Journal of Experimental Psychology:
Learning, Memory, and Cognition
1999, Vol. 25, No. 2, 543–551

0278-7393/99/$3.00

Theories of Automaticity and the Power Law of Practice

Thomas J. Palmeri
Vanderbilt University

In a recent reanalysis of numerosity judgment data from T. J. Palmeri (1997), T. C. Rickard (1999) found that mean response times did not decrease as a pure power law of practice and standard deviations systematically increased and then decreased with practice in some conditions. Rickard argued that these results were consistent with the component power laws (CMPL) theory of strategy shifting (Rickard, 1997), but were inconsistent with instance theory (G. D. Logan, 1988) and the exemplar-based random walk (EBRW) model (R. M. Nosofsky & Palmeri, 1997). In this article, the author demonstrates how a slightly more complex power function fitted the numerosity data nearly as well as the CMPL function, and how race models, such as instance theory and the EBRW, can predict deviations from a pure power law of practice and can predict nonmonotonic changes in standard deviations with practice. Potential limitations of CMPL are also discussed.

One of the most ubiquitous findings in experimental psychology is that the time needed to perform most tasks decreases with practice. Newell and Rosenbloom (1981) observed power-function decreases across nearly every task that showed practice effects and proposed that this functional relationship be elevated to the status of a psychological law. Since then, the *power law of practice* has been viewed as a fundamental benchmark result and has had tremendous influence on the development of theories of automaticity (e.g., Anderson, 1982, 1987, 1992; Cohen, Dunbar, & McClelland, 1990; Ericsson, Krampe, & Tesch-Romer, 1993; Logan, 1988, 1992; MacKay, 1982; Nosofsky & Palmeri, 1997; Palmeri, 1997; Rickard, 1997).

However, recent evidence suggests that the power law may not be as lawful as was once thought (Delaney, Reder, Staszewski, & Ritter, 1998; Heathcote & Mewhort, 1995; Rickard, 1997, 1999). One could argue that if deviations from the power law are observed, then theories that have been explicitly formulated to predict power-law decreases in response time (RT) must surely be wrong. My goal is to discuss whether systematic deviations from the power law pose real difficulties for instance-based theories of automaticity as has been recently claimed by Rickard (1997, 1999).

The Power Law of Practice

Newell and Rosenbloom (1981) observed that the relationship between practice and performance is one in which substantial gains are made early in practice but dramatically diminishing marginal gains are made with increased practice. Formally, the power law of practice is given by

$$RT = A + B \cdot (N + pre)^{-C}, \tag{1}$$

where RT is the response time on trial N, A is the asymptotic RT, B is the difference between initial and final RT, pre is the amount of prior extraexperimental practice, and C is the learning rate parameter that specified how quickly RTs reach asymptote.

Instance Theory and the Exemplar-Based Random Walk (EBRW) Model

One of the most influential theories of automaticity has been instance theory (Logan, 1988). According to the theory, people begin solving a task using general algorithms, strategies, or rules. Every time an instance of that cognitive skill is performed, a trace of that action is obligatorily stored in memory. When a new object must be judged, a race ensues between the completion of the algorithm and retrieval of particular instances from memory, with the winner determining the overt response. Initially, performance is governed solely by the algorithm. However, increases in speed of responding with practice are not due to more efficient application of the algorithm. Rather, memory retrieval gets faster as more repetitions of particular instances are stored in memory. Memory retrieval is assumed to be another race process, with all instances competing to be retrieved in parallel; as the number of runners increases, the expected winning time decreases (see Logan, 1988, 1992). Soon, memory retrieval dominates the race against algorithmic processing. Therefore, qualitative changes in performance are due to shifts from algorithmic processing to memory retrieval.

Logan (1988, 1992; see also Colonius, 1995; Logan, 1995) demonstrated that if memory retrieval is a race

This work was supported by Vanderbilt University Research Council Direct Research Support grants. I thank Peter Delaney, Andrew Heathcote, Gordon Logan, Robert Nosofsky, and Timothy Rickard for providing comments on an earlier version of this article.

For more information, see Thomas J. Palmeri's web address: http://www.vanderbilt.edu/AnS/psychology/cogsci/palmeri/home.html.

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 to Thomas.J.Palmeri@Vanderbilt.edu.

process and if memory retrieval times are distributed as Weibulls,

$$f(t) = 1 - \exp\,[-\,((t - b)/a)^c]\,, \tag{2}$$

then not only do means and standard deviations of retrieval times decrease as a power function of the number of instances stored in memory, but the entire retrieval time distribution decreases as a power function as well. Therefore, quantitative changes in processing speed, characterized by the power law, are caused by increasing numbers of instances in memory that race to be retrieved.

