setting of the linear boundaries across trials of the experiment, but allowing for this criterial variability provided no improvement in

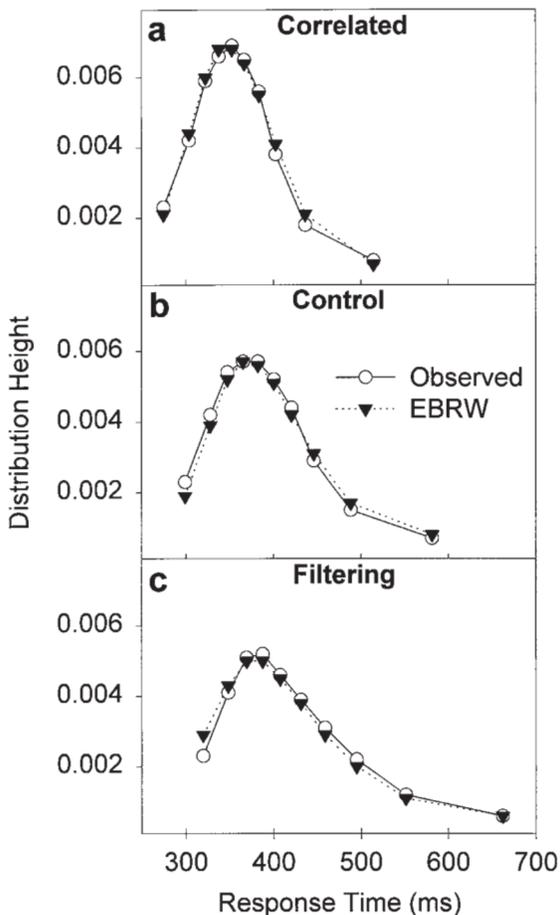


Figure 11. Observed global distributions for the control, correlated, and filtering tasks plotted with the global distributions predicted by the EBRW model with mixed criterion settings.

fit. These analyses do not rule out, of course, the possibility that the participants may have used nonlinear decision boundaries in the different speeded classification tasks.

A straightforward approach to enabling the DBM to fit the interference effects in the filtering data is simply to allow differences in the  $\beta$  parameter across conditions. Recall that  $\beta$  determines the rate at which decision time decreases with distance-from-bound (Equation 6). If  $\beta$  is set at a lower value in the filtering conditions than in the control and correlated conditions, then slower filtering RTs can be well fitted. Indeed, an extended version of the DBM that allowed separate  $\beta$  parameters yielded an outstanding fit to the data [SSD(T) = .048, SSD(RT) = .038, and SSD(A) = .00040]. The best-fitting parameters are reported in Table 5.

Unfortunately, we consider this extended version of the DBM to be unprincipled and completely post hoc. It provides no explanation of the interference observed in the filtering condition but simply redescribes it. (A similarly good fit can be achieved by the EBRW model by allowing changes in the scaling factor  $c$  across conditions.) To achieve a principled explanation, the decision-bound theorist would need to develop some account of how and

why  $\beta$  might vary across conditions (and then this account would need to be subjected to experimental test). Nevertheless, we felt it was important to acknowledge the excellent fits that can be achieved by this extended descriptive model, because it highlights a potential direction for future model development within the decision-bound framework.

A similar criticism regarding post hoc data fitting can be raised about our assumption of varying criterion settings in the extended EBRW model. An important difference, however, is that, in the EBRW model, the criterion setting affects both the RT and the accuracy predictions of the model, so there are independent bases for positing a change in the parameter across conditions. By contrast, in current versions of decision-bound theory, accuracy predictions are completely unaffected by the value of  $\beta$ . Nevertheless, an important direction for future model development within the exemplar-retrieval framework is to develop a richer theory of how the criterion setting is expected to vary across conditions.

