Psychological Review
1997, Vol. 104, No. 2, 266-300

Copyright 1997 by the American Psychological Association, Inc.
0033-295X/97/$3.00

An Exemplar-Based Random Walk Model of Speeded Classification

Robert M. Nosofsky
Indiana University Bloomington

Thomas J. Palmeri
Vanderbilt University

The authors propose and test an exemplar-based random walk model for predicting response times in tasks of speeded, multidimensional perceptual classification. The model combines elements of R. M. Nosofsky's (1986) generalized context model of categorization and G. D. Logan's (1988) instance-based model of automaticity. In the model, exemplars race among one another to be retrieved from memory, with rates determined by their similarity to test items. The retrieved exemplars provide incremental information that enters into a random walk process for making classification decisions. The model predicts correctly effects of within- and between-categories similarity, individual-object familiarity, and extended practice on classification response times. It also builds bridges between the domains of categorization and automaticity.

Models of multidimensional perceptual classification have grown increasingly powerful and sophisticated in recent years, providing detailed quantitative accounts of patterns of classification learning, transfer, and generalization (e.g., Anderson, 1991; Ashby, 1992; Estes, 1986, 1994; Kruschke, 1992; Nosofsky, 1992b; Shanks & Gluck, 1994). However, a fundamental limitation of all the major competing models in the field today is that they offer no processing account of the time course of classification. Because response times provide a window into understanding the nature of cognitive representations and decision processes, it is vital to move in the direction of models that account for this form of data. In this article we propose and test a process-oriented model for predicting response times in tasks of speeded perceptual classification.

Our proposed model follows in the spirit of some leading extant models of categorization by assuming that people represent categories in terms of stored exemplars (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986). Classification decisions are made by retrieving these stored exemplars from memory. In the newly proposed model, retrieved exemplars are used to drive a *random walk process* (e.g., Luce, 1986; Townsend & Ashby, 1983) in which evidence accrues to alternative categories over time. Random-walk models have been successful at accounting for performance in tasks of memory, decision making, sensory discrimination, and unidimensional absolute judgment (e.g., Busemeyer, 1985; Karpiuk, Lacouture, & Marley, in press; Link, 1992; Ratcliff, 1978). Thus, a random walk architecture seemed a promising one to explore in the domain of multidimensional perceptual categorization. Because our model assumes that retrieved category exemplars are used to drive a random walk process, we refer to it as an *exemplar-based random walk (EBRW)* model.

The EBRW integrates and extends two previously developed and well-known exemplar models of cognitive processes. The first is Nosofsky's (1986) generalized context model (GCM), which has had a long history of success in accounting for relations among categorization, identification, old–new recognition, and similarity (Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1984, 1992b). The second model is Logan's (1988, 1990) instance-based model of automaticity, which has been extremely successful at accounting for the development of skilled performance as a function of extended practice. Although this article focuses on how the EBRW accounts for classification response times, another important contribution of the work is that it builds bridges between the domains of categorization and the development of automaticity (Palmeri, 1997). Indeed, we suggest that the EBRW may provide insights into the development of expertise in perceptual classification.

We organize our article by first briefly reviewing Nosofsky's (1986) GCM and Logan's (1988) instance model. We then explain how key components of each are integrated in the EBRW. Next, a formal statement of the EBRW is provided, its key properties are discussed, and analytic predictions are derived. In the empirical section of the article, the EBRW is tested in several experiments involving speeded perceptual classification, and its predictions are compared with those of an alternative descriptive model of classification reaction time. Finally, applications of the EBRW to accounting for the development of skilled performance in tasks of visual numerosity judgment (e.g., Lassaline & Logan, 1993) are briefly illustrated and discussed as we interrelate the domains of categorization and automaticity.

Review of the Component Models

The Generalized Context Model of Classification

According to Nosofsky's (1986) GCM, people represent categories by storing individual exemplars in memory. Classifica-

This work was supported by Grant PHS RO1 MH48494-05 from the National Institute of Mental Health.

Jerome Busemeyer, A. A. J. Marley, Richard Shiffrin, and James Townsend provided extensive commentary and discussion. We also wish to thank John Anderson, Rob Goldstone, John Kruschke, Roger Ratcliff, Roger Shepard, and Trisha van Zandt for their comments, discussion, and advice.

Correspondence concerning this article should be addressed to Robert M. Nosofsky, Department of Psychology, Indiana University, Bloomington, Indiana 47405; or to Thomas J. Palmeri, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240. Electronic mail may be sent via Internet to nosofsky@indiana.edu or palmerit@ctrvax.vanderbilt.edu.

tion decisions are based on summing the similarity of an object to the exemplars of the alternative categories. In the GCM, 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 are assumed to systematically modify the structure of the space in which the exemplars are embedded. In what follows, we review formal aspects of the GCM that underlie the exemplar-based random walk model.

Assume the exemplars reside in an M-dimensional psychological space, and let x_{im} denote the value of exemplar i on psychological dimension m. The distance between exemplars i and j is given by

$$d_{ij} = \sqrt{[\sum_{m} w_m |x_{im} - x_{jm}|^2]}, \quad (1)$$

where w_m ($0 \leq w_m$, $\Sigma w_m = 1$) represents the *attention weight* given to dimension m. The x_{im} psychological coordinate values for the exemplars are generally derived by conducting multidimensional scaling studies or else are assumed to be given by the physical coordinate values used for constructing the stimuli. The attention weights (w_m) are free parameters in the model (although various principles and learning mechanisms have been proposed for predicting the weights a priori, e.g., Nosofsky, 1984, and Kruschke, 1992). Attending selectively to a dimension serves to stretch the space along that dimension and shrink the space along unattended dimensions.

Following Shepard (1987), the distances (d_{ij}) are transformed to similarity measures (η_{ij}) by using an exponential decay function,

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

where c is an overall sensitivity parameter for scaling distances in the space. There is extensive evidence that similarity is related to psychological distance according to this exponential law (Shepard, 1987).

Because of factors such as recency of presentation, exemplars may reside in memory with differing strengths. Let M_j denote the memory strength for exemplar j. The degree to which exemplar j is activated when presented with item i is determined jointly by the exemplar's strength in memory and by its similarity to item i (Nosofsky, 1988, 1991a). Specifically, the activation for exemplar j given presentation of item i (a_{ij}) is given by

$$a_{ij} = M_j \eta_{ij}. \quad (3)$$

Finally, the evidence for Category J given presentation of item i is found by summing the activations for all stored exemplars of Category J. The conditional probability with which item i is classified into Category J is found by dividing this evidence by the summed evidence for all the categories:

$$P(J|i) = \sum_{j \in J} a_{ij} / [\sum_{K} \sum_{k \in K} a_{ik}]. \quad (4)$$

There are numerous demonstrations of the GCM's ability to account accurately for the probability with which individual objects are classified into alternative categories (for reviews, see Nosofsky, 1992a, 1992b). However, the model provides no account of how the exemplar-based similarity comparison process unfolds over time. One of the central goals of this research is to develop such an account.

Logan's Instance-Based Model of Automaticity

The basic idea in Logan's (1988) instance model is that people start with general strategies or algorithms for performing skilled actions. Each time a skilled action is successfully performed, however, it lays down an instance in memory. These instances can later be retrieved and used to perform the task. Skilled performance is conceptualized as a race between executing the algorithm and retrieving any one of the numerous instances stored in memory. As the observer becomes highly experienced, the stored instances begin to dominate and win the race, and performance becomes automatized.

Logan's model provides an elegantly simple account of the development of automaticity in skilled performance. Indeed, it accounts quantitatively for the power law decreases in the mean and standard deviation of response times that are often observed with training (Logan, 1988, 1992; Newell & Rosenbloom, 1981) and deals quantitatively with entire distributions of reaction time data (Logan, 1992). The theory has also shown considerable power in accounting for the development of automaticity in memory search (Strayer & Kramer, 1994), lexical decision (Logan, 1990), alphabet arithmetic (Logan & Klapp, 1991), and repetition priming (Logan, 1990).

Despite these achievements, the current version of the instance model is limited in two important respects. First, it takes no account of graded similarity structure among exemplars. Second, as will be made clear, mechanisms of response competition need to be added to the model.

The Exemplar-Based Random Walk Model

Our elaborated model combines some of the major properties underlying Nosofsky's (1986) GCM and Logan's (1988) instance model. According to the EBRW, people represent categories by storing individual exemplars in memory. As in Logan's model, presenting a test item causes stored category exemplars to race against one another to be retrieved. However, whereas in Logan's model only exemplars that are identical to the presented item enter the race, in our elaborated model all exemplars race against one another with rates that are proportional to the degree to which they are activated by the test item. As we described previously, these activations are determined jointly by the exemplars' strength in memory and by their similarity to the test item. Furthermore, whereas in Logan's model the first retrieved exemplar initiates the action, in the elaborated model we assume that the retrieved exemplars provide incremental information that feeds into a random walk process. Only when sufficient information is obtained to complete the random walk is a categorization response initiated. As we discuss in the section Relations Among Models, extending Logan's single-instance race model by means of the random walk mechanism is critical to achieving plausible predictions of classification response time.

The exemplar-based random walk (EBRW) is illustrated for a two-category case in Figure 1. (More general versions of the model applicable to multiple-category situations are described in a later section.) In the model, there is a random walk counter that accrues information pointing to either Category A or B. The counter has a starting value of zero; positive increments move it in the direction of Category A, and negative increments move it in the direction of Category B. The observer establishes criteria representing the amount of evidence that is needed to execute either a Category A response (+A) or a Category B response (−B). Once the counter reaches either of these criteria, the appropriate categorization response is made.

When Item i is presented, it sets off a race among all the exemplars stored in memory. These race times are random variables with rates that are proportional to the degree to which each exemplar is activated by Item i. For simplicity in getting started, and for reasons of analytic tractability, we assume that the race times are exponentially distributed. Thus, the probability density that Exemplar j completes its race at time t is given by

$$f(t) = a_{ij}\exp(-a_{ij}t), \tag{5}$$

where the a_{ij} activation values are computed as in the GCM (Equations 1–3). Bundesen (1990) and Logan (1997) incorporated similar assumptions concerning object race times in their theories of visual attention. Likewise, Marley (1992; Marley & Colonius, 1992) assumed such an exponential race process as a means of characterizing a wide variety of models of identification, classification, and preference.

The exemplar (x) that wins the race is retrieved and enters into the random walk. If x belongs to Category A, then the random walk counter is increased by unit value, whereas if x belongs to Category B, then the counter is decreased by unit value. If the counter reaches either the Category A criterion (+A) or the Category B criterion (−B), then the appropriate response is executed. Otherwise, a new race is initiated, another exemplar is retrieved, and the process continues.

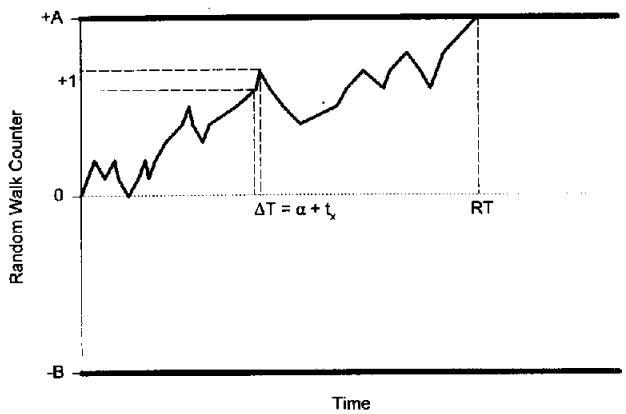

Figure 1. Schematic illustration of the random walk process in the exemplar-based random walk model. ΔT = time taken for each step; α = step-time constant; t_x = time to retrieve exemplar x; +A = Category A criterion; −B = Category B criterion; RT = response time.

The time to take each individual step in the random walk (T_{step}) is given by

$$T_{step} = \alpha + t_w, \tag{6}$$

where α is a constant term associated with each step, and t_w is the time that it takes to retrieve the "winning" exemplar. Note that when an exemplar is retrieved, the observer needs to extract the category label to which it is associated and then accumulate this information on the random walk counter. A psychological interpretation for the α parameter is that it represents the time needed for this category-label extraction and accumulation process.

Formal Properties of the Model

Before proceeding to a discussion of the conceptual and analytic predictions associated with the EBRW, it is useful to briefly review some well-known mathematical properties of exponential distributions (e.g., see Townsend & Ashby, 1983, pp. 36–43). Suppose there are n independent exponential distributions racing in parallel with rates $\lambda_1, \lambda_2, \ldots, \lambda_n$. Consider the process defined as the minimum finishing time among all of these exponentials. This process itself follows an exponential distribution with rate $\lambda_1 + \lambda_2 + \ldots + \lambda_n$. The expected finishing time of an exponential process with rate λ_j is given by $E(T) = 1/\lambda_j$. Thus, the expected minimum finishing time among the n independent exponential processes, $E(T_n)$, is given by

$$E(T_n) = 1/\sum \lambda_k. \tag{7}$$

A final property of the exponential distribution that is relevant for our purposes is that the probability that process j is the first to finish among the n processes (i.e., that it "wins the race") is given by

$$P(j) = \lambda_j/\sum \lambda_k. \tag{8}$$

Consider now the situation described for the EBRW. There are two categories, A and B. Given presentation of item i, each exemplar j races to be retrieved according to an exponential distribution with activation rate a_{ij}. Let S_{iA} denote the summed activation of all exemplars in Category A when item i is presented, that is,

$$S_{iA} = \sum_{j\in A} a_{ij}, \tag{9}$$

and likewise for S_{iB}. Then when item i is presented, the minimum finishing time for all the exemplars that race to be retrieved is distributed exponentially with rate $\lambda = S_{iA} + S_{iB}$. Thus, when item i is presented, the expected minimum finishing time (i.e., the expected retrieval time for the winning exemplar) is given by

$$E(t_w|i) = 1/(S_{iA} + S_{iB}). \tag{10}$$

Therefore, from Equation 6, when item i is presented, the expected time for taking each step in the random walk is given by

$$E(T_{step}|i) = \alpha + 1/(S_{iA} + S_{iB}). \quad (11)$$

It is also straightforward to derive the probability that, on each step of the random walk, the counter moves in the direction of Category A or Category B. We know from Equation 8 that, when item i is presented, the probability that exemplar j is retrieved is given by

$$P(j|i) = a_{ij}/(\sum_K \sum_{k \in K} a_{ik}). \quad (12)$$

The probability of taking a step toward Category A is found by summing the probabilities that any one of the exemplars j from Category A is retrieved, yielding

$$p_i = \sum_{j \in A} a_{ij}/(\sum_K \sum_{k \in K} a_{ik}). \quad (13a)$$

In the present two-category situation (and using our earlier notation), this probability of taking a step toward Category A can be expressed more simply as

$$p_i = S_{iA}/(S_{iA} + S_{iB}). \quad (13b)$$

Likewise, the probability that the random walk takes a step in the direction of Category B is given by

$$q_i = 1 - p_i = S_{iB}/(S_{iA} + S_{iB}). \quad (13c)$$

Conceptual Predictions

Some main conceptual predictions of the EBRW can now be discussed. These predictions follow naturally from the underlying assumption in the EBRW that the duration of the random walk is determined jointly by the total number of steps required to initiate a response and by the speed with which each of the individual steps is made.

The first prediction is that rapid classification decisions should be made for items that are highly similar to exemplars from one category and that are dissimilar to exemplars from the alternative category. Under such conditions, each retrieved exemplar will tend to come from the same category, and the random walk will march in consistent fashion to this category's criterion (Figure 1). For example, an item that is similar only to exemplars of Category A will yield a large value of p_i in Equation 13b, so each step in the random walk will tend to move toward criterion A. By contrast, items that are similar to exemplars from both categories should yield slow response times. The reason is that the random walk counter will tend to wander back and forth, sometimes retrieving exemplars from one category and other times retrieving exemplars from the contrast category.

A second prediction is that increased experience with category exemplars should facilitate performance. As category training increases, more exemplars come to be stored in memory. The greater the number of exemplars that race to enter the random walk, the faster the winning retrieval times will be. Formally, as a greater number of exemplars come to be stored in memory, the exponential processing rate $S_{iA} + S_{iB}$ grows larger, so the expected winning retrieval time decreases (Equation 10). Intuitively, the larger the number of exemplars that participate in the race, the greater is the probability that at least one of the retrieval times will be particularly fast (cf. Logan, 1988). These faster retrieval times result in faster individual steps in the random walk process (Equation 11).

A third prediction is that, all other things being equal, individual item familiarity should also facilitate speeded classification. We say that an item is familiar if it has been presented with high frequency or if it is highly similar to numerous other old exemplars. Note that, according to the proposed model, a highly familiar item will result in fast retrieval times for individual exemplars because numerous exemplars will race with high activation rates. Thus, familiarity should speed the random walk. This prediction about the role of individual item familiarity basically combines our predictions about the roles of similarity and experience in a more specific manner. We test each of these fundamental predictions, as well as some more subtle and fine-grained ones, in quantitative fashion in our subsequent experiments.

Analytic Predictions

The processing assumptions in the EBRW yield a simple random walk from which it is straightforward to derive analytic predictions. Given test item i, then on each step of the random walk there is fixed probability p_i of moving unit value in the direction of criterion +A, and probability $q_i = 1 - p_i$ of moving unit value in the direction of criterion −B. The random walk finishes as soon as either criterion +A or −B is reached. The properties of this type of random walk are extremely well-known (e.g., Feller, 1968, chap. XIV). Because it is extremely difficult to motivate intuitively the derivations of the prediction equations, we simply list the relevant results (see Feller, 1968, chap. XIV).

The expected number of steps in the random walk [$E(N|i)$] is given by

$$E(N|i) = \frac{B}{q_i - p_i} - \frac{A + B}{q_i - p_i}\left[\frac{1 - (q_i/p_i)^B}{1 - (q_i/p_i)^{A+B}}\right], \quad \text{if } p_i \neq q_i, \quad (14a)$$

and

$$E(N|i) = AB, \quad \text{if } p_i = q_i. \quad (14b)$$

Therefore, because each individual step in the EBRW has expected duration $\alpha + 1/(S_{iA} + S_{iB})$, and this duration is independent of which actual step is taken, the expected time of the entire random walk is given by[1]

[1] Equation 15 helps bring out the mathematical importance of the α parameter to the model's predictions. According to the equation, the expected decision time is a weighted sum of the expected number of steps in the random walk, where the weights are α and $1/(S_{iA} + S_{iB})$. If α were zero, then even if the number of steps in the random walk were large, the expected decision time could be quite small if the factor $1/(S_{iA} + S_{iB})$, were small. Note that an item that is highly similar to the exemplars of both categories would produce a small value of $1/(S_{iA} + S_{iB})$, so the $\alpha = 0$ model would tend to predict a fast response time. Such a prediction would be incorrect, however, because an item that is

$$E(T|i) = E(N|i) \cdot E(T_{step}|i)$$
$$= E(N|i) \cdot [\alpha + 1/(S_{iA} + S_{iB})]. \quad (15)$$

The probability of a Category A response given presentation of item i is given by

$$P(A|i) = \frac{1 - (q_i/p_i)^B}{1 - (q_i/p_i)^{A+B}}, \quad \text{if } p_i \neq q_i, \quad (16a)$$

and

$$P(A|i) = \frac{B}{A + B}, \quad \text{if } p_i = q_i. \quad (16b)$$

The probability of a Category B response is given by

$$P(B|i) = \frac{(q_i/p_i)^B - (q_i/p_i)^{A+B}}{1 - (q_i/p_i)^{A+B}}, \quad \text{if } p_i \neq q_i, \quad (17a)$$

and

$$P(B|i) = \frac{A}{A + B}, \quad \text{if } p_i = q_i. \quad (17b)$$

Busemeyer (1982) has derived equations for the expected number of steps in the random walk conditionalized on each category response. The expected number of steps given a Category A response is given by

$$E(N|A, i) = \frac{1}{p_i - q_i}[\theta_1(A + B) - \theta_2(B)], \quad \text{if } p_i \neq q_i, \quad (18a)$$

and

$$E(N|A, i) = \frac{A}{3}(2B + A), \quad \text{if } p_i = q_i, \quad (18b)$$

where

$$\theta_1 = \frac{(p_i/q_i)^{A+B} + 1}{(p_i/q_i)^{A+B} - 1}, \quad \text{and} \quad \theta_2 = \frac{(p_i/q_i)^B + 1}{(p_i/q_i)^B - 1}. \quad (19)$$

For a Category B response, the relevant equations are

$$E(N|B, i) = \frac{1}{q_i - p_i}[\theta_1(A + B) - \theta_2(A)], \quad \text{if } p_i \neq q_i, \quad (20a)$$

and

$$E(N|B, i) = \frac{B}{3}(2A + B), \quad \text{if } p_i = q_i, \quad (20b)$$

where

$$\theta_1 = \frac{(p_i/q_i)^{-(A-B)} + 1}{(p_i/q_i)^{-(A+B)} - 1} \quad \text{and} \quad \theta_2 = \frac{(p_i/q_i)^{-A} + 1}{(p_i/q_i)^{-A} - 1}. \quad (21)$$

similar to the exemplars of both categories would be quite difficult to classify. The α parameter gives an absolute weight to the number of steps in the random walk. Items that are difficult to classify result in a large number of steps, so the model correctly predicts their slow RTs.

