Journal of Experimental Psychology:
Learning, Memory, and Cognition
1997, Vol. 23, No. 2, 324–354

0278-7393/97/$3.00

Exemplar Similarity and the Development of Automaticity

Thomas J. Palmeri
Vanderbilt University

Effects of exemplar similarity on the development of automaticity were investigated with a task in which participants judged the numerosity of random patterns of between 6 and 11 dots. After several days of training, response times were the same at all levels of numerosity, signaling the development of automaticity. In Experiment 1, response times to new patterns were a function of their similarity to old patterns. In Experiment 2, responses to patterns with high within-category similarity became automatized more quickly than responses to patterns with low within-category similarity. In Experiment 3, responses to patterns with high between-category similarity became automatized more slowly than responses to patterns with low between-category similarity. A new theory, the exemplar-based random walk (EBRW) model, was used to explain the results. Combining elements of G. D. Logan's (1988) instance theory of automaticity and R. M. Nosofsky's (1986) generalized context model of categorization, the theory embeds a dynamic similarity-based memory retrieval mechanism within a competitive random walk decision process.

People often make perceptual and conceptual judgments about objects in the world quickly, with little thought or effort and without conscious intention. Considerable progress has been made in understanding the quantitative and qualitative changes that occur during the development of automaticity or the acquisition of cognitive skill, and in delineating the conditions under which it may be acquired (see Kahneman & Treisman, 1984; Logan, 1985; Schneider, Dumais, & Shiffrin, 1984; Shiffrin, 1988).

One highly influential model of the acquisition of cognitive skill, Logan's (1988) instance theory, attributes the development of automaticity to a shift from the use of fairly general algorithmic processes to the retrieval of specific memories for past skilled actions. In contrast to traditional resource-based accounts of automaticity (e.g., LaBerge & Samuels, 1974; Posner & Snyder, 1975; Shiffrin & Schneider, 1977), instance theory argues that automaticity is largely a memory phenomenon, governed by the same principles that govern memory. According to instance theory, memory-based processes and algorithmic processes race in parallel, with the first to complete driving the response. The algorithmic processes are assumed to remain unchanged with experience, but memory retrieval gets faster because of additional instances being stored in memory. As these additional instances are stored, the amount of time required for memory retrieval decreases, causing memory retrieval to win the race over the algorithm; soon, responses are determined entirely by memory retrieval, signaling the development of automaticity.

Instance theory has accounted successfully for a wide variety of data: It predicts power law decreases in the mean and standard deviation of response times with training (Logan, 1988, 1992; Newell & Rosenbloom, 1981) and has also shown considerable power in accounting for the development of automaticity in memory search (Strayer & Kramer, 1990), lexical decision (Logan, 1988, 1990), alphabet arithmetic (Logan & Klapp, 1991; Klapp, Boches, Trabert, & Logan, 1991), numerosity judgments (Lassaline & Logan, 1993), and repetition priming (Logan, 1990). In contrast to some accounts of skill acquisition (e.g., Anderson, 1982, 1987, 1993; Anderson & Fincham, 1994; LaBerge & Samuels, 1974), instance theory also predicts extremely narrow transfer to new objects. That is, automaticity in a task is specific to the particular instances of skilled action that have been experienced and stored in memory.

The highly specific nature of transfer to new objects was demonstrated by Lassaline and Logan (1993; see also Logan & Klapp, 1991) using a numerosity judgment task. In this task, spatial patterns of between 6 and 11 elements were presented (a manipulation of *numerosity*), and people were asked to judge the number of elements as rapidly as possible without making errors. Initially, response times increased roughly linearly with numerosity, suggesting that people counted each element in a pattern. According to instance theory, this initial performance reflects algorithmic processing. After several days of training on a fixed set of patterns, however, response times became the same for all patterns, regardless of numerosity. Such a zero slope is a classic

---

This article is based on a doctoral dissertation submitted to Indiana University. This work was supported by Grant PHS R01 MH48494-05 from the National Institute of Mental Health. Special thanks go to my mentor, Robert M. Nosofsky, for his encouragement and support during my graduate career. I thank the members of my dissertation committee, Robert Goldstone, John Kruschke, Richard Shiffrin, and James Townsend, for their comments and criticisms regarding this work. Gordon Logan provided many useful suggestions and was very supportive throughout all phases of this work. I also thank Jason Bechler, Debra Coomes, and Brenda Graves for their assistance in testing the participants in these studies.

Correspondence concerning this article should be addressed to Thomas J. Palmeri, Department of Psychology, 301 Wilson Hall, Vanderbilt University, Nashville, Tennessee 37240. Electronic mail may be sent via Internet to palmerit@ctrvax.vanderbilt.edu. Thomas J. Palmeri's World Wide Web address is http://www.vanderbilt.edu/AnS/psychology/cogsci/palmeri/home.html.

signature indicating that a task has become automatized (e.g., Schneider & Shiffrin, 1977). These results are consistent with the notion that people relied on memory for the numerosity of those patterns and did not have to count the elements. According to instance theory, this final performance reflects a shift to memory-based processing; memory retrieval eventually wins the race over algorithmic processing as more and more instances are stored in memory.

However, it is also possible that people learned general strategies for judging numerosities of any patterns, rather than remembering the numerosities of specific patterns. To assess the specificity of the automaticity for these numerosity judgments, Lassaline and Logan (1993) presented new patterns after 12 days of training. Response times were again found to increase roughly linearly with numerosity, with the magnitude of the response times nearly identical to what they were before training—that is, no transfer to the new patterns of elements was found. Lassaline and Logan tested the bounds of generalization in a number of ways: Variations irrelevant to the task of judging numerosity, such as changing the shape of each element in a pattern or changing the color of a subset of elements in a pattern, produced little or no effect on response times (i.e., response time slopes remained flat). Changing the orientation of a pattern by 180°, however, produced dramatic effects—the response times were nearly identical to what they were during the first session of the experiment. In keeping with much of the extant automaticity literature, such manipulations of pattern similarity were relatively large in scale.

Empirical Goal

The empirical goal of the present studies was to investigate the effects of more fine-grained manipulations of pattern similarity on the development and transfer of automaticity in numerosity judgments. The general methodology of the present experiments was similar to that used by Lassaline and Logan (1993). People were trained to judge the numerosity of dot patterns with automaticity signaled by a zero slope for the function relating response time and numerosity. The key added manipulation was that similarity between patterns was explicitly controlled at transfer or during training.

The motivation for these studies was the large body of results demonstrating effects of exemplar similarity on perceptual categorization; in the present studies, it is useful to think of each level of numerosity as a unique category. First, new transfer items are classified confidently when they are similar to specific category members stored in memory (e.g., Homa, Dunbar, & Nohre, 1991; Homa & Vosburgh, 1976; Nosofsky, 1991; Shin & Nosofsky, 1992): Experiment 1 investigated transfer of numerosity judgments to new patterns of varying similarity to specific old, "automatized" patterns.

Second, category learning is influenced by how similar category members are to one another (*within-category similarity*) and by how similar they are to members of other categories (*between-category similarity*); in general, increasing within-category similarity facilitates category learning (e.g., Homa, 1984; Homa & Cultice, 1984), whereas increasing between-category similarity impedes category learning (e.g., Homa, 1984; Omohundro & Homa, 1981). Experiments 2 and 3 investigated the effects of within- and between-category pattern similarity, respectively, on the development of automaticity. Specific theoretical motivations and details of these experiments are given later in the article.

One reason for using the numerosity judgment task was that the patterns of elements are the same as stimuli used in the classic "dot pattern" or "prototype distortion" tasks used in innumerable perceptual categorization studies (beginning with Posner & Keele, 1968). Similarity between patterns can easily be parametrically manipulated with statistical distortion algorithms (Posner, Goldsmith, & Welton, 1967). These dot pattern stimuli also have the advantage that they are essentially infinitely variable and have a complex dimensional structure that resists obvious description, so the artificial categories created from these patterns seem, to a certain degree, to mimic the structure of many real-world, natural categories (see Homa, 1984).

Theoretical Goal

The theoretical goal of the present work was to extend instance theory to account for the observed effects of pattern similarity on the development of automaticity. Conceptually, Logan (1988, 1990, 1992) described a pure instance model. Only stored instances identical to the presented item can be retrieved from memory. Notions of exemplar[1] similarity, a critical component of many theories of memory (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Jacoby & Brooks, 1984; Metcalfe-Eich, 1982; Murdock, 1982) and categorization (e.g., Brooks, 1987; Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1984, 1986), are absent (see Logan, 1988). In fact, most automaticity studies have neither manipulated nor measured the similarity of objects (see, however, Feustel, Shiffrin, & Salasoo, 1983). Further development of the instance theory of automaticity must embed the memory retrieval component within richer representational and process models of memory and categorization.

Also, in its present conceptualization, instance theory is a first-instance race process. The first memory retrieved determines the overt response (assuming memory retrieval wins the race against the algorithm). As is explained later, such a race process cannot account for performance in situations in which responses compete as a result of similarities between items associated with alternative categories (see Experiment 3).

A general model of automaticity and categorization, called the exemplar-based random walk model (Nosofsky & Palmeri, in press-a, in press-b; hereinafter referred to as the

[1] Throughout this article, the terms *instance* and *exemplar* refer to the same basic underlying type of memory representation. The term *instance* is generally used within the context of Logan's (1988) instance theory (without similarity-based retrieval). The term *exemplar* is generally used within the context of the GCM and the EBRW (with similarity-based retrieval).

EBRW), is used to understand, explain, and motivate the experiments reported in this article. The EBRW incorporates important elements of both Logan's (1988) instance theory of automaticity and Nosofsky's (1984, 1986) generalized context model (GCM) of categorization. As with Logan's model, algorithmic and memory-based processes compete to produce a response, and automaticity reflects a shift from primarily algorithmic to primarily memory-based processes. But in keeping with the GCM, memory retrieval is similarity based, and responses are determined by the relative similarity of a probe to members of the various response categories. Extending instance theory, memory retrieval is determined by a competitive random walk process rather than a first-instance race process. The following section highlights key aspects of instance theory and the GCM that were incorporated in the development of the EBRW.

Logan's (1988) Instance Theory of Automaticity

Instance theory posits the development of automaticity as a shift from algorithmic or rule-based processing to memory-based processing. It makes three fundamental assumptions: *obligatory encoding*—attention to an object causes it to be committed to memory, at least to some degree; *obligatory retrieval*—attention to an object causes all available information associated with an object to be retrieved from memory; *instance representations*—memory is specific to the particular aspects of the objects that have been experienced (in contrast to abstractionist or prototype accounts).

Three additional assumptions provide a process model of automaticity: First, memory retrieval time is a random variable. Second, performance is determined by the first memory trace to be retrieved. This assumption makes the instance theory a race model, whereby the fastest winning retrieval determines performance (relative to the processing time of the algorithm). Third, all instances are assumed to have the same distribution of retrieval times.

A fundamental prediction of parallel race models is that increasing the number of runners (i.e., memories) in the race decreases the expected winning time. Imagine a race with only 2 runners who have finishing times that are identically distributed random variables. The winner is the person with the fastest (minimum) finishing time. Compare this with a race among 20 runners, each of whom have the same finishing time distributions as above. The winning time tends to be faster with 20 runners than with only 2 runners because of the random nature of the finishing times—there is a higher probability of one very fast finishing time as the number of runners increases.

Within a race model framework, the prediction of decreases in response time with increases in the number of stored instances reduces to the problem of finding the minimum of N samples drawn from the same underlying distribution. Assuming memory retrieval time that is distributed as a Weibull, the minimum time for N samples decreases as a power law function of N (Colonius, 1995; Logan, 1992, 1995). This property of the instance theory is what allows it to account for the fundamental power law decreases in response time observed with training. Furthermore, the shape of the Weibull distribution closely resembles the ex-Gaussian and the gamma distributions, which have provided close approximations to observed response time distributions (Luce, 1986; Ratcliff, 1978; Ratcliff & Murdock, 1976; Townsend & Ashby, 1983).

Although instance theory has successfully accounted for a wide variety of automaticity findings, there are two fundamental limitations in its current process formulation. First, as acknowledged by Logan (1988), there is no notion of similarity-based retrieval (see also Lassaline & Logan, 1993); the only instances entering the race are those that are identical to the presented item. This assumption was largely made for simplicity and mathematical convenience; in fact, Logan (1992) demonstrated that many fundamental predictions of instance theory held up when differential memory retrieval rates (possibly arising from differential similarity between items) were introduced. The vast set of findings on similarity effects in the memory and categorization literatures (see Homa, 1984; Smith & Medin, 1981) suggest that important similarity effects may be found in automaticity tasks as well.

Second, as a race model, only the first instance retrieved drives the response. As such, no possibility exists for a form of *response competition* to emerge, whereby positive evidence for one response causes negative evidence against all other responses. If one allows for similarity-based retrieval, then under many circumstances more than one response could be available when an item is presented, depending on how similar the presented item is to exemplars associated with various responses. For example, to preview results to be presented in Experiment 3, suppose people are asked to judge the numerosity of patterns that are very similar to patterns of a different numerosity. Intuitively, numerosity judgments for these patterns should be relatively slow. With a simple first-instance race model, however, there is no way to predict such a result; rather, the presence of similar patterns of a different numerosity would be expected to speed up numerosity judgment, albeit at the cost of more errors.

The newly proposed EBRW (Nosofsky & Palmeri, in press-b), which I tested in the studies reported in this article, allows for both similarity-based retrieval and response competition, through the use of a random-walk decision process (Link, 1975; Link & Heath, 1975; Luce, 1986; Ratcliff, 1978; Townsend & Ashby, 1983; see also Strayer & Kramer, 1994a, 1994b).

Nosofsky's (1986) Generalized Context Model of Categorization

The GCM is a generalized version of the context model of categorization proposed by Medin and Schaffer (1978). Categories are represented in terms of stored exemplars, and categorization decisions are based on the relative summed similarity of a probe item to the exemplars of each category. Exemplars are represented as points in a multidimensional psychological space (Nosofsky, 1992c).[2] The similarity between exemplars i and j is an exponentially decreasing function of distance in the psychological space (Shepard, 1987) and is given by

$$s_{ij} = exp\ (-c\ d_{ij}), \quad (1)$$

where c is a general sensitivity parameter that scales the psychological distances, d_{ij}, between stimuli.

The probability of classifying item i as a member of Category A, $P(\text{A}|i)$, is given by the relative summed similarity of i to the exemplars of Category A divided by the summed similarity of i to the exemplars of all categories:

$$P(\text{A}|i) = \frac{\sum_{k \in \text{A}} s_{ik}}{\sum_{L} \left(\sum_{k \in C_L} s_{ik} \right)}, \quad (2)$$

where L is the set of all categories and C_L is the set of members of category L, which is the choice rule of Luce (1963) and Shepard (1957).

The GCM has had success in accounting for a wide variety of categorization findings (for reviews, see Nosofsky, 1992a, 1992b). In an example especially pertinent to the present research, Shin and Nosofsky (1992) demonstrated that the GCM could predict effects within the classic prototype distortion paradigm (Posner & Keele, 1968, 1970). The stimuli were random dot patterns generated by statistically distorting a prototype pattern. A multidimensional scaling (MDS; Carroll & Wish, 1974) analysis revealed the underlying psychological coordinates of each of the presented dot patterns. Given this MDS solution, the GCM provided excellent accounts of the data under a wide variety of experimental conditions. In particular, the model accounted for the effects of similarity between prototypes and old exemplars (Homa & Vosburgh, 1976), category size effects (Homa & Chambliss, 1975; Homa, Sterling, & Trepel, 1981), delay of transfer phase (Homa et al., 1981; Homa & Vosburgh, 1976; Posner & Keele, 1970), and effects of individual item frequency (Nosofsky, 1991).

One important shortcoming of the GCM, however, is the lack of a dynamic exemplar retrieval component; that is, the GCM does not specify how long it takes to retrieve exemplars from memory, nor does it specify how long it takes to arrive at a categorization decision. Thus, there is no way for the GCM to predict categorization response times. Surprisingly, few attempts have been made to formalize process models of multidimensional classification response times (see, however, Ashby, Boynton, & Lee, 1994, for one approach). One of the main goals in developing the EBRW was to allow the exemplar-based framework to be extended to predict response times, thus opening up a wide variety of new territories for theoretical investigation.

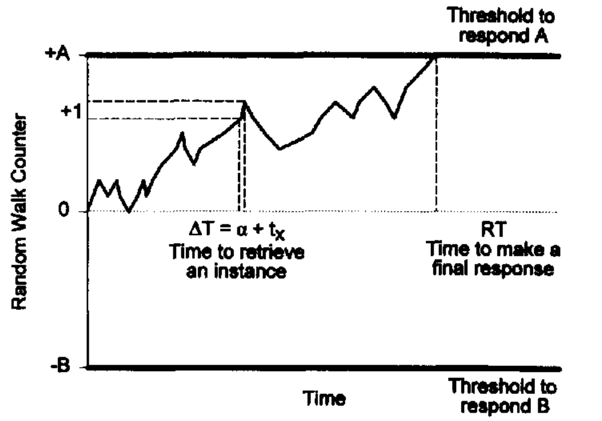

Figure 1. Illustration of the random walk process of the exemplar-based random walk (EBRW) model. Evidence (vertical axis) accumulates over time (horizontal axis) for a Category A or Category B response. A response is made when the random walk counter crosses one of the response boundaries, and the response time (RT) is given by the time when boundary is crossed.