## GENERAL DISCUSSION

### Summary

The central goal of this research was to evaluate two current models, the EBRW model and the DBM, on their ability to account quantitatively for detailed RT distribution and accuracy data observed for individual stimuli in the Garner speeded classification tasks. Because of the fundamental nature of the Garner tasks, we considered them an important testing ground for the development of multidimensional models of speeded perceptual classification. Furthermore, in this investigation, we focused on the performance of relatively inexperienced participants making speeded classifications of integral-dimension stimuli, because it was in this domain that Garner (1974) brought out the systematic and robust facilitation and interference effects that are of greatest challenge to the models.

The baseline version of the EBRW model yielded a reasonably good quantitative account of the RT distribution and accuracy data. In addition to correctly predicting the fundamental qualitative effects of facilitation in the correlated tasks and interference in the filtering tasks, it provided good quantitative fits to the detailed shapes of the RT distributions. A limitation of the model is that it underestimated the magnitude of the facilitation effect in the correlated task. However, an extended version of the model, which made allowance for speed-accuracy trade-offs due to variations in the criterion setting, pinpointed the observed RTs and accuracies across all tasks.

The DBM also yielded a reasonably good global quantitative fit to the data. However, it failed to predict the interference effect observed in the filtering tasks. Past theorizing involving the DBM suggested that the main reasons for filtering interference are a mean-shift integrality in the perceived stimulus locations and an increase in perceptual variance due to uncertainty effects (Ashby & Maddox, 1994; Kingston & Macmillan, 1995).

Although these factors can produce an interference effect in the mean RTs, our present investigation suggests that they are not the only factors operating. In particular, in combination with the other assumptions embodied in current versions of decision-bound theory, these factors produced poor predictions of the detailed structure of the RT distributions in the filtering task. The model also failed to predict the ordering of mean RTs associated with individual stimuli across the tasks. It seems to us, therefore, that the basis for RT interference in the filtering task remains something of a mystery from the perspective of the DBM.

Furthermore, it may not be easy to remedy this shortcoming with only minor modifications of the DBM. Maddox and Ashby (1996) have acknowledged that a limitation of current versions of the DBM is that the exponential function relating decision time to distance-from-bound (Equation 6) places an upper limit on the maximum decision time. Thus, this version of the model may predict too few long RTs. A straightforward idea, therefore, is to use some alternative quantitative function for relating decision time to distance. We believe, however, that for any monotonic decreasing function, the DBM will suffer a fundamental qualitative shortcoming. As illustrated in Figures 3c and 3d, as long as the function relating decision time to distance is monotonic, the DBM predicts that the very fastest RTs in the filtering task will be as fast as or faster than the very fastest RTs in the control tasks. The opposite pattern of results was observed in our data, however. Modifying the exponential RT-distance function might allow the DBM to better fit the skewed right tail of the RT distributions (i.e., the very slowest RTs), but it cannot save the DBM from its incorrect prediction that the very fastest RTs should also occur in the filtering task.

### Future Research Directions

There are numerous avenues that need to be pursued in continued tests of the present models. First, more microlevel analyses are needed that investigate the role of sequential effects in speeded perceptual classification. As currently formalized, neither the EBRW model nor the DBM accounts for the sequential effects known to exist in the Garner tasks. For example, Melara, Mounts, and Yamagishi (1993) and Nosofsky and Palmeri (1997) observed that, in the filtering task, the most rapid responses occurred on trials in which a stimulus was repeated, and the slowest responses occurred on trials in which a new stimulus from the same category was presented. RTs were intermediate on those trials in which a new stimulus was presented from a new category. A variety of mechanisms may be involved in producing such sequential effects. For example, in the EBRW model, decaying memory strength for previously presented exemplars will influence the speed of the random-walk process (see Equations 3–5). Also, there are likely to be systematic criterion shifts from trial to trial that need to be modeled (cf. Purks, Callahan, Braida, & Durlach, 1980). Researchers have also suggested that observers may make

partial use of a “bypass rule” (Fletcher & Rabbitt, 1978; Krueger & Shapiro, 1981) or trial-by-trial “change strategy” (Melara et al., 1993) in their speeded classifications. According to such a rule, if one detects any change in the stimulus from the previous trial, then one responds with the alternative category label. Such strategies could perhaps be modeled in terms of systematic shifts in the starting point of the random walk from trial to trial.