For analogous expressions involving more general random walk processes, see Link and Heath (1975) and Luce (1986). Again, as was the case in Equation 15, the expected time of the entire random walk conditionalized on Response A or Response B is found by multiplying the expressions in Equations 18 and 20 by the factor $[\alpha + 1/(S_{iA} + S_{iB})]$ from Equation 11.

The values of p_i, q_i, S_{iA}, and S_{iB} that enter into the prediction equations are determined by the processing assumptions in the EBRW. These values will vary for each individual test item i depending on its similarity to the exemplars of the alternative categories. The values will also change during the course of learning as more exemplars are stored in memory. However, for any given item i at a particular stage of learning, the relevant values of p_i, q_i, S_{iA}, and S_{iB} can be directly computed (from Equations 1–3, 9, and 13) and then substituted into Equations 14–21 to derive the quantitative predictions of the EBRW.

Relations Among Models

An important property of the EBRW is that it can be viewed as generalizing both Nosofsky's (1986) GCM and Logan's (1988) instance model. In particular, suppose that the category criteria are set at A = 1 and −B = −1. Then if the first exemplar that is retrieved belongs to Category A, a Category A response is made (and likewise for Category B). Thus, the probability of a Category A response is given by $P(A|i) = p_i = S_{iA}/(S_{iA} + S_{iB})$, which is the GCM response rule (Equation 4).[2]

Likewise, suppose that the criteria are set at A = 1 and −B = −1, and that there is zero similarity between nonidentical objects ($c = \infty$ in Equation 2). Then only exemplars that are identical to the presented item participate in the race, and the first one retrieved initiates the response. This process is essentially the one that is formalized in Logan's instance model. An important difference is that whereas we have assumed exponentially distributed races in the EBRW, Logan (1988, 1992) has advanced theoretical arguments for the importance of using a more general Weibull distribution. However, his arguments are based on the assumption that only the first instance retrieved initiates the response, whereas in our model multiple races take place until the random walk reaches a response criterion. It may well be that the predictions of the EBRW can be improved by replacing the exponential-race assumption with the Weibull, but we leave this question as an issue for future research.

The random walk decision mechanism in the EBRW is a

[2] It should be made clear, however, that if the magnitude of the criteria A and B is set at 1, the EBRW does not yield reasonable predictions of classification response time (see main text). Furthermore, with criterion settings of magnitude greater than 1, the EBRW makes predictions of classification accuracy that exceed those predicted by the GCM (assuming all other parameters are held constant). Thus, for the EBRW to roughly match the accuracy predictions yielded by the GCM, modifications in the values of some of the other free parameters would need to be made. It is interesting to note that Maddox and Ashby (1993) found that, when fitting the data of highly experienced individual participants, the GCM does underpredict the observed accuracies, so the EBRW may represent an improvement in this regard. We consider these issues in more depth in the General Discussion.

critical aspect of the model. If Logan's instance model were extended solely by allowing similarity-based retrieval mechanisms, but where the first instance retrieved still drives the response, there would be important situations in which it would be unable to yield reasonable predictions of classification response times. For example, suppose that one added an instance to Category B that was highly similar to an instance from Category A. This manipulation, which increases classification difficulty, would undoubtedly lead to slower response times. A pure single-instance race model predicts the opposite, however. Adding another highly similar instance to the race, even from a contrast category, could only speed the winning retrieval time, albeit at the cost of more errors in performance. In general, a pure single-instance race model with similarity-based retrieval predicts *faster* response times for difficult-to-discriminate objects. The random walk decision process in the EBRW leads the system to obtain additional evidence when discriminations are difficult, thereby trading speed for accuracy.

Domain of Application

The EBRW is intended to model the time course of the classification decision-making stage. A complete account of classification response time also requires modeling of encoding and response execution stages. For simplicity, we assume that the time courses for these other residual stages are identical for all stimuli and are independent of the decision-making stage modeled by the EBRW.

An important question concerns the generality with which exemplar-based memory and retrieval processes operate in classification decision making. Although successful applications of the GCM have been demonstrated in numerous stimulus domains, alternative classification strategies may also operate (e.g., Nosofsky, Palmeri, & McKinley, 1994). We hypothesize that exemplar-based classification processes are most likely to operate in domains involving *integral dimension* as opposed to *separable dimension* stimuli (Garner, 1974; Lockhead, 1972; Shepard, 1964; Shepard & Chang, 1963; Treisman & Gelade, 1980). As discussed by numerous investigators, integral dimension stimuli are ones that tend to be encoded, perceived, and represented as single, unitary wholes. This processing constraint seems conducive to exemplar storage and retrieval strategies. By contrast, highly separable dimension stimuli appear to be processed in terms of their separate dimensions, perhaps rendering the storage of complete exemplars less efficient. Another difficulty involving highly separable dimension stimuli is that their encoding may require serial processing (or limited-capacity parallel processing), thereby further complicating response time predictions. Although the generality with which the EBRW may apply is an empirical question, our initial tests of the model will focus mainly on domains involving integral dimension stimuli.

Model Parameters

The free parameters in the EBRW include the sensitivity parameter (c) for transforming distances to similarities (Equation 2); the set of attention weights (w_m) in the distance function (Equation 1); the category criteria ($+A$ and $-B$) in the random walk; and the time constant (α) in the exemplar retrieval function (Equation 6). Because the predictions of the EBRW are in arbitrary units, simple linear regression is used to rescale the predicted values onto the observed response times. These regressions use a slope parameter, k, which is the scaling factor; and a y-intercept, μ_R, which can be interpreted as the mean of the residual processing stages. More general versions of the EBRW with additional free parameters are also considered in subsequent sections of the article.

Preliminary Considerations: Predicting Reaction Times for Members of Normally Distributed Categories

To illustrate some preliminary successes for the EBRW, we first demonstrate its ability to account for some important classification reaction time data reported by Ashby, Boynton, and Lee (1994). These researchers used an experimental paradigm known as the *general recognition randomization technique* (GRRT). In typical applications of the GRRT, the stimuli vary along two continuous dimensions, and the experimenter establishes two categories defined by bivariate normal distributions. Examples of three different experimental conditions tested by Ashby et al. (1994) are shown in Figure 2. In the figure, each solid-line circle or ellipse represents a *contour of equal likelihood* for a given bivariate normal distribution. All points falling along the contour are equally likely to be produced by the normal distribution. The shape of the contour provides an illustration of the structure of the distribution. The center of each ellipse gives the mean of the normal distribution on each of its dimensions. The expanse of the ellipse along each dimension represents the variability of the distribution along that dimension. The correlation between the dimensions is represented by the angle of orientation of the ellipse. For example, in the leftmost panel of Figure 2, category distributions A and B have equal variance along both dimensions and the dimensions are positively correlated.

Note that any point in the two-dimensional space can be produced by either Distribution A or B. However, certain locations are more probable for one distribution than another, so the degree of overlap between the distributions can be manipulated. In the examples in Figure 2, the distribution means are identical across conditions, and all that varies are the dimension variances and correlations. Ashby et al. (1994) manipulated these parameters to produce a low-overlap, medium-overlap, and high-overlap condition, as illustrated in Figure 2.

In the GRRT paradigm, each bivariate distribution defines a category. Thus, on each trial of the experiment, (a) a category distribution is selected; (b) an exemplar from this distribution is randomly chosen and presented to the observer; (c) the observer judges the exemplar's category assignment; and (d) corrective feedback is then provided. By the time learning is completed, an observer may have experienced hundreds or even thousands of unique training exemplars from each category. Because the distributions are overlapping, it is impossible to perfectly classify all exemplars. However, it is possible to define an optimal decision boundary that maximizes classification accuracy. For the bivariate normal distributions in Figure 2, the optimal decision rule in each condition is the same linear boundary illus-

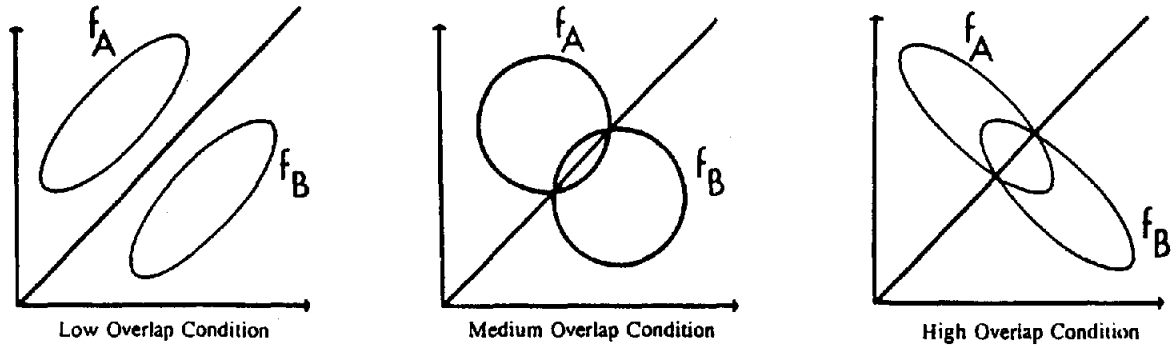

Figure 2. Contours of equal likelihood from the three overlap conditions tested by Ashby, Boynton, and Lee (1994, Experiment 1). f_A = probability density for Category A; f_B = probability density for Category B. Reprinted from "Categorization Response Time with Multidimensional Stimuli," by F. G. Ashby, G. Boynton, and W. W. Lee, 1994, *Perception and Psychophysics, 55,* p. 15. Copyright 1994 by the Psychonomic Society. Reprinted with permission.

trated in each panel of the figure. Anytime an object falls to the upper left of the boundary it should be classified in Category A, and anytime it falls to the lower right it should be classified in Category B. In the numerous experiments reported by Ashby and his colleagues, participants' classification choices are often well described by these optimal decision boundaries or by boundaries with a form very close to optimal (Ashby & Gott, 1988; Ashby et al., 1994; Ashby & Maddox, 1992).

In the Ashby et al. (1994) studies, in addition to recording the classification choice for each stimulus, the experimenters measured the response times. One key finding was that in all conditions, there was a statistically significant (negative) rank-order correlation between the response time for each object and the distance of the object from the classification decision boundary. In other words, in general, the greater the distance of an object from the decision boundary, the faster was the classification response. A second finding was that there was essentially zero correlation between the response times and a measure of familiarity for each stimulus. Familiarity (F_i) was computed by summing the similarity of any given item i to all exemplars in the space, where similarity is computed as in Equations 1 and 2. Thus, objects in dense regions of the space, which are highly similar to numerous other exemplars, have high familiarity, whereas those in isolated regions have low familiarity.

The results obtained by Ashby et al. (1994) were consistent with a descriptive model of classification response time (RT) proposed by Ashby and Maddox (1991; see also Ashby & Maddox, 1994, and Maddox & Ashby, 1996). According to the *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" (Ashby et al., 1994, p. 11). Although not a process-based model, the RT–distance hypothesis is the main current proposal in the field for predicting multidimensional categorization response times. In our article, we compare and contrast the predictions of the RT–distance hypothesis with those of the EBRW.

Our initial question concerns whether or not Ashby et al.'s (1994) results are consistent with the EBRW. To test the model, we generated random samples of 300 exemplars from the category distributions illustrated in Figure 2 (see Ashby et al., 1994, for the parameters that define these distributions). We based our analyses on 300 exemplars because this was the number of training exemplars used by Ashby et al. in their experiments. Assuming that this entire sample of exemplars was stored in memory, we then conducted simulations of the EBRW to generate response time predictions. We used simulations rather than computing the expected values from the analytic model because we needed to mimic the variability inherent in Ashby et al.'s data. In Ashby et al.'s (1994) experiment, only a single response time was obtained for each individual object, so this variability was presumably quite large. Thus, in our analyses, a single response time was simulated for each individual object, and these simulated data were then compared with the objective measures of interest.

Although the precise predictions of the EBRW depend on the values of the free parameters, we discovered that over a wide range of parameter settings the qualitative predictions of the model were essentially the same (details are provided in Appendix A). First, the rank-order correlation between the predicted response times and the distance-from-boundary measure ranged from −.25 to −.55, which is essentially the same range reported by Ashby et al. (1994). Second, as found by Ashby et al. (1994), the rank-order correlation between the predicted response times and the familiarity measure was negligible. Thus, the exemplar-based random walk yields response time predictions that share some key properties with the observed data.[3]

Why does the model predict that response times will generally

[3] Although a detailed explanation goes beyond the scope of this article, we note that the EBRW also predicts some more fine-grained results observed in the Ashby et al. (1994) study. First, it predicts their finding that median response times for correct classification choices were faster than for incorrect ones. Second, it predicts their finding that the difference in response speed for correct versus incorrect choices decreased as one moved from the low-overlap to the high-overlap condition. Third, the EBRW is fully consistent with their finding of null effects of familiar-

be shorter as the distance from the category decision boundary increases? In the GRRT paradigm, an exemplar that is far from the boundary tends to be highly similar only to exemplars from its own category. Thus, on each step of the random walk, exemplars from the correct category are retrieved, and the counter marches in consistent fashion to the appropriate category criterion. By contrast, an exemplar that lies close to the boundary tends to be similar both to exemplars from its own category and to exemplars from the opposite category. The random walk tends to wander back and forth, first in the direction of one category criterion and then in the direction of the other. Thus, more evidence needs to be accrued before a decision can be made.

As discussed previously, the EBRW does predict that there should be a relation between *familiarity* (F_i) and response times because highly unfamiliar exemplars should lead to slow individual steps in the random walk (Equation 6). However, in the GRRT paradigms tested by Ashby et al. (1994), the distance-from-boundary measure and the familiarity measure are themselves negatively correlated. That is, objects lying far from the decision boundary tend to have lower overall familiarity. Because our simulations of the EBRW correctly indicated roughly zero correlation between F_i and the response times, it is apparently the case that the familiarity differences yielded in the Ashby et al. (1994) conditions are not sufficient to undermine the critical contribution of distance-from-boundary. Furthermore, although Ashby et al. (1994) conducted several analyses that revealed no effect of the familiarity factor while attempting to hold constant distance-from-boundary, our simulations with the EBRW indicate that the model correctly predicts the results of these analyses as well (see Footnote 3).

We hypothesize that more extreme manipulations are needed to reveal an effect of the familiarity factor. Indeed, in one experiment conducted by Nosofsky (1991b), participants received initial classification training that was followed by a transfer phase requiring speeded classification responses. Among the transfer stimuli were novel objects that were several standard deviations further away from the category boundary than any other objects previously presented. Response times to these isolated objects were significantly slower than to familiar objects a moderate distance from the category boundary. In a new experiment reported in this article, we demonstrate significant effects of familiarity for objects a fixed distance from the category boundary. In sum, the EBRW predicts the general relation that response times are shorter as distance from the boundary increases but also predicts breakdowns of this relation in situations involving more extreme manipulations.

Experiment 1: Predicting Individual Object Response Times and Speedups in Classification Performance

Our analyses of the response time results in the experiments conducted by Ashby et al. (1994) provide preliminary evidence in favor of the predictions made by the EBRW. However, because response time data are highly variable, and only a single response time was obtained per individual object in this paradigm, the tests of the model thus far are limited to rough rank-order correlations with the data. The main goal of Experiment 1 was to provide more exacting tests of the EBRW by collecting extensive response time data for individual objects in a perceptual classification task and then quantitatively fitting the model to the mean response times for these individual objects. A secondary goal was to study changes in overall categorization response speed as a function of extended practice and to test whether the EBRW could account for this development of skilled classification performance.

The stimuli in our experiment were a set of 12 colors generated on a computer screen. The colors were divided into two categories, each with six members. Three participants engaged in an extended task of speeded perceptual classification. On each trial, one of the colors was randomly selected, and the participant was instructed to classify it into either Category A or B as rapidly as possible without making errors. Feedback was provided on each trial. Participants quickly learned the category assignments and rarely made errors, so the key dependent variable was classification response speed. Each participant was tested over a period of 5 days, with a total of 360 stimulus presentations per day. In addition to the speeded classification task, extensive similarity scaling work was performed to derive individual-participant multidimensional scaling (MDS) solutions for the colors. These MDS solutions were then used in conjunction with the EBRW to predict each participant's mean RT for classifying each color into its category.

Method

Participants. The participants were 3 Indiana University graduate students. They were paid $5.00 per 1-hr session plus up to a $2.00 bonus per session depending on performance. All participants claimed to have normal color vision. None of the participants was familiar with the issues under investigation in this study.

Stimuli. The stimuli were a set of 12 colors generated on computer screens. Extensive pilot work was used to construct a set of stimuli that approximately matched the Munsell color configuration illustrated in Figure 3. (This Munsell configuration was used by Nosofsky [1987, 1988] in previous studies of category learning.) According to the Munsell system, in this configuration the stimuli are of a constant red hue (5R) and vary only in their brightness and saturation. As illustrated in Figure 3, the stimuli were divided by the experimenters into two categories. Stimuli enclosed by squares belong to Category A, and stimuli enclosed by circles belong to Category B. The goal was to generate category structures in which the stimuli varied along two dimensions, where both dimensions were relevant for classifying the objects, and where a relatively smooth curvilinear boundary could separate the objects into categories. As will be seen, because our theoretical analyses of the data make use of individual-participant MDS solutions for the colors, the precise correspondence with the Munsell configuration illustrated in Figure 3 is not critical. The main purpose of aiming for a roughly two-dimensional color space was to reduce the number of free parameters needed for fitting the EBRW to the response-time data.

