

Shapes of Reaction-Time Distributions and Shapes of Learning Curves: A Test of the Instance Theory of Automaticity

Gordon D. Logan
University of Illinois at Urbana-Champaign

The instance theory assumes that automatic performance is based on single-step direct-access retrieval from memory of prior solutions to present problems. The theory predicts that the shape of the learning curve depends on the shape of the distribution of retrieval times. One can deduce from the fundamental assumptions of the theory that (1) the entire distribution of reaction times, not just the mean, will decrease as a power function of practice; (2) asymptotically, the retrieval-time distribution must be a Weibull distribution; and (3) the exponent of the Weibull, which is the parameter that determines its shape, must be the reciprocal of the exponent of the power function. These predictions were tested and mostly confirmed in 12 data sets from 2 experiments. The ability of the instance theory to predict the power law is contrasted with the ability of other theories to account for it.

In studies of skill acquisition and automatization, the learning curve has a characteristic form: The time taken to perform a task decreases as a power function of practice (for a review, see Newell & Rosenbloom, 1981). The power function speed-up is so ubiquitous that it has come to be known as the *power law*. Current theories of skill acquisition and automaticity treat the power law as a benchmark prediction that they must make in order to be taken seriously (e.g., J. R. Anderson, 1982; J. R. Anderson & Milson, 1989; Cohen, Dunbar, & McClelland, 1990; Crossman, 1959; Logan, 1988; MacKay, 1982; Newell & Rosenbloom, 1981; Schneider, 1985). Most theories can account for the power law, in that they can be implemented in such a way as to produce power-function learning. Often, there is little that is fundamental to the theories in the implementation. They could be implemented in other ways to produce learning curves that follow some other function. Few theories actually predict the power law, in the sense that it follows as a necessary consequence of their fundamental assumptions. Similarly, the shape of the learning curve (the exponent of the power function; see the following section) is a free parameter in most theories. There is nothing in the theories to constrain the shape (to constrain the value of the exponent).

In this article, I argue that the instance theory of automaticity (Logan, 1988, 1990) predicts power function learning and that it predicts the shape of the learning curve from the shape of the underlying distribution of memory retrieval times. These predictions are derived from the fundamental

assumptions of the instance theory and new results on the power law are reported: The power law applies to distributions of reaction times, not just the means, and the shape of the learning curve is closely related to the shape of the reaction-time distribution. These predictions were developed and tested in 12 data sets from two experiments.

Power Law

According to the power law,

$$RT = a + bN^{-c}, \quad (1)$$

where RT is reaction time; a is the asymptote, reflecting an irreducible limit on performance; b is the difference between initial and asymptotic performance; N is the amount of practice, measured in sessions or trials per item; and the exponent c is the learning rate. Essentially, a and b are scaling parameters, moving the function into the range of numbers that the data occupy. The shape of the function is determined entirely by the exponent c . Some example power functions are plotted in Figure 1, showing how the shape varies as the exponent varies from 0.25 to 1.0. In real data, the exponent varies over this range, though typically it is less than 1.0 (see Newell & Rosenbloom, 1981).

The power law is ubiquitous. It occurs in virtually every speeded task. Newell and Rosenbloom (1981) reviewed a large variety of experiments conducted over a 50-year span and found the power law fit practice data in all of them. Since 1981, the power law has fit data from an even broader range of tasks, including solving geometry problems (J. R. Anderson, 1982), repeating sentences (MacKay, 1982), typewriting (Gentner, 1983), retrieving facts from memory (Pirolli & Anderson, 1985), performing mental rotation (Kail, 1986), making social judgments (Smith, Branscome, & Bormann, 1988; Smith & Lerner, 1986), making lexical decisions (Logan, 1988, 1990), naming arbitrary shapes (MacLeod & Dunbar, 1988; fits reported in Cohen et al., 1990), learning procedural skills (Woltz, 1988), evaluating logic circuit diagrams (Carlson, Sullivan, & Schneider, 1989), performing Sternberg

This research was supported by Grant BNS 88-11026 from the National Science Foundation. I would like to thank Brian Compton for help with the data analysis and Brian Ross and Ehtibar Dzharfarov for help with the mathematics. I am also grateful to Richard Schmidt, Hal Pashler, and an anonymous reviewer for valuable comments on the article.

Correspondence concerning this article should be addressed to Gordon D. Logan, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820. Electronic mail may be sent to glogan@s.psych.uiuc.edu.

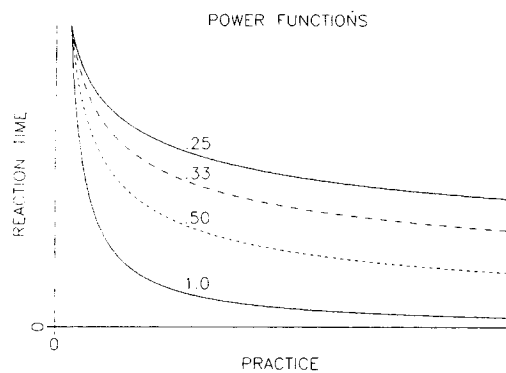


Figure 1. Examples of power functions. (Each function begins at the same point and asymptotes at zero. The functions differ only in their exponents. The value of the exponent appears above each function.)

memory search (Strayer & Kramer, 1990), searching displays for instances of rules (Kramer, Strayer, & Buckley, 1990), making pronunciation decisions (Logan, 1990), and verifying alphabet-arithmetic equations (Logan & Klapp, 1991; fits reported in Logan, 1988).

Nearly all of the power law fits have addressed mean (or median) reaction time. The effects of practice on the distribution of reaction times and parameters other than the mean are largely ignored. This is unfortunate because changes in the distribution can be as important as changes in the mean, both practically and theoretically (e.g., Compton & Logan, 1991). The theory and data presented in this article suggest that the power law applies to the entire reaction-time distribution, not just the mean.

Most of the theories that account for the power law, including the instance theory, fail to make the time-honored distinction between learning and performance. Overt reaction time is a measure of performance and reflects factors—such as motivation, stress, and distraction—other than the learning that is of primary theoretical interest. It is not obvious how to separate factors that affect learning from those that affect only performance, nor is it obvious how performance factors would distort the underlying learning curves. Dramatic changes in motivation, stress, and distraction might disrupt the learning curve, but most learning researchers try to hold these factors constant (see, e.g., the studies reviewed previously; also see Newell & Rosenbloom, 1981). One can hope that performance factors affect primarily the scaling parameters of the power function (i.e., a and b) and have little effect on its shape (i.e., c). One can also hope that variability contributed by performance factors is small in relation to that contributed by the processes of primary interest.

Instance Theory of Automaticity

The instance theory was described in detail in other articles (Logan, 1988, 1990). The theory assumes that performance is automatic when it is based on single-step, direct-access retrieval of solutions from memory and that automatization reflects a transition from performance based on some general algorithm for performing the task to performance based on

memory retrieval. When subjects have no experience on a task, they solve the problems it poses by applying a general algorithm (such as counting in addition tasks). The solutions produced by the algorithm are encoded into memory and retrieved when the problems are encountered again. After sufficient practice, performance will become automatic in that all problems can be solved by memory retrieval.

The instance theory makes three fundamental assumptions:

1. *Obligatory encoding* assumes that attention to an object or event is sufficient to commit it to memory. It may not be encoded well, depending on conditions of attention, but it will be encoded nevertheless.

2. *Obligatory retrieval* assumes that attention to an object or event causes all available information associated with it to be retrieved from memory. Retrieval may or may not be effective, depending on conditions of attention and other factors, but the retrieval process goes on nevertheless.

3. *Instance representation* assumes that each encounter with an object or event is encoded, stored, and retrieved separately as a unique *instance*. This assumption allies the theory with instance or exemplar theories of memory (Hintzman, 1988; Jacoby & Brooks, 1984), categorization (Hintzman, 1986; Medin & Schaffer, 1978), judgment (Kahneman & Miller, 1986), and problem solving (Ross, 1984) and contrasts it with *strength* or *prototype* theories.