In addition to conducting more fine-grained analyses within the control, filtering, and correlated tasks, we need to move in the direction of testing the models on their ability to account for performance in more complex speeded classification tasks. For example, in the condensation task (also known as the divided attention task), participants are required to classify Stimuli A and D into one category and Stimuli B and C into a second category (see Figure 2). It is well known that performance is extremely poor in the condensation task. Nosofsky and Palmeri (1997) found that the EBRW model provided a reasonably good quantitative account of the mean RT and accuracy in the condensation task, as long as assumptions were introduced about criterion shifts involved in a speed-accuracy trade-off across conditions. We need to test the model, however, on its ability to predict the structure of the individual stimulus RT distributions as well.

When “configural” dimensions are used, results of the Garner speeded classification tasks are often quite complicated. For example, if the dimensions are left and right parentheses in either the left or right location, then the four stimuli are  $A = ((, B = (, C = ),$  and  $D = ))$ . Under such conditions, Pomerantz and Garner (1973) found that performance in the condensation task  $[((, ))$  vs.  $(, )]$  is better than in the filtering tasks [e.g.,  $((, ($  vs.  $), ))]$ . We follow Lockhead and King (1977) in interpreting such results. MDS studies conducted by Lockhead and King indicate that similarity relations among the parentheses stimuli predict the speeded classification times quite well. It is not simply the physically manipulated features that determine performance but the emergent or configural psychological dimensions along which the objects are coded. Applications of the EBRW model to speeded classification RTs require either derived MDS solutions for the objects or a close correspondence between the physical space used for constructing the stimuli and the underlying psychological space in which the objects are coded.

One point we wish to emphasize is that, although the present applications of the EBRW model involved single-point representations for each of the stimuli, we acknowledge the importance of the probabilistic representations assumed in decision-bound theory. Because of noise in the sensory/memorial processing systems, the precise internal representation of a stimulus undoubtedly varies from trial to trial. We view the single-point representation as a reasonable approximation in situations involving moderately discriminable stimuli, such as was the case in the present experiment. To apply the EBRW model to model speeded classifications of highly confusable stimuli, however, in which the perceptual/memory

variance is presumably large relative to the distance between the means of the stimulus representations, it will be critical to include parameters to model this variability. Another avenue to improving the predictions of the EBRW model is to incorporate assumptions about how similarities among exemplars may change dynamically during the course of psychological processing. Lamberts (1995, 1996), for example, provided evidence for changes in overall sensitivity (the value of  $c$  in Equation 2) and in the distribution of the attention weights (the values of  $w_m$  in Equation 1) depending on the speed with which participants were required to make their classification judgments.

Finally, in the present research, the fundamental RT distribution data were obtained by pooling over participants. Our reason for initiating the investigation with such a design is that we were primarily interested in modeling the performance of relatively inexperienced participants, who presumably had not yet learned complex decision boundaries. Nevertheless, as we acknowledged at the outset of this article, averaged data can obscure patterns observed at the individual subject level, so it is critical to pursue these investigations by collecting extensive individual-subject data as well. We have already tested 2 individual observers over a period of 16 sessions using the same stimulus set as in the present study. The data of both experienced observers showed the same fundamental qualitative patterns as for the averaged data of our novice observers. Even after 16 sessions of testing, there was marked facilitation in the correlated tasks and marked interference in the filtering tasks. Furthermore, the fastest RTs in the control tasks were faster than the fastest RTs in the filtering tasks, a pattern of results that we believe places important constraints on the DBM. We plan to present a full report of these individual-subject data together with formal modeling analyses in a subsequent article.