The stimuli were generated on CompuAdd 14-in. monitors (Model 51109) by adjusting the red, green, and blue (RGB) color channels on a Dell 486 machine. The RGB values corresponding to each color are

ity in analyses in which only the response times of the 50 exemplars closest to the decision boundary were considered (see Ashby et al., 1994, Figure 4). Finally, it predicts their finding that, among the 50 exemplars whose distance-from-boundary was closest to that of the prototype, the rank ordering of the RT associated with the prototype was approximately 25 (see Ashby et al., 1994, p. 20).

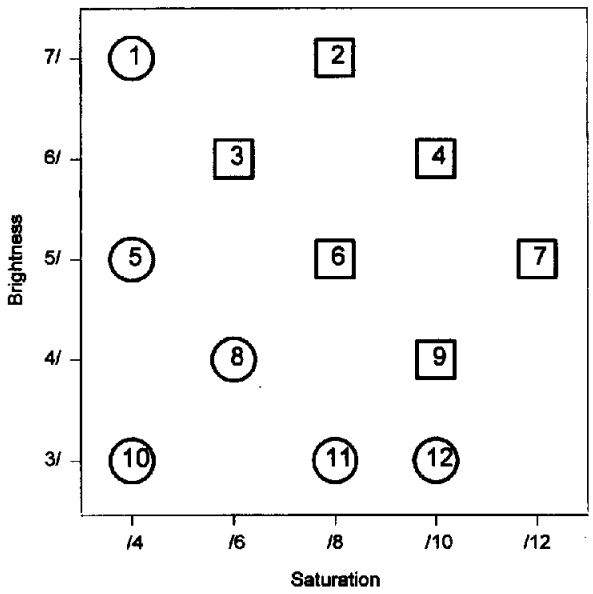

Figure 3. Schematic illustration of the approximate Munsell color configuration used in Experiment 1.

provided in Appendix B. Each color occupied a 6 cm × 6 cm square surrounded by a white background. Participants sat approximately 60 cm from the computer monitor, so the visual angle subtended by each stimulus was approximately 6°. Participants entered responses by pressing appropriate buttons on the computer keyboard. Response times were measured by using the internal ms-accuracy PC timer.

Procedure. The stimulus presentation schedule was organized into blocks of 12 trials, with each color presented once per block. Order of presentation of the colors was randomized within each block. By using this blocked presentation method, we hoped to minimize complex sequential effects resulting from stimulus repetitions or exceedingly long lags between identical stimulus presentations. Each participant completed five sessions of testing, one session per day on contiguous days of the week. There were 30 blocks (360 trials) per session, for a total of 1,800 trials per participant. Thus, each participant classified each individual color 150 times.

Stimulus displays were response terminated. Each stimulus display was followed by 2 s of feedback. Following a 500-ms ISI, the next color was displayed. Participants were instructed to rest their index fingers on the Category A and Category B response buttons throughout the testing session. Participants were instructed to make their responses as quickly and as accurately as possible. They were told that their bonus was determined jointly by the speed and accuracy of their responses.

Following the speeded classification task, each participant completed three sessions of a similarity scaling study, one session per day. There were 6 blocks of similarity judgments per session. On each block, all 66 unique pairs of colors were presented, one pair per trial, in a random order. Thus, there were 396 similarity judgment trials per session. Over the three sessions of testing, each participant provided 18 similarity judgments per color pair. The 6 cm × 6 cm color squares were presented simultaneously on the screen, separated by 3 cm. Left–right placement of each color square was randomized on each trial. Participants judged the similarity between the colors in each pair by using a 10-point scale (1 = *very dissimilar*, 10 = *very similar*) and were urged to use the full range of ratings when making their judgments.

Results and Theoretical Analysis

Multidimensional scaling. MDS solutions for the colors were derived by fitting the INDSCAL model (Carroll & Wish, 1974) separately to each individual participant's similarity ratings. The data input to the INDSCAL program were the averaged similarity rating matrices obtained for each participant on each of the 3 days of testing. Figure 4 displays the two-dimensional scaling solutions obtained for each participant. These scaling solutions accounted for an average of 95.9, 93.6, and 93.9% of the variance in the similarity ratings of Participants 1, 2, and 3, respectively. We judged these fits to be reasonably good and decided to use the two-dimensional solutions for purposes of fitting the EBRW to the speeded classification data.[4]

Speeded classification. The mean response time with which each participant classified the colors is shown as a function of grouped blocks of practice in Figure 5. Each grouped block corresponds to five blocks of training (60 trials).[5] The speedups in performance are well described by power-law functions, as illustrated in the figure. Power-law speedups as a function of practice are well documented in various domains of skilled performance (e.g., Anderson & Fincham, 1994; Logan, 1992; Newell & Rosenbloom, 1981). To our knowledge, however, this study may be the first to document a similar form of facilitation in the domain of speeded multidimensional perceptual classification. Estes (1994, p. 103) also reported speedups in classification performance with increased trials of training; however, the amount of training in his study was limited to a single day and the data were averaged over participants. Because the differences in response speed in Figure 5 are relatively small across Days 2–5 (Blocks 31–150), we decided to use these latter data when computing response times for individual stimuli in our subsequent modeling analyses.

The mean response times and accuracies for the individual stimuli across Days 2–5 are reported in Table 1. As can be seen, the accuracies were close to ceiling, so our analyses focus on the overall mean response times. (In Appendix C we present model-based analyses in which the EBRW is fitted simultaneously to the response time and accuracy data, with reasonable

[4] Improved fits to the similarity ratings data can be achieved by using various methods. First, one can increase the dimensionality of the scaling solutions. Adding a third dimension, for example, increases the percentage of variance accounted for to 97.1, 95.6, and 95.2 for Participants 1, 2, and 3, respectively. Also, instead of using a deterministic scaling solution, in which each color is represented as a single point in the space, one can use probabilistic scaling solutions, in which each color is represented as a probabilistic distribution of points in the space (e.g., Ashby, 1992; Zinnes & MacKay, 1983). Although both methods improve the fit of the scaling model to the similarity data, they do so at a cost of adding extra free parameters. These parameters may be unreliable and can potentially lead one astray when predicting some independent set of data, such as our speeded classifications. Furthermore, adding extra dimensions or using probabilistic representations increases the complexity of fitting the EBRW to the independent set of speeded classification data.

[5] The only exception is the first group, in which we have deleted Block 1. In this block, each color exemplar and its corresponding category feedback were presented for the first time.

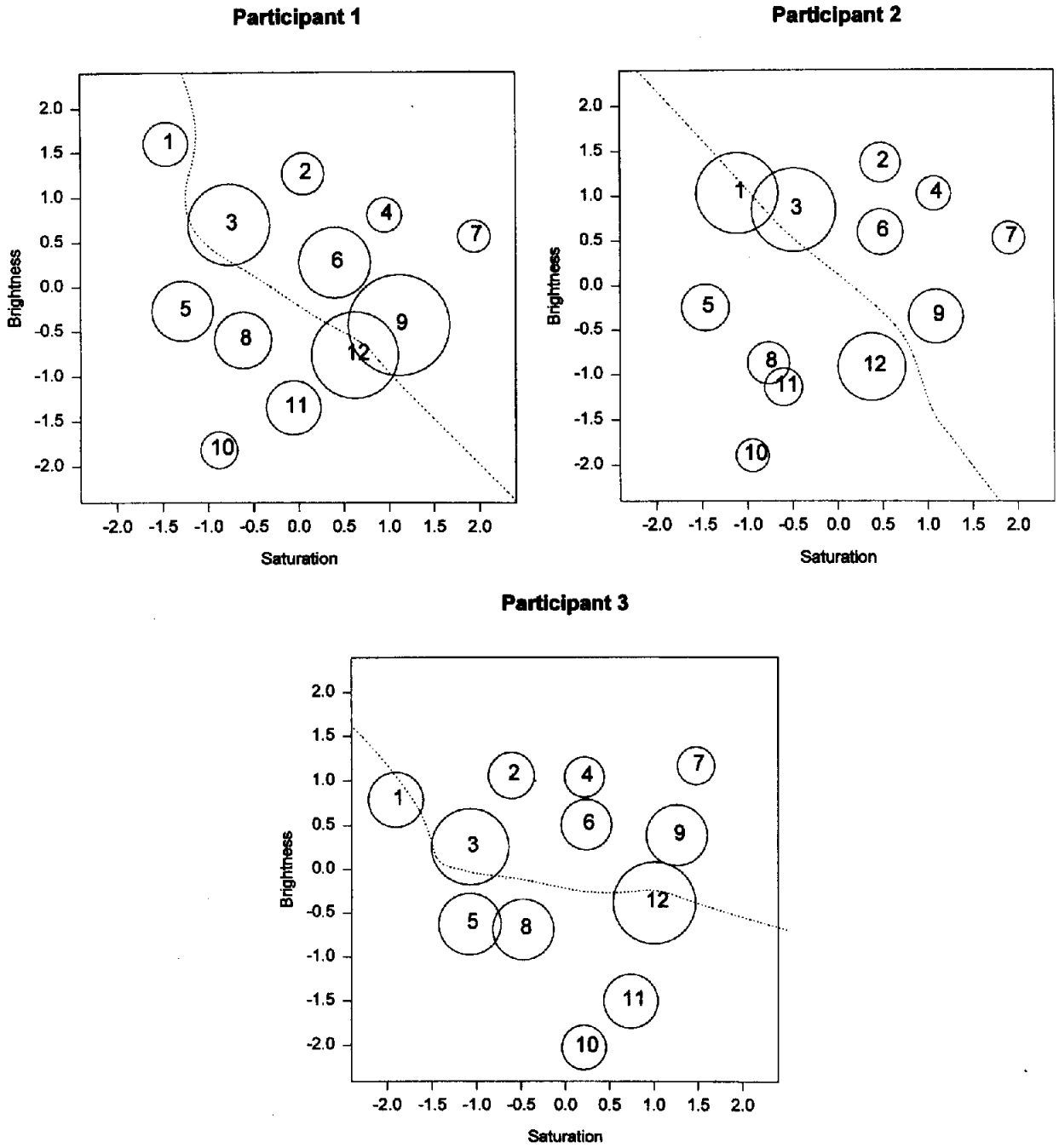

Figure 4. Individual-participant multidimensional scaling (MDS) solutions for the color stimuli used in Experiment 1. The center of each circle represents the MDS coordinate for the color. The diameter of each circle is linearly related to the mean classification response time observed for the color in the speeded classification task. The dashed curve is the boundary of equal summed similarity to the exemplars of each category, as computed in the GCM.

success.) The response times are illustrated graphically in the MDS solutions shown in Figure 4. In the figure, the diameter of the circle enclosing each stimulus is linearly related to the mean response time. The dashed curve in each figure illustrates the exemplar-based *category boundary* predicted by the GCM. (This boundary does not enter into the modeling analyses; it is illustrated simply for descriptive convenience.) Each point on the boundary has equal summed similarity to the exemplars of

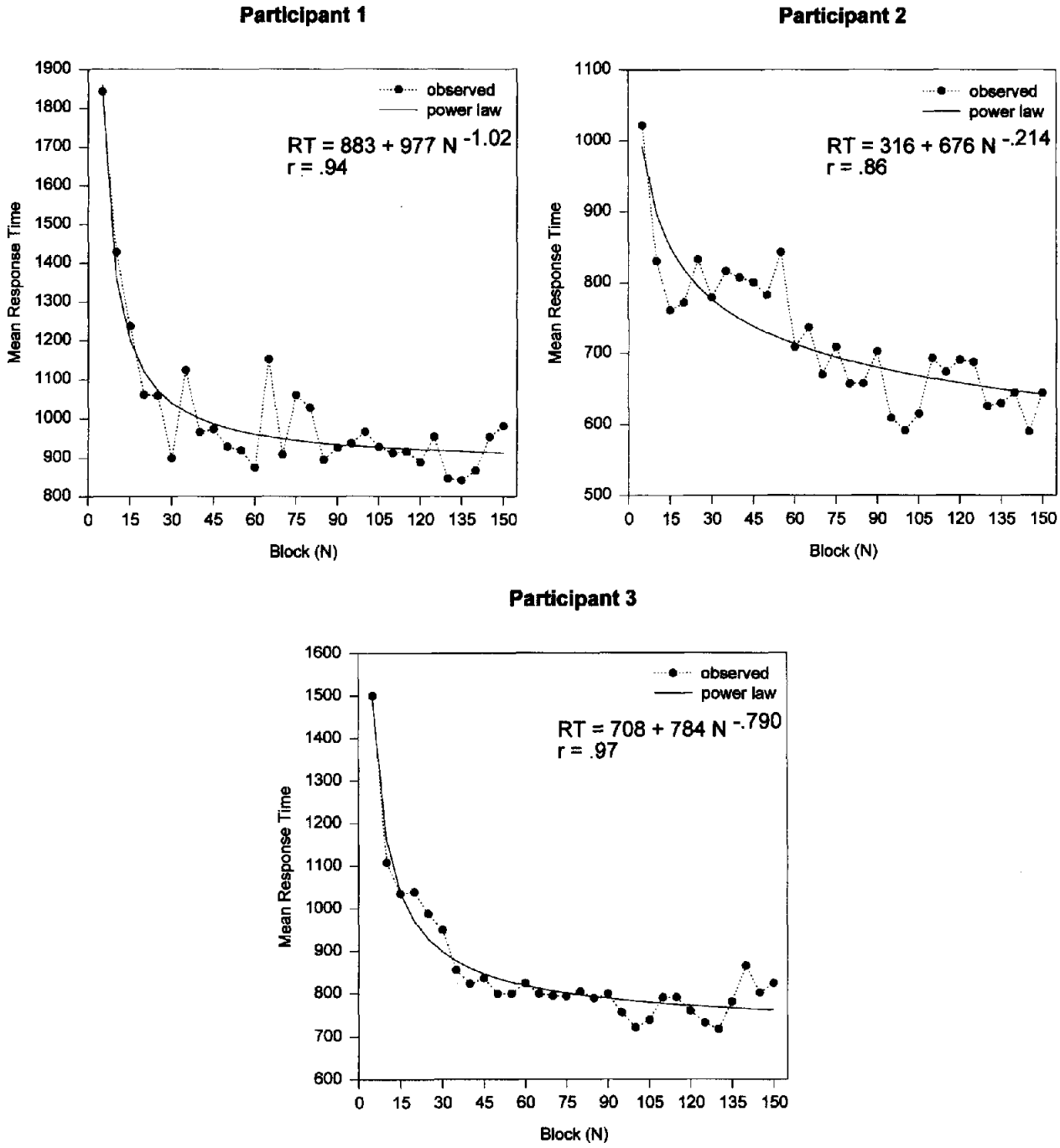

Figure 5. Mean response time (ms) to the colors for each grouped block of 60 trials in Experiment 1. Also shown is the best-fitting power function and the correlation with the observed data. RT = response time.

Category A and Category B. Points falling to the upper right of the boundary have greater summed similarity to Category A, whereas points falling to the lower left have greater summed similarity to Category B. It is evident from inspection that the response times for the individual stimuli are highly regular: In general, the greater the distance of a stimulus from the exemplar-based boundary, the faster is the response time.

Fits of the EBRW. Using the previously derived MDS solutions, we fitted the EBRW simultaneously to the mean response times observed for each individual color and to the aggregated

mean response times observed as a function of grouped blocks of practice. In fitting the model, we assumed that on each block an additional token of each individual exemplar was stored in memory. The first step in the analysis was to use the model to predict the mean response time for each color in each individual block. Then, for each exemplar, these individual block predictions were averaged over Blocks 31–150 (Days 2–5) to obtain the overall predicted mean response times (i.e., to predict the data in Figure 4). Likewise, we averaged the predicted mean response times over all exemplars in each grouped block of practice to predict the speedup curves in Figure 5. A single set of parameters was used to simultaneously predict both sets of data. The free parameters were the sensitivity parameter c, the attention weight w_1 (with $w_2 = 1 - w_1$), the step-time constant α, a single category criterion A (with B = A), and the regression parameters k and μ_R. The model was fitted by minimizing the total sum of squared deviations between predicted and observed mean response times across both data sets (42 data points). The best-fitting parameters and summary fits for each of the three participants are reported in Table 2.

Table 1
Classification Accuracy, Response Time (ms), and Standard Errors for Each Stimulus Averaged Over Sessions 2–5 in Experiment 1

Stimulus	p(C)	M	MSE
	Participant 1		
1	1.000	815	18.1
2	.983	795	17.4
3	.917	1,159	31.7
4	1.000	725	12.4
5	1.000	967	25.6
6	.975	1,068	31.2
7	1.000	706	13.4
8	.975	931	24.8
9	.950	1,346	41.4
10	1.000	744	16.2
11	.992	911	27.7
12	.967	1,208	43.5
	Participant 2		
1	.917	982	50.4
2	1.000	601	11.4
3	.992	1,007	35.8
4	1.000	546	5.9
5	1.000	665	13.0
6	1.000	649	12.2
7	1.000	529	4.5
8	.992	619	14.9
9	.983	734	23.9
10	1.000	530	8.0
11	1.000	577	16.4
12	.950	857	36.6
	Participant 3		
1	.975	780	14.0
2	1.000	709	8.9
3	.956	962	21.8
4	1.000	660	12.0
5	1.000	841	18.9
6	1.000	748	13.6
7	1.000	640	7.7
8	.984	834	20.8
9	.984	834	20.4
10	1.000	697	12.2
11	1.000	779	12.7
12	.934	1,007	25.5

Note. p(C) = probability correct; M = mean response time (all responses); MSE = mean standard error.

Table 2
Best-Fitting Parameters of the Exemplar-Based Random Walk Model, Experiment 1

	Participant		
Parameter	1	2	3
c	2.998	2.295	2.742
w_1	0.450	0.555	0.716
α	0.105	0.136	0.045
A	3	4	6
k	9,411.75	3,271.92	3,518.22
μ_R	100.06	153.28	442.28

Note. c = sensitivity parameter; w_1 = attention weight on Dimension 1; α = step-time constant; A = random walk criterion; k = scaling constant; μ_R = mean residual response time.

The model-fitting results for the individual color RTs are shown in Figure 6. The figure plots, separately for each participant, the observed mean response times for each of the individual colors against the predicted mean response times. The obtained correlations were .886, .990, and .949 for Participants 1, 2, and 3, respectively. We consider these reasonably good fits to provide support for the EBRW.[6] The model predicts these results because objects far from the exemplar-based boundary tend to be similar only to exemplars from their own category and not to exemplars from the contrast category; thus, the random walk marches in consistent fashion to the appropriate response criterion. Objects close to the boundary are similar both to exemplars from their own category and to exemplars from the contrast category; thus, exemplars from both categories tend to be retrieved, and the random walk wanders back and forth.

The fits of the EBRW to the speedup curves are shown for each individual participant in Figure 7. The correlations between predicted and observed RTs were .938, .776, and .963 for Participants 1, 2, and 3, respectively. The fits for Participants 1 and 3 are essentially the same as those achieved by the descriptive power-law functions, and the fit for Participant 2 is only slightly worse (see Figure 5). Given that we are dealing with individual-

[6] We should emphasize that there are only three effective free parameters involved in achieving these correlations, namely, c, w_1, and α. Except for extremely small values of the category criterion (A), the correlations remain essentially unchanged. Likewise, the regression parameters k and μ_R do not influence the correlations; they serve only to rescale the predictions of the EBRW, which are in arbitrary units, into milliseconds. Note also that markedly improved fits to Participant 1's data can be achieved with the addition of an extra free parameter (see Appendix C).

