Effects of similarity and practice on speeded classification response times and accuracies: Further tests of an exemplar-retrieval model

ROBERT M. NOSOFSKY and LEOLA A. ALFONSO-REESE
Indiana University, Bloomington, Indiana

Observers were tested in a perceptual category-learning experiment in which they were instructed to make classification decisions as rapidly as possible without making errors. Nosofsky and Palmeri's (1997b) exemplar-based random walk (EBRW) model of speeded classification was tested for its ability to fit the classification response times and accuracies. The authors demonstrated that the EBRW model provided good quantitative fits to the mean response times and accuracies associated with individual objects as a function of their locations in a multidimensional similarity space and as a function of practice in the task. Preliminary evidence was also obtained that stimulus-specific adjustments in the random walk response criteria may have occurred during the course of learning.

Numerous powerful models of multidimensional perceptual classification exist in the field today, providing detailed quantitative accounts of category learning, performance, and generalization (e.g., Anderson, 1991; Ashby & Lee, 1991; Estes, 1994; Gluck & Bower, 1988; Hintzman, 1986; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986). However, most of these models are limited to predicting the output of classification, such as choice probabilities, confidence ratings, typicality judgments, and so forth. It is only recently that investigators have attempted to extend these models to account for the actual time course of categorization decision making and to predict classification response times (e.g., Ashby, Boynton, & Lee, 1994; Lamberts, 1995, 1998; Maddox & Ashby, 1996; Nosofsky & Palmeri, 1997a, 1997b). Because response times provide a window into understanding the nature of people's category representations and decision processes, it is important to extend current models to account for this form of data.

The purpose of this research was to provide further tests of a recently proposed exemplar-retrieval model for predicting the time course of classification decision making. Nosofsky and Palmeri's (1997b) *exemplar-based random walk* (EBRW) model is an integrated model that combines key elements of Nosofsky's (1986) *generalized context model* of perceptual classification and Logan's (1988) *instance theory of automaticity*. According to the EBRW, people represent categories by storing individ-

ual exemplars of the categories in memory. When a test item is presented to be classified, the stored exemplars race to be retrieved, with rates being determined by how similar they are to the test item. These retrieved exemplars then provide evidence that enters into a random walk process for making classification decisions (see, e.g., Busemeyer, 1985; Link & Heath, 1975; Luce, 1986; Ratcliff, 1978). When sufficient evidence is obtained, the observer makes the appropriate categorization response.

Nosofsky and Palmeri (1997a, 1997b; Palmeri, 1997) demonstrated that the EBRW model provided excellent accounts of categorization response times in a variety of experimental paradigms. In this research, we pursue one of these initial avenues of research in greater detail, so as to provide more extended, rigorous tests of the model. Specifically, we test the ability of the EBRW model to simultaneously predict classification response times and accuracies for individual objects on the basis of their locations in a multidimensional similarity space and as a function of practice in the task. We go beyond previous work by requiring the model to simultaneously fit a rich set of choice probability data in addition to the mean response times and by testing the ability of the model to predict speed-up functions and accuracy improvements for individual objects that occur with learning.

OVERVIEW OF THE EBRW

Because a full presentation of the EBRW model has been provided in several previous articles (Nosofsky, 1997; Nosofsky & Palmeri, 1997a, 1997b; Palmeri, 1997), we only briefly summarize it here. In the EBRW model, each exemplar is represented as a point in a multidimensional psychological space. Let x_{im} denote the value of exemplar i on psychological dimension m. (The x_{im} coordinate values are usually derived by conducting various similarity-scaling studies—see Nosofsky, 1992, for a re-

view.) The distance between exemplars i and j is computed by using a Euclidean metric,

$$d_{ij} = \left[\sum w_m \mid x_{im} - x_{jm} \mid^2\right]^{1/2}, \tag{1}$$

where the w_m $(0 \leq w_m < 1, \sum w_m = 1)$ are free parameters representing the *attention weight* given to each dimension m. The similarity between exemplars i and j (s_{ij}) is an exponential decay function of psychological distance (Shepard, 1987), given by

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

where c is an overall scaling parameter. Each exemplar resides in memory with strength M_j. (The memory strengths may be affected by factors such as recency of presentation, study time, intrinsic salience, and so forth.) When an item is presented to be classified, it causes all exemplars to be *activated*. The activation for exemplar j, given presentation of item i, is given by

$$a_{ij} = M_j \cdot s_{ij}. \tag{3}$$

Thus, the exemplars that are most highly activated are those that have the greatest memory strengths and are highly similar to the test item.

When item i is presented, all the category exemplars stored in memory race to be retrieved. The race times are exponential random variables, with rates being proportional to the degree to which exemplar j is activated by item i (Bundesen, 1990; Logan, 1997; Marley, 1992; Marley & Colonius, 1992). Thus, the probability density that exemplar j completes its race at time t, given presentation of item i, is given by

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

This assumption formalizes the idea that, although the retrieval process is stochastic, the exemplars that tend to race most quickly are those that are most highly activated by the test item.

In addition, *background-noise elements* are assumed to exist in memory at the start of training that are randomly associated with the alternative categories (Estes, 1994; Nosofsky, Kruschke, & McKinley, 1992). The background-noise elements of each category race to be retrieved with rate b, where the rate is independent of the test item that is presented and the amount of learning that has occurred.

Finally, the exemplar (or background-noise element) that "wins the race" on any given step is retrieved and enters into a random walk decision process. In a two-category situation, the random walk process is formalized as follows. First, there is a random walk counter with initial value zero. The observer establishes criteria representing the amount of evidence needed to make either a Category A response $(+A)$ or a Category B response $(-B)$. Suppose that exemplar x wins the race on a given step. If x belongs to Category A, the random-walk counter is increased by unit value in the direction of $+A$, whereas if x belongs to Category B, the counter is decreased by unit

value in the direction of $-B$. If a background element wins the race, the direction of unit change for the counter is chosen randomly. If the counter reaches either criterion $+A$ or criterion $-B$, the appropriate categorization response is made. Otherwise, a new race is initiated, another exemplar is retrieved, and the process continues.

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

$$T_{\text{step}} = \alpha + t_w, \tag{5}$$

where α is a constant term associated with each step, and t_w is the time that it takes to retrieve the "winning" exemplar. A psychological interpretation for the α parameter is that it represents the time needed to extract the category label associated with each exemplar on each retrieval and then accumulate this information on the random walk counter.

Given these processing assumptions, Nosofsky and Palmeri (1997b) showed that, on each step of the random walk, the probability (p_i) that the counter is increased in the direction of Category A is given by

$$p_i = (S_{iA} + b)/(S_{iA} + S_{iB} + 2b), \tag{6}$$

where S_{iA} denotes the summed activation of all the currently stored Category A exemplars given presentation of item i, and likewise for S_{iB}. (The probability that the counter is decreased in the direction of Category B is given by $q_i = 1 - p_i$.) So, for example, as the summed activation of Category A exemplars increases, the probability of retrieving Category A exemplars and thereby moving the counter in the direction of $+A$ increases. In addition, Nosofsky and Palmeri (1997b) showed that the expected time to take each step in the random walk is given by

$$E(T_{\text{step}} \mid i) = \alpha + 1/(S_{iA} + S_{iB} + 2b). \tag{7}$$

The processing assumptions just outlined yield a simple random walk from which it is straightforward to derive analytic predictions of classification choice probabilities and mean response times for each stimulus at any given stage of the learning process. The relevant equations are summarized by Nosofsky and Palmeri (1997b, pp. 269–270).

The main conceptual predictions of the EBRW model follow naturally from the idea that the duration of the random walk is determined jointly by the total number of steps required to initiate a categorization response and the speed with which each of the individual steps is made. One main prediction is that the most rapid and accurate classification decisions should be made for those items that are highly similar to the exemplars of their own category and dissimilar to the exemplars of the alternative category. Under such conditions, each retrieved exemplar will tend to come from the same category, so the random walk will march in consistent fashion to a single criterion. For example, an item that is highly similar to only the exemplars of Category A will result in a large value of p_i in Equation 6, 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 alternative category.

A second main prediction is that increased experience with category exemplars should facilitate performance, both by speeding decision time and by improving accuracy. Note that, as learning proceeds, the summed activations S_{iA} and S_{iB} grow larger and larger in magnitude. The reason is that each time an item is presented to be classified, a new exemplar trace is stored in memory.[1] As the summed activations grow large, relative to the background-noise rate b, the random walk decision process becomes more efficient (see Equation 6). Early in training, before many exemplars are stored in memory, the background-noise elements dominate the retrieval process, so the random walk counter wanders back and forth in a haphazard manner. As training continues, the stored category exemplars begin to dominate the retrieval process, and there are fewer retrievals of background-noise elements. Formally, the value of p_i in Equation 6 moves away from .50 (at the start of training) toward a limit of $S_{iA}/(S_{iA}+S_{iB})$ (with infinite training); so, the random walk counter moves more consistently toward the appropriate category criterion. This process results in increasingly faster response times and more accurate categorization decisions.