Exemplar-Based Random Walk (EBRW) Model

The EBRW makes the same representational assumptions as those underlying the generalized context model: Categories (or response classes) are represented in terms of stored exemplars, exemplars are represented as points in some multidimensional psychological space, and similarities are an exponentially decreasing function of distance in that space.

As with Logan's (1988) instance theory, when an item is presented, exemplars race to be retrieved from memory. However, in the EBRW, all exemplars race to be retrieved with rates proportional to their similarity to the presented item. Furthermore, unlike Logan's model, in which the first retrieved instance drives the response, in the EBRW, each retrieval provides incremental evidence to drive a random walk. Once sufficient evidence is accumulated, then a response is made available. The actual overt response is determined by a race between this memory retrieval process and an algorithmic or rule-based process. (See Nosofsky, Palmeri, & McKinley, 1994, and Palmeri & Nosofsky, 1995, for one possible rule-based model of classification learning.)

As illustrated for a two-category case in Figure 1, the random walk has a counter accruing information over time for a Category A or a Category B response. When the counter crosses one of the response boundaries (+A or −B), a response is made (A or B), and the response time (RT) is given by the time when the boundary is crossed. When item i is presented, all exemplars race to be retrieved from memory, with race times that are exponentially distributed random variables. The exponential function was chosen primarily for its simplicity (it is a special case of the

[2] In the full instantiation of the GCM, selective attention mechanisms operate to "stretch" distances along diagnostic dimensions and "shrink" distances along nondiagnostic dimensions (see Nosofsky, 1984, 1986).

Weibull), its analytic tractability, and because of its history of use in psychological RT models (e.g., Bundesen, 1990; Logan, 1996; Townsend & Ashby, 1983). The exponential probability density function for exemplar j being retrieved from memory at time t, given presentation of item i, is given by

$$f(t) = s_{ij} \exp(-s_{ij} t), \tag{3}$$

where s_{ij} is the similarity between exemplar j and the presented item i. Note, therefore, that the exemplars that are most likely to win the race and be retrieved are those that are most similar to the test item.

Suppose x is the winning exemplar, and let t_x denote the time to retrieve this exemplar. Then the random walk time is increased by

$$\Delta T = \alpha + t_x, \tag{4}$$

where α is a constant time associated with each step. A psychological interpretation of α is that it represents the time needed to extract the category label from the retrieved memory and to accumulate this information in the random walk counter. If x belongs to Category A, then the random walk counter is increased by 1, whereas if x belongs to Category B, then the counter is decreased by 1. A response is made if the counter crosses one of the response boundaries; otherwise another probe of memory is made, upon which a new race ensues, producing a new increase or decrease to the random walk. Each new memory retrieval entails a new race of exemplars. In summary, RTs are determined by the total number of steps it takes to reach one of the response barriers and by the amount of time it takes to make each step.

The EBRW combines elements of both the GCM and instance theory. It is informative to note that both the GCM and instance theories are essentially special cases of the EBRW: If $A = 1$, the EBRW is formally identical to the GCM; if $A = 1$ and zero similarity is assumed between nonidentical exemplars, the EBRW is essentially the same as instance theory (see Nosofsky & Palmeri, in press-b, for details).

At this point, I highlight some key predictions of the EBRW; later, Monte-Carlo simulations are presented that corroborate these fundamental predictions. First, the EBRW predicts that RTs get faster with practice. Every time an item is presented, a new exemplar is stored in memory. Because of the statistical properties of the race underlying memory retrieval within the EBRW, the time to retrieve the "winning" exemplar gets faster as more exemplars are stored (cf. Logan, 1988), thereby driving the random walk at a faster rate.

Second, similarity of an item to exemplars stored in memory influences RT. When an item is similar to other items of the same category, responses get faster. According to the EBRW, exemplars race with rates proportional to their similarity to the presented item. When similarity increases, retrieval rates get faster, leading to faster retrieval times for the winning exemplar, leading to faster RTs.

Third, when an item is similar to items of other categories, RTs get slower. According to the EBRW, when an item is similar to exemplars of more than one category, RTs get slower because the retrieved exemplars from other categories subtract from the random walk, slowing progress toward the correct boundary.

Instance theory could readily be extended to account for the effects of within-category similarity—one needs only to assume that instances race according to their similarity to the presented item. However, the pure race-model formalism of instance theory cannot account for the response competition effects due to manipulations of between-category similarity.

Nosofsky and Palmeri (in press-b) demonstrated the success of the EBRW at predicting RT data in a variety of categorization experiments. For example, in one experiment, people engaged in 5 days of speeded categorization in which they learned to classify colors into two categories. RTs were highly systematic: In general, stimuli close to the "category boundary" were classified more slowly than those far from the category boundary. Basically, the closer a stimulus was to the boundary, the more similar it was to exemplars of both categories. Therefore, when memory was probed, exemplars of either category could be retrieved, causing the random walk to drift back and forth between the response boundaries, leading to relatively slow RTs. It is this response competition that emerges as a result of between-category similarity that is uniquely predicted by the random walk process.

The EBRW is one of the first process models to have been formulated to account for multidimensional perceptual classification RTs. Perhaps of more importance, however, is that beyond its ability to account for perceptual categorization, the EBRW can also account for the development of automaticity in cognitive skills, thereby providing promise as a unified account of performance across these domains.

Monte-Carlo Simulations

In this section, I report Monte-Carlo simulations to corroborate the conceptual predictions just discussed and to illustrate some more detailed quantitative predictions of the EBRW that are tested in the experiments.

First, increasing the number of presentations of an item should cause decreases in RT means and standard deviations in accord with the well-known *power law of practice* (Logan, 1988, 1992; Newell & Rosenbloom, 1981). Indeed, this power law speedup has been taken as a benchmark for all theories of automaticity and skill acquisition (e.g., Anderson, 1982, 1987; Cohen, Dunbar, & McClelland, 1990; Logan, 1988, 1992; Schneider, 1985). According to the power law,

$$RT = A + BN^{-C}, \tag{5}$$

where RT is response time (mean or standard deviation), A is the asymptotic response time, B is the difference between initial and final response time, N is the amount of practice, and C is the learning rate parameter that specifies the shape of the power law function.

Simulations were performed to test whether the EBRW predicted decreases in RTs (means and standard deviations) that were in accord with the power law. Moreover, similarity was manipulated to ascertain how between- and within-category similarity influenced the shape of the power law function.

In all simulations to be reported in this section, two categories[3] were presented containing five exemplars each. With procedures analogous to ones used in previous work by Homa et al. (1981) and Busemeyer, Dewey, and Medin (1984), two parameters were defined to capture the average similarity relations among each pair of exemplars. For simplicity, the similarity, *sw*, between exemplars belonging to the same category (within-category similarity) was assumed to be the same for every pair of exemplars, and the similarity, *sb*, between exemplars belonging to different categories (between-category similarity) was also assumed to be the same for every pair of exemplars.[4] Thus, exemplars of the same category as a presented item raced with rate *sw* (an exemplar identical to the presented item raced with a maximal rate of 1.0), and exemplars of the opposite category from the presented item raced with rate *sb*. In all simulations, bias-free versions of the EBRW were assumed in which the boundaries for each response were the same distance from the starting position; presently, this distance was set to $A = 5$. (In applications of the EBRW to categorization RT data, Nosofsky & Palmeri, in press-b, found best fitting boundary distances of between 3 and 7 units.) In all simulations, the random walk retrieval times were rescaled using a multiplicative term, k.

Overt responses were determined by a race between the random walk memory process and a generic algorithmic process. Processing time for this algorithm was assumed to be normally distributed with mean μ and standard deviation σ (Logan, 1988). A constant additive residual processing time, R, was assumed for basic perceptual and motor components. Each exemplar was presented eight times per block for 20 blocks ($N = 1 \ldots 20$). Because the EBRW is stochastic, in the following simulations, 5,000 iterations were performed per set of parameter values.

Extending Logan's (1988) instance theory, the EBRW predicts faster responses for items that are more similar to exemplars of their own category relative to exemplars of other categories. Adding similar items to the race speeds the expected time of the winning retrieval. To demonstrate this point, simulations were conducted that varied within-category similarity, *sw*, while holding all other parameters constant, with random walk parameters $sb = .05$, $\alpha = 0.05$, $A = 5$, and $k = 200$; algorithm parameters $\mu = 500$ and $\sigma = 100$; and residual time parameter $R = 400$. Within-category similarity, *sw*, was set equal to .15, .25, .35, and .45. As shown in Figure 2A, predicted RTs (black circles) were strongly influenced by within-category similarity—throughout training, exemplars from categories with high within-category similarity were judged more quickly than exemplars from categories with low within-category similarity. Furthermore, as shown in Figure 2B, standard deviations decreased with training and were influenced by within-category similarity in an analogous way—responses from categories with high within-category similarity were less variable than responses from categories with low within-category similarity.

As shown in Figure 2, power law functions (solid lines) were fitted to these simulated data to determine how well the EBRW predicted RT curves that were in accord with the power law. It was also important to examine how within-category similarity influenced the shape of these power curves (as reflected by the value of C). Table 1 displays parameters and fit values for the best fitting power law curves as a function of within-category similarity. The power law functions fitted the simulated RT means and standard deviations quite well (average correlation of .997, .971, respectively).[5] Furthermore, as shown in Table 1, the exponent (C) increases as a function of *sw*. That is, the power law functions were steeper (approached asymptote more quickly) in conditions of high within-category similarity than in conditions of low within-category similarity.

The EBRW also predicts slower RTs when items are similar to exemplars of other categories. When an item is similar to exemplars of both categories, the random walk counter will wander between the two response boundaries, leading to slower RTs. Simulations were conducted varying between-category similarity, *sb*, with random walk parameters $sw = .45$, $\alpha = 0.05$, $A = 5$, and $k = 200$; algorithm parameters $\mu = 500$ and $\sigma = 100$; and residual time parameter $R = 400$. Between-category similarity, *sb*, was set equal to .05, .10, .15, and .25. As shown in Figure 3A, throughout training, exemplars from categories with high between-category similarity were judged more slowly than exemplars from categories with low between-category similarity. In addition, as shown in Figure 3A, RTs were more variable for exemplars from categories with high between-category similarity.

Table 2 displays the best parameters and fit values for the power law curves displayed in Figure 3. As expected, the power law functions fitted the simulated RT means extremely well (average correlation of .999). The power law functions also fitted the simulated RT standard deviations quite well (average correlation of .981).[6] Finally, as shown in Table 2, the EBRW predicts decreasing values of C with

[3] Throughout this article, the term *category* is broadly construed to mean any collection of objects associated with the same response.

[4] The full version of the EBRW assumes objects to be represented as points in a multidimensional psychological space. Nosofsky and Palmeri (in press-a, in press-b) made use of such representations when the physical stimuli remained the same across participants and when the theoretical goal was predicting RTs for individual stimuli. In the present experiments, however, the stimuli varied from person to person, and the goal was to predict RT for classes of stimuli (rather than individual stimuli). The similarity parameters capture the average similarity across a number of stimuli. The variance in these similarities is captured within other stochastic elements in the model (rather than through the introduction of yet another set of parameters).

[5] As is evident in Figure 2B, early blocks of RT standard deviations seem to violate the power law function somewhat (see also Logan, 1992). When the power law was fitted to Blocks 3–20 the fits improved: $sw = .10, .15, .25$, and $.35$; $r = .987, .995, .997$, and .998, respectively.

[6] As is evident in Figure 3B, early blocks of RT standard deviations seem to violate the power law function. When the power law was fitted to Blocks 3–20 the fits improved: $sb = .05, .10, .15$, and $.25$; $r = .998, .997, .997$, and .993, respectively.

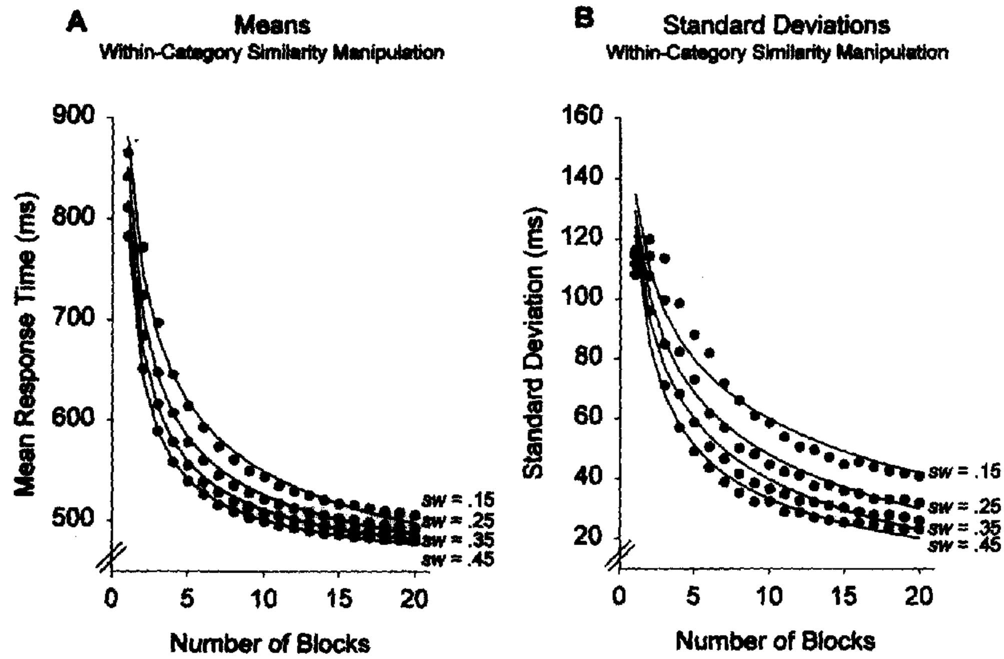

Figure 2. Simulated data from the exemplar-based random walk model. The points display predicted response times as a function of training blocks for different values of within-category similarity (*sw*) with all other parameters held constant. A: mean response times. B: standard deviations (black circles). The solid lines are the best fitting power functions given in Table 1.

increasing values of *sb*. That is, the power law functions were shallower (approached asymptote more slowly) in conditions of high between-category similarity than in conditions of low between-category similarity.

In summary, the EBRW successfully predicted RT curves that were in accord with the power law for both means and standard deviations. As noted by Logan (1988) and others (e.g., Cohen et al., 1990; Newell & Rosenbloom, 1981), the power law for means and standard deviations is a benchmark prediction that any theory of skill acquisition and automaticity must make in order to be taken seriously. Following the lead of instance theory, the EBRW predicts such power law decreases in RT as a result of the underlying race component of memory retrieval. As more instances are added to the race, memory retrieval gets faster. Analyses of the experiments in this article include fits of the power law function to both the observed data and the EBRW predictions.

Table 1

Parameter Values and Measures of Fit for Power Law Functions Fitted to the Simulated Response Time (RT) Means and Standard Deviations Shown in Figure 2 as a Function of Within-Category Similarity

Parameter and measure of fit	Within-category similarity (*sw*)			
	.15	.25	.35	.45
	RT mean			
A	346.0	410.2	434.1	445.2
B	535.1	440.5	382.9	340.8
C	0.421	0.583	0.699	0.790
RMSE	9.12	6.28	4.40	2.51
r	.995	.997	.998	.999
	RT standard deviation			
A	−120.6	−90.2	−58.2	−15.9
B	255.6	219.6	182.3	137.2
C	0.151	0.201	0.270	0.446
RMSE	8.69	6.43	4.85	3.67
r	.940	.969	.982	.991

Note. *A* = asymptotic RT; *B* = the difference between initial and final RT; *C* = the learning rate parameter that specifies the shape of the power law function.