Because the locus of automaticity lies in specific memories for previous instances, instance theory predicts extremely narrow transfer of cognitive skills to new situations. The highly specific nature of transfer was demonstrated by Lassaline and Logan (1993; see also Palmeri, 1997) using a numerosity task in which people judged the number of elements in a pattern as quickly as possible without making errors. Initially, RTs increased as a function of numerosity, indicating that explicit counting dominated performance. With several sessions of training, however, RTs became flat as a function of numerosity, suggesting that memory retrieval dominated performance. To assess specificity, participants judged both old and new patterns in a transfer phase at the end of training. No transfer to new patterns was observed. Whereas RTs to the old patterns were the same as they were at the end of training, RTs to the new patterns were the same as they were at the beginning of training (see also Palmeri, 1997). Such specificity of transfer suggests that training did not cause the algorithm (in this case *counting*) to speed up by any appreciable amount.

However, the memory-retrieval assumptions of instance theory are limited by not taking into account graded similarities among exemplars and by not allowing response competition. The exemplar-based random walk (EBRW) model (Nosofsky & Palmeri, 1997; Palmeri, 1997) dealt with these limitations by combining elements of instance theory with elements of the generalized context model of categorization (Nosofsky, 1986). According to the model, all instances race to be retrieved with rates proportional to their similarity to some presented item. Unlike instance theory, a single memory retrieval does not suffice. Rather, each retrieval provides incremental evidence to a random walk process (Link, 1975; Luce, 1986; Ratcliff, 1978). Once one response exceeds all others by some criterial amount, an output is made. As with instance theory, Palmeri (1997) assumed that the EBRW races against an algorithm. With practice, memory retrieval speeds up, causing the random walk to accumulate evidence more quickly, causing the EBRW to eventually win the race over the algorithm.

In a recent series of experiments, Palmeri (1997) tested participants on numerosity judgment tasks in which the similarities between patterns were systematically manipulated. Consistent with the EBRW,[1] automaticity transferred as a function of similarity to training patterns, increases in within-category similarity facilitated the development of automaticity, and increases in between-category similarity inhibited the development of automaticity. The combination of similarity-based retrieval and a competitive random walk decision process allowed the EBRW to account for these findings.

Component Power Laws (CMPL) Theory of Automaticity

Rickard (1997) recently proposed an alternative account of the development of automaticity called the component power laws (CMPL) theory. Although concurring that automaticity reflects a shift from algorithmic processes to memory retrieval, CMPL differs from instance theory and the EBRW in quite a number of important ways (see Rickard, 1997, for details). However, two fundamental differences were made particularly salient by Rickard (1999). First, CMPL differs from instance theory and the EBRW in that the algorithm *and* memory retrieval cannot be executed in parallel. Rather, on each trial, the algorithm *or* memory retrieval, but not both, are executed. Therefore, at each stage of training, overall RT reflects a mixture of the two components,

$$RT = RT_{\text{alg}} \cdot (1-p) + RT_{\text{mem}} \cdot p\,, \tag{3}$$

where p is the probability that memory retrieval is performed rather than the algorithm on that given trial. Second, unlike instance theory and the EBRW, CMPL assumes that both algorithm completion times *and* memory retrieval times decrease as a power law of practice. Rather than substitute the complete four-parameter power-law function from Equation 1 into Equation 3, Rickard (1997, 1999) found that the following simplifications could reasonably be made:

$$RT_{\text{alg}} = B_{\text{alg}} \cdot (N + pre)^{-C_{\text{alg}}} \tag{4}$$

$$RT_{\text{mem}} = B_{\text{mem}} \cdot N^{-C_{\text{mem}}}\,. \tag{5}$$

Rickard (1999) also assumed the probability that memory retrieval is used on a given trial to be an exponential function of the trial number

$$p = 1 - \exp[\,-\,r \cdot (N - 1)]\,, \tag{6}$$

where r is a rate constant. The resulting CMPL function is given by

$$\begin{aligned} RT = B_{\text{alg}} \cdot (N + pre)^{-C_{\text{alg}}} \cdot \exp[-\,r(N - 1)] \\ + B_{\text{mem}} \cdot N^{-C_{\text{mem}}} \cdot [1 - \exp[-\,r(N - 1)]]\,. \end{aligned} \tag{7}$$

[1] Formally, EBRW is an instance-based model of categorization (Nosofsky & Palmeri, 1997). To account for the development of automaticity in numerosity judgments, Palmeri (1997) assumed that the EBRW raced against an algorithm. While the term *EBRW* was used in that work, and will continue to be used in this work, to refer to this hybrid model (see also Rickard, 1999), the simultaneous race assumption is not fundamental to the EBRW model of categorization per se.