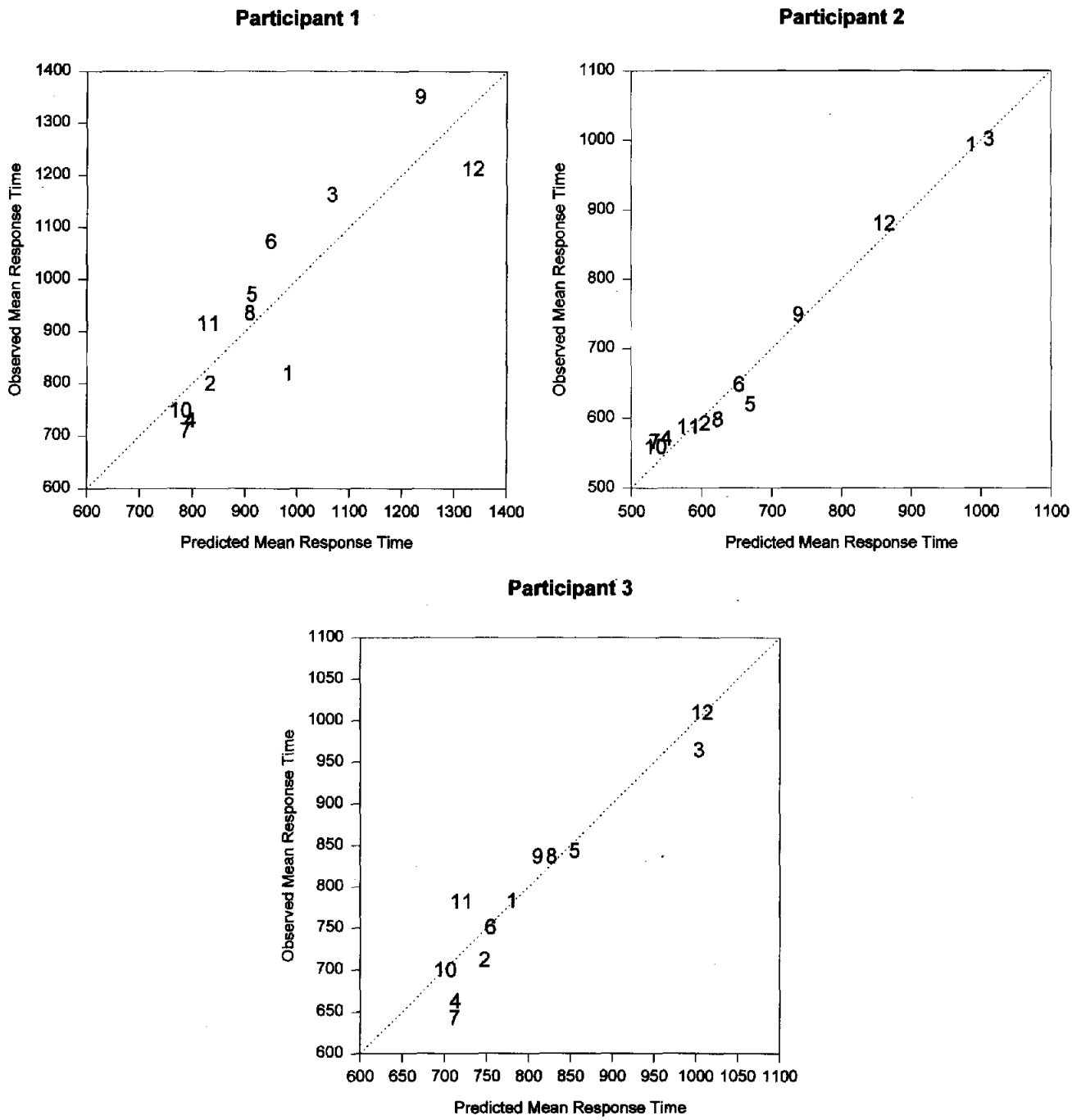

Figure 6. Scatterplot of observed mean response times (ms) for each individual color against the predicted mean response times from the exemplar-based random walk model, Experiment 1.

participant data and that there are bound to be large sources of uncontrolled noise from day to day, we consider the fits of the process model to be reasonably good. (Its main shortcoming is that Participant 2's response times during Day 2, Blocks 31–60, are slower than predicted.) The EBRW predicts these speedup curves for essentially the same reason as in Logan's (1988) model: As practice continues, an increased number of exemplars are stored in memory. The greater the number of exemplars that race to be retrieved, the faster are the winning retrieval times. Thus, the individual steps in the random walk occur more rapidly with increased training.

RT–distance model. As an initial source of comparison for

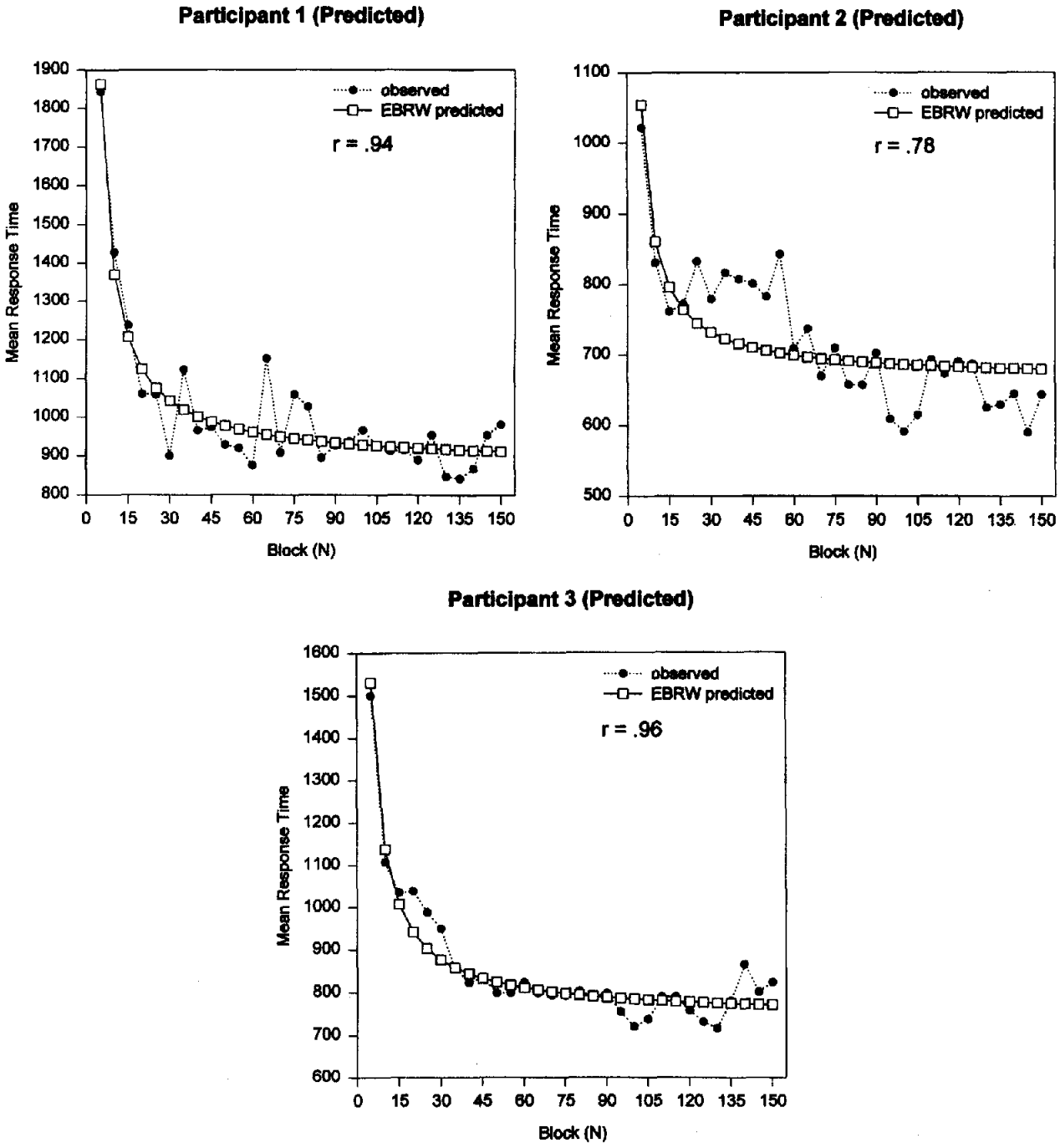

Figure 7. Mean response times (ms) to the colors predicted by the exemplar-based random walk (EBRW) model as a function of grouped blocks of practice in Experiment 1.

the EBRW, we also fitted a version of the RT–distance model to the mean response time data for the individual colors. Because there exists an infinite variety of decision boundaries, it is not possible to test the model in its general form. Following a suggestion from F. G. Ashby (personal communication, October 1995), we assumed that the decision boundary had the same form as the one produced by the EBRW, namely, the locus of points with equal summed similarity to the exemplars of Category A and Category B. This boundary has particular theoretical importance: Given certain parametric assumptions, the boundary of equal summed similarity is the optimal boundary for partitioning the space into categories, in the sense that it maximizes

observers' percentage of correct classification choices (Ashby & Maddox, 1993; Nosofsky, 1990). Following Murdock (1985) and Maddox and Ashby (1996), we further assumed that the mean decision time for classifying an object was given by

$$T = \exp(-\beta \cdot D), \tag{22}$$

where D represents distance-from-boundary, and the β parameter determines the rate at which decision time decreases with distance. The exponential assumption formalized in Equation 22 is the version of the RT–distance model that Maddox and Ashby (1996) found provided the best overall quantitative fits to their sets of speeded classification data. The free parameters in the RT–distance model were the parameter β in Equation 22; regression constants k and μ_R for transforming the predicted decision times into milliseconds; and the parameters c and w_1 that determine the exact shape of the equal summed similarity boundary.[7]

The RT–distance model yielded essentially the same correlations between predicted and observed mean RTs as did the EBRW: .903, .982, and .976 for Participants 1–3, respectively. When the EBRW is fitted to the individual color RT data without the constraint of also fitting the speedup curves, its correlations are .915, .991, and .956 for Participants 1–3, respectively. The fits of the EBRW to the RT data remain essentially the same over a wide range of values for the criterion parameter, A. The value of this parameter exerts an influence only on the accuracy predictions of the model. Thus, the versions of the EBRW and the RT–distance model tested here use the same number of effective parameters.

In sum, the EBRW yields correlations with the data that are as good as those provided by an important representative from the class of RT–distance models. As this class of models provides what is currently the best descriptive approach to modeling multidimensional classification response times, the EBRW appears to be performing fairly well. An important advantage of the EBRW is that it provides an explicit process account of *how* participants might learn decision boundaries with an optimal form, and *why* response time is generally a decreasing function of distance from this boundary. The EBRW also simultaneously accounts for the roughly power-law decreases in mean response time observed as a function of practice, a phenomenon not addressed by current versions of the RT–distance model.

Discussion

In this experiment we demonstrated that the EBRW, when used in conjunction with derived MDS solutions, yielded good quantitative fits to mean classification response times observed for individual objects. The model simultaneously characterized the roughly power-law speedups in classification response speed that were observed as a function of practice. Adding to the impressiveness of these results is that, in all cases, the model was fitted to individual-participant data.

Other factors besides increased exemplar storage may influence the development of response speed as a function of practice. For example, Nosofsky (1987) found evidence of increases in overall sensitivity (the value of c) as a function of classification learning. Increases in c would lead to a speedup because it would increase the tendency for test items to retrieve only their own memory traces, yielding a more consistent walk to the appropriate response criterion. On the other hand, decreases in memory strength due to decay and interference would lower the activation rates with which the exemplars race, thereby slowing the random walk. The baseline version of the EBRW tested in this article provides a parsimonious explanation of the speedup effect, but future research should explore the influence of these other factors.

Experiment 2: The Role of Individual Object Familiarity in Speeded Classification

A fundamental prediction of the EBRW is that large differences in individual object familiarity should influence classification response time. Objects that are unfamiliar should result in relatively slow retrieval times for stored exemplars, thereby slowing the random walk. In Experiment 2 we sought to manipulate individual object familiarity in a speeded classification task while holding fixed the distance of objects from a presumed classification decision boundary. Effects of familiarity on classification RT would thereby support the predictions of the EBRW while placing strain on the RT–distance hypothesis as the sole explanation of classification response time.

The design of Experiment 2 is illustrated schematically in Figure 8. There were eight color stimuli residing in a two-dimensional space. Stimuli 1–3, enclosed by circles, belong to Category 1; whereas Stimuli 4–8, enclosed by squares, belong to Category 2. Stimuli 1–6 were always presented during training, but presentations of Stimuli 7 and 8 were manipulated. In Condition U7 (unfamiliar–7), Stimulus 7 was never presented during training; whereas in Condition U8, Stimulus 8 was never presented during training. Following the initial training phase, a speeded classification test was conducted. Each of the eight colors was presented in a random order in each of eight blocks of testing. The central comparison of interest concerns classification response times for Stimuli 7 and 8. The EBRW predicts that in Condition U7, Stimulus 7 should be classified more slowly than Stimulus 8 but that the reverse should occur in Condition U8.

The RT–distance hypothesis has little basis for predicting differential response times for Stimuli 7 and 8 across conditions. It is possible to estimate an infinite variety of different decision boundaries, and some set must certainly yield high correlations between distance and RT across the conditions. The critical question, however, is what the decision boundary theory predicts a priori. As discussed earlier, most of the a priori predictions

[7] Because there is no analytic method for computing the distance between a point and the boundary of equal summed similarity, we followed a brute force approach. First, we conducted a computer search to find 200 evenly spaced points that spanned the coordinate space illustrated in Figure 4 and that had equal summed similarity to the exemplars of Category A and B, given the parameters c and w_1. Next, we computed the distance between a given exemplar and each of these 200 points. The minimum of these 200 distances was the measure of distance-from-boundary.

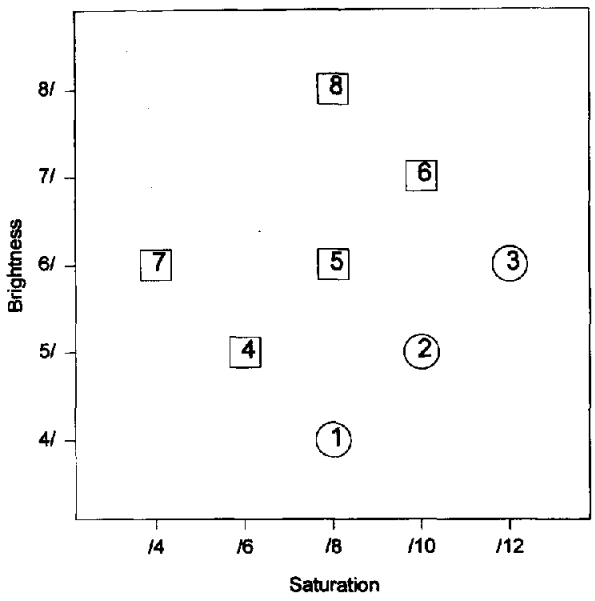

Figure 8. Schematic illustration of the approximate Munsell color configuration used in Experiment 2. Color 7 was not presented during training in Condition U7, whereas Color 8 was not presented during training in Condition U8.

stemming from the decision boundary theory assume that participants adopt optimal boundaries, or at least boundaries that are close to optimal in form (Maddox & Ashby, 1993). Under a variety of plausible assumptions, it is possible to show that the optimal decision boundary for separating the exemplars of Categories 1 and 2 is essentially identical, regardless of whether Stimulus 7 or Stimulus 8 is presented during training. The key intuition here is that Stimuli 1–6, which lie close to the presumed boundary and are relatively difficult to classify, carry the lion's share of the weight in determining the form of the optimal boundary. Stimuli 7 and 8, which lie far from the boundary and are easy to classify, have a minuscule influence. Thus, the pure distance-from-boundary hypothesis predicts no difference in classification RTs between Stimuli 7 and 8 across conditions, whereas the EBRW predicts a crossover interaction based on the familiarity manipulation.

Whereas in Experiment 1 we tested a few individual participants over extended periods of time, in Experiment 2 we examined multiple participants in a single training session. To test for effects of familiarity, multiple observations need to be obtained for the critical transfer stimuli. Repeatedly presenting the same novel stimulus at test, however, would soon render that stimulus familiar. Thus, we needed to rely on data averaged over multiple participants, each of whom experienced the unfamiliar stimuli with limited frequency.

In addition to the speeded classification test, an independent group of participants provided similarity judgments for all pairs of colors in the set. These similarity judgments were used to derive an MDS solution for the colors. The goal was to use the MDS solution in conjunction with the EBRW to generate quantitative predictions of the classification response times across the different familiarity conditions.

Method

Participants. In the speeded classification experiment, there were 40 participants in Condition U8 and 37 participants in Condition U7. There were 43 participants in the similarity scaling study. Participants were undergraduates from Indiana University who received credit toward an introductory psychology course requirement. Small bonuses were paid for good performance. All participants claimed to have normal color vision.

Stimuli. The stimuli were a set of eight colors generated on Compu-Add monitors. Extensive pilot work was used to construct a set of colors that approximately matched the Munsell configuration illustrated in Figure 8. The RGB values used for generating the stimuli are reported in Appendix B. Other aspects of the stimuli and apparatus were the same as in Experiment 1, except the colors were displayed against a medium gray background rather than a white background.

Procedure. In the first part of the experiment, an unspeeded category training phase was conducted. Stimulus presentations were organized into 15 blocks. Each training stimulus was presented once per block in a random order. Color 7 was never presented during training in Condition U7, and Color 8 was never presented during training in Condition U8. In each respective condition in which Colors 7 and 8 were familiar, they were presented once per block along with all of the standard training stimuli. The stimulus displays were response terminated. Category feedback was provided for 2 s following each response. There was a 500-ms ISI.

Following the training phase, a speeded classification transfer phase was conducted. Participants were instructed to classify each stimulus as rapidly as possible without making errors. The transfer phase consisted of two parts. In the first part, only the original training stimuli were presented. There were four blocks of transfer trials, with each stimulus presented once in a random order in each block. These blocks were used to give participants practice in making speeded classification judgments. In the second part, both the training stimuli and the unfamiliar transfer stimulus were presented. There were eight blocks of trials, with each stimulus presented once in a random order in each block. There was no break between the first and second parts of the speeded classification transfer phase. Feedback continued to be presented during Part 1 but was withheld during Part 2 of the transfer phase. Other aspects of the procedure were the same as in Experiment 1.

The similarity scaling study was organized into 10 blocks of trials. Each of the 28 unique pairs of the 8 stimuli was presented once per block in a random order. Thus, there were 10 repetitions of each pair and a total of 280 trials in the experiment. Other aspects of the procedure were the same as in Experiment 1.