Instance theory makes the very specific prediction that the exponents of the power law functions for means and standard deviations should be identical (see Logan, 1988, 1992). As indicated by the above simulations, the EBRW does not make such a prediction. The EBRW makes no strong predictions as to how the shapes of power law curves for means and standard deviations are related (analytic solutions for the EBRW have been derived for mean RTs only; see Nosofsky & Palmeri, in press-b). In the present simulations, when the standard deviation of the algorithmic process was less than the mean of the algorithmic process, the standard deviation power functions were less steep than those for the means. (In fact, in other simulations not reported, when the standard deviation of the algorithmic process was greater than the mean of the algorithmic process, the standard deviation power functions were more steep than those for the means, as observed in data by Logan & Etherton, 1994.) In the theoretical analyses of the

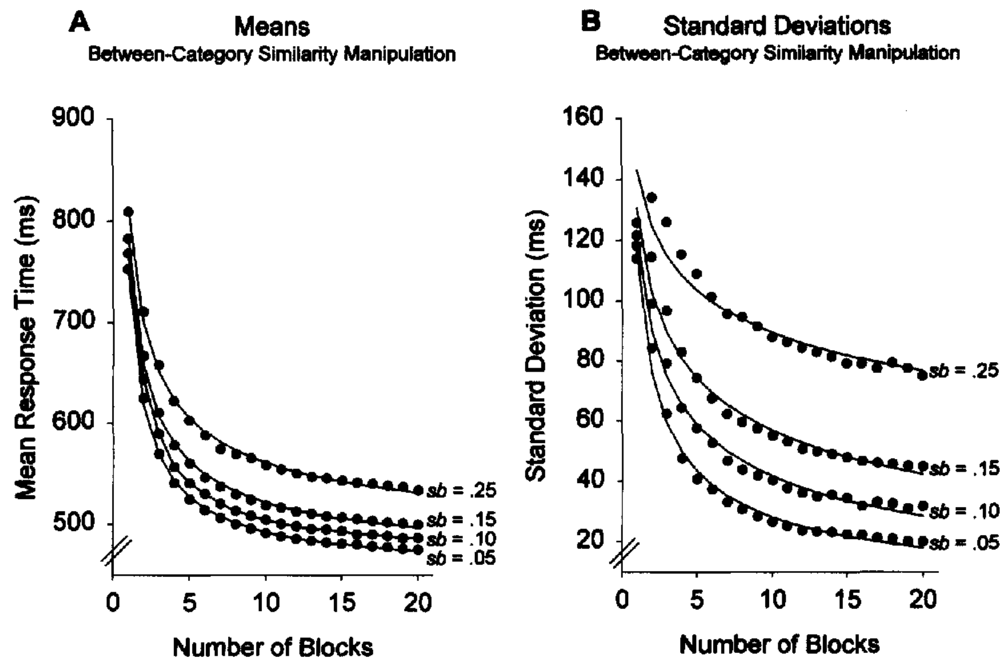

Figure 3. Simulated data from the exemplar-based random walk model. The points display predicted response times as a function of training blocks for different values of between-category similarity (*sb*) with all other parameters held constant. A: mean response times. B: standard deviations (black circles). The solid lines are the best fitting power functions given in Table 2.

experimental results, power law functions are fitted to the means and standard deviations to test the strong prediction of instance theory regarding equal exponents.

Finally, as demonstrated above, the EBRW predicts that the shapes of the power law functions should vary systematically as a function of the similarities among exemplars. Increasing within-category similarity causes steeper power law functions (larger values of *C*), whereas increasing between-category similarity causes shallower power law functions (smaller values of *C*). These fundamental predictions of the EBRW are evaluated as part of the theoretical analyses. In the experimental results and simulations, power law fits are analyzed to test whether the exponential terms vary systematically as a function of *sw* and *sb*.

Table 2
Parameter Values and Measures of Fit for Power Law Functions Fitted to the Simulated Response Time (RT) Means and Standard Deviations Shown in Figure 3 as a Function of Between-Category Similarity

Parameter and measure of fit	Between-category similarity (*sb*)			
	.05	.10	.15	.25
	RT mean			
A	450.2	455.8	458.8	452.9
B	304.6	315.3	327.6	341.6
C	0.858	0.801	0.718	0.590
RMSE	1.95	2.28	2.91	4.06
r	.999	.999	.999	.998
	RT standard deviation			
A	−1.3	−14.1	−28.2	−37.8
B	118.4	136.9	159.0	180.9
C	0.605	0.390	0.271	0.153
RMSE	2.47	2.96	3.92	5.53
r	.995	.992	.985	.951

Note. *A* = asymptotic RT; *B* = the difference between initial and final RT; *C* = the learning rate parameter that specifies the shape of the power law function.

Overview of Experiments

The empirical goal of the present studies was to investigate the effects of fine-grained manipulations of similarity on the development of automaticity in a numerosity judgment task. The theoretical goal was to test the ability of a new model of categorization and automaticity, the EBRW, to account for the observed data.

In each of these studies, participants were trained to judge the numerosity of patterns containing between 6 and 11 dots as rapidly as possible without making errors. Extending previous work by Lassaline and Logan (1993), similarity was manipulated by spatially distorting these dot patterns. Automaticity in these judgments was signaled by a zero slope for the function relating RT and numerosity. In each experiment, of key interest was how both mean RTs and slopes varied as a function of pattern similarity.

Experiment 1 investigated how automatic numerosity judgments generalize to new patterns of varying similarity to the trained patterns. Experiments 2 and 3 investigated how similarity influenced the development of automaticity. Experiment 2 manipulated within-category similarity; on the basis of the above simulations, the central prediction was that automaticity would develop more rapidly for patterns that were similar to other patterns of the same numerosity. Experiment 3 manipulated between-category similarity; the prediction was that automaticity would develop more slowly for patterns that were similar to patterns of a different numerosity. Following each experiment, the EBRW was evaluated on its ability to account for the observed RT data.

Experiment 1

The goal of this experiment was to examine the effects of exemplar similarity on generalizations to new objects following the development of automaticity and to test the ability of the EBRW to account for the observed results. To date, systematic tests of generalization have rarely been done in automaticity studies—people are usually tested on new objects and old objects, with similarity neither measured nor manipulated.

The stimuli used in these studies were dot patterns such as those used in the seminal classification learning studies of Posner and Keele (1968, 1970; see also Homa, 1984; Homa et al., 1981; Shin & Nosofsky, 1992). Participants were trained for 13 sessions on a set of patterns containing between 6 and 11 dots and were required to report how many dots were in each pattern as rapidly as possible without sacrificing accuracy. The development of automaticity was signaled by a zero slope for the function relating RT to numerosity (Lassaline & Logan, 1993; Schneider & Shiffrin, 1977). Following training, tests of generalization were made over 7 transfer sessions. Participants were given old patterns as well as new patterns of varying similarity to the old patterns.

During the transfer sessions, in replication of the predictions of Logan's (1988) instance theory, the EBRW predicts new (dissimilar) patterns to be judged as slowly as they were during the first training session and old patterns to be judged as quickly as they were during the last training session. Because memory retrieval rates in the EBRW are proportional to the similarity between the presented item and the exemplars stored in memory, the EBRW uniquely predicts new similar patterns to be judged with RTs inversely related to their similarity to the old training patterns—moderate-similarity patterns should be judged more quickly than low-similarity patterns.

Method

Participants. Four graduate students from Indiana University participated in 20 experimental sessions. They were each paid $120. Each participant was tested individually. Each session took between 35 and 45 min.

Stimuli. The stimuli were random dot patterns similar to those originally used by Posner and Keele (1968, 1970). Five unique patterns were generated at each level of numerosity (6, 7, 8, 9, 10, and 11). Each pattern was constructed by randomly placing between 6 and 11 dots within the center 30 × 30 of a 50 × 50 square grid, subject to the constraint that the centers of any pair of dots be at least 2 units apart. Each dot had a diameter of 2 mm. Each pattern spanned a maximum of 10 cm × 10 cm on the center of a 14 in. (35.6 cm) computer monitor. Every participant was exposed to a different set of randomly generated patterns. The same set of patterns was shown during all of the 13 training sessions.

New, similar transfer items were created by a statistical distortion algorithm. Basically, this procedure shifts each dot in the pattern some random amount in some random direction. The amount of distortion has traditionally been measured in information-theoretic terms of "bits/dot." The distortion technique was originally introduced by Posner et al. (1967). A grid of 360 cells was defined around each dot in the pattern. The central dot was labeled 0, the surround ring of cells was labeled 1–8, the next ring was labeled 9–24, and so on for a total of 360 cells. Five areas were designated around the central cell consisting of Cells 1–8 (first ring), 9–24 (second ring), 25–80 (third and fourth rings), and 81–360 (fifth–ninth rings). The probability of moving the dot from its central location to one of the surrounding areas is given in Table 3 as a function of the level of distortion in bits/dot. The dot had an equal probability of landing in any one of the cells within the selected area.

Moderate-similarity patterns (moderate-level distortions at 6.0 bits/dot) and low-similarity patterns (high-level distortions at 9.7 bits/dot) were created. New, randomly generated, unrelated transfer items were created by means of the same procedures used to create the original training patterns.

Procedure. At the start of each session, participants were given 60 practice trials to familiarize them with the mapping from numerosity to key response. On every trial, a small cross-hairs appeared at the center of the screen for 500 ms, followed by a spelled out number between *six* and *eleven*. The participant was asked to press the corresponding response key as quickly and as accurately as possible. The number remained on the screen until a response was made. An error message was displayed following an incorrect response. Each level of numerosity (6–11) was represented 10 times in random order (unblocked).

During Sessions 1–13, each of the 5 stimuli at each of six levels of numerosity (30 total stimuli) was presented once per block for a total of 16 blocks (480 total trials per session). Presentation order

Table 3
Probabilities of Moving a Dot to Each Area for the Levels of Distortion Used in These Experiments (From Posner et al., 1967)

Level of distortion	Bits/dot	Area 1	2	3	4	5
Low	3.0	.59	.20	.16	.03	.02
Moderate	6.0	.00	.40	.32	.15	.13
High	9.7	Equally probable in a 29 × 29 square grid				

was randomized for every participant and for every session. On every trial, a small cross-hairs appeared at the center of the screen for 500 ms, followed by a dot pattern. The participant was asked to press the response key corresponding to the number of dots in the pattern as quickly as possible without sacrificing accuracy. The pattern remained on the screen until a response was made. An error message was displayed following an incorrect response to ensure that participants judged true numerosity, rather than apparent numerosity.

During Sessions 14–20, each of the 30 training stimuli was presented once per block for a total of two blocks. Following these two initial blocks, four blocks of transfer trials were presented. On each trial, an old item, a moderate-similarity new item, a low-similarity new item, or an unrelated new item was presented. Each old stimulus was presented once per block and was used twice per block to generate either a moderate-similarity or a low-similarity new item. Five new unrelated patterns at each of six levels of numerosity (30 random patterns) were also presented during each block. Thus, 120 transfer items (30 old, 30 moderate, 30 low, and 30 unrelated) were presented once per block, yielding a total of 480 transfer trials. All other aspects of the procedure were identical to those in Sessions 1–13.

Responses were recorded from the computer keyboard. RTs were measured with the internal millisecond-accuracy timer in the personal computer (PC). The keys *S, D, F, H, J,* and *K* were labeled *6, 7, 8, 9, 10,* and *11,* respectively. Participants rested three fingers from each hand on these keys.

Results

Training data. Because the same qualitative trends were seen in the data from each individual participant, the data were averaged over all participants. Mean accuracies of the numerosity judgments for Sessions 1–13, averaged over numerosity, were .961, .968, .959, .967, .975, .980, .980, .974, .987, .979, .971, .972, and .973, respectively (M = .973). Average accuracy increased slightly across training sessions. Mean accuracies for Numerosities 6–11, averaged over sessions, were .989, .970, .963, .975, .959, and .979, respectively. Accuracy was slightly higher for the endpoints of the response range (6 and 11, M = .984), than the middle of the response range (7, 8, 9, and 10, M = .967). These impressions were confirmed by a 13 (session) × 6 (numerosity) analysis of variance (ANOVA) conducted on the accuracy data from the training sessions (1–13). The main effects of numerosity and session were significant, $F(5, 15) = 2.93$, $MSE = 0.0021$, and $F(12, 36) = 2.27$, $MSE = 0.0007$, respectively. (The alpha level for all statistical tests reported in this article was set at $p = .05$.)

Figure 4A displays mean correct RTs as a function of numerosity during Training Sessions 1–13. During the first few sessions, RTs increased roughly linearly with numerosity but gradually leveled out to be a nearly flat function of numerosity. Linear regression functions were fitted to the RT curves for each session, and the slopes of these regression functions are plotted as a function of session in Figure 5. The initial positive slopes reflect the explicit counting process. The slopes gradually fell to near zero by the end of training, signaling the development of automaticity.

To confirm these impressions, a 13 (session) × 6 (numerosity) ANOVA was conducted on the mean correct RT data from the training sessions (1–13). The main effect of numerosity was significant, $F(12, 36) = 33.07$, $MSE = 155{,}619.02$, reflecting the increase in RTs for patterns with higher numerosity early in training. The main effect of session was also significant, $F(5, 15) = 8.28$, $MSE = 192{,}493.53$, reflecting the decrease in RT over training sessions. The two-way Session × Numerosity interaction was significant, $F(60, 180) = 5.49$, $MSE = 25{,}777.41$, reflecting the greater change in average RTs for high numerosity patterns relative to low numerosity patterns.

Transfer data. To test for the specificity of the memories for these patterns, during the transfer sessions, people were

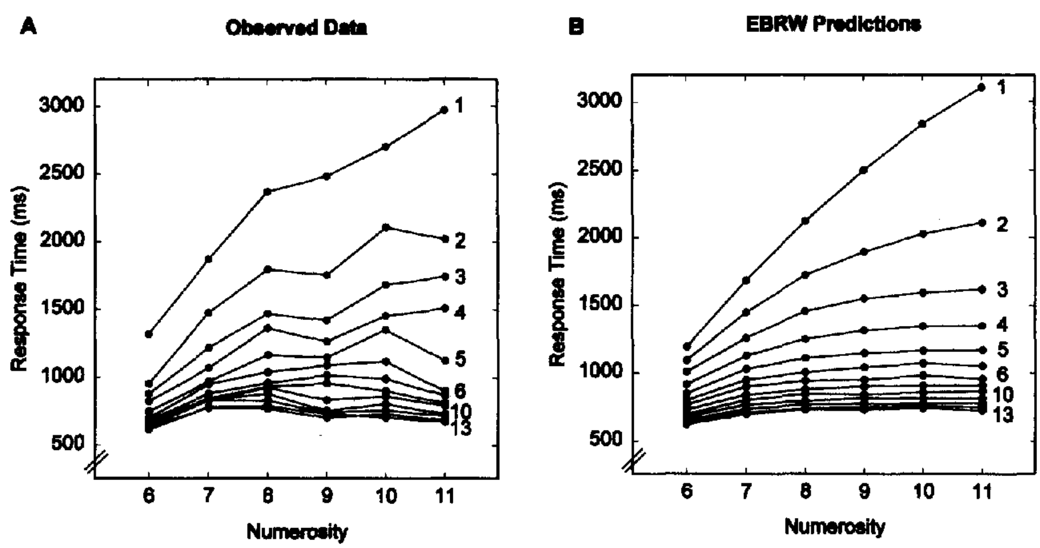

Figure 4. A: Observed response times as a function of numerosity for each training session in Experiment 1. B: Response time predicted by the exemplar-based random walk (EBRW) model as a function of numerosity and training session in Experiment 1.

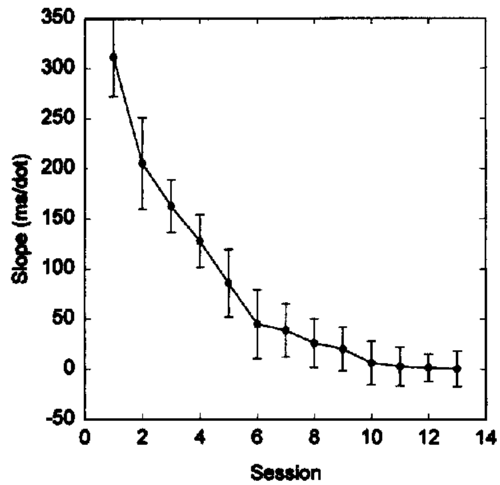

Figure 5. Slopes of linear regression functions fitted to the data in Figure 4A as a function of training session.

presented with four types of patterns: old patterns (old), new patterns of moderate similarity to an old pattern (moderate), new patterns of low similarity to an old patterns (low), and new patterns unrelated to any old patterns (unrelated). The mean accuracy data for each type was .985 for old patterns, .970 for moderate patterns, .939 for low patterns, and .925 for unrelated patterns. A 4 (type) × 6 (numerosity) ANOVA was conducted on the accuracy data from the transfer sessions (14–20). The main effect of type was significant, $F(3, 9) = 6.09$, $MSE = 0.02$, reflecting the increasing accuracy as a function of similarity to the old patterns.

Figure 6A displays mean correct RTs for each of the four types of transfer patterns as a function of numerosity, with Session 1 and Session 13 of training displayed for comparison purposes. Numerosity judgments for new patterns were nearly as slow after 13 sessions of practice as they were during the first session, as can be seen by comparing the solid curve representing the unrelated patterns with the dotted Session 1 curve in Figure 6A. Furthermore, judgments for old patterns were nearly as fast during the transfer sessions as they were at the end of training, as can be seen by comparing the solid curve representing the old patterns with the dotted, Session 13 curve in Figure 6A. These results basically replicate those obtained by Lassaline and Logan (1993). The most important new result is that the moderate patterns were judged more quickly than the low patterns, and the low patterns were judged more quickly than the unrelated new patterns. Thus, as predicted by the EBRW, RTs got faster as the similarity of the transfer patterns to the training patterns increased. The results provide evidence that memory retrieval within a highly automatized task is similarity dependent.