Although its components are power-law functions, this equation is not itself a power law, in general.

Rickard (1997) found evidence consistent with CMPL in a series of pseudoarithmetic tasks similar to others used in the automaticity literature (e.g., Compton & Logan, 1991; Delaney et al., 1998; Logan & Klapp, 1991). Although the power law generally did not account for the overall RT data all that well, the power-law did hold within a given strategy (algorithm or memory retrieval) after the data were suitably partitioned according to the type of strategy used (see also Delaney et al., 1998).

A CMPL Alternative Account of Practice Effects in Numerosity Judgment Tasks

The impetus for the present article was a reanalysis Rickard (1999) performed on data reported in Palmeri (1997). In three experiments, participants judged the numerosity of patterns of between 6 and 11 dots over many sessions as quickly as possible without making errors. Whereas I originally reported these RT data as a function of session, collapsed across training blocks within a session, Rickard (1999) reanalyzed my data as a function of training block instead. His primary goal was to compare fits of the power law with those of CMPL. According to Rickard, significant deviations from a power law might constitute a significant challenge to both instance theory and the EBRW because predictions that were a close approximation to a three-parameter power function were previously used as support for the theories (Logan, 1988, 1992; Palmeri, 1997).

For each of the experiments of Palmeri (1997), Rickard fitted three-parameter power functions at each level of numerosity. The power law fitted the data rather poorly, accounting, on average, for less than 90% of the variance in the observed data. Visual inspection of figures provided by Rickard (1999) confirmed this poor fit—in some cases, there was little overlap between the observed data and the best fitting power-law function, especially for large numerosities. By contrast, excellent fits of constrained versions of CMPL, with fewer free parameters than the power-law functions, were obtained.

Recall, the CMPL theory assumes that algorithm completion times and memory retrieval times decrease as a power law of practice, but the function resulting from the mixture of these two processes is not itself a power law. In his analyses, the power law decrease in algorithm completion times actually contributed fairly little to the success of CMPL. Rather, the power-law decrease in retrieval times coupled with the competition between the algorithm and memory retrieval performed most of the work in accounting for the observed practice effects in these experiments. Therefore, for present purposes, the most important difference between CMPL and the instance-based theories in accounting for the numerosity judgment data given in Palmeri (1997) lies in the strategy-selection process—is it a competition or is it a race?

Very good fits of the power-law function to the numerosity judgment data were previously reported because observed data were averaged across blocks within a session (Palmeri, 1997). Clearly, averaging masks potentially important empirical regularities, particularly in the first few training sessions (Heathcote & Mewhort, 1995); other previously reported successes of the power law may be partially an averaging artifact as well. Unfortunately, until just recently, the power law was generally taken as a benchmark characterization of data—systematic deviations were simply not expected to be found. Rickard's (1999) reanalysis of my data and other recent work (e.g., Delaney et al., 1998; Heathcote & Mewhort, 1995; Rickard, 1997) are very important because they strongly suggest that practice effects may not completely conform to a power law after all.

Rickard's (1999) reanalyses do show that RTs in numerosity tasks do not decrease as a pure power-law function of practice. But does he provide an alternative account of the development of automaticity in numerosity judgment tasks? As discussed earlier, one of the primary goals of my earlier article (Palmeri, 1997) was to examine the effects of pattern similarity on the development of automaticity and to try to account for these effects with the EBRW. As Rickard (1999) conceded, "the current version of the CMPL theory simply cannot account for such effects, and it is an open question whether it can be extended to do so" (p. xx)—the CMPL theory may be too simple to provide a complete alternative account to the one provided by the EBRW.

However, Rickard's (1999) reanalyses do pose potential challenges to instance theory and the EBRW, which I attempt to address in subsequent sections of this article. I begin by comparing the fits of power-law functions with fits of CMPL functions and discuss what can be learned from curve fitting. I next examine whether models that assume a race between algorithm completion and memory retrieval, such as instance theory and the EBRW, can predict deviations from a pure power-law function of practice. Finally, I examine how predicted and observed RT standard deviations change with practice.

What Can Be Learned From Curve Fitting?

Rickard (1999) reported poor fits of power-law functions compared with CMPL functions when applied to the Palmeri (1997) data. However, Rickard tested only a restricted version of the power-law function in which no prior practice was assumed (*pre* = 0). I fitted the three-parameter power law (Equation 1 with *pre* = 0), the four-parameter power law (Equation 1), and the five-parameter CMPL (Equation 7, with $c_{\text{alg}} = c_{\text{mem}}$) to Palmeri's (1997) data. Table 1 provides fit values (r^2) as a function of experimental condition and numerosity; for illustration, Figure 1 displays the three functions and the observed data points from Experiment 1. As Rickard (1999) reported, across nearly every condition, three-parameter power-law functions did not adequately capture the observed decrease in RTs with practice. However, there was actually very little difference in fit between four-parameter power-law functions and five-parameter CMPL functions.[2] As illustrated in Figure 1, the functions

[2] It must be noted that Rickard (1999) significantly reduced the number of free parameters of CMPL by constraining the values of several parameters in meaningful ways.