Results and Theoretical Analysis

Similarity scaling. The INDSCAL model was used to analyze the similarity ratings to derive an MDS solution for the colors. The data input to the program were the averaged similarity rating matrices obtained for each individual participant. The two-dimensional solution, illustrated in Figure 9, accounted for an average of 89.2% of the variance in the 43 individual participants' similarity ratings. The obtained configuration does not correspond precisely to the planned schematic design, mainly because Color 3 is more isolated than anticipated (compare to Figure 8). Nevertheless, the obtained configuration is sufficient to satisfy the goals of the experiment. In particular, consider the

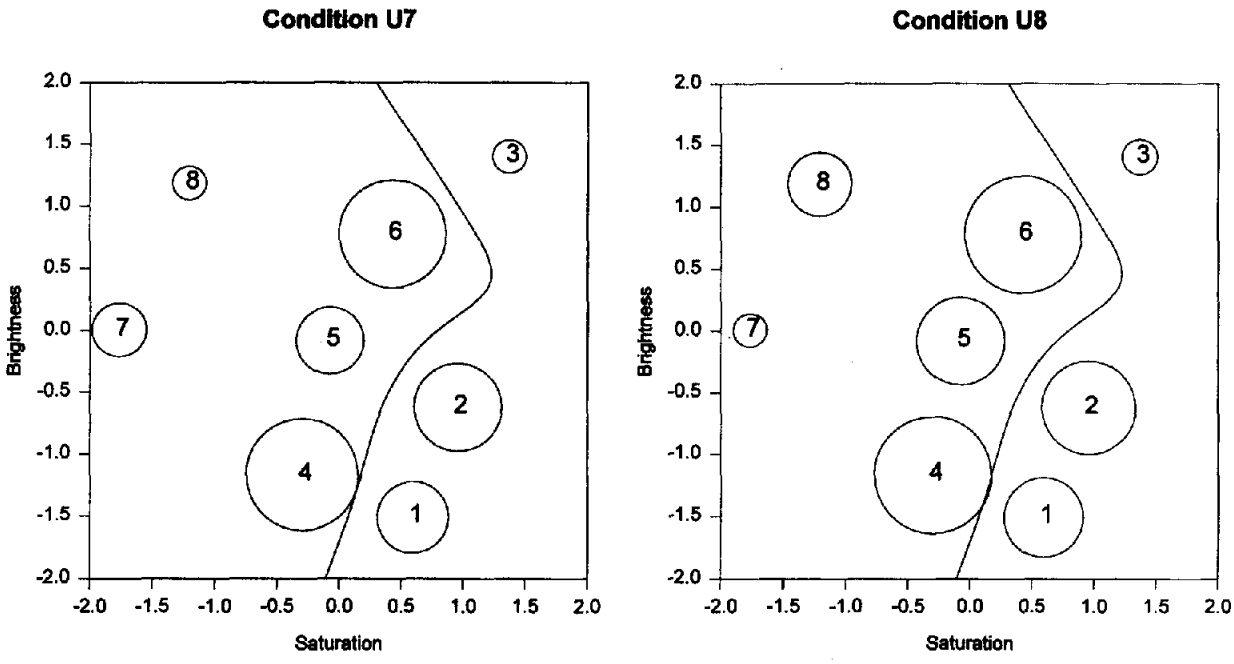

Figure 9. Group multidimensional scaling (MDS) solution for the color stimuli used in Experiment 2. The center of each circle represents the MDS coordinates for the color. The diameter of each circle is linearly related to the mean classification response time observed for the color in each condition (U7 and U8) of the speeded classification task. The solid curve in each panel is the optimal likelihood-based boundary for each familiarity condition (see Appendix D).

optimal decision boundaries (Figure 9). Although these optimal boundaries are highly nonlinear, they are still virtually identical across Conditions U7 and U8, so the RT–distance model has little basis for predicting differential response times based on the familiarity manipulation. The optimal linear boundaries are also virtually identical across conditions, so this version of the RT–distance model also does not predict an effect of familiarity. The methods of deriving the optimal boundaries are discussed in Appendix D.

Speeded classification. The data obtained during Part 1 of the speeded classification transfer phase were analyzed to select participants for inclusion in the theoretical modeling. Because we were interested in modeling only those participants who were highly motivated and performed relatively well, we set a criterion of 80% correct or better during this phase. This criterion led to the removal of 9 participants in Condition U8 and 6 participants in Condition U7. Thus, there were 31 remaining participants in each of the conditions.

The mean correct response times and accuracy rates for each of the stimuli during Part 2 of the speeded classification transfer phase are reported in Table 3. The response times are illustrated graphically in Figure 9, where the diameter of the circle enclosing each stimulus is linearly related to the RT. As expected, Colors 7 and 8, which lie far from the boundary, tended to be classified more rapidly than Colors 1–6, which lie close to the boundary (except for Color 3). The most important result, however, is that Color 7 was classified more rapidly than Color 8 in Condition U8, but the reverse was observed in Condition

Table 3
Classification Accuracy, Response Time (ms), and Standard Errors for Each Stimulus in Conditions U7 and U8 in Experiment 2

Stimulus	*p*(C)	*M*	*MSE*
	Condition U7		
1	.964	750	17.7
2	.927	794	23.8
3	.992	648	11.1
4	.891	859	25.4
5	.988	740	19.5
6	.932	846	26.7
7	.944	703	20.1
8	.996	648	15.9
	Condition U8		
1	.948	795	25.5
2	.948	834	25.4
3	.980	677	16.4
4	.895	897	32.8
5	.972	819	28.0
6	.883	896	31.7
7	.984	672	19.9
8	.923	752	29.3

Note. *p*(C) = probability correct; *M* = mean correct response time (ms); *MSE* = mean standard error.

U7. This observation is confirmed by statistical test. A two-way analysis of variance on the Colors 7 and 8 RT data using Condition (U7 vs. U8) and Stimuli (7 vs. 8) as variables revealed a significant interaction, $F(1, 60) = 9.58$, $MSE = 36{,}290.9$, $p < .005$. This large effect of familiarity is consistent with the predictions of the EBRW and suggests that current versions of decision boundary theory do not provide the sole explanation of classification response time.

Our expectation was that the largest effects of the familiarity manipulation would be seen early in the transfer phase. Once a novel stimulus is classified repeatedly during transfer, it is no longer unfamiliar. Nosofsky (1986), for example, suggested that through internal feedback, novel stimuli that are classified repeatedly during transfer are used to augment the original category representation established at the end of training. The mean correct response times for Colors 7 and 8 are plotted as a function of blocks of transfer in Figure 10. To remove noise from the data, we plot the response times averaged over the conditions in which each stimulus was familiar or unfamiliar. It is clear from inspection that our expectation was confirmed: The largest effects of the familiarity manipulation occurred during the first four transfer blocks. Nevertheless, effects are seen even during Blocks 5–8.

EBRW fits. Our next goal was to quantitatively model the classification response times in terms of the EBRW. Because response times are highly variable, we decided to model the data averaged over all eight blocks of transfer (Table 3). The EBRW was fitted to the data by searching for the free parameters that minimized the sum of squared deviations between predicted and observed response times for each of the individual colors across the two conditions. (Attempts to jointly fit both the response time and accuracy data are reported in Appendix C.)

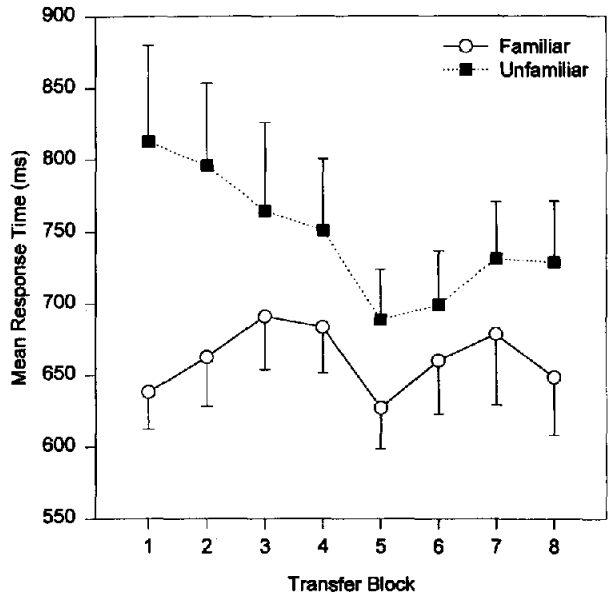

Figure 10. Mean response time (ms) for the critical familiar and unfamiliar test stimuli as a function of blocks of speeded transfer in Experiment 2.

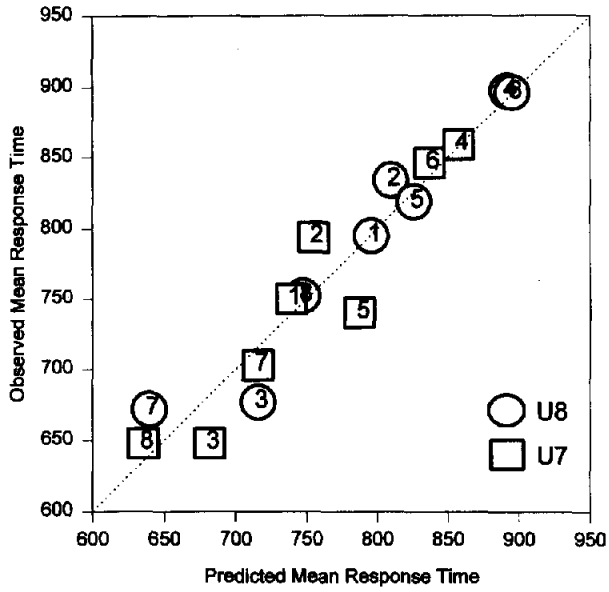

Figure 11. Scatterplot of observed response times (ms) for each color against the predicted response times from the exemplar-based random walk model, Experiment 2. Condition U7 is depicted by squares; Condition U8 is depicted by circles.

The free parameters used for fitting the data were the sensitivity parameter c, the attention-weight w_1, the step-time constant α, the category criteria A and $-$B, and the regression constants k and μ_R. Because overall response times in Condition U7 were somewhat shorter than in Condition U8, we allowed separate sensitivity parameters across the two conditions.

The observed mean response times are plotted against the predicted mean response times in Figure 11. The best-fitting parameters were $c_{U7} = 2.11$, $c_{U8} = 1.90$, $w_1 = .577$, $\alpha = 2.21$, A = 3, B = 4, $k = 40.34$, and $\mu_R = 244.41$. The EBRW achieved a correlation of .959 with the observed data. The model predicts correctly the major qualitative effects of interest. In particular, the model predicts correctly that Colors 7 and 8 are classified more quickly than Colors 1–6 and also predicts correctly the fast response times for Color 3. Most important, the EBRW predicts correctly the fundamental crossover interaction resulting from the familiarity manipulation: Color 7 is classified more quickly than Color 8 in Condition U8, but the reverse occurs in Condition U7.

RT–distance model. The RT–distance model was also fitted to the response time data, using the same assumptions and methods described in Experiment 1. The model yielded a correlation with the data of only .60. An extended version of the model that allowed separate β parameters across Conditions U7 and U8 and that allowed for differential category response bias improved the fit to .84, which is still far worse than the fit yielded by the EBRW ($r = .956$). The main shortcoming of the pure version of the RT–distance model is that it has no mechanism for predicting the large effect of the familiarity manipulation on the response times for Colors 7 and 8.

Discussion

In this experiment we verified the fundamental qualitative prediction of the EBRW that large differences in familiarity should influence classification response time. Holding essentially fixed the distance of two critical transfer stimuli from a theoretically optimal decision boundary, we observed faster classification response times when a transfer stimulus was familiar than when it was unfamiliar. The EBRW also yielded good quantitative predictions of the classification response times for each of the individual stimuli across the different familiarity conditions. Nosofsky (1991b) provided previous evidence that familiarity affects speeded classification response times. In this previous experiment, however, only a single dimension was relevant for classifying the objects, and the unfamiliar stimuli were several standard deviations further away from the decision boundary than any other objects experienced during training. Thus, the present study lends generality to the conclusion that individual object familiarity can play an important role in speeded perceptual classification. Furthermore, it demonstrates that the EBRW is capable of quantitatively predicting the joint effects of distance-from-boundary and familiarity on classification response time.

One question that arises is whether the familiarity effect in this experiment may have reflected a pure "surprise" effect. Of course, the EBRW predicts that the unfamiliar stimuli will be "surprising." However, it posits that the main locus of the surprise effect is in slowing processing during the classification decision-making stage. In a follow-up study intended to address this question, we used the same training conditions as in the present experiment; however, rather than requiring participants to classify the colors during the speeded test phase, they were required simply to detect the colors. An exponentially distributed random foreperiod was used on each trial so that participants would be unable to simply time their responses without actually detecting the colors. In all other respects, the two experiments were identical. Furthermore, roughly the same number of participants were tested. Whereas the familiarity effect in the present classification experiment averaged 67.5 ms during the eight speeded test blocks, it averaged only 7.5 ms in the detection experiment and did not approach statistical significance. Although pure surprise may have played some role in the present classification experiment, the minuscule effect in the detection experiment supports our interpretation that lack of familiarity mainly slowed classification decision making, as predicted by the EBRW. Moreover, it should be remembered that in the present classification experiment, the familiarity effect lasted throughout the entire eight blocks of transfer, which places further strain on the pure surprise hypothesis.

Although extant versions of the decision boundary theory (Ashby et al., 1994; Ashby & Maddox, 1994) do not predict the familiarity effects a priori, various approaches may exist for extending this alternative modeling framework. For example, perhaps instead of assuming optimal decision boundaries, mechanisms may be proposed for how decision boundaries are learned. Depending on which stimuli are experienced during training, decision boundaries with different forms may arise. Alternatively, perhaps the psychological distributions used for representing each stimulus in the decision boundary theory are affected by the familiarity manipulations. We leave these interesting possibilities as issues for future research.

Experiment 3: Modeling Performance in Garner's (1974) Speeded Classification Tests

One of the classic paradigms involving speeded perceptual classification performance is the set of converging tests introduced by Garner and his colleagues for distinguishing between integral and separable dimensions (Garner, 1974, 1976; Garner & Felfoldy, 1970; Gottwald & Garner, 1975; Pomerantz & Garner, 1973). The set-up of the paradigm is illustrated schematically in Figure 12. There are four stimuli varying along two dimensions, with two values per dimension, and the dimension values are roughly equally discriminable. We start by considering three types of tasks of major interest. In all of the 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 is presented that vary along just one dimension. An example is to classify Stimulus A into Category 1 and Stimulus B into Category 2 (see Figure 12). 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 values on the irrelevant dimension are held constant. In the *filtering* task, any one of the four stimuli in the complete set is presented on each trial. 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 a

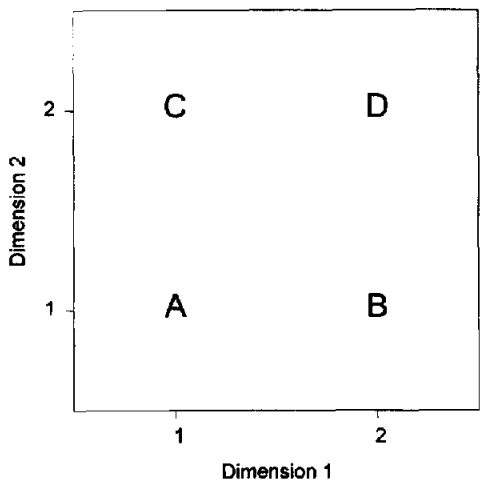

Figure 12. Schematic illustration of the stimulus set used in Garner's (1974) speeded classification tasks.

review). To a first approximation, for highly separable dimension stimuli, response times are basically identical across the control, filtering, and correlated tasks. However, for highly integral dimension stimuli, there is marked interference in the filtering task and marked facilitation in the correlated task. In other words, for integral dimension stimuli, response times in the filtering task are slower than in the control task, whereas response times in the correlated task are faster than in the control task.[8] The interference in the filtering task is interpreted as a failure of selective attention. Intuitively, optimal performance in the filtering task is achieved if the observer attends only to the relevant dimension and ignores variations along the irrelevant dimension. People can attend selectively to integral dimensions to some degree, but the process is far less efficient than occurs for separable dimension stimuli (Foard & Kemler Nelson, 1984; Garner, 1974; Lockhead, 1972; Melara & Marks, 1990; Nosofsky, 1987; Shepard & Chang, 1963).

The results of these speeded classification tasks are among the fundamental converging operations used by Garner (1974, 1976) for distinguishing among alternative dimensional interactions. Until recently, however, no formal theoretical foundations have been offered for why these performance patterns emerge. Ashby and Maddox (1994) recently proposed some theoretical foundations by combining the RT–distance hypothesis with concepts from the general recognition theory of Ashby and Townsend (1986). Maddox and Ashby (1996) successfully fitted models derived from the theory to response time distributions and accuracy data in control, filtering, and correlated tasks involving separable dimension stimuli. In this section, we ask whether the main patterns of results obtained in these speeded classification tasks are consistent with the predictions of the EBRW. We then conduct our own experiment involving these tasks to provide initial tests of the model's ability to quantitatively fit the data.

Applying the EBRW to explain the pattern of results for separable dimension stimuli is straightforward. We assume that participants attend selectively to the single dimension that they are instructed is relevant for classifying the objects. In the extreme case in which all attention weight is placed on this single dimension (Equation 1), the two-dimensional space collapses onto a one-dimensional space, and the structures of the control, filtering, and correlated tasks are identical.

Regarding integral dimension stimuli, why is there facilitation in the correlated task? 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 12). According to the EBRW, 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 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, 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. (For example, 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. The greater the number of these exemplars in memory, the faster the winning retrieval times tend to be, so 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, identical exemplars are presented on only one fourth of the immediately preceding trials. Because the activation rates with which exemplars race are influenced by their memory strengths (Equations 3 and 5), the retrieval of identical exemplars is more efficient in the control task than in the filtering task. Introducing these assumptions about reductions in memory strength also allows the EBRW to explain stimulus repetition effects in the Garner tasks, a point we elaborate on in the *Discussion* section.

In Experiment 3 we collected a battery of response time data in the control, filtering, and correlated tasks by using integral dimension stimuli. To develop further constraints for model fitting, we also tested participants in *stretch-filtering* and *condensation* tasks. In the stretch-filtering task, the amount of variation along the irrelevant dimension is increased. The EBRW predicts somewhat more interference in the stretch-filtering task than in the standard filtering task. Because of its similarity-based retrieval assumptions, the EBRW predicts that, with some probability, a test item will retrieve exemplars to which it is not identical. For example, if the test item is A, Exemplars B and C will be retrieved with some probability. Retrieving an exemplar from the same category moves the random walk counter in the correct direction, whereas retrieving an exemplar from the opposite category moves the counter in the wrong direction. The probability of beneficial same-category retrievals decreases as within-category similarity decreases. Thus, performance in the stretch-filtering task should be worse than in the standard filtering task. This prediction was confirmed recently in experiments by Melara and Mounts (1994).

In the condensation task, Stimuli A and D are classified into one category, and Stimuli B and C are classified into the opposite category (see Figure 12). It is well-known that performance in the condensation task is very poor. Note that both dimensions are needed for purposes of classification in this task, so selective attention to a single dimension is not possible. Furthermore, any given test item is similar to two exemplars from the opposite

[8] Our summary statement regarding separable dimension stimuli is intended only as a first approximation for describing the general pattern of results. For example, facilitation effects in the correlated task are also sometimes observed for separable dimension stimuli, especially under conditions of low discriminability or extensive training (e.g., Garner & Felfoldy, 1970; Maddox & Ashby, 1996).

category. Thus, as will be seen, the EBRW predicts correctly the poor performance in this task.

Method

Participants. The participants were 26 graduate and undergraduate students associated with the Indiana University Department of Psychology. All participants were paid $7.50 plus up to a $4.00 bonus depending on performance.