To confirm these impressions, a 4 (type) × 6 (numerosity) ANOVA was conducted on the mean correct RT data from the transfer sessions. The main effect of type was significant, $F(3, 9) = 65.46$, $MSE = 812{,}539.21$, reflecting the influence of pattern similarity on RTs. Old patterns were judged more quickly than moderate patterns, moderate patterns were judged more quickly than low patterns, and low patterns were judged more quickly than unrelated patterns. The main effect of numerosity was significant, $F(5, 15) = 13.97$, $MSE = 823{,}240.70$, reflecting the greater average RT for higher numerosity patterns, especially for the low and unrelated patterns. The two-way Type × Numerosity interaction was

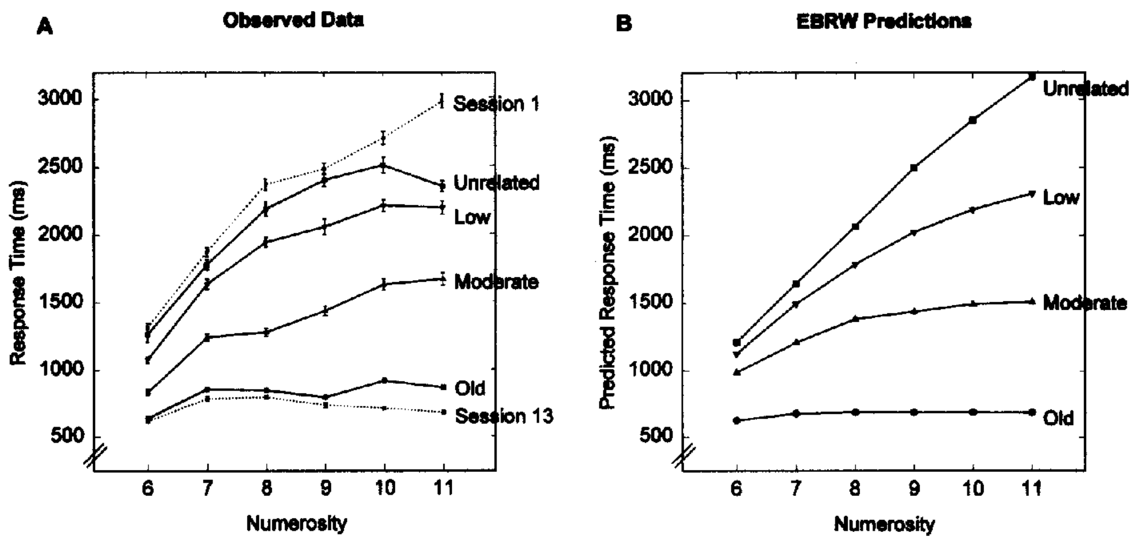

Figure 6. A: observed response times as a function of numerosity for each transfer pattern type: old (Old), moderate similarity (Moderate), low similarity (Low), and unrelated (Unrelated), in Experiment 1. B: exemplar-based random walk (EBRW) model as a function of numerosity for each transfer pattern type. Data from Session 1 and Session 13 are shown by dotted lines in A for comparison purposes.

significant, F (15, 45) = 6.06, MSE = 161,524.30; the increase in RTs as a function of numerosity was steeper for patterns that were less similar to the training patterns.

Theoretical Analyses

The EBRW was next fitted to the data from this experiment. The random walk, with an upper and a lower decision boundary, allows two-choice decisions only. How can the model be extended to situations involving multiple response categories, as is encountered in the numerosity judgment task and many categorization tasks? One fairly simple extension allows multiple counters, one for each category or response. When evidence points to one category, then that response counter is increased. With a traditional counter model, an absolute threshold determines the response (e.g., LaBerge, 1962; Pike, 1973; Townsend & Ashby, 1983)—when any counter exceeds this absolute threshold, a response is made. In the present extension of the EBRW, however, a relative threshold determines the response—when any given counter exceeds all other counters by this relative amount, a response is made. The two-counter version is basically identical to the random walk formalism.

The EBRW was fitted to the observed RT data from the training sessions, shown in Figure 4A, by maximizing the correlation between the predicted and the observed data. An extensive grid search was performed to find the best fitting parameter values. Because the EBRW is stochastic, simulations were performed over 5,000 iterations for each set of parameters.

Recall that overt responses are determined by a race between the random walk memory retrieval process and algorithmic processing. The free parameters for the random walk included the constant time increment, α, the response threshold, A, and the similarities between patterns. Because item similarities were not manipulated during the training phase, within-category similarity and between-category similarity were assumed to be the same for every pair of exemplars, $sw = sb = s$. The random walk retrieval times were rescaled using a multiplicative term, k.

The algorithm is a counting process: For simplicity, it is assumed that every count is a normally distributed random variable with mean μ and standard deviation σ. The time to count a pattern with n dots, therefore, is equal to the sum of n independent and identically distributed normal random variables (truncated at zero). It does not seem reasonable to expect a zero intercept for the counting process, however. Numerosity judgments do not increase linearly throughout the numerosity range; rather they are flat within the *subitizing* range of 1 to 5 items (Mandler & Shebo, 1982; Trick & Pylyshyn, 1993, 1994). Rather than report the intercept term, which is a psychologically meaningless negative number, the mean counting time for a pattern with five elements is reported instead, labeled *subit*, where $Intercept = subit - 5 \cdot \mu$.

Training. The EBRW fitted the training data quite well (r = .986), with random walk parameters s = .015, α = 0.0079, A = 3, k = 7,937, with counting process parameters μ = 584.6, σ = 292.3, $subit$ = 440.4, and with residual time parameter R = 101.8. I start the discussion by pointing out the ability of the EBRW to account for certain important qualitative trends. Figure 4B displays the predicted RT data as a function of numerosity and training session. Comparison with Figure 4A reveals the excellent predictions. The primary shortcoming is the lack of the "bowing" found in the observed data—in particular, observed Numerosity 6 and 11 responses were slightly faster than expected. Part of this difference might have been due to response execution differences, which are not part of the EBRW simulations. That is, it might be easier to execute an extreme response (such as 6 or 11) than an intermediate response (see Klahr & Wallace, 1976, and Mandler & Shebo, 1982, for similar findings of such endpoint effects); in fact, during the initial practice trials, responses to *six* and *eleven* were somewhat faster than responses to the other numbers. Furthermore, this bowing could have been due to differences in the similarity relations for the patterns at the end of the numerosity range relative to those in the middle. If patterns close in terms of numerosity are visually more similar than those far apart in terms of numerosity (e.g., van Oeffelen & Vos, 1982), then the Numerosity 6 and 11 patterns are similar to fewer patterns of a different numerosity—there are no patterns of Numerosity 5 or 12—hence, they are judged more quickly.

Transfer. To fit the transfer data, all of the above parameters were held fixed. Old patterns had similarity s = 1.0 to one of the patterns stored in memory and s = .015 (held fixed from above) to all other patterns; moderate patterns had similarity, sm, to one of the patterns in memory and s = .015 to all other patterns; low patterns had similarity, sl, to one of the patterns in memory and s = .015 to all other patterns; and new patterns had similarity s = .015 to all of the patterns in memory. A two-parameter version of the EBRW, with all other parameters held fixed, fitted the transfer data quite well (r = .962), with sm = .372, and sl = .199. Although no attempt was made to explicitly predict accuracy data, the fits were fairly good. The predicted (observed) accuracies for old, moderate, low, and unrelated patterns were 99.1% (98.5%), 96.9% (97.0%), 94.4% (93.9%), and 85.0% (92.5%), respectively. Figure 6B displays the predicted RT data as a function of numerosity and transfer pattern type. Comparison with Figure 6A reveals the very good account of the transfer data. The primary shortcoming is that the predictions for unrelated patterns are closer to the Session 1 data than to the transfer data. This may indicate a small improvement in the general counting process with practice, or it might also be caused by similarities between some of the randomly generated unrelated patterns and a subset of the training patterns.

Power law analyses. Next, power law functions were fitted to the observed and predicted mean RTs from the training sessions. As shown in Table 4, both the observed and predicted mean RTs were in accord with the power law (average r_{obs} = .996 and r_{pred} = .996, average $RMSE_{obs}$ = 39.05 and $RMSE_{pred}$ = 28.02). Figure 7 displays observed and predicted RT means along with the best fitting power law functions.

Although no attempt was made to explicitly fit the RT standard deviations, as shown in Figure 8, the predicted

Table 4
Measures of Goodness of Fit and Exponent Parameter C for Power Law Fits to Observed and Predicted Response Time Means and Standard Deviations From Experiment 1

Parameter and measure of fit	Numerosity 6	7	8	9	10	11
			Observed means			
C	0.720	0.581	0.448	0.489	0.335	0.480
RMSE	15.19	26.88	31.81	35.38	45.77	79.24
r	.997	.996	.998	.997	.997	.993
			Predicted means			
C	0.198	0.254	0.330	0.435	0.541	0.621
RMSE	29.05	28.54	32.97	29.36	24.21	23.98
r	.986	.995	.997	.998	.999	.999
			Observed standard deviations			
C	0.913	0.321	0.121	0.096	0.099	0.121
RMSE	23.06	31.90	96.52	146.01	155.41	175.58
r	.976	.912	.851	.715	.795	.839
			Predicted standard deviations			
C	0.188	0.184	0.175	0.199	0.216	0.270
RMSE	17.59	25.66	33.63	36.83	36.52	37.35
r	.985	.980	.976	.981	.987	.991

Note. C = the learning rate parameter that specifies the shape of the power law function.

standard deviations were remarkably in line with the observed standard deviations ($r = .819$). Power law functions fitted the noisy observed data fairly well (average $r_{obs} = .848$ $RMSE_{obs} = 104.75$) and fitted the predicted standard deviations very well (average $r_{pred} = .983$ and $RMSE_{pred} = 31.26$). Note that the essentially parameter-free EBRW fit the standard deviations almost as well as the 18-parameter power law functions (.819 vs. .848).

Recall that instance theory makes a very strong prediction that the exponents of the power law functions for means and standard deviations should be identical. However, inspection of Table 4 indicates that the exponents of the standard deviations tend to be smaller than those of the means (average observed $C_\mu = 0.509$ vs. $C_\sigma = 0.279$) and are relatively closer in value to the exponents of the EBRW standard deviations (average predicted $C_\sigma = 0.205$). As corroborating evidence, constrained fits of the power law to the observed standard deviations were slightly better when the exponent was held fixed at the value from the EBRW standard deviations (average $r = .830$ and $RMSE = 111.19$) than when the exponent was held fixed at the value from the observed means (average $r = .803$ and $RMSE = 117.85$). Although these results shed some doubt on the strong predictions made by instance theory regarding equality of exponents for means and standard deviations, further research is clearly needed.

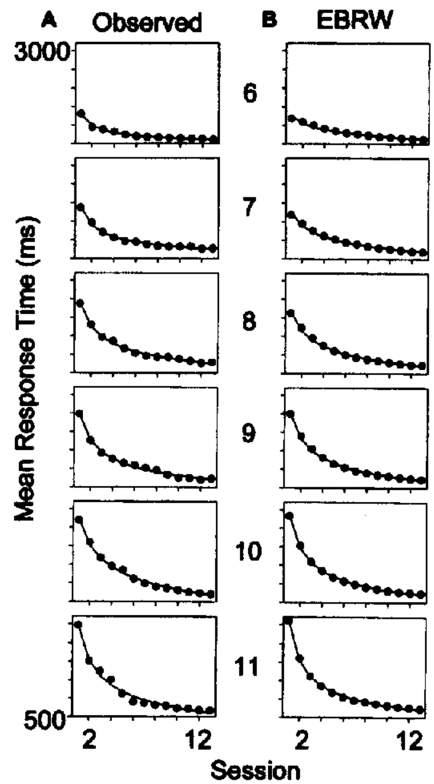

Figure 7. A: observed response time means as a function of training session. B: exemplar-based random walk (EBRW) model predicted response time means as a function of training session. The rows display data and predictions for each level of numerosity from top (numerosity = 6) to the bottom (numerosity = 11). Black circles are observed and predicted means. Solid lines are best fitting power law functions (see Table 4).

Discussion

Replicating the findings of Lassaline and Logan (1993), numerosity judgment times decreased with training, and the slope of the function relating RT and numerosity fell to near zero. During the transfer sessions, old patterns were judged

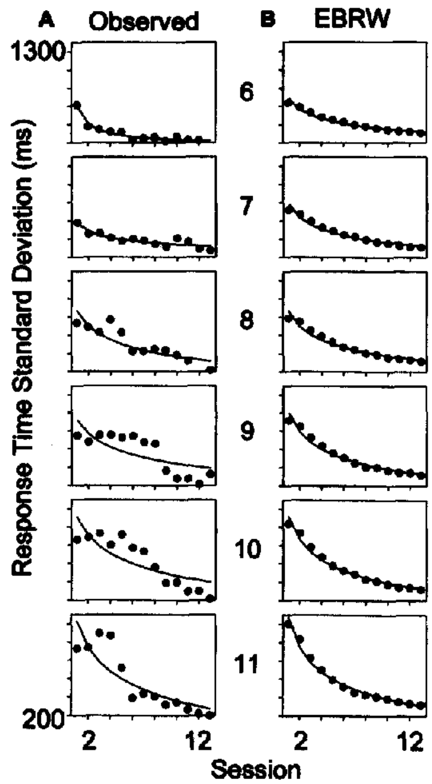

Figure 8. A: observed response time standard deviations as a function of training session. B: exemplar-based random walk (EBRW) model predicted response time standard deviations as a function of training session. The rows display data and predictions for each level of numerosity from top (numerosity = 6) to the bottom (numerosity = 11). Black circles are observed and predicted standard deviations. Solid lines are best fitting power law functions (see Table 4).

about as quickly as they were at the end of training, and new random patterns were judged about as slowly as they were at the start of training. These data are consistent with the notion that people shifted from a primary reliance on explicit counting procedures to a primary reliance on memory for instances. Automaticity in this task seems to reflect the storage of specific instances rather than the development of general procedural abilities (Lassaline & Logan, 1993; Logan, 1988).

Logan and colleagues (Klapp et al., 1991; Lassaline & Logan, 1993; Logan & Klapp, 1991) have previously found instance-specific transfer in a variety of automaticity tasks. A goal of this experiment was to extend these results by demonstrating that transfer would be influenced by the similarity of new patterns to the original training patterns. Fine-grained effects of pattern similarity were observed by giving people new patterns that were moderate- or high-level spatial distortions of the presented patterns (in addition to the old and new patterns that Lassaline & Logan, 1993, had presented). RTs for these patterns decreased as a function of their similarity to the old training patterns: Responses were faster for moderate-similarity patterns than for low-similarity patterns and were faster for low-similarity patterns than for unrelated patterns; RT slopes were flatter for moderate-similarity patterns than for low-similarity patterns and were flatter for low-similarity patterns than for unrelated patterns. These generalization data demonstrate that the specific nature of transfer in automaticity tasks can be influenced by the similarity of patterns to stored exemplars.

To date, instance theory has been formalized with a fairly restricted process model of memory retrieval—only identical instances are retrieved. This is a reasonable simplifying assumption, given that most automaticity studies have neither measured nor manipulated similarity. Clearly, a complete model of automaticity requires a richer process model that allows similarity-based memory retrieval, such as that incorporated into the EBRW. Theoretical analyses demonstrated that the EBRW, with similarity-based memory retrieval, could provide excellent qualitative and quantitative predictions of both the training and the transfer data. The EBRW also accounted for the power law decreases in RT means and standard deviations as a function of practice.

Experiment 2

Experiment 2 examined how the similarity of exemplars presented during training influenced the development of automaticity. A consistent finding in categorization studies is that categories with highly similar members are learned more quickly than categories with less similar members (e.g., Homa & Vosburgh, 1976). By an exemplar-based account, the evidence for an item being a member of any given category is a function of the similarity of that item to the members of that category relative to the similarity of that item to the members of other categories (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984, 1986).

By extension, within an automaticity task, objects learned in conjunction with many similar objects should develop automatic responses more rapidly than objects learned in isolation. In the EBRW, all exemplars in memory race to be retrieved with rates that are proportional to their similarity to the presented item. When many similar items are presented that have the same response, all of these items race to be retrieved with fairly comparable rates. Hence, because of the statistical properties of race models, any one of these items could win the race and would do so more quickly as their number and similarity increased. In contrast, according to the pure version of instance theory, the presence of similar objects should have no effect on how rapidly automaticity is achieved—only the frequency with which any given object is presented should influence how rapidly it can be judged.

In this experiment, people again learned to make rapid judgments of numerosity for a set of dot patterns. At each level of numerosity, moderate-similarity patterns were generated by creating a set of moderate-level distortions from a random prototype pattern; low-similarity patterns were generated by creating a set of high-level distortions of a different random prototype pattern; and unrelated patterns were generated by creating a set of new random patterns. According to the EBRW, the slope of the RT × Numerosity function should approach zero more quickly for moderate-similarity patterns than for low-similarity patterns and should approach zero more quickly for low-similarity patterns than for unrelated patterns. Furthermore, RTs should be faster for moderate-similarity patterns than for low-similarity patterns and should be faster for low-similarity patterns than for unrelated patterns, at all stages of learning.