A second reason that practice facilitates response times is that, as the summed activation $S_{iA}+S_{iB}$ grows larger, the expected retrieval time of the winning exemplar grows smaller (see Equation 7). These faster retrieval times result in faster individual steps in the random walk process. This aspect of the EBRW model is essentially the same as that in Logan's (1988, 1992) instance theory of automaticity. As explained by Logan (1988, 1992), the more instances that simultaneously race to be retrieved, the faster the winning retrieval time tends to be (see also Raab, 1962). Intuitively, the larger the number of instances that participate in the race, the greater is the probability that at least one of the retrieval times will be particularly fast. In summary, according to the EBRW model, practice speeds up response times for two reasons: (1) The retrieved exemplars result in more consistent steps toward the correct criterion, and (2) the time to take each individual step in the random walk gets faster.

EXPERIMENT

The purpose of this experiment was to test the EBRW model for its ability to simultaneously predict classification response times and accuracies for individual stimuli on the basis of their location in a multidimensional similarity space and as a function of practice in the task. The experiment was a partial replication and extension of an earlier one conducted by Nosofsky and Palmeri (1997b, Experiment 1). The stimuli were a set of 12 computer-generated colors. According to the Munsell system, all the colors were of a roughly constant red hue but varied in their brightness and saturation. A schematic illustration of the color set that was used is shown in Figure 1. In the diagram, colors enclosed by circles denote members of Category A, and colors enclosed by squares denote members of Category B. On each trial, a subject was presented with a color and classified it into Category A or B as rapidly as possible without making errors. Corrective feedback was provided on every trial.

In the experiment reported by Nosofsky and Palmeri (1997b), 3 individual observers were tested in this task for an extensive period of time, and the EBRW model was fitted to the data of each individual observer. Nosofsky and Palmeri (1997b) demonstrated that the EBRW model gave good fits to the mean classification response times for each of the colors and also yielded good quantitative predictions of the overall speed-ups in classification response times that were observed as a function of practice in the task.

The earlier experiment was limited in several respects, however. First, because the individual observers were highly experienced and the stimuli were not very confusable, overall accuracy was extremely high in the experiment. Therefore, it was not possible to obtain rigorous tests of the ability of the EBRW model to fit the classification choice probabilities, because all the accuracies were close to ceiling. In the present experiment, our goal was to test the EBRW model on its ability to simultaneously fit both the classification response times *and* the choice probabilities. Second, in the previous experiment, Nosofsky and Palmeri (1997b) modeled only the overall speed-ups in response times averaged over all the stimuli in the task. Because only 3 observers were tested, there were insufficient data to test the model on its ability to predict speed-ups for *individual* stimuli. Our second goal in the present experiment was to model the individual stimulus speed-up functions, as well as to model how accuracy improved for individual stimuli as a function of training.

To meet these goals, in the present experiment, instead of testing only a few highly experienced observers over many sessions, we tested a large number of observers, each of whom participated for only a single session. We then used the EBRW model to fit the data averaged over observers. Because we modeled the early learning data, there was a large number of errors, thereby providing a challenging test of the ability of the EBRW model to jointly predict classification response times and choice probabilities. Also, the large data set allowed us to study how response times and accuracies for individual stimuli changed as a function of practice in the task. Because the parameters across individual subjects may be expected to vary, the fits to the averaged data should be interpreted with some caution. We remark, however, that in their previous tests of the EBRW model, Nosofsky and Palmeri (1997b) found that each of the individual observers showed extremely similar patterns of mean response times for the colors as a function of similarity and
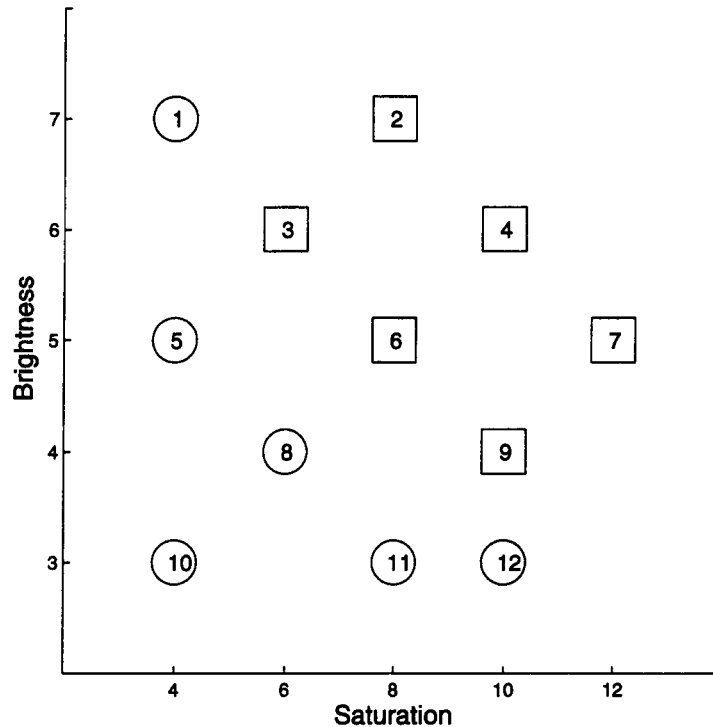
Figure 1. Schematic illustration of the color set that was used in the experiment. In the diagram, colors enclosed by circles denote members of Category A, and colors enclosed by squares denote members of Category B.

practice, and the EBRW model accounted extremely well for the individual-observer response time data. Thus, given our goals, we viewed the present approach as reasonable and as serving as an important complement to the individual observer study already conducted by Nosofsky and Palmeri (1997b).

Method

Subjects. Two groups of subjects were tested. The first group consisted of 145 undergraduates from Indiana University, who participated in partial fulfillment of an introductory psychology course requirement. Because the overall performance levels for many observers in this group were low, we tested a second group of 32 Indiana undergraduates, who competed for monetary rewards. All subjects claimed to have normal color vision.

Stimuli. The stimuli were a set of 12 colors presented on a computer screen. Nosofsky and Palmeri (1997b) used extensive pilot work to construct a set of stimuli that approximately matched the Munsell color configuration illustrated in Figure 1. (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 is illustrated in Figure 1, the stimuli were divided by the experimenters into two categories, A (circles) and B (squares). Because our theoretical analyses of the data make use of a multidimensional scaling (MDS) solution for the colors derived from similarity ratings data, the precise correspondence with the Munsell configuration illustrated in the figure 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 model to the response time and accuracy data.

The stimuli were generated on CompuAdd 14-in. monitors (Model 51109) by adjusting the red, green, and blue (RGB) color channels on Dell 486 computers. The RGB values corresponding to each color were reported by Nosofsky and Palmeri (1997b, Appendix B). Each color occupied a 6 × 6 cm square surrounded by a white background. The subjects sat approximately 60 cm from the computer monitor, so the visual angle subtended by each stimulus was approximately 6°. The subjects entered responses by pressing buttons on the computer keyboard. Response times were measured by using the internal millisecond-accuracy PC timer.

Procedure. The stimulus presentation schedule was organized into 32 blocks of 12 trials, with each color presented once per block. Order of presentation of the colors was randomized within blocks.

Stimulus displays were response terminated. Each stimulus display was followed by 2 sec of feedback. Following a 500-msec interstimulus interval, the next color was displayed. The subjects were instructed to rest their index fingers on the Categories A and B response buttons throughout the testing session and to make their responses as rapidly as possible without making errors. Members of Group 2 were given the following bonus instructions: "In this experiment you have a chance to win $25 if you are one of the top three subjects . . . In order to QUALIFY for the $25 prize, you must give the correct response to over 85% of all the 384 trials. In order to WIN the prize, you must be one of the three fastest subjects."

Following the speeded classification task, each subject was tested in a similarity-scaling study. All 66 unique pairs of colors were presented, 1 per trial, in a random order. The 6 × 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. The subjects 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

Preliminary culling of subjects. Initial inspection of the data from the first group of subjects revealed that many individuals failed to learn the categorization assignments of the colors. Because the 3 observers from Nosofsky and Palmeri's (1997b) earlier study had all performed with extremely high accuracy, we suspected that many of the subjects from this group were insufficiently motivated. Therefore, we tested the second group of observers who competed for a monetary reward. We computed the overall accuracy scores achieved by each individual observer during the course of the entire testing session. After preliminary inspection of these distributions of accuracy scores, we decided to focus our modeling on the data from roughly the top third of performers from Group 1 (45 observers) and the complete set of observers from Group 2 (32 observers). Also, an additional subject from Group 1 was dropped because the subject's overall mean response time on the colors was an extreme outlier with respect to all the remaining subjects. Thus, the main data set consisted of the results from a total of 76 observers.