#### REFERENCES

- ASHBY, F. G., BOYNTON, G., & LEE, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics*, **55**, 11-27.
- ASHBY, F. G., & GOTT, R. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.
- 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. 389-414). New York: Springer-Verlag.
- 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., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.
- BUNDESEN, C. (1990). A theory of visual attention. *Psychological Review*, **97**, 523-547.
- BUSEMEYER, J. R. (1982). Choice behavior in a sequential decision-making task. *Organizational Behavior & Human Performance*, **29**, 175-207.
- DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.
- FLETCHER, B., & RABBITT, P. M. A. (1978). The changing pattern of perceptual analytic strategies and response selection with practice in a two-choice reaction time task. *Quarterly Journal of Experimental Psychology*, **30**, 417-427.
- GARNER, W. R. (1974). *The processing of information and structure*. New York: Wiley.
- GARNER, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, **8**, 98-123.
- GARNER, W. R., & FELFOLDY, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, **1**, 225-241.
- GRAU, J. W., & KEMLER NELSON, D. G. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, **117**, 347-370.
- KINGSTON, J., & MACMILLAN, N. A. (1995). Integrality of nasalization and F1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America*, **97**, 1261-1285.
- KRUEGER, L. E., & SHAPIRO, R. G. (1981). Intertrial effects of same-different judgements. *Quarterly Journal of Experimental Psychology*, **33A**, 241-265.
- LAMBERTS, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, **124**, 161-180.
- LAMBERTS, K. (1996). *The time course of categorization*. Manuscript submitted for publication.
- LINK, S. W., & HEATH, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77-105.
- LOCKHEAD, G. R., & KING, M. C. (1977). Classifying integral stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 436-443.
- LOGAN, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.
- LOGAN, G. D. (1992). Shapes of reaction time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 883-914.
- LOGAN, G. D. (1997). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, **103**, 603-649.
- LUCE, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- LUCE, R. D., GREEN, D. M., & WEBER, D. L. (1976). Attention bands in absolute identification. *Perception & Psychophysics*, **20**, 49-54.
- 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.
- MARLEY, A. A. J. (1992). Developing and characterizing multidimensional Thurstone and Luce models for identification and preference. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 299-333). Hillsdale, NJ: Erlbaum.
- MARLEY, A. A. J., & COOK, V. T. (1984). A fixed rehearsal capacity interpretation of limits of absolute identification performance. *British Journal of Mathematical & Statistical Psychology*, **37**, 136-151.
- 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.
- MELARA, R. D., & MARKS, L. E. (1990). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 398-414.
- MELARA, R. D., & MOUNTS, J. R. W. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception & Psychophysics*, **56**, 73-90.
- MELARA, R. D., MOUNTS, J. R. W., & YAMAGISHI, N. (1993, November). *Why does selective attention fail?* Paper presented at the 34th Annual Meeting of the Psychonomic Society, Washington, D.C.
- MURDOCK, B. B., JR. (1985). An analysis of the strength-latency relationship. *Memory & Cognition*, **13**, 511-521.
- NEWELL, A., & ROSENBLUM, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.
- NOBEL, P. (1996). *Response times in recognition and recall*. Unpublished doctoral dissertation, Indiana University, Bloomington.

- 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. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, **43**, 25-53.
- NOSOFSKY, R. M., & PALMERI, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin & Review*, **3**, 222-226.
- NOSOFSKY, R. M., & PALMERI, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.
- PALMERI, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 324-354.
- POMERANTZ, J. R., & GARNER, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception & Psychophysics*, **14**, 565-569.
- PURKS, S. R., CALLAHAN, D. J., BRAIDA, L. D., & DURLACH, N. I. (1980). Intensity perception: X. Effect of preceding stimulus on identification performance. *Journal of the Acoustical Society of America*, **67**, 634-637.
- RAAB, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, **24**, 574-590.
- RATCLIFFE, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446-461.
- RATCLIFFE, R., & MURDOCK, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, **83**, 190-214.
- 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.
- SHEPARD, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In G. R. Lockhead & J. R. Pomerantz (Eds.), *Perception of structure: Essays in honor of Wendell R. Garner* (pp. 53-71). Washington, DC: American Psychological Association.
- SHEPARD, R. N., & CHANG, J. J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology*, **65**, 94-102.
- THOMAS, R. (1996). *A mathematical model of perceptual interactions in the same-different task: Theory and data*. Unpublished doctoral dissertation, Indiana University, Bloomington.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
- TVERSKY, A., & GATI, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, **89**, 123-154.