Method

Participants. Four graduate students from Indiana University participated in 20 experimental sessions. They were each paid $120 for their participation. All participants were tested individually. Each session took between 35 and 45 min.

Stimuli. The stimuli were random dot patterns constructed by using procedures similar to those used in Experiment 1. At each level of numerosity (6–11), three types of patterns were generated: moderate-similarity patterns (moderate), low-similarity patterns (low), and unrelated patterns (unrelated). A random dot prototype pattern was first created at each level of numerosity, 1 for the moderate patterns and 1 for the low patterns. From these prototype patterns, 4 moderate patterns were created at a moderate distortion level (6.0 bits/dot), and 4 low patterns were created at a high distortion level (9.7 bits/dot), at each level of numerosity. The prototype patterns were never presented. For the unrelated patterns, 4 random dot patterns were created at each level of numerosity. Each person viewed 6 (numerosity) × 3 (types) × 4 (instances) = 72 different patterns. Every participant was exposed to a different set of stimuli. The same set of patterns was shown during each of 20 training sessions.

Procedure. In each day's session, each of the 12 dot patterns at each of six levels of numerosity was presented once per block for a total of eight blocks (576 total trials). All other procedural details were identical to those used in Experiment 1.

Results

Accuracy increased slightly with training session; average accuracy for Sessions 1–5 was .956, for Sessions 6–10 was .967, for Sessions 11–15 was .973, and for sessions 16–20 was .968 ($M = .966$). Accuracy was highest for moderate-similarity patterns (.978), next highest for low-similarity patterns (.963), and lowest for unrelated patterns (.957). A 3 (type) × 20 (session) × 6 (numerosity) ANOVA was conducted on the accuracy data. The main effects of type and session were significant, $F(2, 6) = 6.48$, $MSE = 0.0091$, and $F(19, 57) = 2.65$, $MSE = 0.0015$, respectively.

Mean correct RTs as a function of type, numerosity, and session are reported in Appendix A. As in Experiment 1, RTs initially increased linearly with numerosity but gradually flattened. Slopes of linear regression lines fitted to the RT × Numerosity curves are shown in Figure 9A as a function of session and pattern type. The slopes approached zero more quickly for moderate patterns than for low patterns and approached zero more quickly for low patterns than for unrelated patterns. The difference in slope across conditions was particularly evident during the early sessions. Note that the slopes started out at approximately the same level for all three types (moderate = 377 ms/dot, low = 350 ms/dot, and unrelated = 359 ms/dot; $M = 362$ ms/dot)—during Session 1, people seemed to rely on counting for all types of patterns, regardless of their similarity to other patterns. By the end of training, the slopes were also approximately the same for all three types (moderate = 12 ms/dot, low = 22 ms/dot, and unrelated = 15 ms/dot; $M = 16$ ms/dot)—during Session 20, people seemed to rely on memory for all patterns.

Figure 10 displays RTs as a function of session and pattern type; the same overall pattern of results was observed at each level of numerosity, so RTs were collapsed across numerosity for purposes of illustration. Throughout training, the moderate patterns were judged more quickly than the low patterns, and the low patterns were judged more quickly than the unrelated patterns. It is important to note that this relationship held up even at the end of 20 days of training,

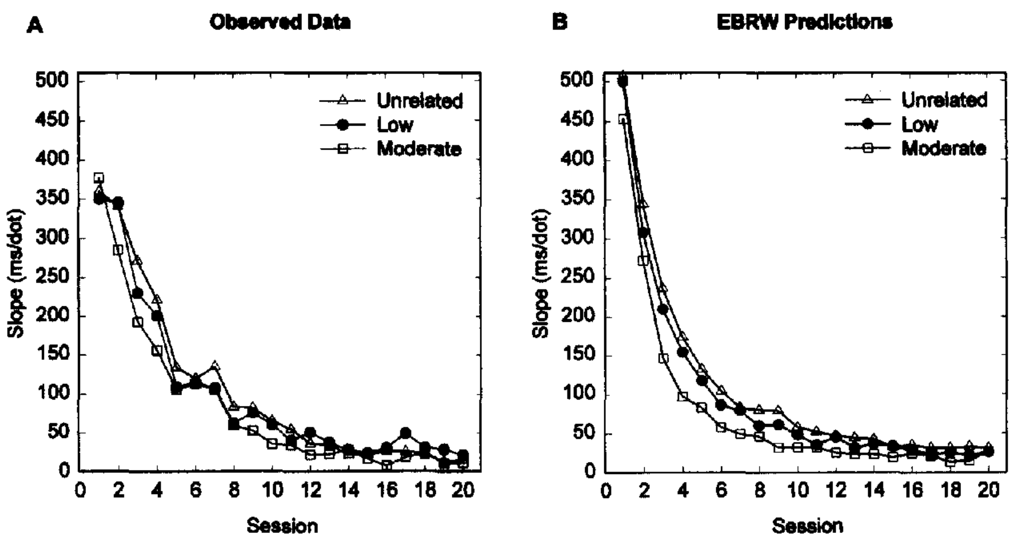

Figure 9. A: the observed slope of the Response Time × Numerosity regression line as a function of session and pattern type in Experiment 2. B: the exemplar-based random walk (EBRW) model predicted slope. Open squares indicate moderate-similarity (Moderate) patterns, filled circles indicate low-similarity (Low) patterns, and open triangles indicate unrelated (Unrelated) patterns.

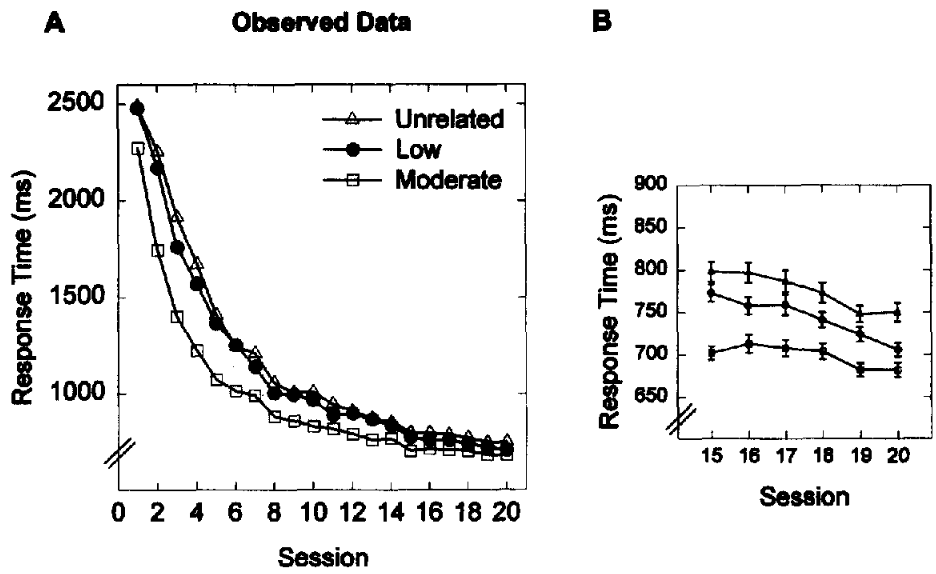

Figure 10. A: response times as a function of session and pattern type, collapsed across numerosity, in Experiment 2. Open squares indicate moderate-similarity (Moderate) patterns, filled circles indicate low-similarity (Low) patterns, and open triangles indicate unrelated (Unrelated) patterns. B: the same data rescaled to enhance the differences in response times between pattern types for Sessions 15–20.

after every pattern had been judged a large number of times, as is shown in Figure 10B.

To confirm these impressions, a 3 (type) × 20 (session) × 6 (numerosity) ANOVA was conducted on the mean correct RT data. The main effect of type was significant, $F(2, 6) = 74.52$, $MSE = 67{,}204.38$, reflecting the ordering of RTs with moderate patterns faster than low patterns and with low patterns faster than unrelated patterns. The main effect of session was significant, $F(19, 57) = 43.17$, $MSE = 375{,}600.31$, reflecting the decrease in RT with training. The main effect of numerosity was significant, $F(5, 15) = 7.96$, $MSE = 1{,}141{,}716.79$, reflecting the slower RTs for higher numerosity patterns, especially early in training. The two-way Type × Session interaction was significant, $F(38, 114) = 5.33$, $MSE = 27{,}925.52$, reflecting the faster decrease in RTs with session for moderate patterns than for either low or unrelated patterns. Finally, the two-way Session × Numerosity interaction was significant, $F(95, 285) = 9.92$, $MSE = 47{,}069.31$, reflecting the change in slope of the RT × Numerosity function with training.

Theoretical Analyses

The EBRW was next fitted to the complete set of observed RT data given in Appendix A, using a grid search to find parameters that maximized the correlation between the predicted and the observed data. The free parameters of the random walk were the similarity between moderate-level (moderate-similarity) distortions, *sm*; the similarity between high-level (low-similarity) distortions, *sl*; the residual similarity between unrelated patterns, *sr*; the constant time increment, α; the response threshold, *A*; and the rescaling term, *k*. The counting process was again assumed to be a sum of *n* normally distributed random variables with mean μ and standard deviation σ and with subitizing term *subit*. As before, a constant residual time parameter, *R*, was also assumed.

The EBRW fitted the 360 data points quite well ($r = .976$), with random walk parameters $sm = .140$, $sl = .055$, $sr = .020$, $\alpha = 0.0023$, $A = 3$, $k = 11{,}621$; with counting process parameters $\mu = 642.6$, $\sigma = 321.3$, $subit = 490.3$; and with residual time parameter $R = 256.6$. The predicted (observed) accuracies for moderate, low, and unrelated patterns were 97.5% (97.8%), 96.8% (96.3%), and 96.2% (95.7%), respectively. I start the discussion by pointing out some of the successful qualitative predictions of the EBRW. Appendix A displays the predicted RT data as a function of pattern type, numerosity, and session. Regression lines were fitted to the predicted RTs as a function of numerosity for each session. Figure 9B displays the slope of the regressions lines as a function of session for the moderate, low, and unrelated patterns. Comparing the predictions in Figure 9B with the observed regression line slopes in Figure 9A reveals the model's ability to predict faster development of automaticity for moderate patterns than for low patterns and faster development for low patterns than for unrelated patterns, as signaled by flattening slopes. Although the EBRW predictions did capture the important qualitative difference between the three types of patterns, the EBRW predicted the values of the slopes to decrease at a faster rate than was observed and predicted initial slopes that were higher than were observed. Overall, though, it is quite impressive that the EBRW was able to account for the critical aspects of such a derived measure.

Figure 11 displays average predicted RTs for the moder-

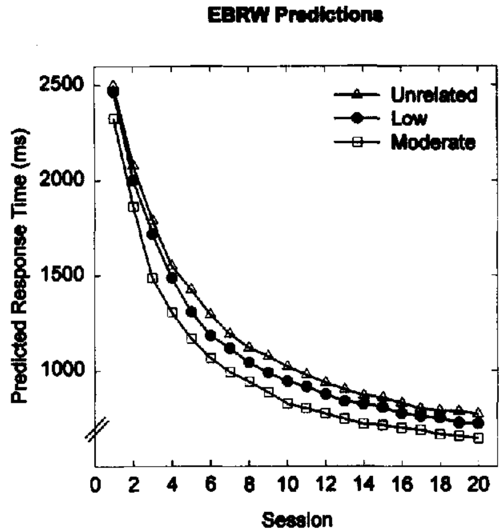

Figure 11. Response times predicted by the exemplar-based random walk (EBRW) model as a function of session and pattern type, collapsed across numerosity, in Experiment 2. Open squares indicate moderate-similarity (Moderate) patterns, filled circles indicate low-similarity (Low) patterns, and open triangles indicate unrelated (Unrelated) patterns.

ate, low, and unrelated patterns as a function of session, collapsed across numerosity. Comparison with Figure 10 reveals that the EBRW predicted the trends in the data quite well (the correlation between average observed and predicted data in Figures 10 and 11 was .993). The EBRW predicted faster responses for moderate patterns than for low patterns and predicted faster responses for low patterns than for unrelated patterns throughout training.

As another measure of how well the EBRW could predict important aspects of the data, power law functions were fitted to the observed and predicted RTs for each level of numerosity for the moderate, low, and unrelated pattern types by using a hill-climbing algorithm that minimized *RMSE*. One reason for fitting power law functions was to determine how the shape of the RT curves, as given by the exponential term, *C*, changed under different experimental conditions. As indicated in Table 5, the observed and predicted mean RT curves for all three types of patterns were well accounted for by the power law functions. Of critical importance is the difference in the exponential term, *C*, between the moderate, low, and unrelated patterns. Overall, as shown in Table 5, the power law functions fitted to the observed data were steeper for moderate than for low patterns and were steeper for low than for unrelated patterns. Recall from the simulations reported earlier that increasing within-category similarity produced steeper power law functions. In these fits, as expected, the EBRW also predicted steeper power law functions (large values of *C*) for the moderate patterns than for the low patterns and predicted slightly steeper functions for the low patterns than for the unrelated patterns. (See Appendix B for analyses of standard deviations.)

Discussion

In summary, the development of automaticity in a numerosity judgment task was influenced by the presence of similar patterns of the same numerosity. Moderate-similarity patterns developed an automatic response, as signaled by flat slopes for functions relating RTs to numerosity, more quickly than low-similarity patterns, and low-similarity patterns developed an automatic response more quickly than unrelated patterns. In addition, responses were quicker for moderate-similarity patterns than for low-similarity patterns, and responses were quicker for low-similarity patterns than for unrelated patterns. This difference in RTs was found throughout training—significant differences were found even after 20 sessions of training on the same set of patterns.

These results are one of the first demonstrations of pervasive similarity effects with extended training on novel stimuli (see also Nosofsky & Palmeri, in press-b). In many categorization tasks, categories with high within-category similarity are learned more quickly than categories with low within-category similarity. However, most studies in the literature have used classification probabilities as the dependent measure and have conducted only a single session of training (see, however, Ashby et al., 1994). This experiment demonstrated pervasive similarity effects after large amounts of training by using RTs as the dependent measure.

The EBRW predicts that increasing within-category similarity should speed the development of automaticity and lead to faster RTs. When an item is similar to many exemplars belonging to the same category (or having the same numerosity), any of those exemplars could be retrieved and would be retrieved relatively quickly, because of the statistical properties of race models. Theoretical analyses revealed that the EBRW provided excellent qualitative and quantitative accounts of the data. Furthermore, the EBRW provided a reasonably good account of how the shapes of the power law functions changed with increases in within-category similarity—power law functions were steeper in conditions of high within-category similarity relative to conditions of low within-category similarity.

Experiment 3

Logan's (1988) instance theory is a pure race model. When an item is presented, instances race to be retrieved from memory, with the winner of this single race driving the response. Although this model is elegant in its simplicity, the present experiment aimed to demonstrate limitations of the single race conception and to suggest instead that competitive response processes, such as those embedded within a random walk framework, are necessary.

The effects of exemplar similarity on the development of automaticity, demonstrated in the first two experiments, could be predicted by an extended version of Logan's (1988) instance theory that allows all exemplars to race with rates proportional to their similarity to the presented item. This assumption works, however, only when the similarity relations are cooperative in nature—that is, when similar exemplars are all associated with the same response. If any

Table 5
Measures of Goodness of Fit and Exponent Parameter C for Power Law Fits to Observed and Predicted Response Time Means at Each Level of Numerosity for Patterns From Experiment 2

Mean type, parameter, and measure of fit	Numerosity 6	7	8	9	10	11
Moderate-similarity patterns						
Observed means						
C	0.609	0.605	0.495	0.904	0.363	0.516
RMSE	21.72	40.06	46.83	63.80	69.11	77.21
r	.989	.987	.990	.992	.993	.991
Predicted means						
C	0.202	0.210	0.339	0.468	0.545	0.663
RMSE	32.36	54.99	36.33	42.98	47.67	50.54
r	.981	.981	.996	.997	.997	.997
Low-similarity patterns						
Observed means						
C	0.370	0.196	0.499	0.215	0.258	0.408
RMSE	20.26	69.02	54.24	99.50	90.48	154.75
r	.995	.985	.994	.982	.988	.980
Predicted means						
C	0.173	0.215	0.319	0.358	0.438	0.528
RMSE	39.24	46.93	49.40	48.67	44.62	48.92
r	.970	.986	.993	.996	.998	.998
Unrelated patterns						
Observed means						
C	0.153	0.180	0.360	0.245	0.240	0.280
RMSE	36.29	45.95	101.74	79.95	133.74	141.05
r	.989	.991	.976	.988	.980	.983
Predicted means						
C	0.174	0.213	0.268	0.300	0.364	0.470
RMSE	38.50	46.23	49.02	55.07	66.12	45.38
r	.970	.987	.993	.995	.995	.998

Note. *C* = the learning rate parameter that specifies the shape of the power law function.

of these similar exemplars is the winner of the race, then the same response is produced. What happens, however, when similar items are associated with different responses?