These three assumptions imply a learning mechanism: When people perform the same task repeatedly, obligatory encoding causes instance representations of the same act to be stored in memory. The more repetitions, the more instances are stored. Obligatory retrieval causes information to become available when familiar situations are encountered once again. The more instances there are in memory, the more will be retrieved (i.e., the response from memory will be stronger). The assumption of instance representation allows one to model the retrieval process as a race in which the fastest trace determines performance (i.e., performance is based on the first instance to be retrieved from memory).

Race Model

The race model requires three additional assumptions: First, the time to retrieve solutions from memory is a random variable. This is a plausible assumption because few would believe the alternative assumption, that retrieval time is constant. Second, performance is determined by the first trace to be retrieved. This assumption makes the model a race model. Retrieval times vary randomly, and the instance with the fastest retrieval time determines performance. Intuition may suggest that the retrieval time of the fastest instance may decrease as more runners are added to the race and that the race model may produce something approximating the power law. (The more instances there are in the race, the greater the chances of randomly sampling an extremely fast value; this produces the speed-up. The more extreme the value, the less likely it is to sample one that is more extreme; this produces the negative acceleration characteristic of power functions.)

The third assumption plays a crucial role in allowing one to go beyond intuition and prove mathematically that the race model predicts a power function. The theory assumes that all instances have the same distribution of retrieval times

and are stochastically independent of each other. The assumption that the retrieval-time distribution is the same for all instances may strike many readers as implausible, but it may well be approximately correct. The power of this assumption is that the race model reduces to the problem of finding the minimum of N samples from the same distribution and learning how the minimum behaves as N increases. This is a well-studied problem in the statistics of extreme values (see, e.g., Gumbel, 1958), and it is possible to prove mathematically that the entire distribution of minimum values (of *minima*) decreases as a power function of N .

If the assumption is violated—if the distributions are not identical—then the proofs do not apply. However, I investigated the consequences of violating the assumption using simulation and mathematical analysis and found that it is not very important. The distributions of retrieval times can have different parameter values (i.e., they are all Weibull distributions, as discussed later, with different values of a , b , and c); as long as they have the same form, the distribution of minima appears to decrease as a power function of N . This analysis is presented in Appendix A.

Before proceeding to the power-function proof, it is important to note that the instance theory assumes there are two races going on. One, described above, is between the various instances in memory. The other is between memory and the algorithm that supports initial performance on the task. In essence, the algorithm races with the fastest instance retrieved from memory, and the winner of this race determines performance. An important question is whether the race between the algorithm and memory retrieval distorts the power function predictions. Formally, it must. Although it may be plausible to assume that all memory instances have the same distribution of retrieval times, it stretches credulity to assume that the finishing-time distribution of an arbitrary algorithm will be the same as the retrieval-time distribution. Consequently, the power-function proofs cannot apply, and there is no guarantee that the full race model will produce power-function learning. However, three considerations mitigate this difficulty: First, the algorithm will drop out as practice progresses. Eventually, there will be so many instances in the race that the algorithm will have no chance of winning. At that point, the power-function proofs will apply. Second, Logan (1988) used simulation to investigate the consequences of having different distributions for the algorithm and memory retrieval and found that power functions fitted the simulated data very well (also see Strayer & Kramer, 1990). The algorithm poses problems in principle that may not be very important in practice. Third, the analysis presented in Appendix A suggests that the power function predictions are not compromised much if the distributions come from the same family but have different parameter values. The distribution of finishing times from the algorithm may have a similar form to the distribution for memory retrieval, in which case the power-function proof should apply to the whole data set.

Weibull Distribution

The statistics of extreme values suggest that reaction times from a race model should follow the Weibull distribution if there are sufficient runners in the race. The Weibull distri-

bution is important because it is the third asymptotic distribution of extreme values. Many readers will be familiar with the concept of asymptotic distributions through their knowledge of the normal distribution and the central limit theorem. According to the central limit theorem, distributions of sums or averages will conform to the normal distribution as sample size increases. The normal distribution is *asymptotic* for sums and averages in that sums and averages taken from any parent distribution (with finite variance) will be distributed normally when sample size is sufficiently large.

There are three distributions that are asymptotic in this sense for extreme values (minima and maxima). Which distribution applies in a particular case depends on very general properties of the parent distribution, namely, whether high and low values are bounded or infinite (see Gnedenko, 1943; Gumbel, 1958; Leadbetter, Lindgren, & Rootzen, 1983). For distributions that are bounded at zero at the low end and extend toward positive infinity at the right, the asymptotic distribution of minima is the Weibull. Distributions of reaction times must belong to this family if they belong to any at all.¹ Reaction times cannot be smaller than zero, and in principle they can be infinitely large.

The Weibull distribution is related to the exponential distribution, which is used commonly in mathematical psychology. The distribution function for the exponential is

$$F(x) = 1 - \exp(-x).$$

The Weibull distribution is an exponential distribution in which the independent variable is raised to some power (for details, see Johnson & Kotz, 1970, chap. 20). Its distribution function is

$$F(x) = 1 - \exp(-x^c). \tag{2}$$

The generalized Weibull has three parameters: the exponent c and two scaling parameters, a and b . The generalized distribution function is

$$F(x) = 1 - \exp[-((x-b)/a)^c]. \tag{3}$$

¹ Gnedenko (1943) showed that the necessary and sufficient conditions for minima sampled from an initial distribution, $F(x)$, to converge asymptotically on the Weibull distribution were (also see Leadbetter et al., 1983; Luce, 1986) $x_F > -\infty$ and

$$\lim_{h \downarrow 0} \frac{F(hx - x_F)}{F(h - x_F)} = x^c,$$

where $c > 0$, $x > 0$, and $x_F = \inf\{x; F(x) > 0\}$. The variable c turns out to be the exponent of the asymptotic Weibull distribution (see Equations 2, 3, and 6). The variable x_F is the smallest possible value of x (i.e., the smallest value for which $F(x) > 0$). In reaction-time distributions, x_F must be greater than or equal to zero, because reaction times cannot be negative (anticipatory errors and prescience notwithstanding). Exponential and gamma distributions, which are commonly used to describe empirical reaction-time distributions, fulfill these conditions, as does the Weibull distribution. Technically speaking, the ex-Gaussian distribution will not fulfill these conditions because the Gaussian component of the distribution is not bounded at zero but extends to negative infinity (i.e., $x_F = -\infty$). This technicality also limits the application of the ex-Gaussian to reaction-time data. Most applications assume the ex-Gaussian is truncated at zero, however.

The scaling parameters, a and b , serve primarily to bring the distribution into the range of numbers that occur in the data set. The shape of the Weibull is determined by the exponent, c . The effect of the exponent on the shape of the distribution is illustrated in Figure 2, which displays Weibull density functions with exponents that vary from 1.0 to 4.0. When the exponent is 1.0, the Weibull becomes the exponential. When the exponent is 3.6, the Weibull is approximately normal. With exponents between these values, the Weibull is shaped like typical reaction-time distributions, truncated at the low end with a long upper tail. Its shape closely resembles the shape of distributions produced by the convolution of a normal and an exponential distribution, which provides an excellent quantitative description of reaction-time distributions (Ratcliff & Murdock, 1976). The Weibull is compared with the convolution of normal and exponential distributions in Appendix B.

The fact that Weibull distributions with appropriate exponents are shaped like reaction-time distributions is important in two respects. First, the instance theory predicts that reaction times from well-practiced subjects will conform to the Weibull. This follows from the application of the statistics of extreme values to the race model: The Weibull is the asymptotic distribution of minima sampled from the same distribution, and hence, it is the asymptotic distribution predicted by the race model. Second, with appropriate exponents, the Weibull may be used to approximate reaction times at all stages of practice. This makes the development of mathematical proofs easier and motivates fitting Weibull distributions to empirical data.