Table 1
Fits (r^2) of Three-Parameter Power-Law Function, Four-Parameter Power-Law Function, and Five-Parameter CMPL Function to Numerosity Judgment Data From Palmeri (1997)

Experiment and function	Condition	Numerosity 6	7	8	9	10	11	All
Experiment 1								
Power (3)	—	.928	.902	.874	.908	.843	.828	.873
Power (4)	—	.959	.948	.969	.980	.978	.962	.961
CMPL (5)	—	.942	.937	.962	.966	.968	.968	.967
Experiment 2								
Power (3)	Moderate	.885	.897	.893	.882	.824	.868	.876
Power (4)	Moderate	.886	.923	.920	.928	.931	.941	.939
CMPL (5)	Moderate	.889	.931	.921	.949	.936	.955	.949
Power (3)	Low	.915	.833	.899	.789	.824	.799	.841
Power (4)	Low	.932	.929	.952	.935	.942	.925	.942
CMPL (5)	Low	.933	.940	.959	.950	.952	.946	.955
Power (3)	Unrelated	.842	.837	.837	.846	.784	.813	.833
Power (4)	Unrelated	.913	.925	.929	.942	.935	.949	.945
CMPL (5)	Unrelated	.920	.934	.956	.957	.949	.968	.960
Experiment 3								
Power (3)	Friends	.879	.904	.891	.909	.913	.919	.911
Power (4)	Friends	.923	.934	.959	.973	.947	.967	.960
CMPL (5)	Friends	.928	.936	.974	.973	.969	.976	.969
Power (3)	Enemies	.887	.882	.891	.890	.878	.870	.879
Power (4)	Enemies	.925	.933	.944	.957	.951	.944	.951
CMPL (5)	Enemies	.928	.933	.944	.963	.973	.968	.964

Note. The number of parameters in each function is indicated within parentheses. CMPL = component power laws.

were nearly coincident across most experimental conditions. On average, the one additional parameter of CMPL captured approximately 1% additional variance in the observed data. Moreover, the best fitting power-law functions usually had A equal to zero (see also Heathcote & Mewhort, 1995; Logan, 1988); in such cases, the power law is a special case of CMPL, and it would be mathematically impossible for the power law to fit the data any better than CMPL.[3]

Certainly, I do not want to suggest the four-parameter power-law function as any kind of serious psychological process model. Rather, my purpose in conducting these fits was to counter Rickard's (1999) strong claim that finding poor fits of the three-parameter power function might be enough to rule out theories of automaticity that assumed concurrent execution of algorithmic and memory-retrieval strategies. In fact, on their own, the results of this curve fitting are insufficient to rule out even fairly broad classes of theories. In this case, a fairly simple power function, which could well characterize the predictions of any number of *single-process* models of automaticity, provided a good account of the observed practice curves (see also Van Zandt & Ratcliff, 1995). Generally, great care must be taken in using any single pattern of data to make strong inferences about underlying psychological structures (e.g., Townsend, 1990; Van Zandt & Ratcliff, 1995). Theories must be evaluated on their ability to account for a broad spectrum of empirical results.

Can Race Models Predict Deviations From the Power Law of Practice?

However, there still is the question of whether race models, such as instance theory and the EBRW, can predict deviations from the pure power law of practice. A critical assumption made by Rickard (1997) was that instance-theory predictions were equated with power-law predictions—a failure to fit the power law meant a failure of instance theory. Although this assumption was somewhat relaxed by Rickard (1999), it still implicitly underlies his discussion.

As mentioned earlier, if retrieval times are distributed as Weibulls and if memory retrieval is a race process, then memory retrieval times decrease as a power law (Logan, 1988, 1992). However, memory retrieval races against algorithm completion, which has its own temporal characteristics. Does this additional race distort the power-law predictions? As Logan (1988, 1992) has pointed out repeatedly, it must. However, power-law predictions made by

[3] If $A = 0$ in the power-law function (Equation 1), then CMPL (Equation 7) can always mimic this power-law function by simply setting r equal to zero. Therefore, it would be impossible to find a set of data for which CMPL could not provide an equal or better fit than the power-law function (without A). In this sense, using the terminology of hierarchical model testing, the power-law function (without A) is a special case of CMPL. Hence, it should come as no surprise that CMPL provided a somewhat better fit.