Compare two situations: First, an item is similar to $n - 1$ exemplars of its own category and is dissimilar to exemplars of any other category; this yields n exemplars that have a high probability of being retrieved. In this case, the similarity relations are cooperative in nature; for ease of exposition, let us call this condition the "friends." Second, an item is similar to $n - 1$ exemplars of its own category but is also similar to n exemplars of another category; this yields $2 \cdot n$ exemplars that have a high probability of being retrieved. In this case, the similarity relations are not cooperative in nature; for ease of exposition, let us call this condition the "enemies."

According to the extended similarity-based version of instance theory, all items race to be retrieved, now with rates proportional to their similarity, with the winning retrieval driving the response. For the enemies, this winning exemplar could be associated with the correct response (that associated with the presented item), or it could also be associated with an incorrect response. If the time for the winner of the race gets faster when there are more runners in the race, then this version of instance theory predicts that the enemies could produce faster RTs than the friends, because there are more enemies than friends, albeit with many more errors.

In contrast, intuition suggests and the EBRW predicts that friends should produce faster responses than enemies. According to the EBRW, when an item is similar to exemplars of just a single category, then only those exemplars will tend to be retrieved, causing the random walk counter to accumulate information in only a single direction and leading to relatively fast RTs. When an item is similar to exemplars of more than one category, however, then any of those exemplars can be retrieved, causing the random walk counter to be increased on some retrievals and decreased on other retrievals and leading to relatively slow RTs. These suggestions were confirmed earlier during the discussion of Monte-Carlo simulations involving the EBRW. Basically, as between-category similarity increased, RTs got slower.

This experiment used the same basic numerosity judgment task as in the previous two experiments. Between-category similarity (where *category* is equated with *numer-*

osity) was manipulated by creating statistical distortions of prototype patterns, but with some of the distortions having an additional dot added to them. For example, six distortions were made from a prototype pattern containing eight dots. Three of these new distortions then had an additional dot added, creating three patterns with nine dots. Presumably, these nine-dot patterns were highly similar to the eight-dot pattern from which they were created. Therefore, at each level of numerosity, people judged the numerosity of some patterns that were similar to patterns of a different numerosity (enemies). For comparison, at each level of numerosity, people also judged the numerosity of some patterns that were similar only to patterns of the same numerosity (friends). The EBRW predicts that RTs should be faster for the friends than for the enemies, as signaled by a zero slope for the function relating RTs to numerosity. A pure race model predicts the opposite.[7]

Method

Participants. Five graduate students from Indiana University participated in 20 experimental sessions. They were each paid $120 for their participation. All participants were tested individually. Each session took between 35 and 45 min.

Stimuli. The stimuli were random dot patterns constructed by using procedures similar to those used in Experiments 1 and 2. Two types of patterns were generated: Friends were patterns that were similar only to patterns of the same numerosity; enemies were patterns that were similar to patterns of the same numerosity and that were similar to patterns of a different numerosity.

For the friends, at each level of numerosity, two different prototype patterns were generated. From each of these two prototypes, three moderate-level distortions (6.0 bits/dot) were generated. The prototype patterns were never presented.

For the enemies, at each of the even levels of numerosity (6, 8, and 10), two different prototype patterns were generated. From each of these two prototype patterns, three moderate-level distortions were generated. Three other moderate-level distortions of each prototype were generated, and then an additional dot was randomly placed within the pattern, subject to the constraint that the new dot was at least 4 mm away from any existing dot. This yielded three patterns of a numerosity that were one greater than the initial numerosity. The prototype patterns were never presented.

Thus, there were three patterns with 6 dots that were similar to one another and that were similar to three other patterns with 7 dots, and there were three other patterns with 6 dots that were similar to one another and that were similar to three other patterns with 7 dots. Similarly, there were three patterns with 8 dots that were similar to one another and that were similar to three other patterns with 9 dots, and there were three other patterns with 8 dots that were similar to one another and that were similar to three other patterns with 9 dots. There were comparable patterns with 10 and 11 dots. In summary, at each level of numerosity, the friends were similar only to patterns that had the same numerosity; the enemies were similar to patterns of the same numerosity but were also similar to patterns that had different numerosities.

Each participant viewed 6 (numerosity) × 2 (friends or enemies) × 6 (instances) = 72 different patterns. Every participant was exposed to a different set of randomly generated stimuli. The same set of patterns was shown during each of 20 training sessions.

Procedure. In each day's session, each of the 12 dot patterns at each of six levels of numerosity was presented once per block for a total of eight blocks (576 trials). All other procedural details were identical to those used in Experiments 1 and 2.

Results

The accuracy data showed the following results: First, the friends were judged more accurately than the enemies (.974 vs. .956, respectively). Second, accuracy increased slightly with training—accuracy for Sessions 1–5 was .958, for Sessions 6–10 was .964, for Sessions 11–15 was .967, and for Sessions 16–20 was .970. Third, the ends of the numerosity response continuum tended to be judged somewhat more accurately than the middle of the response continuum—accuracy for Numerosities 6–11 were .977, .971, .947, .958, .973, and .963, respectively. To confirm these impressions, a 2 (type) × 20 (session) × 6 (numerosity) ANOVA was conducted on the accuracy data. The main effect of type was significant, $F(1, 4) = 9.40$, $MSE = 0.010$, reflecting the higher accuracy for friends than for enemies. The main effect of session was significant, $F(19, 76) = 2.37$, $MSE = 0.0007$, reflecting the overall increase in accuracy with training. The main effect of numerosity was significant, $F(5, 20) = 4.28$, $MSE = 0.0056$, reflecting the lower accuracy at the middle of the response range (Numerosities 8 and 9).

Mean correct RTs as a function of numerosity, session, and type (friends or enemies) are given in Appendix C. As in Experiments 1 and 2, linear regression functions were fitted to the RT × Numerosity curves for every session. Figure 12A displays the slopes of the regression lines as a function of session for both the friends and the enemies. The difference in slope between the friends (212) and the enemies (286) during Session 1 was likely due, in part, to averaging over eight blocks of trials. During the first block of Session 1, people had to count the dots in the patterns explicitly, so there is no reason to expect a difference in slopes initially. Clearly, however, automaticity must have begun to develop so quickly for the friends that there were appreciable slope differences by the end of Session 1. By the end of training, both the friends and the enemies eventually reached a slope near zero (−5 and −3, respectively). Note that this zero slope was reached more quickly for the friends than for the enemies; therefore, automaticity developed more quickly for the friends than for the enemies.

Figure 13 displays RTs as a function of session and type (friends or enemies), collapsed across numerosity. Throughout training, the friends were judged more quickly than the enemies. As shown in the right panel of Figure 13, this difference held up even at the end of 20 days of training.

To confirm these impressions, a 2 (type) × 20 (session) × 6 (numerosity) ANOVA was conducted on the mean correct RT data. The main effect of type was significant, $F(1, 4) =$

[7] Clearly, this prediction depends on the relative number of friends and enemies. The critical point is that it is impossible for the race model to predict the friends to be *faster* than the enemies in the present design.

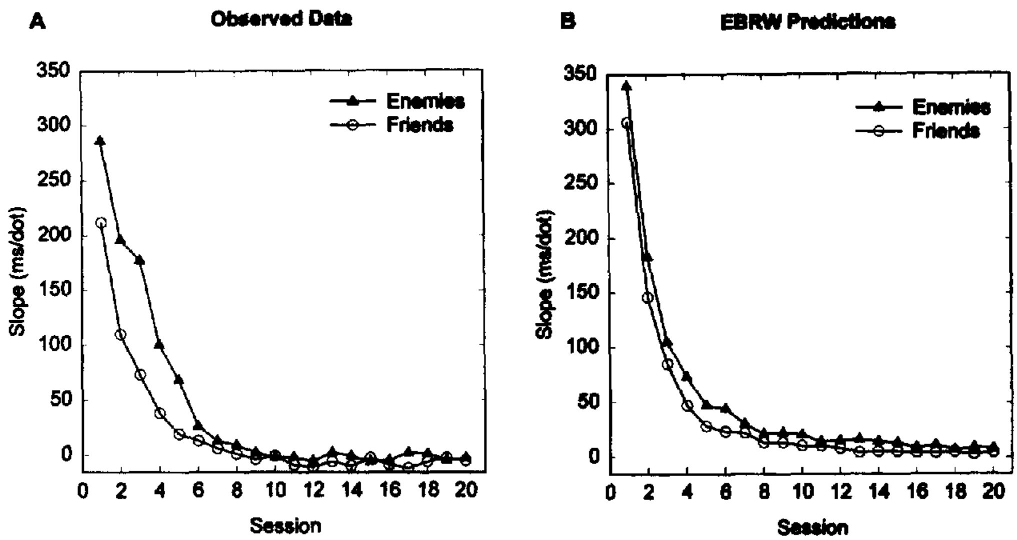

Figure 12. A: the observed slope of the Response Time × Numerosity regression line as a function of session and pattern type in Experiment 3. B: the exemplar-based random walk (EBRW) model predicted slope. Open circles indicate friends and filled triangles indicate enemies.

11.40, MSE = 366,971.53, reflecting the faster RTs for friends than for enemies. The main effect of session was significant, $F(19, 76)$ = 39.07, MSE = 246,727.36, reflecting the decrease in RT with training. The main effect of numerosity was significant, $F(5, 20)$ = 8.37, MSE = 298,978.72, reflecting the slower RTs with increased numerosity, especially early in training. The two-way Session × Numerosity interaction was also significant, $F(95, 380)$ = 6.77, MSE = 28,389.94, reflecting the change in the slope of the RT × Numerosity curves with training.

Theoretical Analysis

The EBRW was next fitted to the entire set of observed RT data given in Appendix C. The free parameters of the random walk were the similarity between moderate-level distortions, sm; the similarity between unrelated patterns, sr; the constant time increment, α; the response threshold, A; and the rescaling term, k. The counting process was again assumed to be a sum of n normally distributed random variables with mean μ and standard deviation σ, with

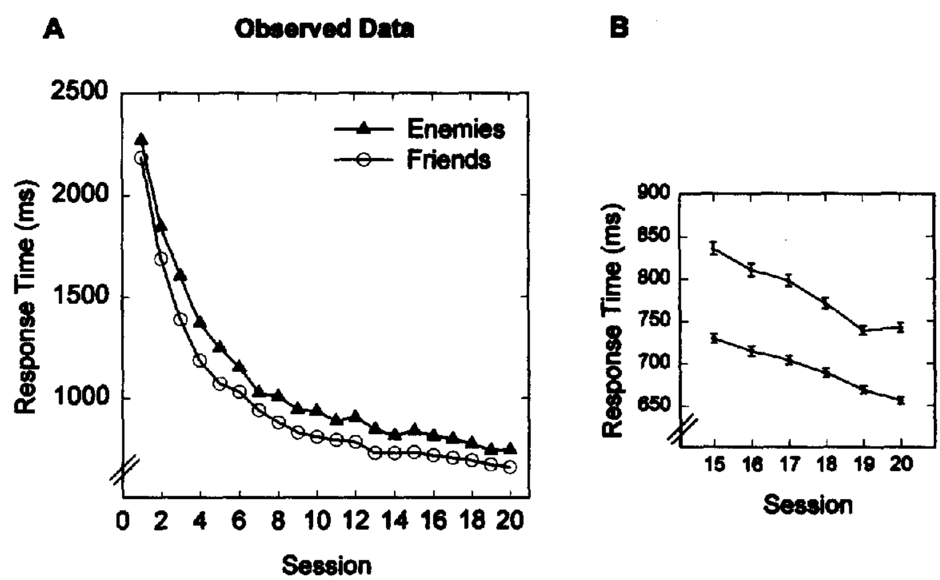

Figure 13. A: response times as a function of session and pattern type, collapsed across numerosity, in Experiment 3. Open circles indicate friends and filled triangles indicate enemies. B: the same data rescaled to show the differences in response times between pattern types for Sessions 15–20.

subitizing term *subit*. As before, a constant residual time parameter, R, was also assumed.

For the friends, at each level of numerosity, there were three patterns with similarity sm to one another and similarity sr to everything else. For the enemies, at each level of numerosity, there were three patterns with similarity sm to one another, similarity sm to three patterns of a different numerosity, and similarity sr to everything else.

The EBRW fitted the data quite well ($r = .975$), with random walk parameters $sm = .100$, $sr = .018$, $\alpha = 0.008$, $A = 3$, and $k = 7{,}694$, with counting process parameters $\mu = 544.3$, $\sigma = 199.8$, $subit = 462.8$, and residual time parameter $R = 188.3$. The predicted (observed) accuracies for friends and enemies were 99.1% (97.4%) and 94.2% (95.6%), respectively.

I begin the discussion by pointing out certain important qualitative predictions. Appendix C displays the predicted RTs as a function of type (friends or enemies), numerosity, and session. Regression lines were fitted to the predicted RTs as a function of numerosity for each session for the friends and the enemies. Figure 12B displays the slope of the regression lines as a function of session for the friends and the enemies. Mirroring the observed data shown in Figure 12A, the friends achieved automaticity more quickly than the enemies. Although accounting for the qualitative trends, a quantitative shortcoming of the predictions was that the model overpredicted the rate at which the slopes reached zero, and underpredicted the magnitude of the slope differences.

Figure 14 displays predicted RTs for the friends and the enemies as a function of session, collapsed across numerosity. Comparison with Figure 13 reveals the excellent quantitative as well as qualitative fits to the data. (The correlation between the average data shown in Figure 13 and the average predictions shown in Figure 14 was excellent, $r = .998$.)

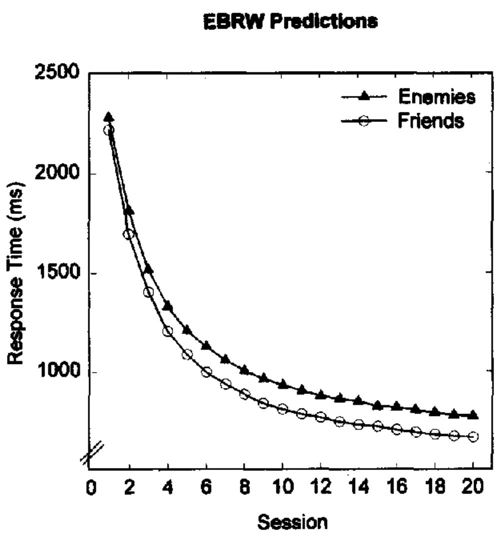

Figure 14. Response time predicted by the exemplar-based random walk (EBRW) model as a function of session and pattern type, collapsed across numerosity, in Experiment 3. Open circles indicate friends and closed triangles indicate enemies.

As additional analyses, power law functions were fitted to the observed and predicted RTs for both the friends and the enemies at each level of numerosity. As shown in Table 6, power law functions fitted the observed means and the predicted means quite well, as expected. Of critical importance, as predicted by the EBRW, the power law functions tended to be steeper for the friends than for the enemies. (See Appendix D for analyses of standard deviations.)

Discussion

In this experiment, the development of automaticity was influenced by the similarity of patterns having different numerosities. Patterns with high between-category similarity, the enemies, developed automatic responses more slowly than did patterns with low between-category similarity, the friends. Throughout training, the friends were judged more quickly than the enemies. Also, best fitting power law functions were steeper for the friends than for the enemies. Each of these important findings was accounted for by the EBRW.

Recall that even an extended version of Logan's (1988) instance theory, allowing for similarity-based memory retrieval, cannot account for the results of this experiment. Because responses are based on the first retrieved instance, the theory predicts the enemies to be judged more quickly than the friends, contrary to the data. As currently formulated, this version of Logan's instance theory also produces accuracy results that are inconsistent results with the task instructions. People were told to respond as quickly as possible without making errors, but instance theory predicts many errors to be made on the enemies. Clearly, some alteration of the theory is needed. These results suggest that competitive response processes, such as those incorporated into the EBRW, are necessary in generalizing the instance theory of automaticity.

General Discussion

The present research investigated some acknowledged, but largely unexplored, parallels between automaticity and categorization. Although the automatic nature of perceptual classification and perceptual judgments is pervasive, little theoretical or empirical research has addressed how such levels of performance are achieved. Part of the reason for this state of affairs is that research on automaticity and research on categorization have had quite distinct histories. Whereas categorization has often been studied in conjunction with concept formation and memory, automaticity has traditionally been viewed as a special topic in the study of attention (e.g., Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin, 1988; Shiffrin & Schneider, 1977; see also Logan, 1988).