Stimuli and apparatus. The stimuli were tones varying in pitch and loudness. Previous work indicates that such stimuli lie toward the integral side of the integral–separable dimension continuum (Grau & Kemler Nelson, 1988; Melara & Marks, 1990). The stimulus set is illustrated schematically in Figure 13. Most of the tasks involved only Stimuli A–D, which were created by combining orthogonally frequency values of 900 and 950 Hz and intensity values of 60 and 70 dB. Previous work reported by Melara and Marks (1990) and Melara and Mounts (1994) indicates that the loudness and pitch differences used here are roughly equally discriminable. The stretch-filtering tasks also used Stimuli E–F (frequency of 1000 Hz) and G–H (intensity of 80 dB; see Figure 13). The stimuli were square waves generated by a SoundBlaster AWE32 soundboard and presented through SONY Digital Reference MDR CD350 stereo headphones. Each tone was presented for 500 ms. Intensity levels were measured with a Bruel and Kjoer Precision Sound Level Meter Type 2203. The experiments were controlled by IBM-compatible 486 computers. Participants entered responses by pressing appropriate buttons on the computer keyboard. Response times were measured by using the internal ms-accuracy PC timer.

Procedure. The experiment was organized into 13 conditions as shown in Table 4.[9] 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 2 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. Participants initiated trials by pressing the space bar. Following an error response, the feedback, *Wrong*, was displayed on the computer screen for 2 s. There was no feedback following correct responses.

Table 4
Experimental Conditions Tested in Experiment 3

Condition	Stimuli
Control pitch 1	A vs. B
Control pitch 2	C vs. D
Control loudness 1	A vs. C
Control loudness 2	B vs. D
Correlated 1	A vs. D
Correlated 2	B vs. C
Filter pitch	A, C vs. B, D
Filter loudness	A, B vs. C, D
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

At the start of each condition, instructions were presented on the computer screen regarding the category distinctions for that task. Participants were then allowed to listen to the stimuli associated with each response by pressing the Category 1 and Category 2 keys. They were allowed to engage in this preliminary training as long as they wished. Participants pressed the space bar when they were ready to start the task. At the end of each condition, participants were informed as to their accuracy and mean response time.

The experiment was conducted in a single session that lasted 60–90 min. The experiment was self-paced and participants were allowed to take breaks between conditions. Participants were instructed to respond as quickly as possible without making errors and were told that their bonuses were determined jointly by the speed and accuracy of their responses.

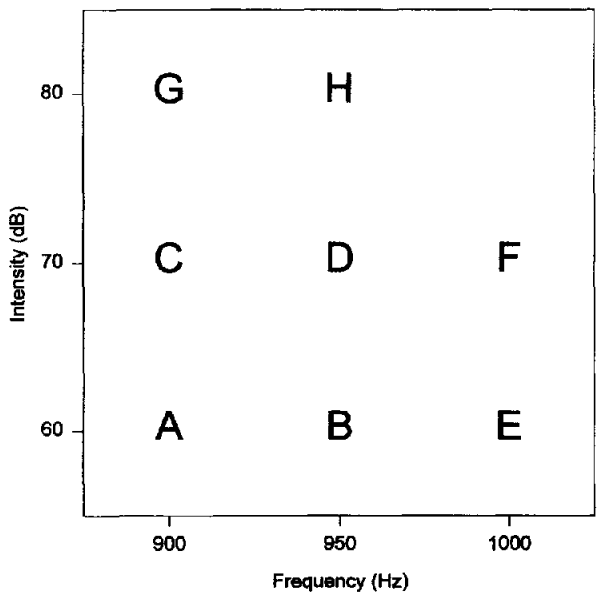

Figure 13. Pitch and loudness values of the auditory stimuli used in Experiment 3.

Results and Theoretical Analysis

The first 24 trials of each condition were considered practice. Only the final 72 trials of each condition were included in the analyses. Response times greater than three standard deviations above the mean were eliminated, which led to the removal of less than .1% of the total observations.

Observed data. The mean correct response times and accuracy rates for each type of task are summarized in Table 5. Although pitch turned out to be slightly more discriminable than loudness, for simplicity in our theoretical analyses we model the data averaged over pitch and loudness conditions. These averaged data for the control, correlated, filtering, stretch-filtering, and condensation tasks are reported in Table 6. As expected, mean response time in the correlated condition was significantly faster than in either control condition, indicating facilitation: pitch, $t(25) = 4.93$, $p < .001$, and loudness, $t(25) = 5.24$, p

[9] Because of certain theoretical complexities involving asymmetric placement of the response criteria and speed–accuracy trade-offs, we do not analyze the data in the two focusing conditions, however. The model can fit the mean response time data, but there are nearly as many free parameters added as there are degrees of freedom in the data sets.

Table 5
Accuracy, Response Times (ms), and Standard Errors for Each Condition in Experiment 3

Condition	p(C)	M	MSE
Control pitch	.968	397	2.79
Filter pitch	.959	435	3.32
Stretch pitch	.941	466	3.72
Control loudness	.966	420	2.07
Filter loudness	.952	446	3.21
Stretch loudness	.955	466	3.82
Correlated	.981	370	1.53
Condensation	.876	1,066	14.46

Note. p(C) = probability correct; M = mean correct response time (ms); MSE = mean standard error.

$< .001$. Mean response time in the filtering conditions was significantly slower than in the control conditions, indicating interference: pitch, $t(25) = 3.73$, $p < .005$, and loudness, $t(25) = 2.39$, $p < .05$. For pitch, mean response time in the stretch-filtering condition was significantly slower than in the standard filtering condition, indicating greater interference as within-category similarity decreased, $t(25) = 2.81$, $p < .05$. The results were marginally significant for increased interference in the stretch-loudness condition, $t(25) = 1.78$, $p = .088$. Finally, mean response time in the condensation task was far slower than in any other condition. These results replicate well-known findings already reported in the literature, but they provide a data source suitable for quantitative fitting because all conditions were tested under the umbrella of a single experiment.

Modeling analyses. Because our goal in these initial tests of the EBRW was to fit the main quantitative trends in the averaged response time and accuracy data, we assumed for simplicity that Stimuli A–D resided along the four corners of a unit square: A = (0, 0), B = (1, 0), C = (0, 1), and D = (1, 1). We further assumed a linear relation between psychological and physical distances, meaning that the additional stimuli used in the stretch-filtering conditions had the following psychological coordinates: E = (2, 0), F = (2, 1), G = (0, 2), and H = (1, 2). Finally, for ease of the ensuing discussion, we assume that the relevant dimension in each of the unidimensional classification tasks (control, correlated, filtering, stretch-filtering) is Dimension 1.

Rather than doing extensive parameter searching, we set most of the parameters at values deemed reasonable on the basis of model fitting conducted in the earlier experiments. We set $c = 1.5$ and $\alpha = .30$. In the unidimensional conditions, in which perfect classification can be achieved by attending to just Dimension 1, we set the attention weight at $w_1 = .75$ (with $w_2 = .25$). This intermediate value of w_1 is intended to represent the imperfect selective attention allowed by these integral dimension stimuli. In the condensation condition, in which both dimensions are needed and are equally relevant for purposes of classification, we set the attention weights at $w_1 = w_2 = .50$.

Although not needed for fitting the main data sets reported in Tables 5 and 6, we assumed for reasons of theoretical plausibility, as well as to potentially account for certain sequential effects, that memory strengths for previously stored exemplars decreased with each subsequent test item. Following previous research in which exemplar models were fitted to sequential learning data (Estes, 1994; Estes & Maddox, 1995; Nosofsky, Kruschke, & McKinley, 1992), on the trial on which an individual exemplar was initially presented and stored, its memory strength was set at one; on each subsequent trial, the memory strength was multiplied (decreased) by the factor γ ($\gamma < 1$). We arbitrarily set γ at .95.

Holding these parameters fixed, we conducted model fits by varying the value of only the response criterion (A and −B in Figure 1). Because of the symmetry in the category structures, we assumed A = B. The predictions are based on averaging over the results of 1,000 randomly generated 96-trial sequences (with the first 24 trials deleted, as was done when computing the observed data). We searched for a value of the response criterion parameter that yielded reasonable predictions of the accuracies in the four unidimensional conditions and then plotted the relation between the predicted and observed response times. In this sense, the predicted ordering of response times is parameter free.

The results of the model fits with the response criterion set at A = 4 are shown in Figure 14 (solid symbols). Consider first the four unidimensional tasks. The EBRW is in the right range as far as predicting the accuracy data (see parenthesized values by each response time point). More critically, the model also correctly orders the mean response times, predicting facilitation in the correlated task, interference in the filtering task, and increased interference in the stretch-filtering task. Our model-fitting explorations suggest that these predictions of mean response time are robust over a wide range of parameter settings.

The EBRW makes the correct qualitative prediction that the condensation task is more difficult than any of the unidimensional tasks. However, inspection of Figure 14 reveals that the model has some quantitative shortcomings, with observed performance in this task even slower than predicted. Inspection of the figure also reveals, however, that with the present parameter settings, the model underpredicts the accuracy rate obtained in this task: predicted, p(C) = .83; observed, p(C) = .88. Thus, a sensible explanation of these results is that a speed–accuracy trade-off may have occurred: Because of the extreme difficulty of the condensation task, participants needed to expand their response criteria, gathering more evidence so as to perform with

Table 6
Accuracy, Response Times (ms), and Standard Errors for Each Condition, Averaged Across Pitch and Loudness, in Experiment 3

Condition	p(C)	M	MSE
Correlated	.981	370	1.53
Control	.967	409	1.38
Filter	.956	441	2.31
Stretch	.948	466	2.66
Condensation	.876	1,066	14.46

Note. p(C) = probability correct; M = mean correct response time (ms); MSE = mean standard error.

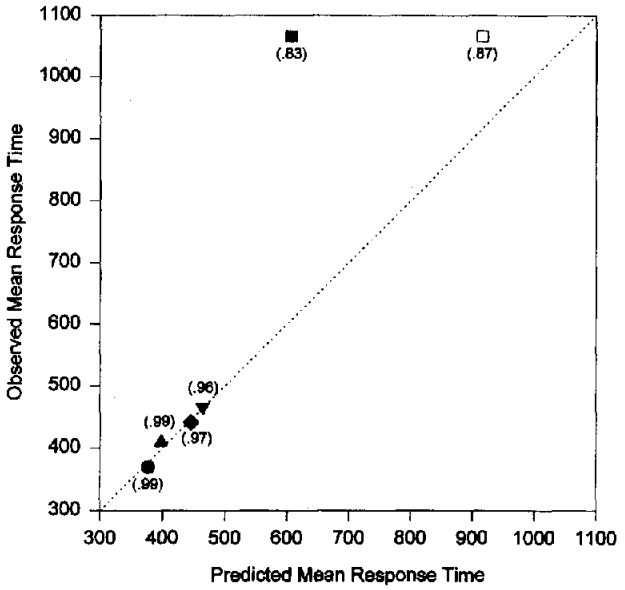

Figure 14. Scatterplot of observed mean correct response times (ms) in each type of speeded classification task against the predicted mean correct response times from the exemplar-based random walk (EBRW) model, Experiment 3. Values in parentheses are the predicted proportions of correct responses in each task, according to the EBRW. The solid square for the condensation task is the prediction with the response criterion set at A = 4, whereas the open square for the condensation task is the prediction with the response criterion set at A = 6. Circle = correlated task; triangle (up) = control task; diamond = filtering task; triangle (down) = stretch-filtering task; square = condensation task.

acceptable accuracy. We searched for a value of the response criterion parameter that yielded predicted accuracy close to the observed accuracy for the condensation condition. The results with A = 6 are shown in Figure 14 (open square). Now, the quantitative prediction of the response time in the condensation task is fairly good.

Discussion

The main purpose of this section was to demonstrate that the EBRW accounts for some fundamental patterns of results involving the classic Garner speeded categorization tasks. In summary, for stimuli varying along integral dimensions, the model predicts facilitation in the correlated task relative to the control task; interference in the filtering task relative to the control task; increased interference in the filtering task as within-category similarity decreases; and performance in the condensation task that is far worse than in all of the other tasks.

The EBRW is consistent with some other important findings involving the Garner speeded classification tasks. For example, researchers have sometimes tested rotated stimulus sets, in which the configuration in Figure 12 is rotated, for example, 45°. Filtering performance is better for unrotated sets than for rotated ones (Grau & Kemler Nelson, 1988; Melara & Marks, 1990). The interpretation of this result is that selective attention can be oriented along the direction of primary psychological dimensions but not in arbitrary directions in the psychological space. The EBRW models such a result because the attention weights (the set of w_m parameters in Equation 1) can stretch the psychological space along the relevant, attended psychological dimensions but not in oblique directions in the space (Kruschke, 1993; McKinley & Nosofsky, 1996).

The memory decay assumption in the EBRW allows the model to predict the robust stimulus repetition effects that exist in the Garner tasks. For example, in the filtering task, classification of Stimulus A on Trial n is most rapid if the preceding stimulus on Trial $n - 1$ was also Stimulus A. (In the present filtering experiments, the mean response time on stimulus repetition trials was 404 ms, whereas the mean response time on nonrepetition trials was 454 ms.) Indeed, Garner (1974, pp. 139–145) suggested that the main cause of interference in the filtering task relative to the control task is that there are twice as many repetition trials in the control task. The stimulus repetition effect is predicted by the EBRW because a test item is most likely to retrieve exemplars to which it is identical, and the memory strength for an identical exemplar is greatest when it was presented on the immediately preceding trial.

Development of Automaticity in Skilled Performance

The focus of our article has been on applications of the EBRW to predicting response times in tasks of speeded classification. A virtue of our proposed model, however, is that it promises also to provide detailed quantitative accounts of the development of automaticity and the acquisition of cognitive skills (Anderson, 1982; Logan, 1988; Shiffrin, 1988). Indeed, by integrating and extending Nosofsky's (1986) GCM and Logan's (1988) instance model, the EBRW builds bridges between the domains of perceptual categorization and automaticity. In this section we briefly describe some of these applications. A full report is provided by Palmeri (1997).

Our example of a task of skilled cognitive performance is the visual numerosity judgment task, which has been studied, for example, by Lassaline and Logan (1993). The stimuli are random dot patterns consisting of between 6 and 11 dots. On each trial, a dot pattern is presented, and the participant judges its numerosity as quickly as possible without making errors. There is extensive training, and the same dot patterns are judged multiple times.

Lassaline and Logan (1993) found that early in training, response times increased linearly with the numerosity of the patterns. The straightforward interpretation was that participants engaged in a counting process, with each extra dot adding a noisy increment to the total response time. By the end of training, however, after participants had seen each dot pattern on numerous occasions, mean response time was essentially a flat function of numerosity. Lassaline and Logan's (1993) interpretation was that multiple instances of the dot patterns had been stored in memory, and participants simply retrieved memories of the correct answers associated with the instances to perform the task. This interpretation of the development of automaticity is supported by the instance-specific nature of participants' performance: When completely new dot patterns were presented

during a transfer stage, response times were as slow as at the start of training.

A critical prediction of the EBRW, however, is that some generalized automaticity to new dot patterns should occur if there are fine-grained manipulations of similarity to the original patterns. Palmeri (1997) replicated Lassaline and Logan's (1993) task, except that at time of transfer, in addition to presenting old training patterns and unrelated new patterns, he also presented objects that were moderate distortions or high distortions of the original training patterns. [The classic statistical distortion paradigm introduced by Posner, Goldsmith, and Welton [1967] was used for creating these patterns.]

Palmeri's (1997) observed transfer data, collected following 13 days of numerosity judgment training, are shown in the top panel of Figure 15. The figure plots mean response times as a function of numerosity for the old patterns and the moderate-distortion, high-distortion, and unrelated new patterns. For purposes of comparison, the figure also plots the response time function observed at the start of the training phase (Session 1). The major aspects of Lassaline and Logan's (1993) study were replicated, with the response time function for unrelated new patterns being essentially the same as at the start of training, and the response time function for old patterns being close to flat. The important new finding is that the response time functions for the moderate-distortion and high-distortion new patterns lie systematically intermediate between these extremes, providing clear evidence of generalized automaticity.

Palmeri (1997) used the EBRW to quantitatively fit these data. First, following Lassaline and Logan (1993), he assumed that on each trial participants engaged in a counting process. The time for each additional count was a normally distributed random variable with mean μ and variance σ^2. The EBRW process took place simultaneously with the counting algorithm, and the first process to finish determined the response. Applying the EBRW to the visual numerosity judgment task is straightforward. A counter is established for each numerosity category. Anytime an exemplar from numerosity category j is retrieved, the j'th counter is incremented. The process finishes as soon as one of the counters is at least A units greater than all other counters. This model reduces to the form of the EBRW illustrated in Figure 1 when there are just two categories. (A more in-depth discussion of variants of the EBRW applied to multiple-category tasks is provided by Palmeri [1997] and Nosofsky [in press].)

The parameters used for simulating the EBRW included the step-time constant α, interexemplar similarity parameters for moderate-level and high-level distortions (s_m and s_h), and a residual similarity parameter for unrelated patterns (s_r). The response criterion parameter A was set at 3. In addition, the mean and variance parameters for the dot-counting process (μ and σ^2) were estimated. Palmeri (1997) started by fitting the model to the set of numerosity judgment training data obtained in the experiment. Then, holding all parameters fixed except for s_m and s_h (which could not be estimated from the training data), he used the model to fit the transfer data. The predictions of the model are illustrated in the bottom panel of Figure 15. The two-parameter model provides an excellent fit to the data, achieving a correlation of .961 with the observed numerosity judgment

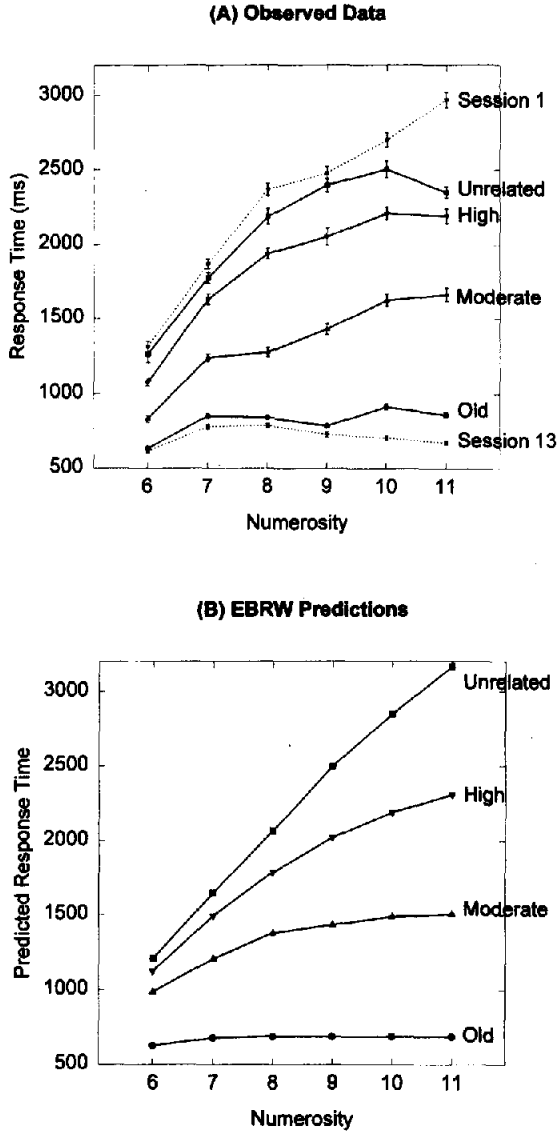

Figure 15. In A, the observed mean response times (ms) as a function of numerosity and type of item (old, moderate distortion, high distortion, unrelated new) during the transfer phase of Palmeri's (1997) experiment. The dashed curve gives the mean response times as a function of numerosity during the first session of training. In B, predicted mean response times from the exemplar-based random walk (EBRW) model. Adapted from "Exemplar Similarity and the Development of Automaticity," by T. J. Palmeri, 1997, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* p. 334. Copyright 1997 by the American Psychological Association.

response times. The model predicts these results because similarities of the moderate and low distortions to the old instances cause these training instances to be retrieved, allowing for automatic numerosity judgments.