Power Function Proof

According to the statistics of extreme values, the distribution function, $F_1(x)$, for minima drawn from n independent samples from any distribution, $F(x)$, is

$$F_1(x) = 1 - [1 - F(x)]^n. \quad (4)$$

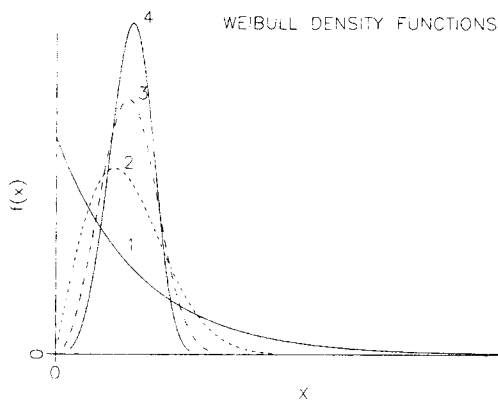


Figure 2. Examples of Weibull probability density functions. (Each function has the same scaling parameters. The functions differ only in their exponents. The value of the exponent appears above each function.)

The distribution function for minima drawn from n samples from a Weibull distribution can be obtained by substituting Equation 2 into Equation 4:

$$\begin{aligned} F_1(x) &= 1 - \{\exp[-x^c]\}^n \\ &= 1 - \exp[-n(x^c)] \\ &= 1 - \exp[-(n^{1/c}x)^c] \\ &= F(n^{1/c}x). \end{aligned} \quad (5)$$

A more general result can be obtained by substituting Equation 3, which represents the generalized Weibull distribution, into Equation 4 to yield

$$F_1(x) = 1 - \exp[-(n^{1/c}(x-b)/a)^c]. \quad (6)$$

Equations 5 and 6 demonstrate that the distribution of minima sampled from a Weibull remains a Weibull with the scale reduced by a factor of $n^{-1/c}$. This is an important result. The fact that the distribution of minima sampled from a Weibull remains a Weibull means that the Weibull is *stable* with respect to the minimum (see Gumbel, 1958). This means that once minima of samples drawn from any parent distribution become Weibull, they will remain Weibull thereafter. After that point, the Weibull will provide an accurate description of practice data. If one assumes that the distribution of retrieval times and the distribution of algorithm finishing times are both distributed as Weibull (i.e., the Weibull distribution applies throughout practice), Equation 6 should provide a good approximation to practice data from the beginning to the end of the experiment. The fits reported later rest on this assumption.

The fact that the distribution of minima remains a Weibull with its scale reduced by a power function is also very important. It implies that

$$\mu_1 = n^{-1/c}\mu$$

and

$$\sigma_1 = n^{-1/c}\sigma.$$

The mean and the standard deviation of the distribution of minima both decrease as a power function of practice with the same exponent, $-1/c$. This prediction was tested and confirmed by Logan (1988) and Kramer et al. (1990).

The power function reduction in mean and standard deviation is beginning to be taken seriously theoretically: First, Cohen et al. (1990) were able to account for this result with a connectionist model by choosing parameters appropriately, but it did not follow as a necessary consequence of their assumptions. They noted that some parameter values produced different exponents for means and standard deviations. Second, J. R. Anderson (1992) noted that his ACT* theory predicts a similar result because it assumes an exponential distribution of reaction times and the mean of the exponential distribution equals the standard deviation. So the standard deviation will decrease as a power function of practice just as the mean does. However, Anderson's theory necessarily makes two predictions that are not borne out by data: It predicts that the mean will equal the standard deviation, and it predicts that the reduction in the mean will equal the reduction in the standard deviation. In virtually every data

set the mean is larger than the standard deviation (typically 3 to 5 times as large), and the reduction in the mean is much larger than the reduction in the standard deviation (see, e.g., Kramer et al., 1990; Logan, 1988).² This article goes beyond means and standard deviations to predict a power function reduction in the entire distribution of reaction times.

The fact that the distribution of minima remains the same with its scale reduced is important because it means that the instance theory is able to predict the power function without adding new parameters to the distribution. The generalized distribution of minima in Equation 6 has only three parameters, just like the generalized Weibull distribution in Equation 3. The power function enters the equation as an independent variable (n) raised to an exponent ($1/c$) that is the reciprocal of the exponent of the Weibull distribution (c). There is no need to add free parameters. In this article, I take seriously the fact that the exponent of the power function is the reciprocal of the exponent of the Weibull distribution. This is the central result in the article: *The shape of the power function is determined by the same parameter that determines the shape of the reaction-time distribution.* Thus, the instance theory predicts power function learning, and it predicts the shape of the power function.

These predictions can be illustrated by examining Figures 1 and 2. Minima sampled from a Weibull distribution with exponent c in Figure 2 will speed up as a power function of practice with exponent $1/c$ in Figure 1. The constraint is easy to see: For example, the minimum of samples drawn from the Weibull with the exponent of 1 in Figure 2 will approach asymptote after relatively few samples, producing a sharply inflected learning curve described by the power function with the exponent of 1 in Figure 1. By contrast, the minimum of samples drawn from the Weibull with the exponent of 4 in Figure 2 will approach asymptote more slowly, producing a more gradual learning curve described by the power function with the exponent of 0.25 in Figure 1. The next three sections describe tests of these assumptions in real data.

Testing the Theory

The instance theory predicts that (1) the entire distribution of reaction times should decrease as a power function of the number of training trials, (2) reaction times at all stages of practice will be distributed as a Weibull whose scale reduces as a power function of practice, and (3) the exponent for the power function reduction will be the reciprocal of the exponent for the Weibull distribution. These predictions address learning rather than performance. They assume that memory retrieval is the only source of variability in the data. They assume that residual processes such as sensory registration and motor execution, which are usually treated as intercept parameters, contribute no variability to the data. Strictly speaking, the predictions could hold if such processes took constant amounts of time and so added no variability. However, the predictions should still be valid if the variability produced by intercept processes is small in relation to the variability produced by memory retrieval.³

The first prediction was tested in two ways. First, separate power functions (i.e., Equation 1) were fitted to quantiles of

reaction-time distributions, and their exponents were compared. According to the instance theory, the exponent should be the same for each quantile. Second, constrained power functions were fitted to all of the quantiles simultaneously. Each quantile was allowed to have a separate asymptote (a) and multiplicative parameter (b), but each quantile was required to share a common exponent (c). The quality of this fit was compared with the quality of the separate fits. The instance theory predicts no difference.

The second prediction was also tested in two ways. First, separate Weibull distributions (i.e., Equation 3) were fitted to the empirical distributions at each stage of practice. The instance theory predicts that the scaling parameters of the Weibull distributions (a and b) should decrease as power functions of practice and that the exponent (the shape parameter c) should remain constant. Power functions (Equation 1) were fitted to the parameters to test the former prediction, and the Weibull exponents were compared to test the latter. Second, constrained Weibull distributions were fitted to all the practice data simultaneously. Each stage of practice was allowed to have separate scaling parameters, but every stage was required to share a common exponent. The quality of this fit was compared with the quality of the separate fits. Instance theory predicts no difference.

The third prediction was tested in two ways as well. First, the reciprocals of power-function exponents generated in tests of the first prediction were compared with the Weibull distribution exponents generated in tests of the second prediction. The instance theory predicts close agreement. Second, the entire data set was fitted by a single Weibull distribution whose scale decreased as a power function of practice (i.e.,

² J. R. Anderson (1992) suggested that his model may be generalized to produce gamma distributions of reaction times. A gamma distribution is the sum of n identical exponential distributions. It has two parameters: n , the number of exponential distributions contributing to the sum, and λ , the rate parameter for the exponential distributions. Its mean is n/λ , and its standard deviation is \sqrt{n}/λ (for further details, see Johnson & Kotz, 1970, chap. 17). The coefficient of variation—the ratio of the standard deviation to the mean—is \sqrt{n}/n . It depends only on the number of exponentials contributing to the sum and not at all on the rate parameter. This is an advantage for Anderson, because his strengthening mechanism influences λ and not n . Consequently, strengthening will affect the mean and the standard deviation proportionally. However, the real data again present a problem. The coefficient of variation in real data is typically 0.2 to 0.4. In the counting data analyzed later, for example, the mean coefficient of variation was 0.342. In order for a gamma distribution to produce that coefficient of variation, n would have to equal 8 or 9. If n is interpreted as the number of steps or stages underlying the reaction process, this is not an acceptable value in Anderson's theory. In his theory, production composition reduces the number of steps or stages to 1. In order to adapt his theory to the constraints of the data, Anderson would have to adopt a different interpretation for n .