The overall mean response time for the subjects from Group 2 was faster than that for the top third of the subjects from Group 1 (753.2 vs. 843.7 msec), whereas the Group 1 subjects displayed slightly higher percentage correct scores (90.6 vs. 88.7). Thus, the top Group 1 observers probably placed relatively more emphasis on accuracy than on speed, as compared with the Group 2 observers. Importantly, however, the pattern of response times and accuracies was extremely similar across the two groups. For the 12 colors, the correlation between the mean response times across the two groups was .991, and the correlation between the accuracies was .986. Similar speed-up functions were also observed for the two groups as a function of practice. Given the extremely similar pattern of results, we decided to collapse across the two groups in conducting our primary data analyses and formal model fits. Because a large number of observers from Group 1 were removed from these initial analyses, we emphasize that the ensuing results and conclusions pertain to observers who we believe are highly motivated (or who are able to perform well even with low motivation). In a later section of our article, we consider the data from those observers who did not perform as well.

Multidimensional scaling. An MDS solution for the colors was derived by fitting the standard Euclidean model to the matrix of averaged similarity ratings ob-
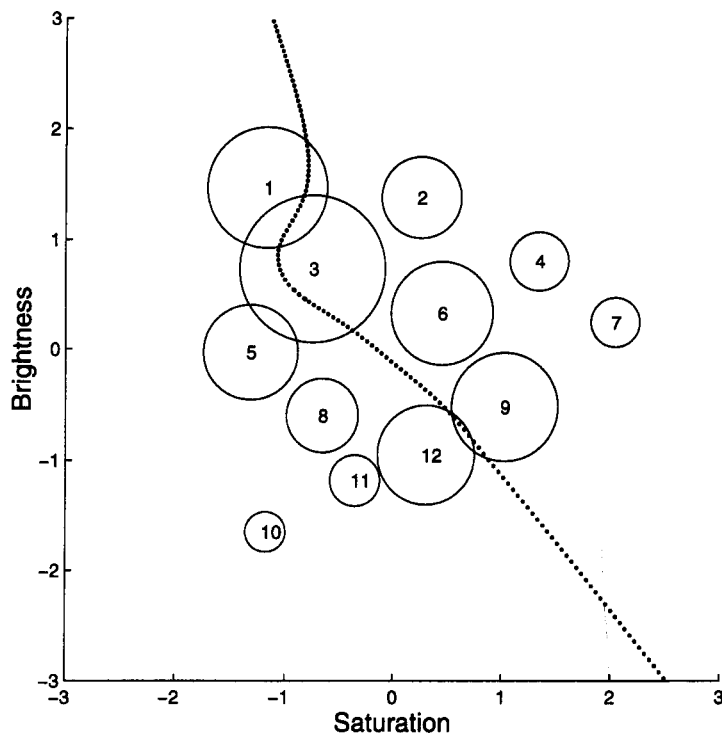

Figure 2. Mean response times, averaged over all blocks of learning, for the 12 colors. The diameter of the circle enclosing each color is linearly related to the mean response time. The center of each circle represents the MDS coordinates of each color derived from the similarity ratings. The dashed curve gives the locus of points that have equal summed similarity to the exemplars of Categories A and B, according to the exemplar-based random walk model.

tained from the 76 observers. A two-dimensional scaling solution yielded a stress of .027 and accounted for 99.5% of the variance in the ratings. Adding a third dimension provided only small improvements in fit (stress = .022, percent variance = 99.7). Thus, for simplicity in the model fitting, we used the two-dimensional solution. The MDS solution is shown graphically in Figure 2, where the center of each circle represents the MDS coordinates of each color. Because the orientation of the Euclidean scaling solution is arbitrary, the solution was rotated so as to place it into maximal correspondence with the Munsell scaling configuration (Figure 1) that served as a model for generating the colors.

Speeded classification. Data were discarded from all trials in which response times were greater than 5,000 msec or less than 100 msec. Fewer than 0.2% of the trials were discarded.

To analyze the speeded classification performance and create a data set suitable for quantitative model fitting, we combined the 32 individual blocks into 4 main grouped blocks: Group 1 = Blocks 1–8, Group 2 = Blocks 9–16, Group 3 = Blocks 17–24, and Group 4 = Blocks 25–32. We then computed the mean response time and accuracy for each color in each of the 4 grouped blocks. These data are reported in Tables 1A and 1B, as are the overall mean response times and accuracies averaged over all blocks.

The mean classification response times for each of the individual colors, averaged over all blocks, are illustrated graphically in Figure 2. In the figure, the diameter of each circle is linearly related to the mean response

Table 1A
Predicted and Observed Mean Response Times
for Each of the 12 Colors During Grouped Blocks 1–4

Color	Grouped Block				
	1	2	3	4	Average
1	1,089.34	1,007.57	986.65	977.33	1,015.22
	1,050.04	1,057.67	920.34	892.26	980.08
2	887.59	750.46	728.76	719.59	771.60
	919.72	798.48	726.07	688.29	783.14
3	1,115.51	1,078.23	1,067.36	1,062.45	1,080.89
	1,177.56	1,167.11	1,090.11	1,029.48	1,116.06
4	770.94	635.72	617.06	609.26	658.24
	807.96	641.49	617.80	627.96	673.80
5	908.30	780.43	759.61	750.77	799.78
	949.05	853.13	815.70	784.54	850.61
6	953.58	848.08	830.52	823.03	863.80
	1,033.96	908.68	809.34	825.16	894.29
7	778.89	624.07	602.38	593.33	649.67
	745.59	591.97	570.27	595.38	625.80
8	832.68	717.14	700.24	693.12	735.79
	828.34	735.37	686.78	727.15	744.41
9	1,037.01	943.74	925.26	917.25	955.81
	988.89	936.86	862.60	872.49	915.21
10	707.47	556.73	537.16	529.04	582.60
	657.48	557.44	559.58	541.96	579.12
11	765.09	644.35	627.75	620.80	664.50
	701.93	616.47	611.11	603.93	633.36
12	979.51	875.74	857.48	849.65	890.60
	928.63	890.58	836.81	828.17	871.05

Note—Top row, predicted; bottom row, observed.

Table 1B
Predicted and Observed Accuracies
for Each of the 12 Colors During Grouped Blocks 1–4

Color	Grouped Block				
	1	2	3	4	Average
1	64.81	78.46	80.42	81.24	76.23
	56.34	77.30	80.10	80.99	73.68
2	88.29	96.01	96.64	96.89	94.46
	87.79	97.36	99.01	98.68	95.71
3	56.54	67.79	69.51	70.23	66.02
	54.60	63.16	70.89	77.69	66.58
4	94.21	98.85	99.11	99.21	97.85
	94.07	99.18	100.00	99.51	98.19
5	86.47	94.35	95.09	95.38	92.83
	84.79	94.56	94.08	97.36	92.70
6	83.84	91.70	92.52	92.85	90.23
	81.52	92.90	92.93	95.38	90.68
7	93.52	99.04	99.31	99.41	97.82
	93.73	99.18	99.83	99.51	98.06
8	91.28	96.56	97.00	97.17	95.50
	92.74	98.68	98.35	97.03	96.70
9	74.67	85.28	86.57	87.10	83.40
	73.64	82.01	86.30	87.81	82.44
10	95.55	99.57	99.71	99.76	98.65
	96.21	99.51	99.67	99.67	98.76
11	94.19	98.34	98.62	98.73	97.47
	95.05	99.18	98.85	99.17	98.06
12	80.81	89.48	90.46	90.86	87.90
	80.89	86.51	89.80	90.43	86.91

Note—Top row, predicted; bottom row, observed.

time. The dashed curve in Figure 2 gives the locus of points that have equal summed similarity to the exemplars of Categories A and B, according to the EBRW model—see the subsequent theoretical analyses. (This boundary plays no direct role in the formal modeling; it is illustrated only for descriptive convenience.) It is evident from inspection that the mean classification response times for the colors follow a highly regular pattern. In general, the further a color is from the boundary of equal summed similarity, the faster is its mean response time. Furthermore, the correlation between the overall accuracies and the mean response times for the colors was $r = -.92$. Thus, in general, the faster the mean response time for a color, the higher was its associated accuracy.