In other experiments, Palmeri (1997) demonstrated that automaticity in visual numerosity judgment develops more rapidly when patterns have high within-category similarity and that such

automaticity develops more slowly when there is high between-category similarity. Again, the EBRW provided an excellent account of the time course of the development of automaticity as a function of these fundamental variables. Because, in the pure version of Logan's (1988) instance model, only instances that are identical to a test item are retrieved from memory, the similarity-based retrieval processes formalized in the EBRW are critical for explaining these data.

Furthermore, Palmeri's (1997) finding that increased *between*-category similarity slowed the numerosity judgment response times demonstrated that it is not enough only to add similarity-based retrieval to Logan's (1988) model. Because Logan's model is a pure, first-instance race model, it predicts faster response times as between-category similarity increases, albeit with more errors. The reason is that the more similar instances there are participating in the race, the faster any given instance would be retrieved, regardless of whether it belongs to the target category or a contrast category. A decision process such as the one in our random walk model, which causes exemplars from competing categories to counteract one another, is able to explain such results (see also Strayer & Kramer, 1994, for additional evidence bearing on this point).

General Discussion

Summary

In this article we proposed and tested an exemplar-based random walk (EBRW) model of response times in tasks of speeded multidimensional perceptual classification. The EBRW integrates and extends two well-known exemplar models of cognitive processes: Nosofsky's (1986) generalized context model (GCM) of perceptual categorization, and Logan's (1988) instance-based model of automaticity. As is the case in the GCM, in the EBRW exemplars are represented as points in some multidimensional psychological similarity space. Selective attention processes systematically modify the structure of the space in which the exemplars are embedded. Classification decisions are based on the similarities of objects to these stored exemplars. The EBRW goes beyond the GCM by specifying a dynamic process by which these exemplar-similarity comparisons unfold over time. As is the case in Logan's instance model, test items cause stored exemplars to race among one another to be retrieved from memory. However, whereas in Logan's model only exemplars that are identical to the test item race to be retrieved, in the EBRW all exemplars race, with rates determined by their similarity to the presented item. Furthermore, whereas in Logan's model the first retrieved exemplar drives the response, in the EBRW the retrieved exemplars provide incremental information that feeds into a random walk decision process.

To our knowledge, the EBRW may be the first rigorously formalized process model to provide quantitative predictions of response times in diverse tasks of speeded multidimensional perceptual classification. Our initial tests demonstrated some broad-based support for the model. First, we demonstrated that the EBRW predicts the distance-from-boundary effects reported by Ashby et al. (1994) in their experiments involving bivariate normal category distributions. Next, in Experiment 1, we demonstrated that the EBRW is capable of yielding accurate quantitative predictions of classification response times for individual objects based on their locations in a multidimensional similarity space. The model simultaneously characterized the roughly power-law decreases in mean classification response times that were observed as a function of extended practice in this task. In Experiment 2, we demonstrated that individual object familiarity can play a major role in speeded classification, as predicted by the EBRW. Again, the EBRW was able to provide reasonably good quantitative predictions of response times for individual objects as a function of these familiarity manipulations. In Experiment 3, we demonstrated that the EBRW goes a long way toward explaining patterns of performance in the classic Garner speeded categorization tasks: For stimuli varying along integral dimensions, the model predicts correctly facilitation in the correlated task; interference in the filtering task; increased interference in the filtering task as within-category similarity decreases; and very poor performance in the condensation task. It also predicts the stimulus repetition effects observed in this paradigm. Finally, beyond its ability to account for speeded perceptual classification, an added virtue of the EBRW is its ability to account for similarity and response competition effects in the development of automaticity in tasks of skilled performance.

Limitations and Extensions

In this section we discuss some limitations of the EBRW and avenues of extension that will guide future research.

RT distributions. Our initial tests of the EBRW were limited to predicting mean response times for individual objects or tasks. To provide more rigorous tests of the model, in future work we need to fit it to entire distributions of response time data. Conducting these more rigorous tests will require us to extend the model by specifying properties of additional stages beyond the classification decision-making stage (which is modeled by the EBRW). Such stages include, for example, encoding and response execution stages. Nosofsky (in press) and Nosofsky and Palmeri (in press) have already demonstrated some significant progress along these lines: They fitted, with reasonably good success, an extended EBRW model to the detailed response time distribution data observed for individual stimuli in the color classification task and the Garner tasks described in Experiments 1 and 3 of this article. Among the many issues that remain to be examined is the extent to which the EBRW can account for power-law reductions in the entire distribution of RTs as a function of practice (e.g., Logan, 1992), as predicted by the pure single-instance version of Logan's theory.

Speed–accuracy trade-off. It will also be important in future work to model speed–accuracy trade-off functions. A straightforward approach to using the EBRW to model speed–accuracy trade-offs is to assume variations in the locations of the response criteria. As the criteria move closer to the starting point of zero, response times get faster but accuracy should decline.

Other more complex factors may also be involved in modeling speed–accuracy trade-offs, however. For example, Lamberts (1995a) used an extended version of the generalized context

model to account for patterns of classification performance under time pressure. He discovered that patterns of generalization depended systematically on whether participants were required to respond by short or long deadlines. The data could be well modeled by assuming that under short deadlines, overall sensitivity (c in Equation 2) was low and the attentional weights (the w_m values in Equation 1) were determined primarily by the perceptual saliences of the dimensions. Without deadlines, however, overall sensitivity was high and the attention weights depended primarily on the formal category structure. Thus, using the EBRW to provide a full account of speeded classification performance may require extending the model to account for dynamic changes in the values of the sensitivity and attention weight parameters over time (see Lamberts, 1995b, for a formal proposal along these lines).

Probabilistic representations. In our applications of the EBRW in this article, each object was represented as a single point in the space. Numerous researchers have argued forcefully for the importance of probabilistic representations, in which each object is represented as a probability distribution of points (for reviews, see Ashby, 1992). We view the single-point representation in the EBRW as a simplification. Undoubtedly, the precise perceptual and memorial representation of an object varies from trial to trial, so a complete account of speeded classification will require modeling these fluctuations. It is straightforward to extend the EBRW to include assumptions about probabilistic representations. Nosofsky (in press), for example, proposed an extension of the EBRW with probabilistic memory representations that can reproduce some of the major phenomena observed in tasks of unidimensional absolute judgment.

Response accuracy and training. Although the EBRW predicts facilitation in response times with increased training, the version of the model presented in this article does not predict changes in response accuracy. There are various approaches to extending the model to account for improvements in response accuracy. Following Nosofsky et al. (1992) and Estes (1994), one idea is that background-noise elements exist in memory at the start of training. Retrieving a background element is assumed to move the random walk counter in a random direction toward either Category A or B. The background elements race exponentially with rate b, independent of the test item that is presented. Early in learning, before many category exemplars are stored in memory, the background elements would have a high probability of winning the races and being retrieved, which would result in relatively poor classification performance. As more category exemplars are stored in memory, the probability of background elements winning the races declines, so response accuracy would improve. Another process that would lead to improvements in response accuracy is increases in the value of the overall sensitivity parameter (c) as a function of experience, evidence for which was provided by Nosofsky (1987).

Probabilistic Versus Deterministic Response Rules and Overall Accuracy

A shortcoming of the GCM is that, in designs involving probabilistic assignments of exemplars to categories, the model basically predicts *probability matching* behavior (for more detailed discussion of this point, see Nosofsky et al., 1992). By probability matching, we mean that if an exemplar receives Category A feedback, for instance, with probability .70, then the observer tends to classify that exemplar into Category A with probability .70. Although probability matching is often observed in probabilistic classification designs (Estes, 1976), research conducted by Ashby and his colleagues indicates that highly experienced individual observers often respond more deterministically (i.e., with probabilities closer to zero or unity) than predicted by a probability matching rule (Ashby & Gott, 1988; Ashby & Maddox, 1992). The EBRW predicts probability matching behavior in the special case in which the response criteria are set at magnitude one. (Recall that it is under such circumstances that the EBRW is formally identical to the GCM.) As the criteria are extended further away, and more and more exemplar evidence is recruited before a decision is made, the model predicts behavior that is more deterministic.

More precisely, suppose that in the EBRW, the response criteria +A and −B are set an equal magnitude K from the starting point of zero, A = B = K. Then it is straightforward to show that the response probability predictions of the EBRW (see Equations 16, 9, and 13) can be expressed as

$$P(\mathrm{A}|\mathrm{i}) = S_{\mathrm{iA}}^{\mathrm{K}}/(S_{\mathrm{iA}}^{\mathrm{K}} + S_{\mathrm{iB}}^{\mathrm{K}}). \qquad (23a)$$

When K = 1, this EBRW response rule is the same as in the GCM (Equation 4), whereas when K > 1, the model predicts responding that is more deterministic than that predicted by the GCM. Indeed, Maddox and Ashby (1993) proposed an extended version of the GCM with precisely this type of response rule to allow the model to potentially account for the deterministic responding exhibited by highly experienced individual participants.[10] McKinley and Nosofsky (1995) found that this extended response rule yielded fits to Maddox and Ashby's (1993) data that were far better than those of the GCM and as good as important representatives from the class of decision boundary models. Thus, at least in situations in which models are fitted to individual-participant accuracy data from highly experienced observers, the EBRW appears to provide a major improvement over the GCM.

In situations in which models are fitted to averaged participant data of relatively inexperienced observers, however, the GCM has generally yielded extremely good predictions of classification response probabilities (e.g., Nosofsky, 1987). Of course, the EBRW can match these predictions if the response criterion parameter is set at K = 1. Unfortunately, however, if K = 1, the EBRW becomes a pure single-instance race model, which we have argued often produces implausible predictions of classification response times.

A critical question, therefore, is whether or not an EBRW

[10] In the response rule proposed by Maddox and Ashby (1993), which did not involve any processing considerations pertaining to classification response time, the K parameter was real-valued, whereas in the EBRW it is integer-valued. In the EBRW, if one allows probability mixtures in the setting of K across trials, the real-valued version of the response rule can be well approximated.

model with $K > 1$ can match the response probability predictions yielded by the GCM (when the goal is to fit averaged data collected from inexperienced participants). Because the EBRW with $K > 1$ predicts accuracies that are higher (i.e., more deterministic) than those predicted by the GCM, one idea is to add sources of processing noise to the model to bring back down the predicted accuracies. The assumption that background-noise elements participate in the exemplar retrieval process, which we discussed in the previous section, provides a straightforward possibility along these lines. Combining the assumptions about background elements introduced by Nosofsky et al. (1992) and Estes (1994) with the processing machinery proposed in this article, the response rule in the EBRW becomes:

$$P(A|i) = (S_{iA} + b)^K / [(S_{iA} + b)^K + (S_{iB} + b)^K], \qquad (23b)$$

where b is a free parameter representing the rate at which the background elements race to be retrieved.

In preliminary tests, we fitted the Equation 23b model to six sets of categorization response probability data reported by Nosofsky (1987, Tables 4 and 5), using the same methods as described previously for the GCM in that article. We set $K = 3$ and searched for the value of b that produced the best fits to each data set in combination with the other model parameters (c and w_1 from Equations 1 and 2). The EBRW yielded fits to these data sets that were slightly better overall than those produced by the GCM (see Appendix E). An additional free parameter (b) was involved, so we are not arguing that the EBRW is an improvement over the GCM in this regard. Rather, we are simply demonstrating that versions of the EBRW with $K > 1$ can produce fits to accuracy data as good as those produced by the GCM in previously reported studies. An important direction for future research is to test more rigorously the ability of the EBRW to simultaneously model how accuracy and response time data evolve with increased classification experience.

Perceptual Expertise

The GCM has sometimes been criticized on grounds that it is highly implausible that in making classification decisions, observers would sum the similarity of an object to *all* category exemplars ever experienced. While retaining the exemplar-based category representation of the GCM, the EBRW requires that only a relatively small subset of exemplars be retrieved for any given classification decision. Furthermore, it predicts that this process should become more and more efficient as learning proceeds because a greater number of exemplars race to be retrieved from memory.

This conception of the efficient retrieval of multitudes of exemplars stored in memory may shed light on the nature of expert perceptual classification. A paradigm case of our ideas comes from the work on chess mastery conducted by De Groot (1965) and Chase and Simon (1973). This research suggested that differences in chess skill emerge not only from differences in the quality of operational thinking but also from differences in memory and perception. Apparently, the chess master stores a vast warehouse of meaningful chunks and patterns that are immediately recognized: The chess master just "sees" the right move. Indeed, investigations involving a simulation program known as the Memory-Aided Pattern Perceiver led Simon and Gilmartin (1973) to estimate that master players have between 10,000 and 100,000 chess chunks stored in long-term memory.

The role of exemplar similarity in expert perceptual classification has also been well documented by the work of Brooks, Norman, and Allen (1991; Allen, Norman, & Brooks, 1992). Using expert dermatologists as participants, these researchers presented slides of dermatological cases for initial diagnoses and then, several weeks later, presented new slides that were similar or dissimilar to the originals. The experts correctly diagnosed the similar slides significantly more often than the dissimilar ones, even though a simple rule was available for all cases.

We can well imagine that expertise in all forms of perceptual classification is closely related to extensive exemplar-based experience. Thus, the expert radiologist, rather than relying solely on conscious, analytic, rule-based strategies such as may be taught in a medical textbook, may just "see" the proper diagnosis. The vast warehouse of prior examples that this expert classifier has stored in memory all race to be retrieved and drive his or her categorizations.

Finally, although our emphasis in this article has been on exemplar retrieval processes, we expect that the human observer makes extensive use of simple analytic rules as well (e.g., Nosofsky et al., 1994). Recall that in the domain of skilled performance, Logan (1988) hypothesized that simple algorithms are executed in parallel with the instance retrieval process, with the first to finish determining the response. Analogously, in the domain of perceptual categorization, simple analytic rule systems may coexist with the exemplar retrieval system. Our hypothesis is that exemplar retrieval processes come to play a more dominant role as the human observer gains increased expertise in a given perceptual domain. Achieving a comprehensive model of perceptual categorization, however, will require a deeper understanding of both rule-based and exemplar-based knowledge representations, and of how they interact with one another in novice and expert observers.

References

Allen, S. W., Norman, G. R., & Brooks, L. R. (1992). Experimental studies of learning dermatologic diagnosis: The impact of examples. *Teaching and Learning in Medicine, 4,* 35–44.

Anderson, J. R. (1982). Acquisition of a cognitive skill. *Psychological Review, 89,* 369–406.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1322–1340.

Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449–483). Hillsdale, NJ: Erlbaum.

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, and Cognition, 14,* 33–53.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categoriza-

tion from identification. *Journal of Experimental Psychology: General, 120,* 150–172.

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. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 50–71.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology, 37,* 372–400.

Ashby, F. G., & Maddox, W. T. (1994). A response time theory of separability and 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.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278–287.

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 and Human Performance, 29,* 175–207.

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 538–564.

Carroll, J. D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 57–105). San Francisco: W. H. Freeman.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81.

De Groot, A. D. (1965). *Thought and choice in chess.* The Hague: Mouton.

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

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18,* 500–549.

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

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1075–1095.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Volume I.* New York: Wiley.

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

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.

Gottwald, R. L., & Garner, W. R. (1975). Filtering and condensation tasks with integral and separable dimensions. *Perception & Psychophysics, 18,* 26–28.

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.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Karpiuk, P., Jr., Lacouture, Y., & Marley, A. A. J. (in press). A limited capacity, wave equality, random walk model of absolute identification. In A. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce.* Mahwah, NJ.: Erlbaum.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science, 5,* 3–36.

Lamberts, K. (1995a). Categorization under time pressure. *Journal of Experimental Psychology: General, 124,* 161–180.

Lamberts, K. (1995b). The time course of categorization. Manuscript submitted for publication.

Lassaline, M. E., & Logan, G. D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 561–581.

Link, S. W. (1992). *The wave theory of difference and similarity.* Hillsdale, NJ: LEA.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40,* 77–105.

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

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms. *Cognitive Psychology, 22,* 1–35.

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, and 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.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 179–195.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

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

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification–speeded classification relationship. *Journal of Experimental Psychology: Human Perception and 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: LEA.

Marley, A. A. J., & Colonius, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *Journal of Mathematical Psychology, 35,* 1–20.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 128–148.

McKinley, S. C., & Nosofsky, R. M. (1996). Selective attention and

the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 294–317.

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

Melara, R. D., & Marks, L. E. (1990). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 398–414.

Melara, R. D., & Mounts, J. R. W. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception and Psychophysics, 56,* 73–90.

Murdock, B. B. (1985). An analysis of the strength–latency relationship. *Memory and Cognition, 13,* 511–521.

Newell, A., & Rosenbloom, 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.

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

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87–109.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65.

Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology, 34,* 393–418.

Nosofsky, R. M. (1991a). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 3–27.

Nosofsky, R. M. (1991b). Typicality in logically-defined categories: Exemplar-similarity versus rule instantiation. *Memory and Cognition, 19,* 131–150.

Nosofsky, R. M. (1992a). Exemplar-based approach to relating categorization, identification and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363–393). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M. (1992b). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43,* 25–53.

Nosofsky, R. M. (in press). An exemplar-based random walk model of speeded classification and unidimensional absolute judgment. In A. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce.* Mahwah, NJ: Erlbaum.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 211–233.

Nosofsky, R. M., & Palmeri, T. J. (in press). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Perception and Psychophysics.*

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354.

Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception and Psychophysics, 14,* 565–569.

Posner, M. I., Goldsmith, R., & Welton, K. E., Jr. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology, 73,* 28–38.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Shanks, D. R., & Gluck, M. A. (1994). Tests of an adaptive network model for the identification and categorization of continuous-dimension stimuli. *Connection Science, 6,* 59–89.

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., & Chang, J. J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology, 65,* 94–102.

Shiffrin, R. M. (1988). Attention. In R. A. Atkinson, R. J. Hernstein, G. Lindzey, & R. D. Luce (Eds.), *Steven's handbook of experimental psychology* (pp. 739–811). New York: Wiley.

Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5,* 29–46.

Strayer, D. L., & Kramer, A. F. (1994). Strategies and automaticity: I. Basic findings and a conceptual framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 318–341.

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

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

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

Appendix A

Applying the EBRW to Experiment 1 of Ashby, Boynton, and Lee (1994)

As described in the text, the EBRW was used to simulate data from the Experiment 1 design of Ashby et al. (1994). Tables A1–A3 present the obtained rank-order correlations between the predicted response times of the EBRW and the distance-from-boundary and familiarity measures described by Ashby et al. (1994). The distance-from-boundary measure calculates the distance between an item and the optimal population classification decision boundary. The familiarity measure calculates the summed similarity between an item and all 300 exemplars in the two category distributions. Unfortunately, some complications arise because the summed-similarity computation depends on the value of the sensitivity parameter (c) in Equation 2. Ashby et al. (1994) searched for the value of c that yielded the most negative rank-order correlations between the familiarity measure and the response times for each individual participant. These estimated values differed widely across conditions and individual participants. The familiarity correlations reported in Tables A1–A3 assume the median values of c that Ashby et al. (1994) estimated for each overlap condition. Note that the values of c used for computing familiarity were not the same as the value of c used to simulate the EBRW.