#### NOTES

1. As discussed previously by Nosofsky and Palmeri (1997, p. 271), this simplifying assumption is most likely to be approximately satisfied for stimuli varying along integral dimensions, which are encoded and represented as unitary wholes. More complex assumptions regarding encoding processes are needed for modeling speeded classification of separable-dimension stimuli, where serial processing (or limited capacity parallel processing) of individual dimensions is likely involved. The present tests, however, focused on speeded classification of integral-dimension stimuli.

2. Maddox and Ashby (1996) also considered a power model for relating decision time to distance, but they found that the exponential

model yielded better overall quantitative fits. We observed the same pattern in the fits to our data; therefore, in this article, we report only the exponential-model fits.

3. To the extent that selective attention is imperfect in the filtering task in a situation involving separable-dimension stimuli, then the EBRW model begins to predict the same interference effects as observed for integral-dimension stimuli. Likewise, to the extent that an experienced observer learns to attend simultaneously to both dimensions in the correlated task, the EBRW model will predict facilitation in this task. The basis for these predictions is explained in the text in the section in which the EBRW model is applied to integral-dimension stimuli. Note that, in the speeded classification tasks, it would be optimal for an observer to attend selectively to only the relevant dimension in the control and filtering tasks but to attend to both dimensions in the correlated task (where both dimensions are relevant). Thus, assuming that highly experienced observers are able to learn to distribute attention in an optimal manner across tasks, performance patterns involving separable-dimension stimuli that show a lack of interference in the filtering task yet facilitation in the correlated task (cf. Maddox & Ashby, 1996) seem quite explicable in terms of the EBRW model.

4. The lowest and highest intervals are unbounded. Solely for purposes of presenting visual displays of the distributions, we used the following arbitrary procedure for computing the width of these intervals. (The procedure has no influence on how the fits of the models are computed.) For the lowest interval (Bin 1), we defined an imaginary lower cutoff to be the same distance from the Bin 1 mean as the Bin 1 and Bin 2 means were from one another. Note that this procedure causes the lower cutoff to be twice as far from the Bin 1 mean as is the upper cutoff, so as to give a more accurate visual impression of what is truly an unbounded interval. The width of the lowest interval was then given by the distance between the upper cutoff and this imaginary lower cutoff. An analogous procedure was used for computing the width of the highest interval.

5. Because our main concern was with the RT data in this study, and the error proportions across all tasks were quite low, we decided in our main modeling analyses to fit only the averaged error data across tasks. Note, of course, that each model was allowed to predict that individual stimuli within each task had different proportions of errors; it was simply that our *criterion of fit* considered only the averages computed over the relevant stimuli. In an alternative set of modeling analyses, the accuracy SSD [SSD(A)] was computed by summing over the squared deviations between predicted and observed error probabilities for each *individual* stimulus in each task. [Because there were 20 such individual error probabilities, and 200 RT data points, the SSD(A) was weighted by 10 when computing the SSD(T) in this analysis.] This modeling analysis yielded the same pattern of results as the main one we report in our article, and none of our conclusions were changed.

6. An exception for the EBRW model is that we assumed shifting patterns of selective attention across conditions in which different dimensions were relevant. (However, this shifting pattern was modeled in terms of a single attention-weight parameter,  $w_{att}$ , which was held fixed.) This assumption of shifting selective attention has always been a cornerstone of the exemplar-based classification model (Nosofsky, 1984, 1986). An exception for the DBM is that the variances of the stimulus distributions were allowed to increase in the filtering conditions, as modeled by the standard-deviation multiplier  $\kappa$ . This potential increase in variance due to uncertainty effects is a central hypothesis stemming from the DBM. Of course, different decision bounds were also assumed to operate across the different conditions, but no free parameters were involved in the estimation of these bounds.

(Manuscript received May 6, 1996;  
revision accepted for publication September 17, 1996.)