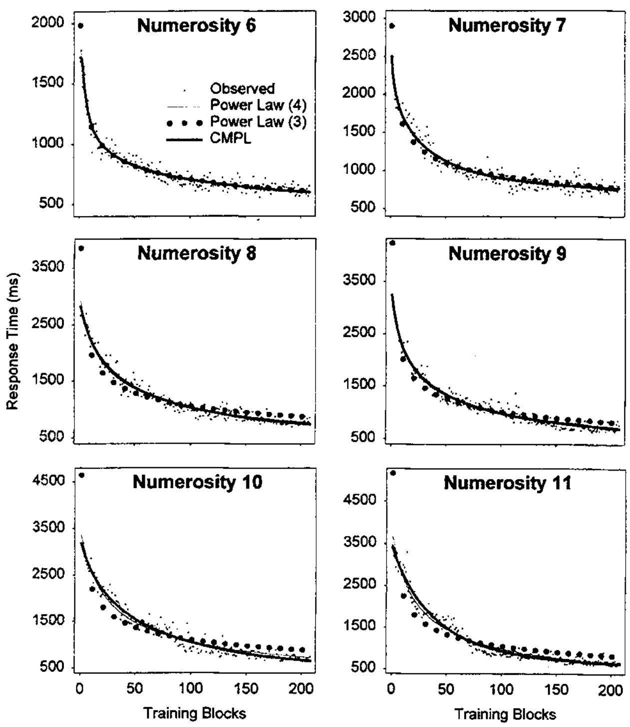

Figure 1. Analyses of numerosity judgment data from Experiment 1 of Palmeri (1997). Dots indicate observed mean RTs as a function of training block, fine lines indicate best fitting four-parameter power-law functions, circles indicate best fitting three-parameter power-law functions, and thick lines indicate best fitting five-parameter component power laws (CMPL) functions.

instance theory "are not compromised much by the race with the algorithm, *provided that the mean for the algorithm is reasonably close to the mean for memory retrieval [of individual instances]*" [italics added] (Logan, 1988, p. 522). Moreover, any distortions that may occur are limited to the early portions of learning. Once memory retrieval dominates performance, the power law governs further decreases in RT with practice. For simplicity, much of Logan's (1988, 1992; Compton & Logan, 1991) work has assumed that the algorithm does not appreciably distort the predictions of the model; Logan (1988) previously reported how the presence of an algorithm with temporal characteristics similar to those of memory retrieval did not distort the power-law predictions by very much.

Using Monte Carlo simulations, I explored how the presence of a competing algorithm with temporal characteristics different from those of memory retrieval might cause a simple race model to predict deviations from a pure power-law function of practice. For simplicity, in these simulations, a pure instance-based memory retrieval component was assumed (Logan, 1988), rather than the EBRW (the main conclusions from this section hold for the EBRW as well). Algorithm completion and memory retrieval race, with the winner determining the response. The effects of presenting between 1 and 400 training instances were simulated 5,000 times. Memory retrieval times for each stored instance were distributed as Weibulls, with $a_{\text{mem}} = 4{,}000$, $b_{\text{mem}} = 500$, and $c_{\text{mem}} = 2$. Assuming that stored instances race to be retrieved from memory, then the winning memory retrieval time decreases as a power law of the practice. Algorithm completion times were normally distributed, with μ_{alg} and σ_{alg} varying systematically; an additional base time of 500 ms was assumed for the algorithm.

Figure 2 displays the predicted mean RTs as a function of practice for four different combinations of μ_{alg} and σ_{alg}. Three-parameter power-law functions fitted to the predicted RTs are displayed as well. When the algorithm completion times were quite fast ($\mu_{\text{alg}} = 500$ and $\sigma_{\text{alg}} = 50$), the transition from algorithmic processes to memory processes was quite slow, and the power function fitted the predicted RTs quite poorly. By contrast, when the algorithm completion times were quite slow ($\mu_{\text{alg}} = 5{,}000$ and $\sigma_{\text{alg}} = 500$), the transition was quite fast, and the power function fitted the predicted RTs almost perfectly. Clearly, a race model can predict significant deviations from the pure power law. Little or no deviations from the pure power law were observed only when the algorithm was as slow as, or slower than, the mean retrieval time for a single instance (see Logan, 1988).