Part of the reason for this lack of communication also seems to be due to methodological differences. Automaticity studies usually involve extended training over many sessions and use fairly unitized or highly familiar stimuli, such

Table 6
Measures of Goodness of Fit and Exponent Parameter C for Power Law Fits to Observed and Predicted Response Time Means at Each Level of Numerosity for Friends and Enemies From Experiment 3

Mean type, parameter, and measure of fit	Numerosity 6	7	8	9	10	11
			Friends			
Observed means						
C	0.314	0.446	0.413	0.381	0.836	0.575
RMSE	25.05	35.74	65.89	27.82	49.59	47.34
r	.991	.993	.989	.998	.993	.995
Predicted means						
C	0.255	0.331	0.463	0.568	0.694	0.709
RMSE	22.70	27.25	29.40	33.37	24.92	25.78
r	.993	.996	.997	.997	.999	.999
			Enemies			
Observed means						
C	0.368	0.235	0.288	0.410	0.409	0.488
RMSE	27.92	33.57	29.68	30.76	94.96	99.81
r	.987	.993	.997	.998	.985	.983
Predicted means						
C	0.233	0.301	0.393	0.503	0.575	0.698
RMSE	21.27	29.20	26.84	37.53	34.85	28.86
r	.992	.995	.997	.996	.998	.999

Note. *C* = the learning rate parameter that specifies the shape of the power law function.

as numbers, words, or letters (see, however, Lightfoot & Shiffrin, 1992; Shiffrin, Czerwinski, & Lightfoot, 1991). Most categorization studies involving novel stimuli and formal modeling, on the other hand, usually involve a single training session, use less unitized or unfamiliar stimuli, such as schematic faces or random dot patterns, and often measure or manipulate stimulus similarity directly.

Logan's (1988) instance theory of automaticity offers a starting point for relating these two domains. According to the theory, automaticity is largely a memory phenomenon. Tasks are automatic to the extent they rely on past instances stored in memory. However, the theory lacks similarity-based retrieval mechanisms that are central in categorization and memory models, and it is constrained by its instantiation as a first-instance race process. Nosofsky's (1984, 1986) generalized context model (GCM) provides a rich description of exemplar similarity and categorization processes. However, the theory lacks a mechanism to allow RT predictions.

A new theory of automaticity and perceptual classification, the EBRW (Nosofsky & Palmeri, in press-b), combines elements of Logan's (1988) instance theory of automaticity and Nosofsky's (1986) GCM of categorization. The model embeds a dynamic similarity-based memory retrieval mechanism within a competitive random walk decision process. Nosofsky and Palmeri (in press-b) explained a wide variety of classification RT data with the EBRW. The goal of the present experiments was to test predictions of the model within the context of the development of automaticity.

Summary of Results

The goal of Experiment 1 was to assess generalizations to new objects following the development of automaticity. People were trained to judge the numerosity of dot patterns as quickly and as accurately as possible. Following training, they were asked to judge the numerosity of old patterns and new patterns of varying similarity to the old patterns. Mirroring classic categorization results (e.g., Homa & Vosburgh, 1976; Posner & Keele, 1968; Shin & Nosofsky, 1992; see also Estes, 1994; Smith & Medin, 1981) and consistent with the predictions of the EBRW, judgment RTs were determined by the similarity of the transfer patterns to the old training patterns. Old patterns were judged just as quickly as they were at the end of training, and new patterns were judged just as slowly as were new patterns at the start of training, replicating the results of Lassaline and Logan (1993). More important, new moderate-similarity and low-similarity distortions were judged with an intermediate RT, in accord with their similarity to the old patterns.

The goal of Experiment 2 was to investigate the influence of exemplar similarity on the development of automaticity. At each level of numerosity, people were trained on three types of patterns: moderate-similarity patterns were moderate-level distortions of a prototype pattern, low-similarity patterns were high-level distortions of a prototype pattern, and unrelated patterns were randomly generated patterns. Analogous with traditional categorization studies, this design manipulated within-category similarity. In accord with the predictions of the EBRW, numerosity judgments became automatized more quickly—as indicated by a zero slope for the function relating RT and numerosity—for moderate-similarity patterns than for low-similarity or unrelated patterns. Similarly, throughout training, the moderate-similarity patterns were judged more quickly than the low-similarity or unrelated patterns.

Finally, the goal of Experiment 3 was to further investigate the influence of exemplar similarity on the development of automaticity by manipulating between-category similarity. Distortions of a prototype pattern were created, as before, but now an additional dot was added to half of the patterns. In this way, there were patterns of different numerosities that were similar to one another; these patterns were called the enemies. For comparison, patterns were created that were similar only to other patterns of the same numerosity; these patterns were called friends. In accord with the predictions of the EBRW, the friends developed automatic judgments more quickly than the enemies, and the friends were judged more quickly than the enemies throughout training.

The EBRW successfully accounted for the major qualitative trends in these experiments and provided reasonably good quantitative fits. The model also successfully predicted power law decreases in RT means and standard deviations. The shape of the power law functions for means changed systematically as a function of the similarity relations between patterns. In Experiment 2, power law functions were steeper for patterns with high within-category similarity than for patterns with low within-category similarity. In Experiment 3, power law functions were steeper for patterns with low between-category similarity than for patterns with high between-category similarity.

Instance theory predicts that the exponents of the power law functions for means and standard deviations should be identical. However, the power law exponents of the EBRW standard deviations provided a better fit to the observed standard deviations than did the power law exponents of the observed means. Although these results were not overwhelming in terms of quantitative measures of fit, the advantage for the EBRW was found across all three experiments.

Although these experiments suggested limitations in the pure single race version of Logan's (1988) instance theory, the fundamental notion of automaticity as largely a memory phenomenon was fully supported. Both Logan's instance theory and the EBRW are exemplar-based models. The EBRW simply extends Logan's model by incorporating a similarity-based memory retrieval process and response competition in the form of a random walk decision process.

Automaticity, Categorization, and Perceptual Expertise

The present research is consistent with some emerging notions of the development of perceptual expertise and automaticity in classification. In a variety of domains, such as diagnostic radiology, biological taxonomy, or paleontology, experts spend many years learning to classify objects on the basis of subtle perceptual cues. Novices in these domains are often presented with detailed sets of analytic rules that can be used to classify objects. These rules are often distilled from experts asked to formalize their thought processes—the implicit assumption has been that experts develop and use intricate sets of classification rules. Expert performance in these domains, however, may be more than merely committing more and more rules to memory. Rather, certain aspects of expertise may reflect the vast memory for examples the expert has experienced.

One of the foundations for the work on expertise comes from the studies of chess masters by de Groot (1965) and Chase and Simon (1973). These seminal studies suggested that one important key to achieving chess mastery seemed to lie in improved perceptual processing of the layout of chess pieces, rather than more rapid evaluation of legal chess moves. This perceptual skill results from years of practice. Whereas novices seem to rely on slow, conscious, deductive reasoning, experts seem to rely on fast, relatively unconscious processing—the chess master "sees" the right moves. Chase and Simon explained chess mastery in terms of the size of perceptual structures (or chunks) that experts use relative to novices. In the present context, these chunks could have their basis within a rich exemplar memory system. Furthermore, Chase and Simon (see also de Groot, 1965) found that when confronted with random patterns of chess pieces, chess masters' performance on memory and perception tasks using such configurations was reduced to that of novices (see, however, Gobet & Simon, 1996). Mirroring the results of the present studies using artificial dot pattern stimuli, transfer to new patterns of chess pieces that have not been seen before (because they were not valid positions according to the rules of chess) was very poor.

Recent research on the development of expertise in medical diagnosis has also begun to recognize shifts from highly analytic, conscious reasoning processes to nonanalytic, perceptual, and memory-based processes (e.g., Groen & Patel, 1988). For example, in a study using procedures similar to those of Chase and Simon (1973), Coughlin and Patel (1987) gave normal cases and randomly ordered cases to both physicians and medical students. Mirroring the results with chess masters, memory for the randomly presented case studies was as bad for the physicians as it was for the medical students (see also Myles-Worsley & Johnston, 1988). Physicians and medical students formulate information from clinical cases differently—the physicians recognize patterns of familiar problems but the students generally do not.

Recall that in the present experiments, effects of exemplar similarity were found even after 20 days of training. In a series of studies using expert dermatologists, Brooks, Norman, and Allen (1991; Allen, Norman, & Brooks, 1992) presented slides of dermatologic cases for initial diagnoses and then several weeks later presented new slides that were similar or dissimilar to the original slides. The slides similar to the original slides were correctly diagnosed significantly more often than those dissimilar to the original slides. Exemplar similarity seems to have pronounced effects even after years of exposure to highly similar cases.

Much of the work in automaticity and categorization has assumed that exemplar representations remain relatively unchanged with experience, apart from changes in how dimensions are selectively attended (Kruschke, 1992; Nosofsky, 1984, 1986). In contrast, Lesgold et al. (1988) demonstrated fairly dramatic changes in how X-ray images were perceived as a function of expertise. Significant features in chest films showing cancerous lung tumors were illustrated by physicians and medical students by outlining areas on the X-ray images. Little correspondence was observed between the features delineated by the expert physicians and the novice medical students. Not only must an expert learn what features are diagnostic, they must also learn what the relevant set of features are (see also Hock, Webb, & Cavedo, 1987; Schyns, Goldstone, & Thibaut, 1995; Schyns & Murphy, 1991).

In another domain, Biederman and Shiffrar (1987) studied the perceptual expertise of chick sexers. These people had spent upwards of 50 years classifying male and female genitalia of newborn chicks for the poultry industry. By one estimate, the chick sexer studied by Biederman and Shiffrar may have classified over 50 million chicks. Through interviews with this expert, a fairly simple rule was discovered that could correctly classify over 80% of the chicks; however, numerous rare exceptions to this simple rule existed. Because of the economics of the poultry industry, the experts were required to classify 1,000 chicks per hr with over 98% accuracy, so the simple rule was inadequate. In fact, although the expert was able to communicate the simple rule in some form, it was not used to classify the chicks. Rather, in keeping with the present discussion, expert chick sexers seemed to rely on their vast memory for common and rare configurations of genitalia.

Clearly, unlike chick sexing, medical diagnosis (as well as other highly skilled domains) is not purely a matter of perceptual expertise. Training in medical diagnosis consists of learning a great deal of highly structured domain-specific knowledge, often presented in the form of rules or lists of diagnostic criteria. Experts clearly have ready access to this structured knowledge when called upon to justify a given diagnosis or when asked to pass knowledge along to physicians in training. The main issue here is that the initial (automatic?) diagnosis may reflect the use of nonanalytic, seemingly unconscious, presumably memory-based processes. Just as the chess master may "see" the right move, the expert medical practitioner may "see" the right diagnosis. Current research suggests that the role of such highly developed perceptual classification skills has been underappreciated.

Conclusion

The present findings offer additional strong support for exemplar-based views of memory (Gillund & Shiffrin, 1984; Jacoby & Brooks, 1984), automaticity (Logan, 1988), and categorization (Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1986). The EBRW provides a unified account of both perceptual categorization and automaticity. As stated by Newell (1990),

> a unified theory will unify our existing understanding of cognition. It will not be a brand-new theory that replaces current work at every turn. Rather, it will put together and synthesize what we know. On the other hand, it can't be just a pastiche, in which disparate formulations are strung together with some sort of conceptual baling wire. The parts must work together. (p. 16)

The EBRW is not the all-encompassing theory of cognition envisioned by Newell—it does take steps in the right direction, however. It is not a brand-new theory—rather, it synthesizes and inherits the successes of the GCM of categorization (Nosofsky, 1984, 1986) and the instance theory of automaticity (Logan, 1988, 1990). The integration of these two underlying theories retains key elements of each, as evidenced by the fact that both are essentially special cases of the EBRW (see Nosofsky & Palmeri, in press-b)—the parts do work together. In the spirit of Newell's maxim on unified theories of cognition, an aim of future work is to further expand the explanatory scope of EBRW.

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. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94,* 192–210.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

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., Boynton, G., & Lee, W. W. (1994). Categorization response times with multidimensional stimuli. *Perception & Psychophysics, 55,* 11–27.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 640–645.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: The ecological and intellectual factors in categorization* (pp. 141–147). Cambridge, England: Cambridge University Press.

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., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 638–648.

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* (pp. 57–105). San Francisco: Freeman.

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

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Colonius, H. (1995). The instance theory of automaticity: Why the Weibull? *Psychological Review, 102,* 744–750.

Coughlin, L. D., & Patel, V. L. (1987). Processing of critical information by physicians and medical students. *Journal of Medical Education, 62,* 818–828.

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

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

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112,* 309–346.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Gobet, F., & Simon, H. A. (1996). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review, 3,* 159–163.

Groen, G. L., & Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 287–310). Hillsdale, NJ: Erlbaum.

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

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hock, H. S., Webb, E., & Cavedo, L. C. (1987). Perceptual learning in visual category acquisition. *Memory & Cognition, 15,* 544–556.

Homa, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 18, pp. 49–94). New York: Academic Press.

Homa, D., & Chambliss, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 351–359.

Homa, D., & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 83–94.

Homa, D., Dunbar, S., & Nohre, L. (1991). Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 444–458.

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

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 322–330.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 1–47). New York: Academic Press.

Kahneman, D., & Treisman, A. M. (1984). Changing views of attention and automaticity. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.

Klahr, D., & Wallace, J. (1976). *Cognitive development: An information-processing view.* Hillsdale, NJ: Erlbaum.

Klapp, S. T., Boches, C. A., Trabert, M. L., & Logan, G. D. (1991). Automatizing alphabet arithmetic: II. Are there practice effects after automaticity is achieved? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 196–209.

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

LaBerge, D. (1962). A recruitment theory of simple behavior. *Psychometrika, 27,* 375–396.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 7,* 495–531.

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.

Lesgold, A., Glaser, R., Rubinson, H., Klopfer, D., Feltovich, P., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.

Lightfoot, N., & Shiffrin, R. M. (1992). On the unitization of novel, complex visual stimuli. *Proceedings of the 14th annual conference of the Cognitive Science Society* (pp. 277–282). Hillsdale, NJ: Erlbaum.

Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology, 12,* 114–135.

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

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39,* 367–386.

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. (1995). The Weibull distribution, the power law, and the instance theory of automaticity. *Psychological Review, 102,* 751–756.

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103,* 603–649.

Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1022–1050.

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. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103–190). New York: Wiley.

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

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111,* 1–22.

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

Metcalfe-Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

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

Myles-Worsley, M., & Johnston, W. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 553–557.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

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. (1991). 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. (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). Exemplars, prototypes, and similarity rules. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *Essays in honor of William K. Estes* (Vol. 1, pp. 149–167). Hillsdale, NJ: Erlbaum.

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

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

Nosofsky, R. M., & Palmeri, T. J. (in press-b). An exemplar-based random walk model of speeded classification. *Psychological Review.*

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

Omohundro, J., & Homa, D. (1981). Search for abstracted information. *American Journal of Psychology, 94,* 267–290.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 548–568.

Pike, A. R. (1973). Response latency models for signal detection. *Psychological Review, 80,* 53–68.

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.

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

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

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

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

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 190–214.

Schneider, W. (1985). Toward a model of attention and the development of automatic processing. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 475–492). Hillsdale, NJ: Erlbaum.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1984). Automatic and control processing and attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 1–27). New York: Academic Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1–66.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. -P. (1995). The development of features in object concepts. *Indiana University Cognitive Science Technical Report No. 133,* Bloomington: Indiana University.

Schyns, P. G., & Murphy, G. L. (1991). The ontogeny of part representation in object concepts. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 31, pp. 305–354). San Diego, CA: Academic Press.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22,* 325–345.

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

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

Shiffrin, R. M., Czerwinski, M., & Lightfoot, N. (1991). Automatization and training in visual search. *American Journal of Psychology, 105,* 271–315.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General, 121,* 278–304.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Strayer, D. L., & Kramer, A. F. (1990). An analysis of memory-based theories of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 291–304.

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

Strayer, D. L., & Kramer, A. F. (1994b). Strategies and automaticity: II. Dynamic aspects of strategy adjustment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 342–365.

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

Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 331–351.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101,* 80–102.

van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics, 32,* 163–170.