³ I performed simulations to evaluate the effects of variability of the intercept processes on the instance theory predictions. The results of those simulations are described briefly in the *Discussion* sections of the alphabet-arithmetic data and the counting data. In general, the predictions were fairly robust with respect to variability in the intercept processes.

Equation 6). The exponent for the power function was constrained to be the reciprocal of the exponent for the Weibull distribution. The instance theory predicts that this constrained fit should be as good as the less-constrained fits generated in the tests of the first and second prediction.

The different tests involve vastly different numbers of parameters. At one extreme, the unconstrained power function fits require three parameters for each quantile, and the unconstrained Weibull fits require three parameters for each stage of practice. The more quantiles analyzed and the more stages of practice that are distinguished, the greater the number of parameters. At the other extreme, the constrained Weibull power-function fits require only three parameters no matter how many quantiles or stages of practice there are. These fits, predicted by the instance theory, offer a remarkably parsimonious description of the data.

The predictions were tested in data from two experiments, one on alphabet arithmetic and one on dot counting. Each experiment provided six separate data sets.

Alphabet Arithmetic

Alphabet-Arithmetic Task

I tested the predictions first using distributions of reaction times from an alphabet-arithmetic task. The means, standard deviations, accuracy scores, and ancillary data were reported by Compton and Logan (1991). The focus here is on the distributions.

In the alphabet-arithmetic task, subjects were asked to verify equations of the form $A + 2 = C$ and $B + 3 = F$. In essence they were asked whether C is 2 letters down the alphabet from A (it is) and whether F is 3 letters down the alphabet from B (it is not). Subjects performed this task initially by counting through the alphabet beginning with the first letter for a number of steps determined by the digit addend and then comparing the computed answer with the presented one. Their reaction times increased linearly with the magnitude of the digit addend (which determines the number of counting steps) with a slope of 400 ms to 500 ms per count. With practice, however, subjects came to remember which equations were true and which were false, and they relied on retrieving solutions from memory to perform the task rather than counting through the alphabet (Compton & Logan, 1991; Logan & Klapp, 1991). Thus, practice at alphabet arithmetic produced the transition from (counting) algorithm to memory retrieval that the instance theory identifies as automatization.

In Compton and Logan's (1991) experiment, 36 subjects served in one 432-trial session. They saw 12 stimuli altogether (2 Examples \times 3 Digit Addends \times True vs. False). Each stimulus was presented 36 times, 6 times in each of 6 blocks. Previous research suggested that 36 presentations would be enough to produce automaticity with 12 stimuli (see Logan & Klapp, 1991, Experiment 3). The mean reaction times, presented in Figure 3 as a function of digit addend and practice block, confirm this suggestion. In the first block, reaction time increased linearly with digit addend with a slope

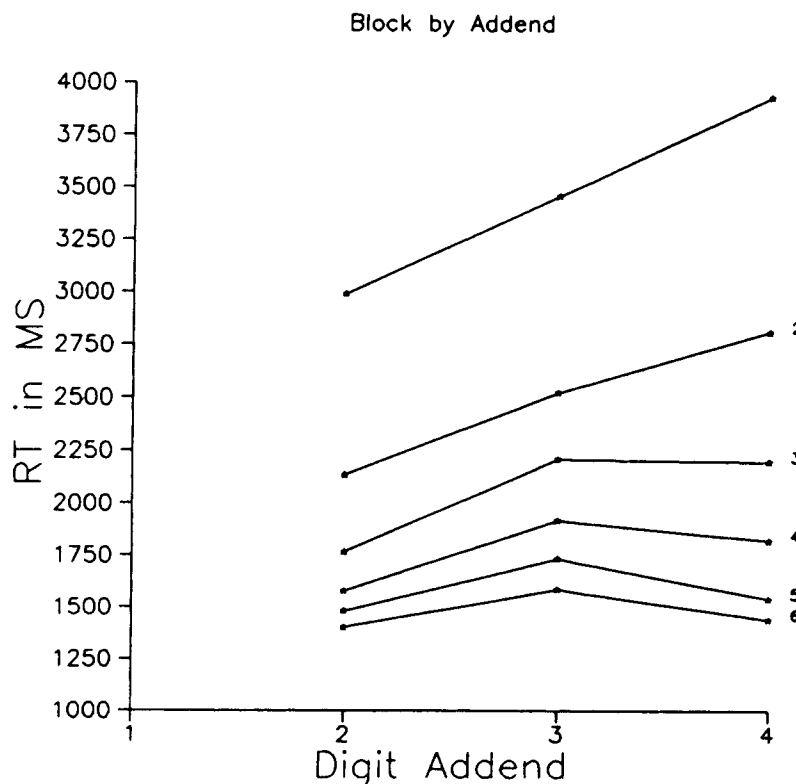


Figure 3. Mean reaction times for each practice block as a function of the magnitude of the digit addend in the alphabet-arithmetic task. (The data are from Compton & Logan, 1991.)

of 469 ms per count. By the sixth block, reaction time decreased considerably, and the slope was only 17 ms per count (for further details, see Compton & Logan, 1991).

Reaction-time distributions were constructed for each combination of digit addend and true versus false response for each practice block. This resulted in six distributions at each of six levels of practice, providing six data sets in which to test the instance theory's predictions. Each subject contributed 12 trials to each distribution at each level of practice. The distributions were averaged over subjects by calculating quintiles (i.e., the value of the 10th, 30th, 50th, 70th, and 90th percentiles) for each subject and averaging the quintiles over subjects (see Ratcliff, 1979; Thomas & Ross, 1980).

Power function fits. The first prediction was that the entire distribution of reaction times should decrease as a power function of practice. This implies that the different quantiles of the distribution should all be well fit by power functions and that the exponent, *c*, should be the same for each quantile. Power functions (i.e., Equation 1) were fitted to each quintile of the practice data in each combination of digit addend and true versus false equation using STEPIT (Chandler, 1965). Measures of goodness of fit are presented in Table 1. The predicted and observed values are plotted in Figure 4.

The data were well fit by power functions, confirming the first part of the prediction. This is to be expected. The 30 points in each panel were fitted by 15 parameters, 2 scaling parameters, and 1 exponent for each of the five quintiles. This is a large number of parameters for a small number of data points, so the fits should be good. The question is whether the power functions for the different quintiles had the same exponent. The values averaged over digit addend and true versus false were $-.576$, $-.640$, $-.641$, $-.660$, and $-.883$ for Quintiles 1 to 5, respectively. The three middle quintiles were roughly equal, but the lowest and highest quintiles were more discrepant.

Table 1
Measures of Goodness of Fit and Exponent Parameter (*c*) for Fits of Equation 1 to Quintiles of Reaction-Time Distributions from Alphabet-Arithmetic Data of Separate and Constrained Power Functions

Measure	True			False			<i>M</i>
	Digit addend			Digit addend			
	2	3	4	2	3	4	
Separate power functions							
<i>rmsd</i>	19	48	30	30	25	50	34
<i>r</i> ²	.999	.998	.999	.999	.999	.998	.999
<i>c</i>	.881	.613	.607	.731	.593	.647	.679
Constrained power functions							
<i>rmsd</i>	54	103	151	71	68	168	95
<i>r</i> ²	.996	.991	.993	.994	.996	.983	.992
<i>c</i>	.597	.459	.471	.549	.415	.466	.493

Note. Separate power functions have separate exponents and scaling parameters; constrained power functions have the same exponent but are allowed different scaling parameters; *rmsd* = root-mean-square deviation between predicted and observed values.