The results of the EBRW simulations shown in Tables A1–A3 were based on the following parameter settings: $c = .01$, $w_1 = w_2 = .50$, and $A = B = 15$. Similar qualitative results are obtained for a wide range of parameter settings. The step-time constant α in Equation 6 was varied in increments of .1 from 0.0 to 2.0.

As can be seen in the tables, except for values of α near zero, the following results were obtained. First, the rank-order correlation between predicted response times and the distance-from-boundary measure ranged between −.25 and −.55. A similar range of rank-order correlations was observed by Ashby et al. (1994, Table 6). In the moderate-overlap and high-overlap conditions, the rank–order correlations between predicted response times and the familiarity measure ranged between 0 and .15, whereas the familiarity correlations were slightly negative in the low-overlap condition. Again, similar results were observed by Ashby et al. (1994, Table 8).

Table A1
EBRW-Predicted Correlations Between Response Time and Distance-From-Boundary Measure (r^D) and Familiarity Measure (r^F) as a Function of Step-Time Constant for Low-Overlap Condition of Ashby et al. (1994)

α	r^D	r^F
0.000	−0.254	−0.349
0.100	−0.297	−0.229
0.200	−0.260	−0.202
0.300	−0.252	−0.062
0.400	−0.323	−0.021
0.500	−0.338	−0.089
0.600	−0.279	−0.115
0.700	−0.292	−0.132
0.800	−0.204	−0.118
0.900	−0.257	−0.222
1.000	−0.284	−0.135
1.100	−0.360	−0.116
1.200	−0.394	−0.132
1.300	−0.355	−0.081
1.400	−0.260	−0.150
1.500	−0.296	−0.111
1.600	−0.310	−0.094
1.700	−0.360	−0.041
1.800	−0.379	−0.013
1.900	−0.319	−0.105
2.000	−0.362	−0.166

Note. EBRW = exemplar-based random walk; α = step-time constant.

Table A2
EBRW-Predicted Correlations Between Response Time and Distance-From-Boundary Measure (r^D) and Familiarity Measure (r^F) as a Function of Step-Time Constant for Medium-Overlap Condition of Ashby et al. (1994)

α	r^D	r^F
0.000	−0.236	−0.155
0.100	−0.433	0.095
0.200	−0.473	0.179
0.300	−0.474	0.115
0.400	−0.465	0.146
0.500	−0.516	0.026
0.600	−0.460	0.117
0.700	−0.484	0.166
0.800	−0.415	0.083
0.900	−0.461	0.133
1.000	−0.547	0.148
1.100	−0.501	0.130
1.200	−0.503	0.134
1.300	−0.441	0.090
1.400	−0.454	0.111
1.500	−0.547	0.162
1.600	−0.490	0.209
1.700	−0.530	0.180
1.800	−0.437	0.154
1.900	−0.491	0.166
2.000	−0.557	0.169

Note. EBRW = exemplar-based random walk; α = step-time constant.

(Appendixes continue on next page)

Table A3
EBRW-Predicted Correlations Between Response Time and Distance-From-Boundary Measure (r^D) and Familiarity Measure (r^F) as a Function of Step-Time Constant for High-Overlap Condition of Ashby et al. (1994)

α	r^D	r^F
0.000	−0.027	0.083
0.100	−0.520	0.031
0.200	−0.521	0.149
0.300	−0.510	0.113
0.400	−0.509	0.019
0.500	−0.525	0.113
0.600	−0.499	0.104
0.700	−0.509	0.150
0.800	−0.514	0.134
0.900	−0.504	0.050
1.000	−0.472	0.069
1.100	−0.468	0.095
1.200	−0.476	0.087
1.300	−0.502	0.152
1.400	−0.477	0.146
1.500	−0.584	0.054
1.600	−0.433	0.078
1.700	−0.580	0.112
1.800	−0.523	0.014
1.900	−0.502	0.054
2.000	−0.552	0.172

Note. EBRW = exemplar-based random walk; α = step-time constant.

Appendix B

RGB Values

The RGB values used to construct the colors in Experiments 1 and 2 are reported in Tables B1 and B2, respectively.

Table B1
Red (R), Green (G), and Blue (B) Pixel Values for Each Stimulus in Experiment 1

Stimulus	R	G	B
1	236	172	172
2	252	140	140
3	220	136	136
4	252	108	108
5	188	124	124
6	224	104	104
7	252	44	44
8	180	96	96
9	212	60	60
10	128	64	64
11	160	24	24
12	184	0	0

Table B2
Red (R), Green (G), and Blue (B) Pixel Values for Each Stimulus in Experiment 2

Stimulus	R	G	B
1	144	36	28
2	184	32	20
3	252	20	0
4	156	68	60
5	200	76	64
6	228	68	56
7	180	112	108
8	252	120	108

Appendix C

Joint Fits of the EBRW to the Response Time and Accuracy Data in Experiments 1 and 2

The EBRW makes predictions of accuracy as well as response time. A natural question, therefore, is whether or not it can simultaneously account for the response time and accuracy data observed for individual stimuli in our experiments. Our goal in these initial tests of the EBRW was to focus on the time course of categorization. Therefore, we used readily discriminable stimuli and extensive training. As a result, most accuracies are near ceiling (see Tables 1 and 3), and there is very little variability in the data. Under such circumstances, even rare idiosyncratic events and mental processes not part of the EBRW (e.g., lapses of attention, fast guessing, etc.) can account for the lion's share of the variance in the accuracy data, so it is unclear how well one should expect the main model to perform.

Nevertheless, with these caveats in mind, we refit the EBRW simultaneously to the individual color RT and accuracy data in Experiments 1 and 2. Because the response times and accuracy data are measured on different scales and display vastly different amounts of variability, any attempt to evaluate the joint fit requires some arbitrary method for combining the separate component fits. After preliminary exploration, the criterion of fit we chose was to minimize the total weighted sum of squared deviations between the predicted and observed mean response time and the predicted and observed percentages of correct choices, where the percentage correct data were given 100 times the weight of the response time data. Otherwise, the method of fitting the EBRW was the same as described previously.

Experiment 1

The predicted response times and accuracies for each participant are shown alongside the observed data in Table C1, and the best-fitting parameters are reported in Table C2. The fits to the response time data are essentially the same as reported in the main text. About all that we can say about the accuracies is that the predicted data are certainly in the ballpark of the observed data, which we believe is all that one can reasonably expect, given the close-to-ceiling performance.

It is worth noting that markedly improved fits to Participant 1's data can be achieved with the addition of an extra free parameter. Participant 1's performance on Colors 1 and 12 was better than predicted by the

Table C1

Predicted and Observed Mean Response Times and Accuracies From the Exemplar-Based Random Walk Model in Experiment 1

	Mean response time		Percentage correct			Mean response time		Percentage correct	
Stimulus	Pre	Obs	Pre	Obs	Stimulus	Pre	Obs	Pre	Obs
	Participant 1, baseline model					Participant 2			
1	971.31	815.00	99.91	100.00	1	995.36	982.00	94.49	91.70
2	831.36	795.00	100.00	98.30	2	587.56	601.00	99.98	100.00
3	1,040.01	1,159.00	99.78	91.70	3	980.50	1,007.00	94.93	99.20
4	797.84	725.00	100.00	100.00	4	564.17	546.00	100.00	100.00
5	907.66	967.00	99.97	100.00	5	614.85	665.00	99.95	100.00
6	945.15	1,068.00	99.94	97.50	6	649.22	650.00	99.88	100.00
7	779.56	706.00	100.00	100.00	7	552.68	529.00	100.00	100.00
8	908.97	931.00	99.97	97.50	8	596.60	619.00	99.98	99.20
9	1,257.20	1,346.00	98.79	95.00	9	744.79	734.00	99.34	98.30
10	767.66	744.00	100.00	100.00	10	545.57	530.00	100.00	100.00
11	826.12	911.00	100.00	99.20	11	585.13	577.00	99.99	100.00
12	1,341.97	1,208.00	98.10	96.70	12	880.43	857.00	97.44	95.00
% Var	78.4		−34.3		% Var	98.1		55.2	
	Participant 1, elaborated model					Participant 3			
1	852.39	815.00	99.72	100.00	1	766.33	780.00	99.84	97.50
2	843.11	795.00	99.75	98.30	2	740.77	709.00	99.93	100.00
3	1,182.00	1,159.00	95.92	91.70	3	985.17	962.00	94.67	95.60
4	784.50	725.00	99.92	100.00	4	706.54	661.00	99.99	100.00
5	970.68	967.00	98.95	100.00	5	837.17	841.00	99.12	100.00
6	1,014.31	1,068.00	98.50	97.50	6	753.02	749.00	99.88	100.00
7	747.21	706.00	99.97	100.00	7	692.40	641.00	100.00	100.00
8	972.75	931.00	98.92	97.50	8	827.37	834.00	99.26	98.40
9	1,303.63	1,346.00	92.99	95.00	9	786.71	834.00	99.69	98.40
10	726.78	744.00	99.99	100.00	10	684.81	697.00	100.00	100.00
11	837.91	911.00	99.77	99.20	11	710.44	779.00	99.99	100.00
12	1,179.29	1,208.00	95.99	96.70	12	1,003.30	1,007.00	93.78	93.40
% Var	95.3		59.9		% Var	90.2		81.0	

Note. Pre = predicted; Obs = observed; % Var = percentage of variance accounted for.

(*Appendixes continue on next page*)

Table C2
Best-Fitting Parameters for the EBRW When It Is Fitted Simultaneously to the Response Time and Accuracy Data in Experiments 1 and 2

	Experiment 1				
Parameter	P1	P1*	P2	P3	Experiment 2
c	2.855	2.212	1.957	2.196	$c_{U8} = 1.396$, $c_{U7} = 1.582$
w_1	0.408	0.426	0.542	0.716	0.508
α	1.906	1.889	1.080	0.242	2.097
A	4	3	4	5	3
B	—	4	—	—	4
k	70.00	100.01	69.15	125.07	27.71
μ_R	213.08	100.00	211.30	491.07	354.56
δ	—	1.376	—	—	1.300

Note. P1–P3 = Participants 1–3, respectively. Column P1* gives the best-fitting parameters for the extended version of the EBRW fitted to Participant 1's data. EBRW = exemplar-based random walk. c = sensitivity parameter; w_1 = attention weight on Dimension 1; α = step-time constant; A = Category A criterion; B = Category B criterion; k = scaling constant; μ_R = mean residual response time. δ = stimulus-specific sensitivity multiplier.

model. Perhaps this participant devoted extra attention to these difficult-to-classify objects during the extensive test phase, or perhaps these colors were simply more discriminable from the remaining objects in the set than was revealed by the MDS analysis. To formalize this idea, we allow a stimulus-specific sensitivity parameter for Colors 1 and 12. In this elaborated model, the similarity of Color i (i = 1 and i = 12) to each remaining Color j is given by $s(i, j) = \exp(-\delta \cdot c \cdot d_{ij})$, where δ is a stimulus-specific sensitivity multiplier (cf. Kruschke, 1992; Nosofsky, 1991a). The predicted response times and accuracies for this elaborated model are shown in Table C1, with the best-fitting parameters and summary fits given in Table C2.

Experiment 2

Achieving good fits to the accuracy data in Experiment 2 required use of a stimulus-specific sensitivity parameter for Color 3. Otherwise, the free parameters and methods of fitting the data are the same as described previously. The predicted mean response times and accuracies from this version of the EBRW are reported alongside the observed data in Table C3, with the best-fitting parameters and summary fits reported in Table C2. The fit to the response time data is as good as reported in the main text, and the model appears to give a good overall account of the accuracy data. A shortcoming is that the unfamiliar stimuli have somewhat more errors than predicted.

Table C3
Predicted and Observed Mean Response Times and Accuracies From the EBRW in Experiment 2

	Mean response time		Percentage correct	
Stimulus	Pre	Obs	Pre	Obs
	Condition U8			
1	832.04	795.00	93.11	94.80
2	854.77	834.00	91.38	94.80
3	683.27	677.00	99.13	98.00
4	872.08	897.00	92.48	89.50
5	808.19	819.00	95.93	97.20
6	861.96	896.00	93.12	88.30
7	649.22	672.00	99.87	98.40
8	741.07	752.00	98.57	92.30
	Condition U7			
1	775.63	750.00	96.29	96.40
2	807.04	794.00	94.68	92.70
3	654.88	648.00	99.55	99.20
4	855.98	859.00	93.53	89.10
5	778.24	740.00	97.18	98.80
6	811.62	846.00	95.83	93.20
7	705.01	703.00	99.37	94.40
8	639.12	648.00	99.93	99.60
% Var	92.6		27.8	

Note. EBRW = exemplar-based random walk; Pre = predicted; Obs = observed; % Var = percentage of variance accounted for.

Appendix D

Optimal Classification Decision Boundaries

Following Ashby and Lee (1991), we derive the optimal classification decision boundaries by assuming that there is a distribution of perceptual effects associated with each individual exemplar. Each distribution is bivariate normal. The means of the distributions along each dimension are given by the multidimensional scaling coordinates associated with each exemplar. All distributions are assumed to have the same variance, σ^2, along both dimensions, and zero covariance. The likelihood of observation x given Category A (B) is found by summing the likelihood of x given each of the exemplar distributions associated with A (B), weighted by their probability of occurrence (Ashby & Lee, 1991, p. 155). Using Bayes's theorem, one then computes the likelihood of each category given observation x. A computer search algorithm is used to locate a set of points (x) for which the likelihood of Category A equals the likelihood of Category B. It is well-known that these likelihood-based boundaries are the optimal decision boundaries for partitioning the space into response regions in the sense that using such boundaries will maximize observers' percentage of correct classification choices. These estimated sets of points constitute the boundaries illustrated in Figure 9. They were computed while setting σ^2 arbitrarily equal to .25. Essentially the same boundaries arise for a wide range of values of σ^2. Their shape changes substantially only when σ^2 is extremely large relative to the distance between the distribution means. (Moreover, essentially the same boundaries arise if one simply computes the set of points that have equal summed similarity to the exemplars of Category A and B, as we discussed in Experiment 1.) Note that these computations made the simplifying assumption that all distributions had the same variance and zero correlation between dimensions. Slightly different shaped boundaries would arise if individual exemplar distributions had unique variances and correlations. The critical point, however, is that these complicating factors would be unlikely to result in different shaped boundaries across Conditions U7 and U8.

Suppose instead that participants use linear decision boundaries. Again, we assume that each exemplar is represented by a bivariate normal distribution, as discussed previously. The probability that a participant classifies an exemplar into Category A is found by integrating over the portion of the exemplar distribution that falls in the Category A response region defined by this linear boundary. The overall percentage of correct classifications is then found by averaging over the percentage of correct classifications for each individual exemplar. One then conducts a computer search to find the slope and y-intercept of the linear boundary that maximizes this overall percentage correct. These optimal linear boundaries are essentially identical across Conditions U7 and U8.

Appendix E

Fitting the EBRW to Nosofsky's (1987) Color Categorization Data

In this appendix we briefly describe how the EBRW with background noise (Equation 23b) is fitted to the color categorization data reported by Nosofsky (1987, Table 4). Recall that the term S_{iA} in Equation 23b is given by $S_{iA} = \Sigma M_j s_{ij}$, where the sum is over all exemplars that belong to Category A, and likewise for S_{iB}. The exemplars assigned to each category in Nosofsky's (1987) experiments, and that therefore enter into these sums, are shown in Figure 5 of Nosofsky (1987). To compute the similarity of item i to exemplar j, s_{ij}, one makes use of the MDS solution for the colors reported by Nosofsky (1987, p. 95, Table 3) and then applies Equations 1 and 2. Applying these equations requires estimating two free parameters for each condition, c and w_1. Two versions of both the GCM and the EBRW were fitted to the data. In the first version, all memory strengths in the equations for S_{iA} and S_{iB} were set at $M_j = 1$. In the second version, the memory strengths were set equal to *stimulus bias* terms estimated in a separate experiment reported by Nosofsky (1987, Table 3). We present this second version because in Nosofsky's (1987) original article, the stimulus bias terms were included when fitting the GCM. In fitting each version of the EBRW, we set K = 3 in Equation 23b and searched for the values of c, w_1, and the background-noise constant b that gave a best fit of the model to the data in each condition. In fitting the GCM, we set K = 1 and b = 0 in Equation 23b and searched for the best-fitting values of c and w_1 in each condition.

Following Nosofsky (1987), the EBRW and the GCM were fitted to the color categorization response probability data by searching for the free parameters that maximized the log-likelihood of the data (or, equivalently, that minimized the negative of the log-likelihood). The results are shown in Table E1, where a smaller value of $-\ln L$ indicates a better fit. As auxiliary measures, we also report the root-mean-squared deviation between predicted and observed classification probabilities as well as the percentage of variance accounted for in each condition. (The best-fitting parameters are available from the first author on request.) Both models provide excellent quantitative fits to the data. Regardless of the specific version of the models tested (with or without stimulus bias terms), the EBRW gives a slightly better overall fit to the data across the six conditions than does the GCM. Again, as explained in the text, we are not claiming that the EBRW represents an improvement over the GCM in this regard because an additional free parameter (b) is involved. The purpose of this analysis is simply to demonstrate that versions of the EBRW can produce good fits to response probability data in situations in which the GCM has previously been successful.

(*Appendixes continue on next page*)

Table E1
Fits of the EBRW and the GCM to Nosofsky's (1987) Color Categorization Response Probability Data

	GCM			EBRW		
Condition	−ln L	*RMSD*	% Var	−ln L	*RMSD*	% Var
			Models with stimulus bias terms			
Saturation A	44.6	.055	97.8	38.8	.041	98.8
Saturation B	71.0	.050	98.0	49.6	.026	99.5
Brightness	56.1	.023	99.7	56.6	.022	99.7
Criss-cross	43.4	.035	99.1	46.4	.042	98.7
Pink–brown	58.0	.035	99.1	49.0	.026	99.5
Diagonal	66.9	.046	97.8	73.4	.048	97.6
			Models without stimulus bias terms			
Saturation A	46.0	.059	97.5	41.1	.046	98.5
Saturation B	58.8	.041	98.7	43.4	.018	99.7
Brightness	60.3	.026	99.7	45.1	.017	99.9
Criss-cross	50.4	.042	98.7	52.9	.043	98.6
Pink–brown	70.9	.045	98.6	70.6	.045	98.6
Diagonal	99.6	.064	95.8	106.7	.068	95.2

Note. EBRW = exemplar-based random walk; GCM = generalized context model; ln L = log-likelihood; *RMSD* = root mean squared deviation; % Var = percentage of variance accounted for.

Received June 26, 1995
Revision received April 29, 1996
Accepted June 26, 1996 ■

New Editor Appointed for *Contemporary Psychology*: 1999-2004

The Publications and Communications Board of the American Psychological Association announces the appointment of Robert J. Sternberg (Yale University) as editor of *Contemporary Psychology*, for a 6-year term beginning in 1999. The current editor, John H. Harvey (University of Iowa), will continue as editor through 1998.

All reviews are written by invitation only, and neither the current editor nor the incoming editor receives books directly from publishers for consideration. Publishers should continue to send two copies of books for consideration, along with any notices of publication, to PsycINFO Services Department, APA, Attn: *Contemporary Psychology* Processing, P.O. Box 91700, Washington, DC 20090-1700 or (for UPS shipments) 750 First Street, NE, Washington, DC 20002-4242.