Figure 3 plots the mean classification response times for the individual colors as a function of the four grouped blocks of practice. It is evident from inspection that, in general, the response times for the colors got faster with each subsequent grouped block of testing. (Two clear exceptions to this pattern, which we will discuss subsequently, were that the mean response times for Colors 1 and 3 were almost the same across Grouped Blocks 1 and 2.) Likewise, the accuracies for each of the individual colors increased regularly as a function of the grouped blocks of practice, as we illustrate graphically in Figure 4. The critical question now concerns the ability of the EBRW model to fit the mean response times, accuracies, learning functions, and speed-up functions associated with the individual colors.
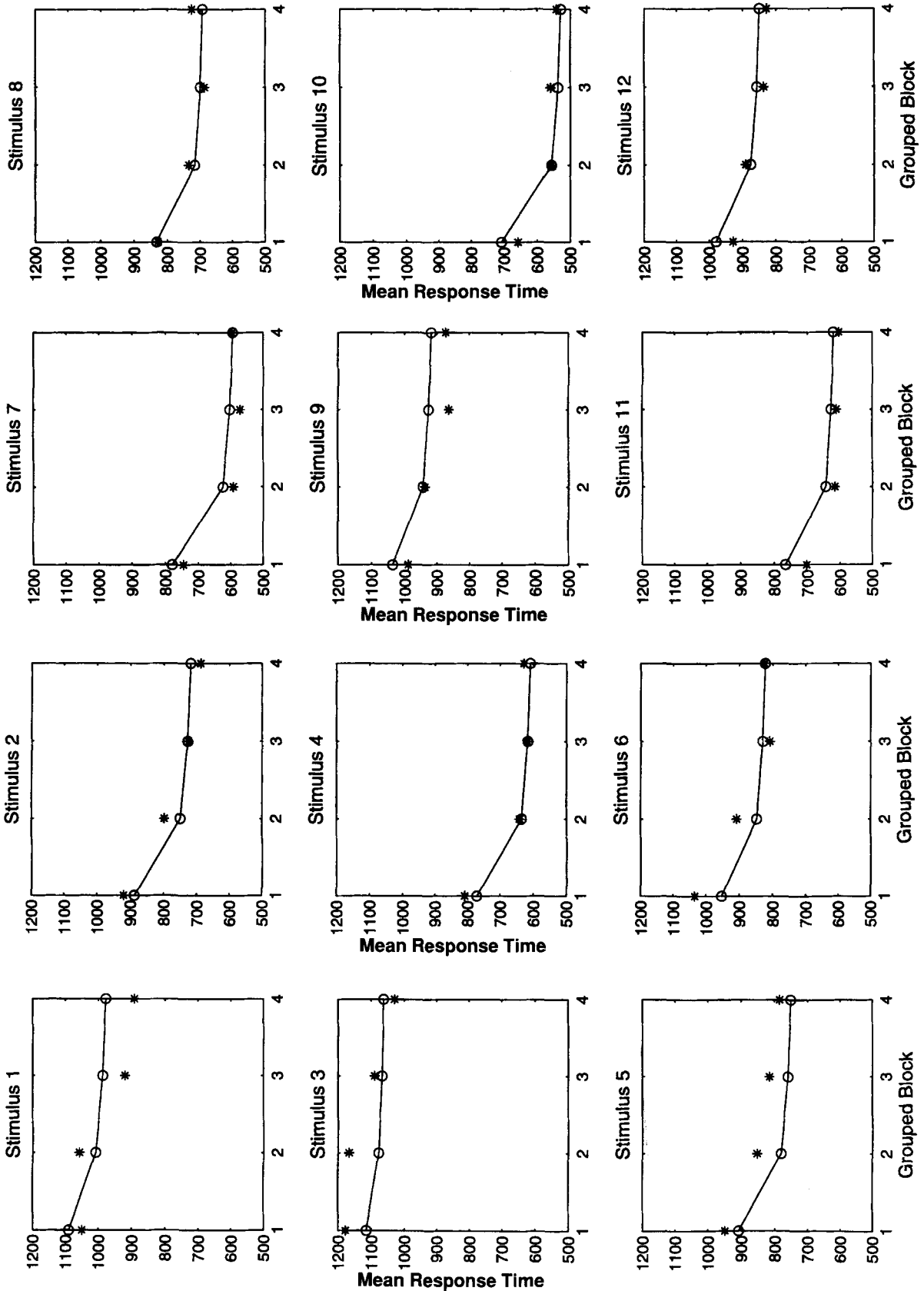
Figure 3. Observed mean classification response times for the individual colors as a function of the four grouped blocks of practice (asterisks). The solid curve in each panel gives the predicted mean response times from the exemplar-based random walk model.
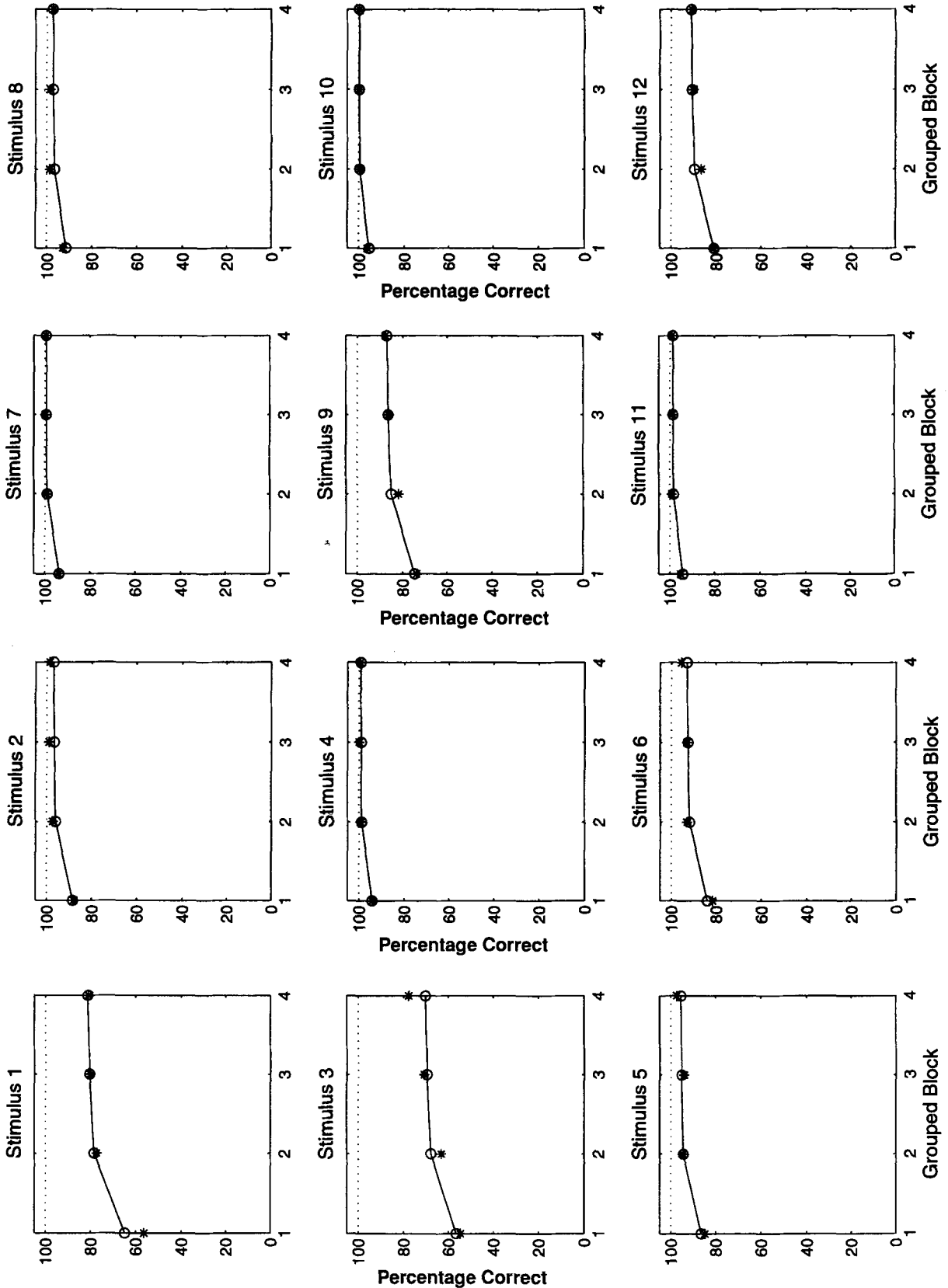
Figure 4. Observed mean percentage of correct classification responses for the individual colors as a function of the four grouped blocks of practice (asterisks). The solid curve in each panel gives the predicted percentages from the exemplar-based random walk model.