Are these differences significant? STEPIT offers no measure of the standard error of the values of the fitted parameters, so the significance of these differences could not be tested directly. Instead, a one-way analysis of variance (ANOVA) was performed on the exponents with quintile as the independent variable and the six combinations of digit addend and true versus false as subjects. The main effect of quintile approached significance, $F(4, 25) = 2.70$, $p < .06$, $MS_e = .0307$, which suggests that the differences between quintiles may be reliable, contrary to the predictions of the instance theory.

Another way to test the significance of the differences between the exponents is to see whether the fit is worse when the exponent for each quintile is constrained to be the same. Consequently, the data were fitted again. This time the power functions for each quintile were allowed to have separate scaling parameters (i.e., *a* and *b*), but each quintile was constrained to have the same exponent. The fits of these constrained functions are plotted in Figure 5. Measures of goodness of fit and the exponents of the fitted functions are presented in Table 1.

The data were well fit by constrained power functions. Inspection of the fits reveals no systematic deviations. The fits are not as good as those produced by unconstrained power functions. The squared product-moment correlations between predicted and observed values were slightly lower, and the root-mean-square deviations between predicted and observed values increased by a factor of nearly 3, but the fits were still quite good. Moreover, these fits required fewer parameters than the unconstrained fits. The constrained fits in each panel of Figure 5 required 11 parameters, 2 scaling parameters for each function, and 1 common exponent, whereas the unconstrained fits in Figure 4 required 15. The constrained fits provide a more economical description of the data.

The important point to be taken from Figure 5 is that in each panel, the five functions all have the same shape. They differ only in scale. All five functions share the same exponent, *c*; they differ only in the scaling parameters, *a* and *b*.

Weibull distribution fits. The second prediction was that changes in reaction time with practice could be described by a single Weibull distribution whose scale decreased as a power function of practice. Separate Weibull distributions (Equation 3) were fitted to the quintiles from each practice block in each combination of digit addend and true versus false equation using STEPIT. The fits were unconstrained in that each distribution was allowed to have a separate exponent and separate scaling parameters in each practice block. Measures of goodness of fit are presented in Table 2. Observed and predicted values are plotted in Figure 6.

The fits to the unconstrained Weibull distributions were good. The 30 points in each panel were fitted by 18 parameters, 1 exponent, and 2 scaling parameters for each of the six blocks of practice. With so many parameters and so few data points, good fits are to be expected. Again, the question is whether the Weibull distributions retained the same shape but contracted as a power function of practice. The shape is determined by the exponent; if the shape remained constant across practice, then the exponent should remain the same. The average exponents across digit addends and true versus false equations were 1.316, 1.392, 1.222, 1.159, 1.069, and

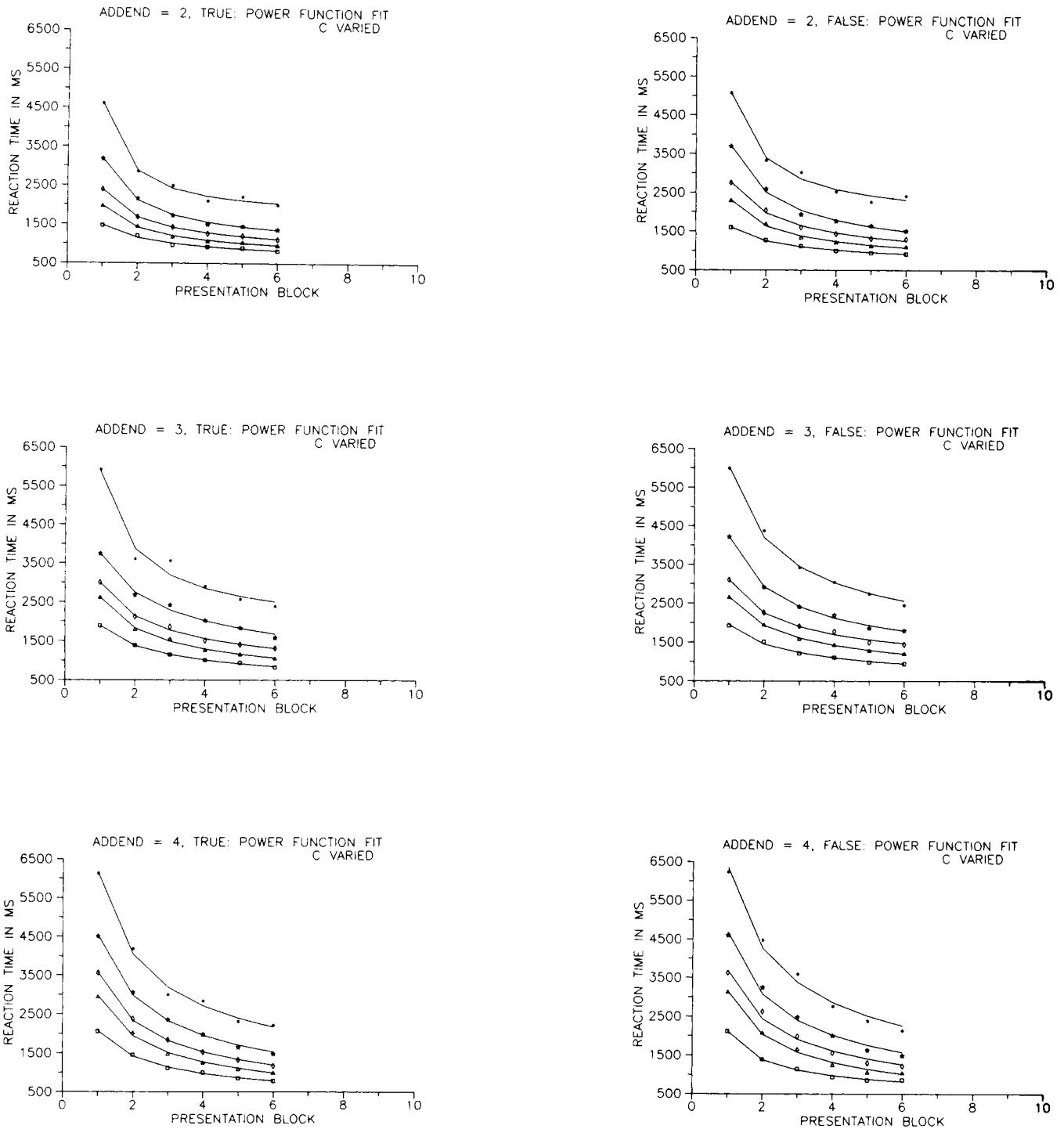


Figure 4. Power functions from Equation 1 (lines) fitted to quintiles of the distributions (dots) of alphabet-arithmetic reaction times as a function of practice for each combination of digit addend and true versus false equation. (The five power functions in each panel were allowed to have separate exponents and scaling parameters.)

1.009 for Practice Blocks 1 to 6, respectively. The significance of these differences was assessed with a one-way ANOVA with practice block as the independent variable and the six combinations of digit addend and true versus false as subjects.

The main effect of practice block approached significance, $F(5, 30) = 2.45, p < .06, MS_e = .0520$, which suggests that the shape changed over practice, contrary to the predictions of the instance theory.