(Appendixes follow on next page)

Appendix A

Observed Mean Response Times (in Milliseconds) and EBRW Predictions (in Parentheses) as a Function of Type (Moderate, Low, and Unrelated), Numerosity, and Session in Experiment 2

Session	Numerosity 6	7	8	9	10	11
			Moderate similarity			
1	1,218 (1,159)	1,608 (1,575)	2,077 (2,148)	2,871 (2,686)	2,966 (3,010)	2,886 (3,364)
2	944 (1,044)	1,377 (1,537)	1,531 (1,751)	2,001 (2,129)	2,357 (2,302)	2,259 (2,413)
3	912 (942)	1,017 (1,338)	1,515 (1,564)	1,275 (1,643)	1,921 (1,696)	1,769 (1,739)
4	826 (969)	924 (1,164)	1,224 (1,322)	1,225 (1,449)	1,706 (1,466)	1,446 (1,460)
5	785 (881)	888 (1,078)	1,083 (1,187)	1,112 (1,249)	1,310 (1,290)	1,263 (1,319)
6	697 (813)	782 (1,066)	1,065 (1,056)	1,034 (1,127)	1,371 (1,188)	1,140 (1,133)
7	712 (790)	766 (930)	961 (1,053)	1,087 (1,073)	1,364 (1,071)	1,065 (1,047)
8	665 (762)	763 (910)	933 (981)	947 (960)	1,118 (1,035)	865 (1,009)
9	665 (747)	718 (879)	958 (900)	936 (940)	1,042 (906)	843 (944)
10	668 (696)	710 (786)	933 (860)	938 (872)	984 (885)	751 (858)
11	663 (648)	708 (788)	920 (839)	898 (862)	966 (839)	753 (839)
12	687 (655)	709 (759)	876 (801)	883 (823)	826 (821)	768 (786)
13	622 (647)	702 (735)	880 (760)	833 (760)	804 (812)	728 (760)
14	645 (616)	663 (688)	877 (771)	807 (749)	911 (753)	705 (741)
15	621 (617)	644 (714)	767 (730)	762 (754)	725 (738)	690 (730)
16	646 (594)	659 (702)	799 (707)	777 (715)	746 (739)	653 (740)
17	621 (608)	652 (669)	769 (718)	768 (696)	729 (716)	707 (733)
18	603 (596)	639 (669)	754 (688)	804 (680)	715 (683)	711 (680)
19	624 (609)	640 (623)	732 (686)	736 (681)	693 (668)	669 (687)
20	624 (552)	623 (616)	758 (663)	737 (664)	669 (681)	681 (699)
			Low similarity			
1	1,361 (1,152)	2,111 (1,682)	2,618 (2,317)	2,547 (2,779)	2,841 (3,228)	3,387 (3,632)
2	1,098 (1,081)	1,820 (1,550)	2,121 (1,992)	2,395 (2,289)	2,588 (2,456)	2,998 (2,637)
3	1,003 (1,079)	1,602 (1,429)	1,578 (1,719)	2,062 (1,917)	2,125 (2,014)	2,204 (2,157)
4	903 (999)	1,468 (1,289)	1,357 (1,490)	1,897 (1,641)	1,793 (1,697)	2,003 (1,804)
5	910 (916)	1,371 (1,159)	1,286 (1,337)	1,495 (1,454)	1,671 (1,455)	1,450 (1,532)
6	828 (861)	1,154 (1,115)	1,164 (1,195)	1,442 (1,299)	1,659 (1,295)	1,273 (1,332)
7	762 (842)	973 (1,015)	1,058 (1,141)	1,454 (1,234)	1,447 (1,230)	1,151 (1,247)
8	714 (845)	900 (968)	1,048 (1,039)	1,234 (1,119)	1,184 (1,134)	942 (1,151)
9	717 (777)	862 (932)	994 (982)	1,188 (1,065)	1,212 (1,089)	999 (1,092)
10	685 (747)	895 (895)	970 (976)	1,176 (1,043)	1,186 (997)	892 (1,009)
11	694 (760)	819 (907)	926 (933)	1,020 (969)	1,085 (950)	796 (969)
12	673 (694)	850 (870)	928 (877)	994 (902)	1,055 (949)	890 (959)
13	694 (687)	824 (837)	879 (864)	948 (888)	1,048 (872)	807 (880)
14	660 (697)	821 (785)	912 (837)	899 (863)	923 (866)	798 (896)
15	633 (678)	779 (780)	767 (842)	857 (832)	880 (843)	722 (876)
16	625 (670)	735 (725)	747 (840)	829 (823)	872 (820)	741 (804)
17	600 (643)	695 (760)	733 (788)	830 (790)	893 (811)	805 (770)
18	596 (640)	697 (760)	773 (765)	812 (768)	857 (778)	711 (801)
19	600 (632)	669 (705)	743 (752)	800 (750)	855 (766)	679 (747)
20	597 (609)	693 (706)	717 (748)	751 (761)	806 (742)	673 (767)
			Unrelated			
1	1,495 (1,177)	2,024 (1,743)	2,381 (2,347)	2,680 (2,768)	3,002 (3,192)	3,364 (3,771)
2	1,350 (1,107)	1,684 (1,575)	2,237 (1,972)	2,401 (2,387)	2,877 (2,627)	2,983 (2,798)
3	1,236 (1,039)	1,552 (1,458)	1,791 (1,828)	1,927 (2,012)	2,414 (2,155)	2,593 (2,243)
4	1,057 (987)	1,400 (1,335)	1,556 (1,549)	1,789 (1,689)	2,070 (1,914)	2,158 (1,826)
5	1,036 (981)	1,268 (1,271)	1,256 (1,443)	1,582 (1,562)	1,632 (1,611)	1,687 (1,679)
6	942 (953)	1,073 (1,157)	1,144 (1,319)	1,349 (1,405)	1,502 (1,430)	1,479 (1,498)
7	844 (903)	1,030 (1,093)	1,104 (1,204)	1,335 (1,310)	1,482 (1,341)	1,470 (1,313)
8	805 (825)	964 (1,012)	1,017 (1,167)	1,100 (1,255)	1,267 (1,220)	1,191 (1,246)
9	771 (800)	916 (994)	929 (1,075)	1,108 (1,175)	1,146 (1,194)	1,171 (1,216)
10	809 (800)	959 (964)	922 (1,033)	1,088 (1,136)	1,187 (1,083)	1,094 (1,113)
11	781 (780)	898 (884)	849 (1,054)	1,008 (1,043)	1,195 (1,075)	953 (1,032)
12	728 (730)	893 (893)	940 (1,000)	974 (997)	1,028 (1,005)	891 (994)
13	702 (709)	860 (887)	851 (917)	994 (980)	985 (935)	843 (976)

Appendix A (*continued*)

Session	Numerosity 6	7	8	9	10	11
14	692 (710)	876 (849)	881 (876)	967 (917)	874 (901)	829 (961)
15	676 (732)	775 (847)	824 (851)	890 (906)	849 (905)	773 (914)
16	638 (693)	794 (796)	827 (830)	919 (883)	807 (909)	796 (862)
17	657 (688)	771 (764)	788 (815)	904 (847)	812 (835)	790 (857)
18	654 (666)	737 (763)	794 (813)	886 (829)	803 (845)	764 (828)
19	645 (667)	762 (742)	780 (822)	816 (817)	751 (813)	735 (855)
20	650 (623)	723 (770)	799 (786)	852 (830)	748 (818)	730 (804)

Note. EBRW = exemplar-based random walk.

Appendix B

Exemplar-Based Random Walk (EBRW) Predictions and Power Analyses of Standard Deviations From Experiment 2

Predictions of standard deviations were made by using the parameters that best fit the means (observed and predicted standard deviations are available from me). The correlation between the observed and predicted standard deviations was good (r = .807), especially given that the observed standard deviations are quite noisy and that no attempt was made to explicitly fit the observed standard deviations. As shown in Table B1, power law functions fitted the noisy observed data fairly well (average r_{obs} = .821 $RMSE_{obs}$ = 118.06) and fitted the predictions quite well (average r_{pred} = .971 $RMSE_{pred}$ = 46.16).

Recall that instance theory predicts that the exponent of the power law functions will be the same for observed RT means and standard deviations (Logan, 1988, 1992). Simple inspection of Tables 5 and B1 reveals that, in most cases, the exponent of the observed standard deviations is quite a bit smaller than the exponent of the observed means. For the most part, the exponents of the standard deviations are closer to those of the EBRW predicted standard deviations. As a more systematic comparison, constrained power law functions were fitted to the observed standard deviations by setting the exponent equal to that of the observed means or equal to that of the predicted standard deviations. Constrained fits were somewhat better with the EBRW standard deviation exponent (for moderate, average r = .839, $RMSE$ = 106.82; for low, average r = .801, $RMSE$ = 131.12; for unrelated, average r = .789, $RMSE$ = 127.34) than with the observed mean exponent (for moderate, average r = .806, $RMSE$ = 118.5; for low, average r = .776 $RMSE$ = 138.42; for unrelated, average r = .773, $RMSE$ = 131.32).

Table B1
Measures of Goodness of Fit and Exponent Parameter C for Power Law Fits to Observed and Predicted Response Time Standard Deviations at Each Level of Numerosity for Patterns From Experiment 2

Standard deviation type, parameter, and measure of fit	Numerosity 6	7	8	9	10	11
	Moderate-similarity pattern					
Observed standard deviations						
C	0.619	0.443	0.180	0.149	0.123	0.165
RMSE	55.99	67.44	82.16	78.59	175.48	156.53
r	.789	.867	.826	.901	.820	.891
Predicted standard deviations						
C	0.183	0.184	0.236	0.225	0.230	0.264
RMSE	28.72	35.74	27.84	37.55	47.04	62.92
r	.973	.974	.989	.985	.983	.976

(*table continues*)

Table B1 (*continued*)

Standard deviation type, parameter, and measure of fit	Numerosity 6	7	8	9	10	11
	Low-similarity pattern					
Observed standard deviations						
C	0.343	0.128	0.113	0.105	0.093	0.129
RMSE	47.63	106.54	119.79	164.40	156.30	166.46
r	.901	.854	.820	.757	.715	.844
Predicted standard deviations						
C	0.156	0.190	0.182	0.166	0.190	0.215
RMSE	32.62	33.35	37.27	54.50	55.92	64.08
r	.958	.974	.977	.965	.973	.976
	Unrelated pattern					
Observed standard deviations						
C	0.184	0.163	0.154	0.103	0.075	0.114
RMSE	58.25	60.48	82.53	124.20	220.70	201.69
r	.912	.898	.876	.745	.548	.813
Predicted standard deviations						
C	0.159	0.183	0.171	0.158	0.162	0.194
RMSE	32.19	31.29	40.28	62.44	77.90	70.03
r	.960	.976	.971	.953	.949	.971

Note. *C* = the learning rate parameter that specifies the shape of the power law function.

Appendix C

Observed Mean Response Times (in Milliseconds) and EBRW Predictions (in Parentheses) as a Function of Type (Friends or Enemies), Numerosity, and Session in Experiment 3

Session	Numerosity 6	7	8	9	10	11
	Friends					
1	1,326 (1,349)	1,993 (1,762)	2,355 (2,167)	2,513 (2,449)	2,393 (2,722)	2,538 (2,855)
2	1,178 (1,225)	1,486 (1,482)	2,021 (1,732)	1,887 (1,842)	1,739 (1,917)	1,822 (1,966)
3	1,029 (1,093)	1,294 (1,329)	1,604 (1,414)	1,592 (1,514)	1,236 (1,506)	1,578 (1,564)
4	961 (1,018)	1,146 (1,170)	1,285 (1,236)	1,451 (1,276)	993 (1,260)	1,286 (1,281)
5	857 (956)	1,119 (1,068)	1,160 (1,131)	1,310 (1,120)	903 (1,107)	1,090 (1,131)
6	819 (896)	1,145 (990)	1,065 (1,017)	1,247 (1,016)	869 (1,025)	1,039 (1,034)
7	776 (857)	1,023 (911)	1,003 (940)	1,114 (956)	796 (974)	935 (971)
8	765 (825)	920 (885)	935 (898)	1,077 (900)	751 (896)	842 (908)
9	723 (781)	889 (842)	883 (852)	969 (844)	752 (858)	760 (861)
10	697 (757)	860 (809)	890 (833)	905 (817)	728 (835)	769 (812)
11	697 (738)	894 (791)	841 (792)	867 (800)	734 (795)	727 (800)
12	707 (731)	857 (769)	849 (773)	839 (780)	709 (776)	723 (784)
13	674 (712)	772 (756)	762 (766)	793 (750)	665 (748)	698 (750)
14	691 (713)	789 (724)	755 (729)	761 (748)	669 (750)	696 (729)
15	676 (691)	771 (731)	765 (721)	791 (735)	665 (731)	712 (726)
16	660 (682)	792 (706)	749 (721)	764 (699)	647 (715)	679 (711)
17	682 (671)	761 (696)	735 (696)	732 (712)	651 (699)	662 (695)
18	649 (657)	712 (689)	754 (687)	734 (690)	627 (690)	660 (689)
19	634 (669)	701 (663)	702 (682)	687 (677)	634 (678)	653 (679)
20	620 (646)	692 (670)	693 (668)	676 (679)	612 (676)	640 (677)

Appendix C (*continued*)

Session	Numerosity 6	7	8	9	10	11
	Enemies					
1	1,332 (1,335)	1,859 (1,797)	2,299 (2,188)	2,694 (2,522)	2,706 (2,787)	2,749 (3,049)
2	1,222 (1,236)	1,664 (1,577)	1,796 (1,791)	1,949 (2,011)	2,244 (2,092)	2,215 (2,164)
3	1,044 (1,132)	1,397 (1,401)	1,596 (1,546)	1,755 (1,636)	1,889 (1,710)	1,957 (1,668)
4	942 (1,062)	1,278 (1,264)	1,425 (1,360)	1,583 (1,408)	1,577 (1,466)	1,428 (1,440)
5	915 (1,011)	1,215 (1,174)	1,363 (1,238)	1,350 (1,283)	1,327 (1,280)	1,328 (1,267)
6	890 (961)	1,173 (1,074)	1,335 (1,188)	1,272 (1,154)	1,144 (1,156)	1,099 (1,225)
7	830 (930)	1,043 (1,041)	1,203 (1,079)	1,144 (1,080)	954 (1,095)	986 (1,105)
8	806 (900)	1,049 (999)	1,182 (1,024)	1,117 (1,038)	973 (1,033)	925 (1,023)
9	781 (880)	971 (931)	1,112 (989)	1,065 (989)	845 (990)	882 (998)
10	778 (841)	972 (910)	1,134 (973)	1,034 (950)	829 (945)	867 (966)
11	750 (828)	923 (918)	1,047 (910)	968 (914)	826 (938)	809 (914)
12	790 (810)	954 (868)	1,055 (884)	960 (902)	814 (896)	852 (894)
13	725 (783)	886 (856)	971 (882)	873 (895)	824 (880)	793 (877)
14	733 (798)	829 (834)	921 (853)	869 (865)	761 (854)	779 (885)
15	784 (776)	866 (818)	939 (830)	837 (827)	774 (832)	817 (861)
16	726 (763)	874 (826)	895 (838)	845 (827)	762 (826)	767 (828)
17	706 (770)	849 (786)	880 (820)	810 (813)	760 (827)	785 (824)
18	692 (752)	809 (789)	863 (809)	769 (813)	722 (803)	773 (794)
19	699 (732)	761 (779)	808 (794)	758 (804)	696 (772)	713 (797)
20	689 (739)	768 (773)	815 (784)	759 (782)	694 (784)	727 (791)

Note. EBRW = exemplar-based random walk.

Appendix D

Exemplar-Based Random Walk (EBRW) Predictions and Power Analyses of Standard Deviations From Experiment 3

Predictions of standard deviations were made using the parameters that best fit the means (observed and predicted standard deviations are available from me). The correlation between the observed and predicted standard deviations was quite impressive ($r = .878$). As shown in Table D1, power law functions were fitted to the standard deviations of the observed and predicted RTs. Both the observed (average $r_{obs} = .899$, $RMSE_{obs} = 106.04$) and predicted (average $r_{pred} = .971$, $RMSE_{pred} = 31.33$) RTs were well in accord with the power law.

Assessing the strong power law prediction of instance theory requires comparing the power law exponent of the observed standard deviations to that of the observed means and to that of the EBRW predicted standard deviations. Simple inspection of Tables 6 and D1 does not reveal an advantage for either comparison. As a more systematic analysis, constrained power law functions were fitted to the observed standard deviations by setting the exponent equal to the exponent of the observed means or the exponent of the EBRW predicted standard deviations. Constrained fits were slightly better with the EBRW exponent (for friends, average $r = .926$, $RMSE = 95.15$; for enemies, average $r = .851$, $RMSE = 121.18$) than with observed mean exponent (for friends, average $r = .920$, $RMSE = 100.14$; for enemies, average $r = .838$, $RMSE = 127.59$). Again, although these results may question the strong instance theory predictions, the findings were not overwhelming. Given the inherent problems in obtaining clean standard deviation data, it may prove difficult to systematically test the standard deviation power law predictions.

(*Appendix continues on next page*)

Table D1

Measures of Goodness of Fit and Exponent Parameter C for Power Law Fits to Observed and Predicted Response Time Standard Deviations at Each Level of Numerosity for Friends and Enemies From Experiment 3

Standard deviation type, parameter, and measure of fit	Numerosity 6	7	8	9	10	11
Friends						
Observed standard deviations						
C	0.378	0.780	0.145	0.197	0.505	0.375
RMSE	56.42	64.39	126.54	85.27	109.81	105.52
r	.934	.943	.885	.955	.936	.950
Predicted standard deviations						
C	0.175	0.156	0.192	0.218	0.299	0.375
RMSE	14.23	27.85	32.38	37.11	34.84	38.43
r	.980	.959	.972	.976	.986	.988
Enemies						
Observed standard deviations						
C	0.150	0.149	0.151	0.178	0.203	0.223
RMSE	48.66	84.00	147.36	98.76	174.42	171.35
r	.900	.860	.682	.913	.857	.899
Predicted standard deviations						
C	0.150	0.149	0.151	0.178	0.203	0.279
RMSE	12.27	22.55	35.83	39.28	46.02	35.13
r	.972	.954	.948	.963	.971	.988

Note. *C* = the learning rate parameter that specifies the shape of the power law function.

Received October 26, 1995
Revision received July 22, 1996
Accepted July 25, 1996 ■