Theoretical Analysis

Fits of the EBRW model. Using the previously derived MDS solution, we fitted the EBRW model simultaneously to the mean response times and accuracies observed for each individual color as a function of the four grouped blocks of practice (i.e., we fitted the model to the data in Tables 1A and 1B). In fitting the model, we assumed that on each individual block of testing, an additional token of each exemplar was stored in memory. The first step in the analysis was to use the model to predict the mean response time and accuracy for each color in each individual block. These individual block predictions were then averaged to create the grouped block predictions that were compared with the observed data.[2]

The free parameters in the model included: the sensitivity parameter c (Equation 2); the attention weight w_1 (Equation 1); the step-time constant α (Equation 5); the background-noise constant b (Equation 6); the random walk criteria $+A$ and $-B$;[3] and regression constants μ and k that transform the predictions of the EBRW model, which are in arbitrary units, into milliseconds. (The y-intercept μ can be interpreted as the mean response time associated with all residual stages of processing, such as encoding and response execution. The EBRW models only the decision-making stage. The parameter k is a scaling constant that transforms the decision time predictions into milliseconds.) Note that these eight free parameters are being used to predict 96 data points, 48 mean response times, and 48 accuracies (see Tables 1A and 1B).
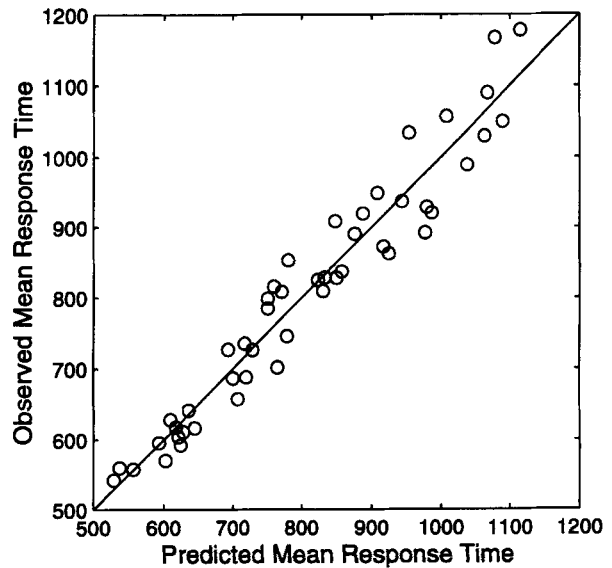
Any method for combining the fits to the response time and accuracy data is arbitrary. Following preliminary exploration, we fitted the EBRW model to the data by conducting a computer search for the free parameters that minimized the weighted sum of squared deviations (SSD) between predicted and observed response times

a) EBRW Model Fit (complete data set)
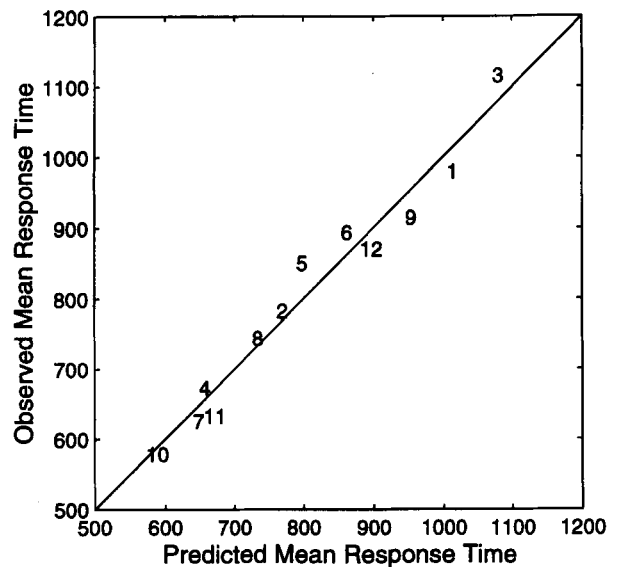

b) EBRW Model Fit (averaged data)

Figure 5. Panel A: Scatterplot of the complete set of observed against predicted mean response times from the exemplar-based random walk model. Panel B: Scatterplot of observed against predicted mean response times averaged over the four grouped blocks.

Table 2
Best-Fitting Parameters and Summary Fits of the EBRW to the Mean Response Times and Accuracy Data

Best-Fitting Parameters	
c	1.556
w_1	0.467
b	1.296
α	3.504
A	3.175
B	2.799
μ	100.000
k	32.343
Summary Fits	
RT SSD	82,293.3
RT RMSD	41.4
RT % Var.	93.9
Acc. SSD	221.7
Acc. RMSD	2.1
Acc. % Var.	96.4

Note—The value $\mu = 100$ was the minimum setting allowed for the mean residual time in the parameter search. See note 3 for a discussion regarding the real-valued settings of the response criteria parameters A and B. RT, response time; SSD, sum of squared diviations; RMSD, root-mean squared deviation; Acc., accuracy; % Var., percentage of variance accounted for.

and predicted and observed accuracies, where the accuracy data received 100 times the weight of the response time data. (This procedure resulted in response time and accuracy [weighted] SSDs of roughly the same order of magnitude.) The best-fitting parameters and summary fits are reported in Table 2.

Predictions of the mean response times are shown along with the observed data in Table 1A and are illustrated graphically in Figure 3 (solid line curves). The EBRW

model does a reasonably good job of predicting the over-all mean response times for the colors, as well as the fa-cilitation in response times that occurred with practice. Figure 5A shows a scatterplot of the complete set of ob-served against predicted mean response times. The model accounted for 93.9% of the variance in these data. Fig-ure 5B shows the same scatterplot after averaging over the observed and predicted mean response times from

a) EBRW Model Fit (complete data set)

b) EBRW Model Fit (averaged data)

Figure 6. Panel A: Scatterplot of the complete set of observed against predicted mean accuracies from the exemplar-based ran-dom walk model. Panel B: Scatterplot of observed against pre-dicted mean accuracies averaged over the four grouped blocks.

the four grouped blocks. The model accounted for 96.4% of the variance in these data.

The predictions of the accuracy data are shown along with the observed accuracies in Table 1B and are illus-trated graphically in Figure 4 (solid line curves). Just as was the case for the response time data, the EBRW model yields good quantitative predictions of the overall accu-racies and the individual stimulus learning functions. Figure 6A shows a scatterplot of the complete set of ob-served against predicted mean accuracies. The model ac-counted for 96.4% of the variance in these data. Figure 6B shows the same scatterplot after averaging over the ob-served and predicted accuracies from the four grouped blocks. The model accounted for 99.0% of the variance in these data.

In summary, the EBRW model gives simultaneous good quantitative fits to the mean response times, accuracies, speed-up functions, and learning functions associated with the individual colors in this speeded classification task.

Figure 7 plots the averaged predicted and observed mean response times for all 12 colors as a function of grouped blocks of practice. This figure reveals some sys-tematic shortcomings in the EBRW model, with the model predicting too great an overall speed-up from Grouped Block 1 to Grouped Block 2 and too small a speed-up from Grouped Block 2 to Grouped Block 3. At least part of the problem can be traced to the results for Colors 1 and 3 (see Figure 3). As was noted previously, the mean response times for these colors were either un-changed or increased slightly from Grouped Block 1 to Grouped Block 2. There is no mechanism in the present version of the model that can predict such an effect. A potential explanation of the effect is that it might involve a speed–accuracy tradeoff. Colors 1 and 3 are the most difficult objects in the stimulus space to classify. The ob-servers may have recognized that they were making a substantial number of errors on these colors. A reason-able strategy, therefore, would be to take extra care when classifying objects from this region of the stimulus space by increasing the magnitude of the response criteria gov-erning the random walk process. Such a strategy would lead to improved accuracy for these stimuli, but at the cost of slower response times. Note that implementing such a strategy requires a form of preprocessing in which the observer learns to adjust the magnitude of the response criteria contingent on the general region of the similarity space in which a stimulus is presented. In future work, we will need to formalize a mechanism that allows for this type of stimulus-specific adjustment of the response criteria.[4]