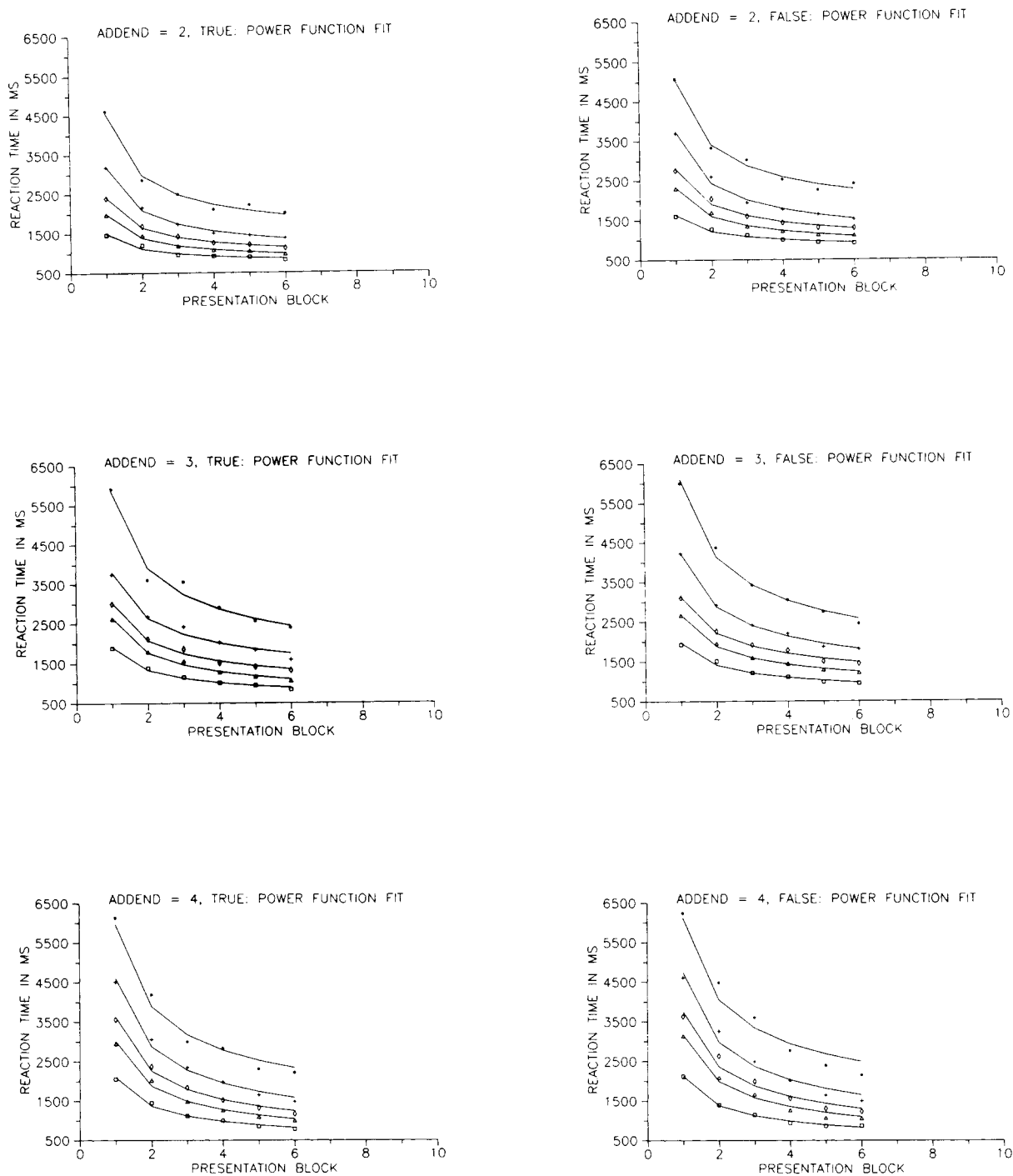


Figure 5. Power functions from Equation 1 (lines) fitted to quintiles of the distributions (dots) of alphabet-arithmetic reaction times as a function of practice for each combination of digit addend and true versus false equation. (The five power functions in each panel were allowed to have separate scaling parameters but were constrained to have the same exponent.)

Did the scaling parameters decrease as a power function of practice, as the instance theory predicts? The values averaged over digit addend and true versus false are plotted in the top panel of Figure 7. Both the a and b values decreased over practice. Both reductions were well fit by power functions

(Equation 1, fitted by STEPIT): For a and b , respectively, r^2 was .9931 and .9997 and $rmsd$ was 45 ms and 4 ms. These fits confirm the prediction of the instance theory.

Next, the data were fitted to the Weibull distributions (Equation 3) that were allowed to have separate scaling param-

Table 2
Measures of Goodness of Fit and Exponent Parameter (c) for Fits of Equation 3 to Quintiles of Reaction-Time Distributions from Alphabet-Arithmetic Data of Separate and Constrained Weibull Distributions

Measure	True			False			M
	Digit addend			Digit addend			
	2	3	4	2	3	4	
Separate							
<i>rmsd</i>	11	28	16	18	21	23	20
<i>r</i> ²	.999	.999	.999	.999	.999	.999	.999
<i>c</i>	1.035	1.113	1.346	1.147	1.178	1.350	1.195
Constrained							
<i>rmsd</i>	34	70	53	69	58	61	58
<i>r</i> ²	.998	.996	.998	.995	.997	.998	.997
<i>c</i>	1.104	1.141	1.536	1.266	1.281	1.515	1.307

Note. Separate Weibull distributions have different exponents and scaling parameters; constrained Weibull distributions have different scaling parameters but a common exponent; *rmsd* = root-mean-square deviation between predicted and observed values.

eters for each practice block but were constrained to have a common exponent. Observed and predicted values are plotted in Figure 8. Measures of goodness of fit are presented in Table 2.

The fit was almost as good as the unconstrained fit. The squared product-moment correlation between observed and predicted values was nearly identical, and the root-mean-square deviations between predicted and observed values increased by 38 ms, on average. Constraining the shape of the distributions to be the same across practice had little effect on the goodness of fit. Again, fewer parameters were required for the constrained fits (13 parameters including 2 scaling parameters for each practice block and 1 exponent) than for the unconstrained fits (18 parameters). Apparently, the variation in exponents in the unconstrained fits was not very important.

The important point to be taken from Figure 8 is that the relative spacing of the lines representing each practice block is constrained to be the same. For example, the ratio of the distance from the bottom to the top line to the distance from the bottom to the middle line is the same for each practice block. Thus, the shape of the fitted distribution (i.e., the parameter *c*) is constrained to be the same, differing only in scale (i.e., *a* and *b*).

The scaling parameters of the constrained Weibulls, averaged over digit addend and true versus false, are plotted in the bottom panel of Figure 7. Both *a* and *b* decreased as power functions of practice; *R*² was .995 and .997 for *a* and *b*, respectively; *rmsd* was 35 and 15 for *a* and *b*, respectively. The close fits of the power functions are consistent with the predictions of the instance theory.

Constrained power-function Weibull fits. The third prediction was that the exponent of the power function should be the reciprocal of the exponent of the Weibull distribution. This was tested in three ways. First, the reciprocals of the

exponents from the unconstrained power-function fits (*M* = 1.502) were compared with the exponents from the unconstrained Weibull fits (*M* = 1.195). The significance of the difference was assessed with a *t* test for paired observations that compared the reciprocal of the average power function exponent in each combination of digit addend and true versus false with the average value of the Weibull exponent for the corresponding combination. The result was significant, *t*(5) = 4.00, *p* < .05, *MS_e* = .0769.

Second, the reciprocals of the exponents from the constrained fits were compared (mean reciprocal of power function exponent = 2.058; mean Weibull exponent = 1.307). Here, the reciprocal of the constrained power function exponent in each combination of digit addend and true versus false was compared with the constrained Weibull exponent for the corresponding combination. Again, the result was significant, *t*(5) = 6.48, *p* < .01, *MS_e* = .1158.

Third, Equation 6 was fitted to the data from each combination of digit addend and true versus false. Equation 6 uses only three parameters to capture simultaneously the shape of the reaction-time distribution and the shape of the learning curve. Two of the three parameters are scaling parameters, bringing the predicted values into the range of the observed values. The third parameter determines the shape of the distribution, and its reciprocal determines the shape of the learning curve. Thus, the quality of the fit of Equation 6 bears on the prediction that the exponent of the power-function learning curve should be the reciprocal of the shape-determining exponent of the Weibull distribution.