But, is it plausible for the memory component to be slower than the algorithmic component? Plausibility is most definitively demonstrated using the EBRW as the memory component rather than the conceptually simpler instance theory (although it would be possible to provide a plausibility argument for instance theory as well). Recall that the memory component of the EBRW requires multiple retrievals before a given response is chosen. In the numerosity task, accuracy was highly stressed, so the decision criterion was set quite high (see Palmeri, 1997); that is, evidence for the winning response needed to exceed all other responses by a significant amount. Although individual retrievals might have occurred quickly, a large number of retrievals were necessary before one of the potential responses exceeded all others by a sufficient amount. Moreover, if there was noise in the memory-retrieval process (e.g., Nosofsky & Alfonso-Reese, in press), then it is quite plausible that no response would reach criterion before even a lengthy algorithm was completed. Because RT distributions generated by a random walk can be approximated by a Weibull (e.g., Luce, 1986), and because the EBRW predicts that means and standard deviations for memory retrieval decrease as a power law (Palmeri, 1997), the above simulations nicely summarize the first-order predictions of the EBRW as well.

How unique are the predicted practice curves generated from a race model versus a competition model? The assumption made by CMPL that algorithm completion and memory retrieval compete implies that the observed practice curves will deviate from the power law in a particular way, following the component power law of practice. Rickard (1999) used good fits of the CMPL function as evidence in favor of a competitive strategy-selection process and as evidence against a race process. The five-parameter CMPL was fitted to the instance-theory predictions in Figure 2. In every case, the CMPL function was able to capture well over 99% of the variance in the race model predictions. There-

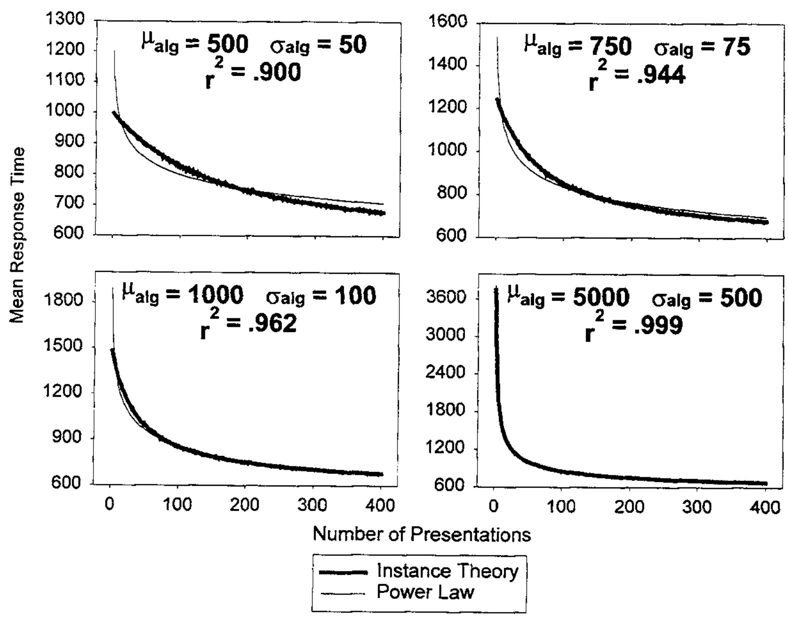

Figure 2. Instance-theory predictions (thick lines) and best fitting three-parameter power-law functions (fine lines) as a function of number of presentations. Fit values (r^2) for the power-law functions are reported. Each panel displays predictions for a different combination of algorithm parameters.

fore, simply fitting the power-law function and the CMPL function to observed data may be insufficient to reveal whether the underlying strategy-selection process is a competition (Rickard, 1997) or a race (Logan, 1988; Palmeri, 1997).

These analyses do reveal a potential problem with instance theory and the EBRW. As originally reported by Rickard (1999), deviations from the pure power law of practice in the Palmeri (1997) data was more pronounced in higher numerosity conditions (e.g., see Figure 1). If we assume that the memory-retrieval time distribution for each pattern is essentially the same, regardless of numerosity (as in Palmeri, 1997), and if we assume that algorithm completion (counting) times increase as a function of numerosity, then a race model will predict faster transitions from algorithmic processes to memory processes for higher numerosity patterns. Therefore, instance theory and the EBRW would predict more significant deviations from the pure power law of practice for patterns with low numerosity than for patterns with high numerosity, opposite of what was found.