Converging evidence for this hypothesis about the role of stimulus-specific criterion shifts is provided by ex-amining separately the mean response times associated with correct and error responses. These data are reported in Table 3 for Grouped Block 1. (Subsequent grouped blocks had too few errors to allow for a meaningful analy-sis.) It is clear from an examination of Table 3 that, in general, error response times were longer than correct
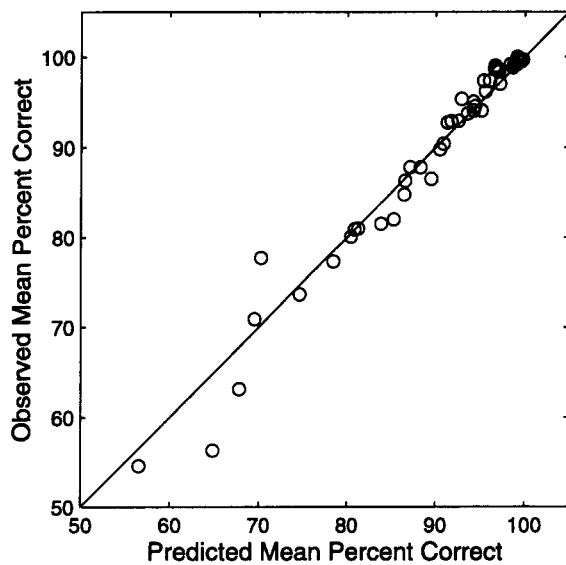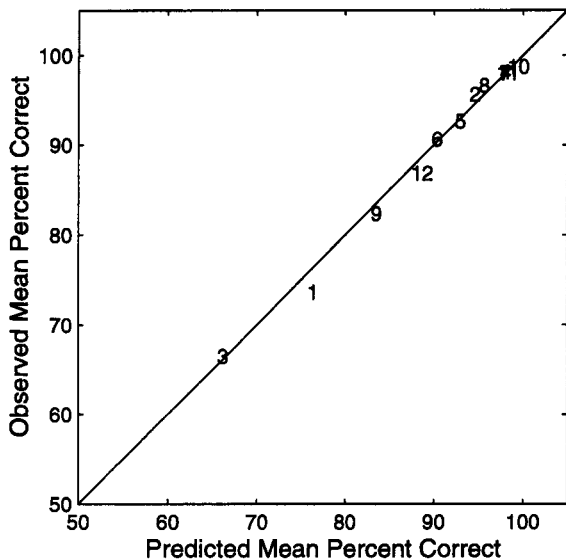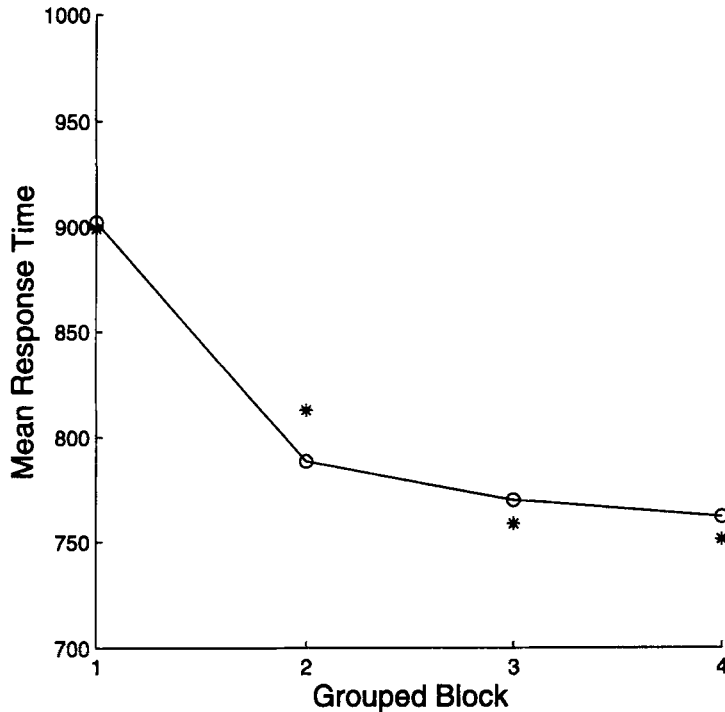
Figure 7. Observed mean response times for all 12 colors as a function of the four grouped blocks of practice (asterisks). The solid curve gives the predictions from the exemplar-based random walk model.

response times, often times substantially so. (Crucially, the only exceptions to this pattern occurred for Colors 1 and 3, a result that we will discuss subsequently.) The finding that correct response times were, on average, faster than error response times is unsurprising and is to be expected on the basis of numerous factors. One factor, for example, is that the data are averaged over observers, and highly motivated observers would be expected to show faster mean response times and more accurate responding than poorly motivated observers. Another factor is that the data are averaged over trials of learning, and as learn-

Table 3
Mean Response Times Associated With
Correct and Error Responses in Grouped Block 1

Color	Correct		Error	
	M	n	M	n
1	1,095.0	342	992.0	265
2	903.5	532	1,036.7	74
3	1,218.8	332	1,128.0	276
4	790.6	571	1,083.2	36
5	923.5	513	1,091.4	92
6	979.1	494	1,276.2	112
7	726.9	568	1,025.4	38
8	797.7	562	1,220.3	44
9	950.3	447	1,096.8	160
10	632.3	584	1,298.0	23
11	684.4	576	1,038.8	30
12	867.7	491	1,186.5	116

Note—Values for n give the number of observations on which each mean response time is based.

ing proceeds, responding gets faster and more accurate. In general, any factors that lead to variability in the values of p_i and q_i across observers and/or trials cause the EBRW model to predict faster correct responses than error responses in the averaged data. (As explained previously, when the magnitude of p_i is large, the EBRW model predicts fast and accurate responding, whereas when the magnitude of p_i is small [close to .50], the EBRW model predicts slow and inaccurate responding—see Ratcliff, Van Zandt, & McCoon, in press, for a more extended discussion regarding the role of variability in random walk rate parameters in the modeling of correct and error response times.) Among the numerous factors that may contribute to variability in the random walk rate parameters are individual differences across observers, learning, lapses of attention, momentary forgetting of exemplar assignments, experimental disruptions, sensory/perceptual noise associated with the stimulus presentations, and so forth.

Conversely, to the extent that there is variability in the settings of the response criteria (+A and −B) across trials, error responses are expected to be faster than are correct responses (Ratcliff et al., in press). On trials in which the criteria are set at a low magnitude, responses will tend to be fast but inaccurate, whereas on trials in which the criteria are set at a high magnitude, responses will tend to be slow but accurate. Averaging data across such trials will result in error responses that are faster than correct responses. Interestingly, the only two stim-

uli that showed faster error responses than correct responses in the averaged data were Colors 1 and 3 (see Table 3). It is precisely these two stimuli for which we previously hypothesized that large shifts in the response criteria may have occurred during the learning process. Thus, both the shape of the individual object speed-up functions and the pattern of mean decision times for correct and error responses support the view that there were shifts in the placements of the response criteria during the learning process when observers classified Colors 1 and 3.

Subgroup analyses. In the primary modeling analyses reported thus far in our article, we considered the data from only the top performers in our experiment. In this section, we consider the data from those observers who did not perform as well.[5] Our goal was both to test the ability of the EBRW model to model the performance of observers displaying lower accuracy levels and to use the model to help interpret the reasons for the performance differences among groups.

We conducted analyses in which the data from all the observers who participated in the experiment were divided into four subgroups. Subgroup 1 consisted of the complete set of observers from Group 2 of our experiment—namely, the observers who had competed for a monetary award (a total of 32 observers). Subgroup 2 consisted of the roughly top third of observers from the Group 1 condition that we had included in our previous modeling analyses (a total of 49 observers). These observers had all scored at least 87.5% correct responses during the 384 trials of training.[6] Subgroup 3 consisted of roughly the middle third of observers from the Group 1 condition. These observers had all scored between 81.5% and 87.5% correct responses during training (a total of 51 observers). Finally, Subgroup 4 consisted of the bottom third of observers from Group 1. All had scored less than 81.5% correct (a total of 44 observers).

The summary data from Subgroups 1–4 are reported in Tables 4–6. Tables 4 and 5 report the mean response times and accuracies for each individual color, averaged

Table 4
Predicted and Observed Mean Response Times
for Each of the 12 Colors in Each Subgroup, Averaged Across Blocks of Training

| | Subgroup | | | | | | | |
| | 1 | | 2 | | 3 | | 4 | |
Color	Pre.	Obs.	Pre.	Obs.	Pre.	Obs.	Pre.	Obs.
1	934.82	911.73	1,033.80	998.30	1,025.04	1,020.48	979.74	939.32
2	730.76	756.45	792.97	794.25	856.74	850.67	916.46	907.33
3	988.93	1,050.58	1,110.86	1,127.83	1,050.65	1,047.03	994.03	987.70
4	630.50	637.58	678.69	690.90	748.89	777.07	853.86	891.27
5	749.66	792.55	819.67	867.85	846.59	876.90	876.46	909.70
6	810.36	818.02	886.44	930.45	932.06	983.97	951.08	986.03
7	623.92	585.27	668.20	642.05	741.54	707.65	856.28	793.07
8	690.11	678.42	760.18	788.40	791.88	784.30	839.20	866.68
9	894.07	840.73	970.31	945.58	998.96	964.58	981.37	980.68
10	552.13	549.50	608.54	591.73	624.86	634.85	729.77	682.70
11	624.81	587.92	691.66	662.40	720.87	711.83	792.12	786.20
12	821.06	802.80	921.89	903.50	912.46	886.20	909.57	922.68

Note—Pre., predicted; Obs., observed.