The observed and predicted values are plotted in Figure 9. Measures of goodness of fit and parameters of Equation 6 are presented in Table 3. The fits were quite good. Root-mean-square deviation was much larger than in the previous fits, and the product-moment correlation was lower, although it was still quite high. These differences are remarkable considering that only 3 parameters were required for the fits in each panel of Figure 9, in comparison with the 11 to 18 parameters required for the previous fits. Three parameters are not many to predict 30 data points.

The important point to be taken from Figure 9 is that in each panel, the shape of the five functions is constrained in two dimensions. As with the power function fits, each of the functions has the same shape over blocks, differing only by a change in scale. As with the Weibull distribution fits, the spacing within blocks is constrained to be the same over blocks except for a change in scale. But unlike the previous fits, the same factor—the exponent *c*—constrains both the shape and the spacing of the functions.

Close inspection of Figure 9 reveals some systematic deviations from predictions, particularly for Addend = 4 data. In the bottom two panels, some of the empirical functions seem to decrease faster over blocks than the theoretical functions. For example, in the Addend = 4, False data, the observed values for the 70th percentile fall close to predicted values for the 70th percentile in the first practice block but end up close to predicted values for the 50th percentile in the sixth practice block.

I suspect these discrepancies are due to contamination from the counting algorithm used in the first few blocks. Compton

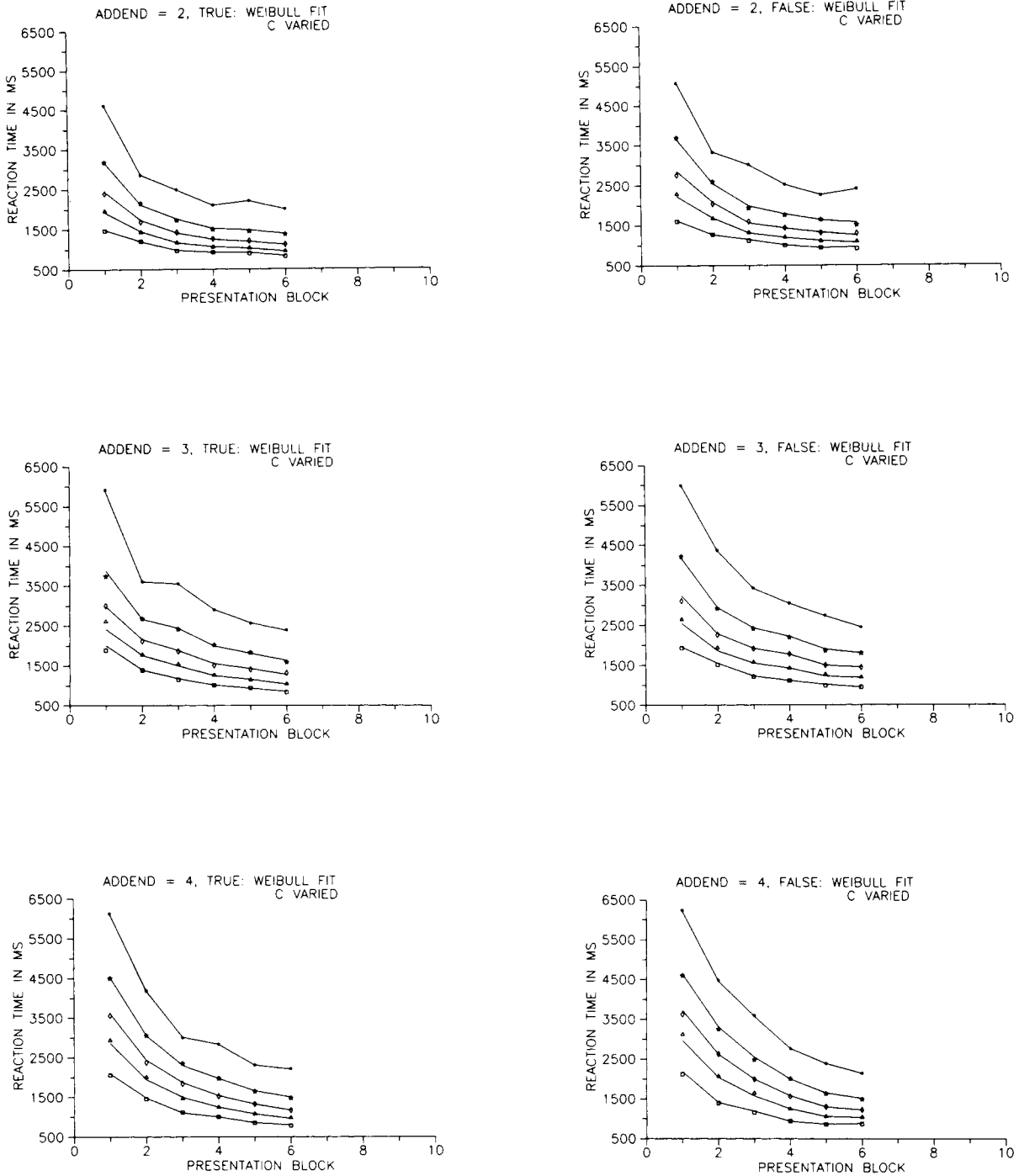


Figure 6. Weibull distributions from Equation 3 (lines) fitted to quintiles of the distributions (dots) of alphabet-arithmetic reaction times as a function of practice for each combination of digit addend and true versus false equation. (The five functions in each panel were allowed to have separate scaling parameters and separate exponents.)

and Logan (1991) probed subjects on one sixth of the trials and asked them to report how they performed the task. In the first block, subjects reported counting 46% of the time. This value dropped to 7% by the sixth block. Counting is slower

than memory retrieval, particularly when the digit addend is large, so the tendency to count early in practice may artificially inflate the early portion of the learning curve. Because there were so few data points, there was not much to be done about

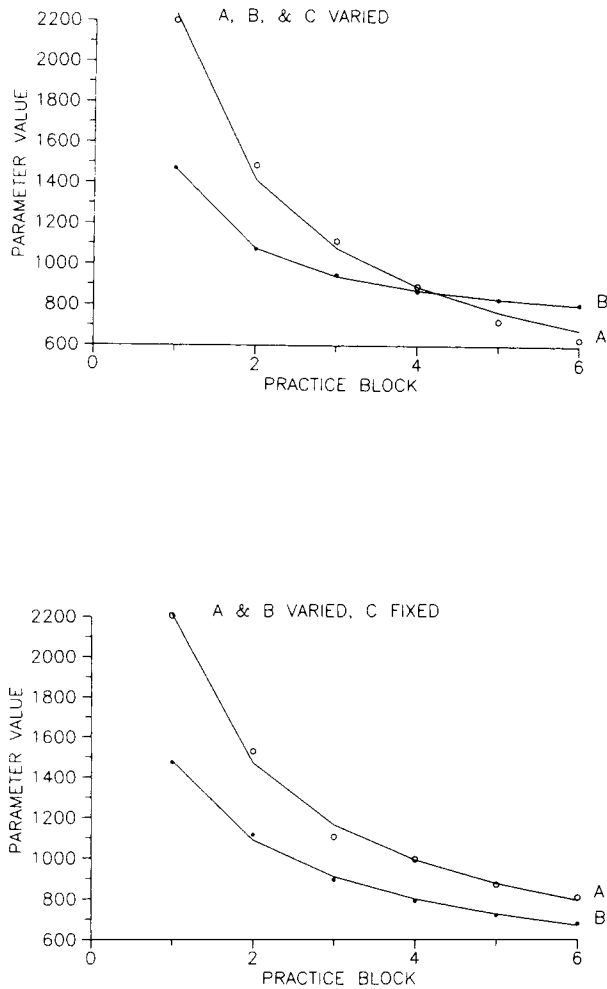


Figure 7. The a and b parameters from (top) unconstrained and (bottom) constrained Weibull distribution (Equation 3) fits to alpha-numeric data as a function of practice block. (Lines are predictions from fitted power functions; points are parameter values.)

this in this data set. I attempted to remove data that were dominated by the algorithm in the next data set.