There are a number of potential solutions to this problem. Each conspires to cause low-numerosity patterns to be judged via memory retrieval relatively more quickly than high-numerosity patterns, especially early in training. One possibility is that there are many prior extraexperimental memories for low-numerosity patterns. Certainly some patterns of 6 elements, such as a die face, have been seen and enumerated many times. It is less likely that patterns of 10 or 11 elements have ever been enumerated enough times to have formed enduring memory traces prior to the experiment. Another possibility is that patterns of low numerosity may be more similar to one another than patterns of high numerosity. Different patterns of 6 elements may be more similar to one another than are different patterns of 11 elements. The EBRW assumes that the presence of similar instances causes memory retrieval to speed up more rapidly (Palmeri, 1997). Conversely, patterns of low numerosity may be more discriminable from one another than patterns of high numerosity. Patterns of 10 elements may be relatively similar to patterns of 11 elements, whereas patterns of 6 elements may be relatively dissimilar to patterns of 7 elements. The EBRW assumes that the presence of similar instance from another category causes the random walk process to require more steps to reach a decision (Palmeri, 1997). Finally, patterns of high numerosity may simply be more difficult to learn than patterns of low numerosity. If memory retrieval is contaminated by background noise (e.g., Nosofsky & Alfonso-Reese, in press), and if high-numerosity patterns produce "weaker" traces than low-numerosity patterns, then it may require many more repetitions of high-numerosity patterns for them to exert any influence. For example, if each element of a pattern is encoded probabilistically, then high-numerosity patterns would have a lower probability of being completely encoded in memory than low-numerosity patterns. Future empirical research is needed to fully explore each of these alternatives and to determine their effects on the quantitative predictions of the EBRW.

So, CMPL and instance theory (and the EBRW) agree that memory retrieval dominates later trials. However, depending on how effectively memories are stored when first learning a task, many trials may be required before memory retrieval begins to influence behavior in any appreciable way. That is, memory retrieval is likely to be quite noisy,

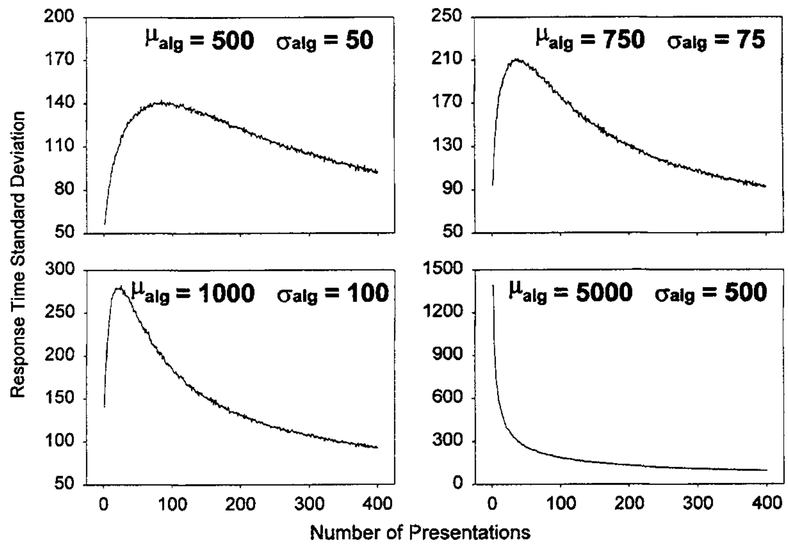

Figure 3. Instance-theory standard deviation predictions as a function of number of presentations. Each panel displays predictions for a different combination of algorithm parameters.

with several repetitions required before people adequately learn the response associated with particular objects. Also, CMPL and instance theory (and the EBRW) agree that algorithm use dominates early trials. This leaves a potentially small transition window when either algorithm use or memory retrieval could govern performance. Determining unequivocally whether these relatively few trials are governed by a competition or a race could prove quite challenging.

What About Deviations From Power-Law Decreases in Standard Deviations?

Another essential component of Rickard (1999) was analyses of RT standard deviations. A probability mixture model, such as CMPL, assumes that RTs are determined by a combination of two parent distributions. The variance of this mixture is given by (Townsend & Ashby, 1983)

$$\sigma^2_{\mathrm{CMPL}} = (1-p)\cdot\sigma^2_{\mathrm{alg}} + p\cdot\sigma^2_{\mathrm{mem}} + p\cdot(1-p)\cdot(\mu_{\mathrm{alg}} - \mu_{\mathrm{mem}})^2. \quad (8)$$

The third term of the equation, $p\cdot(1-p)\cdot(\mu_{\mathrm{alg}} - \mu_{\mathrm{mem}})^2$, first increases and then decreases as p goes from 0 to 1. This can allow total variance to increase and then decrease as a function of training (Compton & Logan, 1991; Logan, 1988), although this "bubble" is not required for every set of parameters (Rickard, 1999). A nonmonotonic change in RT standard deviations would provide support for a mixture model such as CMPL. By contrast, monotonically decreasing RT standard deviations have been taken as evidence in support of race models, such as instance theory (Compton & Logan, 1991; Logan, 1988, 1992): "Race models such as the instance theory are mathematically required to predict a monotonic reduction on average in the overall *SD*" (Rickard, 1999, p. 540), "the race model predicts that variability will decrease with training, and never increase" (Compton & Logan, 1991, p. 152).