Table 5
Predicted and Observed Percentage of Correct Responses
for Each of the 12 Colors in Each Subgroup, Averaged Across Blocks of Training

| | Subgroup | | | | | | | |
| | 1 | | 2 | | 3 | | 4 | |
Color	Pre.	Obs.	Pre.	Obs.	Pre.	Obs.	Pre.	Obs.
1	75.45	69.50	77.30	76.75	65.63	60.00	61.06	56.00
2	93.68	94.00	94.74	96.75	88.18	93.75	75.75	78.75
3	65.92	65.25	65.25	65.75	57.41	52.75	51.58	46.00
4	97.41	97.50	98.04	98.25	94.40	95.75	83.35	82.75
5	92.11	92.25	93.24	93.50	87.86	91.25	80.01	76.00
6	89.21	88.75	90.54	92.00	81.55	83.25	69.75	65.25
7	97.38	97.75	98.04	98.50	94.59	96.25	82.94	88.00
8	95.05	95.50	95.68	97.00	91.39	95.00	83.87	83.25
9	81.65	80.25	84.54	83.75	72.24	70.50	61.03	57.00
10	98.47	98.50	98.74	98.50	97.54	98.00	91.28	94.50
11	97.20	97.75	97.55	98.00	94.73	95.75	87.67	89.75
12	87.62	87.50	87.70	86.25	82.39	81.50	75.84	74.50

Note—Pre., predicted; Obs., observed.

Table 6
Predicted and Observed Mean Response Times
and Accuracies in Each Grouped Block of Practice
for Each Subgroup, Averaged Across the 12 Colors

Grouped Block	Mean Response Time		Percentage Correct	
	Pre.	Obs.	Pre.	Obs.
		Subgroup 1		
1	840.55	854.18	82.88	81.58
2	739.29	746.54	90.71	89.67
3	722.25	701.12	91.56	91.17
4	714.95	702.01	91.90	92.42
M	754.26	750.96	89.26	88.71
		Subgroup 2		
1	923.34	912.59	84.13	83.00
2	811.31	844.29	91.51	91.25
3	793.63	779.51	92.26	93.17
4	786.12	778.02	92.56	94.25
M	828.60	828.60	90.12	90.42
		Subgroup 3		
1	919.77	914.59	78.10	77.08
2	843.05	859.15	85.25	84.92
3	829.85	828.41	86.13	87.58
4	824.18	813.02	86.49	88.33
M	854.21	853.79	83.99	84.48
		Subgroup 4		
1	949.54	962.95	66.15	64.00
2	888.95	912.48	76.40	73.42
3	866.40	854.94	78.86	78.92
4	855.09	820.74	79.97	80.92
M	890.00	887.78	75.35	74.32

Note—Pre., predicted; Obs., observed.

across the four grouped blocks of training. Table 6 reports the mean response times and accuracies in each grouped block of training, averaged across the 12 individual colors. Although the overall accuracy levels and mean response times differ systematically among the subgroups, the patterns of data within each subgroup are remarkably similar. For all four subgroups, colors far from the boundary of equal summed similarity (see Figure 2) tend to have fast mean response times and high accuracies, whereas colors close to the boundary tend to have slow mean response times and low accuracies (see Tables 4 and 5). Likewise, all four subgroups showed similar speed-ups in mean response times as a function of the four grouped blocks of practice and similar improvements in overall accuracy (Table 6).

We fitted the EBRW model to the complete sets of mean response times and accuracies of each separate subgroup by using the same basic procedure as that described previously. However, whereas we had allowed all parameters to vary freely in the previous model fits, in the present analysis we decided to hold the step-time constant α and the scaling constant k fixed at those values that had provided best fits to the Table 1 data. One of our goals in the present analyses was to use the EBRW model to help interpret the performance differences among the four subgroups of observers by noting systematic changes in the values of its best-fitting parameters. Un-

fortunately, various of the free parameters tend to trade off with one another in determining the model's predictions, so it is difficult to pinpoint each individual parameter's "true" value. Although holding α and k fixed at their previously estimated values limits slightly the ability of the EBRW model to fit the data from the four subgroups, this procedure yields important gains in our ability to interpret the reasons for the performance differences.

The predicted mean response times and accuracies for each of the 12 colors, averaged across the four grouped blocks of practice, are shown along with the observed data in Tables 4 and 5. Likewise, the predictions of the mean speed-up and learning functions are shown along with the observed data in Table 6. Summary fits of the EBRW model to the *complete* sets of mean response time and accuracy data (i.e., the data analogous to those reported in Tables 1A and 1B) are shown separately for each subgroup in Table 7, as are the best-fitting parameters from the model. Figures 8A and 8B present scatterplots of the observed against predicted mean response times and accuracies for each of the 12 colors (averaged across blocks of training) for each subgroup taken separately. The EBRW model achieves reasonably accurate quantitative predictions, although it systematically overpredicts the percentage of correct classifications associated with the very difficult colors from the low-performing subgroups.

An examination of the best-fitting parameters in Table 7 suggests the following tentative conclusions. First, according to the model, the main reason that Subgroups 3 and 4 had lower overall accuracies and slower mean response times than did Subgroups 1 and 2 is that these observers had lower levels of overall sensitivity (c) in discriminating among colors, and greater background noise (b) in memory. Estimates of the attention-weight parameter (w_1), the response criteria A and B, and the mean re-

Table 7
Best-Fitting Parameters and Summary Fits of the
EBRW Model to the Mean Response Times and
Accuracy Data of Each of the Subgroups

	Subgroup			
	1	2	3	4
	Best-Fitting Parameters			
c	1.57	1.57	1.15	0.95
w_1	0.46	0.48	0.44	0.42
b	1.36	1.18	1.45	8.67
A	3.05	3.16	3.22	3.22
B	2.64	2.81	2.60	2.43
μ	100.00	130.10	100.00	100.00
	Summary Fits			
RT SSD	90,188.0	94,849.3	70,776.8	106,335.9
RT RMSD	43.3	44.5	38.4	47.1
RT % Var.	92.5	93.0	92.3	82.0
Acc. SSD	354.8	313.4	804.3	1,042.0
Acc. RMSD	2.7	2.6	4.1	4.7
Acc. % Var.	94.9	94.9	93.2	91.7

Note—RT, response time; SSD, sum of squared deviations; RMSD, root-mean squared deviation; % Var., percentage of variance accounted for; Acc., accuracy.

a) Model Fit of Subgroup RT Data

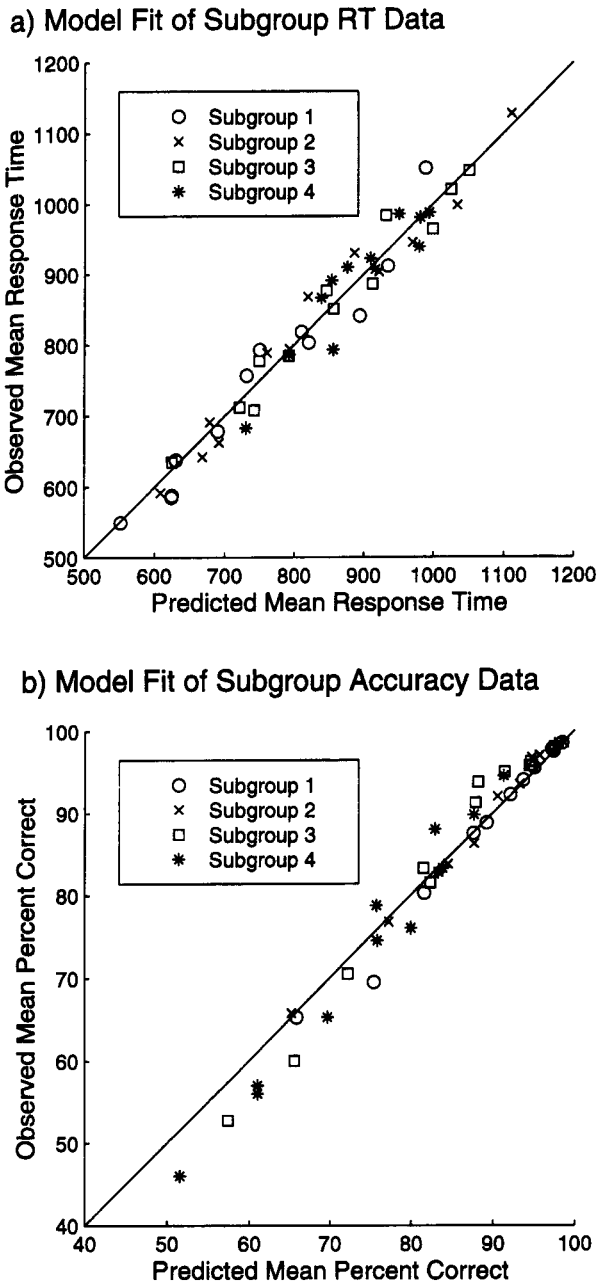## b) Model Fit of Subgroup Accuracy Data

Figure 8. Panel A: Scatterplot of observed against predicted mean response times for the 12 colors, averaged across the four grouped blocks, for each subgroup. Panel B: Scatterplot of observed against predicted mean accuracies for the 12 colors, averaged across the four grouped blocks, for each subgroup.

sidual time (μ) are fairly similar across the four subgroups. Earlier in this article, we noted that Subgroup 1 (those observers motivated with a monetary award) had slightly faster mean response times but slightly lower accuracies than did Subgroup 2 (the top performers from the un-motivated group). This speed–accuracy tradeoff is captured in part by the slightly higher settings of the response criteria A and B for Subgroup 2 than for Subgroup 1.

In summary, in addition to accounting fairly well for the mean response time and accuracy data for each of the subgroups, the EBRW modeling analysis provides straight-forward interpretations for the reasons underlying the differing levels of performance.

GENERAL DISCUSSION

This research provided further support for Nosofsky and Palmeri's (1997b) recently proposed EBRW model of speeded perceptual classification. In previous research involving the collection of extensive individual observer data, Nosofsky and Palmeri (1997b) demonstrated that the EBRW model provided good quantitative fits to the mean response times for a set of colors on the basis of their locations in a multidimensional similarity space and also predicted well the overall speed-up functions observed as a function of practice in the task. A limitation of this earlier work, however, is that, because the observers were so highly practiced, they performed with accuracies close to ceiling, so the research provided little information concerning the ability of the EBRW model to jointly fit response time and choice probability data. Furthermore, in this earlier work, there were insufficient data to model the speed-up and learning functions for individual stimuli in the task.

In this study, we addressed these limitations by modeling the averaged early learning data obtained from a large number of observers. The key achievement was that the EBRW model provided good quantitative fits not only to the mean response times, but to the choice probability data as well. Furthermore, the EBRW model accounted well for the speed-up functions in mean response time observed for the individual stimuli, as well as for the improvements in response accuracy that occurred with practice in the task. The results also provided some preliminary evidence that stimulus-specific adjustments in the random walk response criteria may have taken place, which is an important issue to investigate in future work. Finally, the EBRW model was able to characterize differences in overall performance levels among alternative subgroups of observers in terms of systematic changes in the values of its free parameters.

Currently, the major competing models for predicting multidimensional classification response times and accuracies derive from the decision boundary theory of Ashby, Maddox, and their associates (Ashby et al., 1994; Maddox & Ashby, 1996). The key idea in such models is that the observer establishes a decision boundary for partitioning a multidimensional space into response regions. If a percept falls in Region A of the space, the observer makes a Category A response. Furthermore, according to these models, response time is a decreasing function of distance of the percept from the decision boundary. To date, however, there are no published accounts of how such decision boundaries are learned, why response time decreases monotonically with distance from the boundary, and what the processes are that lead to practice effects

on response time. Quantitative tests between the EBRW model and decision boundary theory in this domain of early category learning will, therefore, have to await the further development of this latter class of models.

REFERENCES

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

ASHBY, F. G., BOYNTON, G., & LEE, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics*, **55**, 11-27.

ASHBY, F. G., & LEE, W. L. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, **120**, 150-172.

BARSALOU, L. W., HUTTENLOCHER, J., & LAMBERTS, K. (1998). Basing categorization on individuals versus events. *Cognitive Psychology*, **36**, 203-272.

BUNDESEN, C. (1990). A theory of visual attention. *Psychological Review*, **97**, 523-547.

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, & Cognition*, **11**, 538-564.

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

GLUCK, M. A., & BOWER, G. H. (1988). Evaluating an adaptive network model of human learning. *Journal of Memory & Language*, **27**, 166-195.

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

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

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

LAMBERTS, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 695-711.

LINK, S. W., & HEATH, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77-105.

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

LOGAN, G. D. (1992). Shapes of reaction time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 883-914.

LOGAN, G. D. (1997). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, **103**, 603-649.

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

MADDOX, W. T., & ASHBY, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 795-817.

MARLEY, A. A. J. (1992). Developing and characterizing multidimensional Thurstone and Luce models for identification and preference. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 299-333). Hillsdale, NJ: Erlbaum.

MARLEY, A. A. J., & 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.

MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

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

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

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

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

NOSOFSKY, R. M. (1997). An exemplar-based random-walk model of speeded categorization and absolute judgment. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 347-365). Mawhaw, 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, & Cognition*, **18**, 211-233.

NOSOFSKY, R. M., & PALMERI, T. J. (1997a). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Perception & Psychophysics*, **59**, 1027-1048.

NOSOFSKY, R. M., & PALMERI, T. J. (1997b). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.

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

RAAB, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, **24**, 574-590.

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

RATCLIFF, R., VAN ZANDT, T., & MCCOON, G. (in press). Comparing connectionist and diffusion models of reaction time. *Psychological Review*.

SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.

NOTES

1. This multiple trace version of the exemplar model is only one possibility and is adopted for simplicity. Alternative models that assume single traces that increase in strength (or in the completeness of their representation) with increasing repetitions would produce similar predictions. Future research is needed to distinguish among these more fine-grained possibilities (cf. Barsalou, Huttenlocher, & Lamberts, 1998).

2. The precise model-fitting method involved the following assumptions. Note that on the first block of testing, each color exemplar was classified for the first time. In modeling response times for the first block, however, it is inappropriate to assume that only background-noise elements are stored in memory. The reason is that, when classifying any given color on the first block, an observer will have experienced some subset of *other* category exemplars on the previous trials during that block. (The precise subset will vary across observers and trials, depending on the random order in which the colors were presented.) For simplicity, and as a reasonable approximation, we assumed that on Block 1, all the exemplars were stored in memory *except* for the specific color being classified on that trial. Likewise, on Block n, there were n tokens of each exemplar stored in memory, but only $n - 1$ tokens of the specific color being classified on that trial. The memory strength for each individual exemplar token (see Equation 3) was set at 1.

3. In the EBRW model, the random walk criteria $+A$ and $-B$ are integer valued. However, we found that conducting a parameter search in integer-valued steps led to numerous local minima. To allow for a more continuous parameter search that was not as prone to local minima, we treated the criteria $+A$ and $-B$ as real-valued parameters. Our modeling investigations indicate that the predictions yielded by the analytic equations with $+A$ and $-B$ real valued can be extremely well approximated by allowing probabilistic mixtures of integer-valued settings of the criterion parameters. For example, the predictions with $+A = 3.5$ can be well approximated by assuming that, on some proportion of trials, $+A$ is set at 3, and on the remaining proportion of trials, $+A$ is set at 4.

4. Not surprisingly, when the magnitudes of the response criteria were allowed to be free parameters for Colors 1 and 3 in Grouped Block 2, clear improvements in the fit of the EBRW model were obtained. Our hypothesis about stimulus-specific criterion shifts for Colors 1 and 3 can be described more intuitively as follows. Colors 1 and 3 are quite similar to one another and are aptly labeled as *tan*. With learning, the

observers eventually realized that when a tan color appeared on the screen, they needed to take special care to distinguish which particular tan was on display. Thus, we hypothesize that a type of rapid gross-level categorization stage may have preceded the operation of the standard random walk process.

5. At least some of these observers are likely to have performed poorly for reasons that go outside the scope of the model, including failure to understand instructions, lack of motivation, experimental disruptions or other sources of error, and so forth. Thus, the modeling analyses for the very low performers should be interpreted with some caution.

6. There were a couple of minor differences in the procedure used for summarizing the composite data in Tables 1A and 1B and the subgroups data discussed here. First, in forming the original composite data set in Tables 1A and 1B, we had intended to include the complete set of 49 observers who scored at least 87.5% correct but discovered later that we had inadvertently not included 4 observers who scored exactly 87.5%. Second, whereas we had used a cutoff of 5,000 msec for including a response time in the earlier composite data, we used a cutoff of 4,000 msec in the present analyses. We do not expect that these minor procedural differences have any bearing on the central conclusions reached in this article.