However, race models in general, and instance theory in particular, predict monotonically decreasing RT standard deviations only if the mean and standard deviation for the algorithm are reasonably close to the mean and standard deviation for memory retrieval of a single instance. Recall the simulations that were reported earlier. Figure 2 displays mean RTs as a function of training for various combinations of μ_{alg} and σ_{alg}. Figure 3 displays the predicted RT standard deviations from these same simulations as a function of practice. Depending on particular parameter settings, both a mixture model and a race model can predict nonmonotonic changes in RT standard deviations as a function of practice.

Summary

Both CMPL and the instance-based theories (instance theory and the EBRW) assume that the development of automaticity reflects a shift from algorithmic processes to memory retrieval. The present article described two fundamental ways in which the two classes of theories differ: First, CMPL allows algorithm completion times to decrease as a power law of practice. Largely for simplicity, the instance-based theories have assumed that algorithm completion times remain constant throughout training. Certainly, it seems reasonable that algorithm use, and possibly other task components,[4] might improve with practice. It must also be

[4] For example, in the numerosity judgment task (Palmeri, 1997), the participant began each training session with several blocks of practice trials to learn the mapping between numerosities and response keys. On each trial, the name of a number, such as "seven," would appear in the center of the screen, and the

noted that it would be impossible to find incontrovertible evidence that disproved CMPL in favor of the extant instance-based theories with respect to their assumptions about algorithm speedups. In any given paradigm, an absence of algorithm speedup could always be accounted for by CMPL by simply setting c_{alg} equal to zero or *pre* equal to some relatively large number (Rickard, 1999, reported one case in which *pre* was larger than the number of experimental blocks). It becomes a statistical argument whether assuming algorithm speedup contributes significant improvements to the fit of the model. At least in the case of numerosity judgments, it appears that algorithm speedups play a surprisingly minor role in the development of automaticity. Rickard's (1999) own analyses suggest that algorithm speedups in numerosity judgments contribute significantly less to the overall practice effect than do speedups in memory retrieval. Moreover, transfer to new patterns is nearly absent in this task (Lassaline & Logan, 1993; Palmeri, 1997).[5]

The second fundamental difference between CMPL and the instance-based theories centers around the strategy-selection mechanism. Whereas CMPL assumes a competition between algorithm completion and memory retrieval, instance theory and the EBRW assume a race. The analyses reported in this article demonstrate that distinguishing between these two candidate processes may not be as straightforward as has been suggested in previous work (Compton & Logan, 1991; Logan, 1988; Rickard, 1997, 1999). Both a competition model and a race model can predict significant deviations from the pure power-law function of practice. Both a competition model and a race model can predict nonmonotonic changes in RT standard deviations with practice. Of the two possibilities, the race model provides a conceptually simpler account of strategy selection. In a race model, the choice between a retrieval strategy or an algorithmic strategy is entirely determined by the relative amount of time it takes to complete each of the two components. By contrast, in a competition model, without further specified processing constraints, strategy choice is completely independent of completion times. In fact, the strategy selection process may need to be hand tailored to specific domains (see the appendix of Rickard, 1999). In the CMPL equation, the transition from algorithm to memory retrieval is entirely determined by a free parameter, *r*. In my opinion, without greater specification of how the competition between various strategies is resolved, CMPL may be too underconstrained to provide a reasonably testable psychological model.

---

participant would be required to press the corresponding response key as quickly as possible. The speed with which participants could perform this very simple task increased by a few hundred milliseconds over the first four or five sessions. Therefore, some of the practice effects reported by Palmeri (1997) may reflect a speedup in the response-mapping component of numerosity judgments, independently of any speedups in algorithm completion or memory retrieval.

[5] It should be noted that in fairly complex multistep computation tasks (e.g., Delaney et al., 1998; Rickard, 1997), speedups in algorithm completion have been observed.

References

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. (1992). Automaticity and the ACT* theory. *American Journal of Psychology, 105,* 165–180.

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.

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition, 19,* 151–158.

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

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applied by strategy within task. *Psychological Science, 9,* 1–7.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406.

Heathcote, A., & Mewhort, D. J. K. (1995, November). *The law of practice.* Paper presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles.

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

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

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

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

MacKay, D. G. (1982). The problem of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review, 89,* 483–505.

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

Nosofsky, R. M., & Alfonso-Reese, L. (in press). Effects of similarity, practice, and familiarity on speeded classification response times and accuracies: Further tests of an exemplar-retrieval model. *Memory & Cognition.*

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

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

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

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126,* 288–311.

Rickard, T. C. (1999). A CMPL alternative account of practice effects in numerosity judgment tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 532–542.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1,* 46–54.

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

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin & Review, 2,* 20–54.

Received February 26, 1998
Revision received July 14, 1998
Accepted August 10, 1998 ■