Discussion

Each of the three predictions of the instance theory received some support. First, the distribution of reaction times decreased as a power function of practice. Each quintile was well fit by a power function (Equation 1), as the theory predicts. The exponents of power functions fitted to quintiles separately were not identical, but the fit was reasonably good when all quintiles were constrained to have the same exponent. Second, the distribution of reaction times in each practice block was well fit by a Weibull distribution (Equation 3). The theory predicts this for extended practice and assumes it for earlier stages of practice. The exponents of Weibull distributions fitted to each practice block separately were not identical, but the fit was good when all blocks were constrained to have the same exponent. Moreover, the scaling

parameters decreased as power functions of practice, as the theory predicts. This was true for both the constrained and unconstrained fits. Third, the reciprocals of the exponents of the power functions differed from the exponents of the Weibull distributions in both the constrained and unconstrained fits, contrary to the instance theory prediction. However, reasonable fits were obtained when all quintiles and all practice blocks were fitted simultaneously with the constraint that the reciprocal of the exponent of the power function equal the exponent of the Weibull (Equation 6). This is consistent with the instance theory prediction.

How is one to interpret the cases in which instance theory predictions were not confirmed? Do they falsify the theory, or do they fail to test it appropriately? I would like to believe the latter. The predictions that failed to be confirmed all involved comparisons of exponents over quintiles or practice blocks or both. The problem may be that the exponents were estimated from too few data points to be reliable. Power function exponents were estimated from six data points; Weibull exponents were estimated from five. It is easy to fit three-parameter functions to five or six data points, but the best-fitting parameters may be affected by random fluctuations in the data as much as the underlying processes.

Some evidence for this interpretation is available in the analyses presented so far: The predictions of the theory were confirmed more readily when the fits were constrained than when they were unconstrained. The constrained fits brought more data points to bear on the estimation of parameters than did the unconstrained fits (i.e., 30 instead of 5 or 6). Further evidence will be obtained in the next set of experimental data. Unconstrained power functions will be based on 12 data points rather than 6, and unconstrained Weibull distributions will be based on 10 data points rather than 5. If the paucity of data points was responsible for the failed predictions in the current data set, the predictions should be confirmed more readily in the next one.

The instance theory predictions may also fail because the theory fails to take into account "intercept" processes, such as those involved in perceptual registration and motor execution, that may affect the distribution of observed reaction times. In its present form, the theory assumes that intercept processes are negligible or constant in duration, and that is not likely to be true. Although a theory of perceptual and motor processes is beyond the scope of this article, I performed some simulations to see whether variability in intercept processes could account for the failures of the instance-theory predictions. The memory process was represented as a race between N traces (N varied from 1 to 12). N samples were taken from the same Weibull distribution, and the smallest one of them was chosen to represent the retrieval time. This retrieval time was then added to a sample from a single (different) Weibull distribution, which represented the duration of the intercept processes, to produce a simulated reaction time. This procedure was performed 10 times for each value of N . The simulated reaction times were rank ordered to estimate deciles of the reaction-time distribution. This process was replicated 1,000 times so that each decile was based on 1,000 observations. Then separate power functions (Equation 1) were fitted to each of the deciles separately, and separate

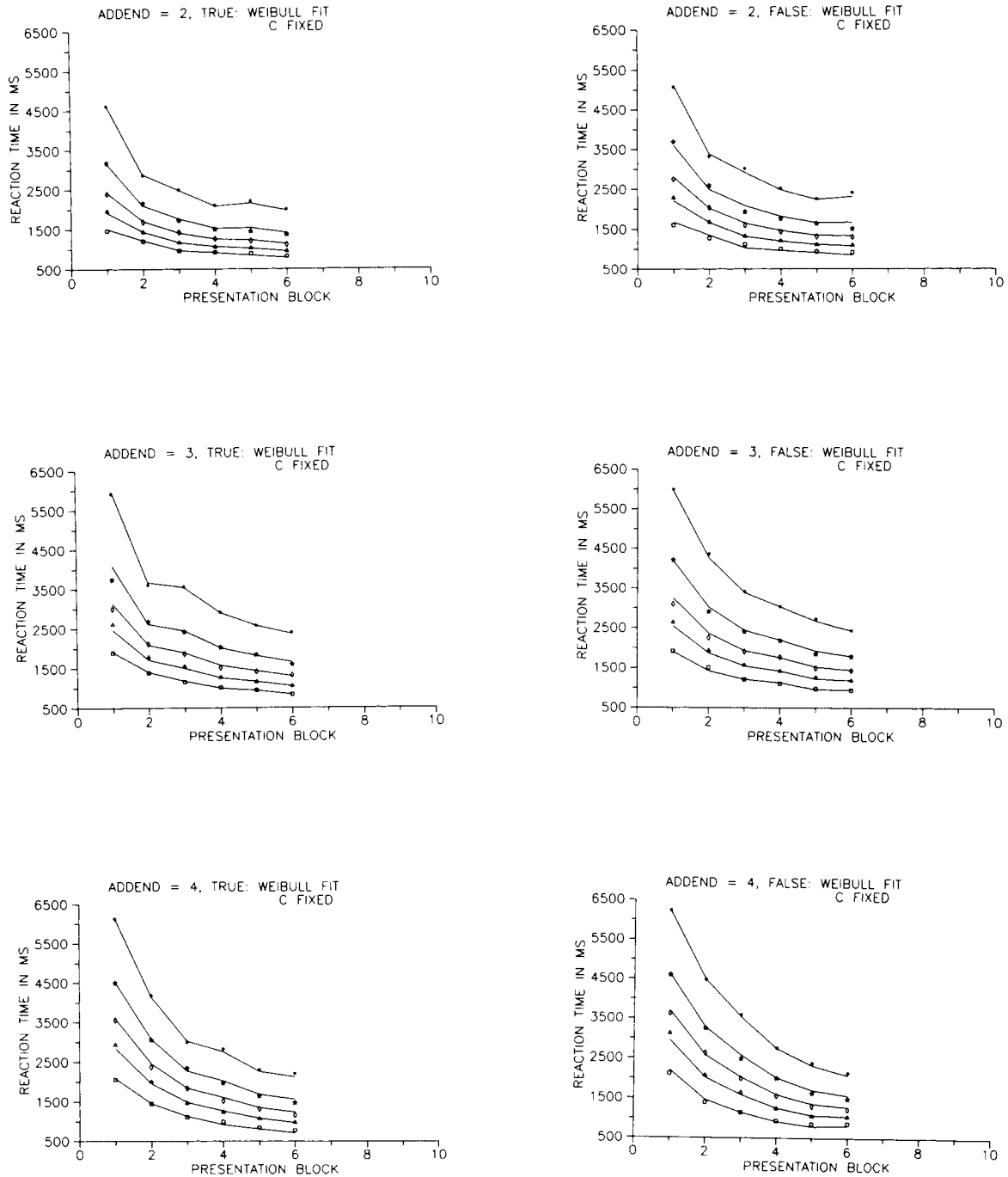


Figure 8. Weibull distributions from Equation 3 (lines) fitted to quintiles of the distributions (dots) of alphabet-arithmetic reaction times as a function of practice for each combination of digit addend and true versus false equation. (The five functions in each panel were allowed to have separate scaling parameters but were constrained to have the same exponent.)

Weibull distributions (Equation 3) were fitted to each of the 12 levels of practice.

I performed several simulations, varying the parameters of the memory distribution and the intercept distribution, to try to produce the failures of instance-theory predictions observed

with the alphabet-arithmetic data. The parameter space was not searched systematically, so it is hard to say how characteristic the results were, but some combinations of parameter values produced failures of prediction as observed. In particular, the Weibull exponents decreased with practice, the power

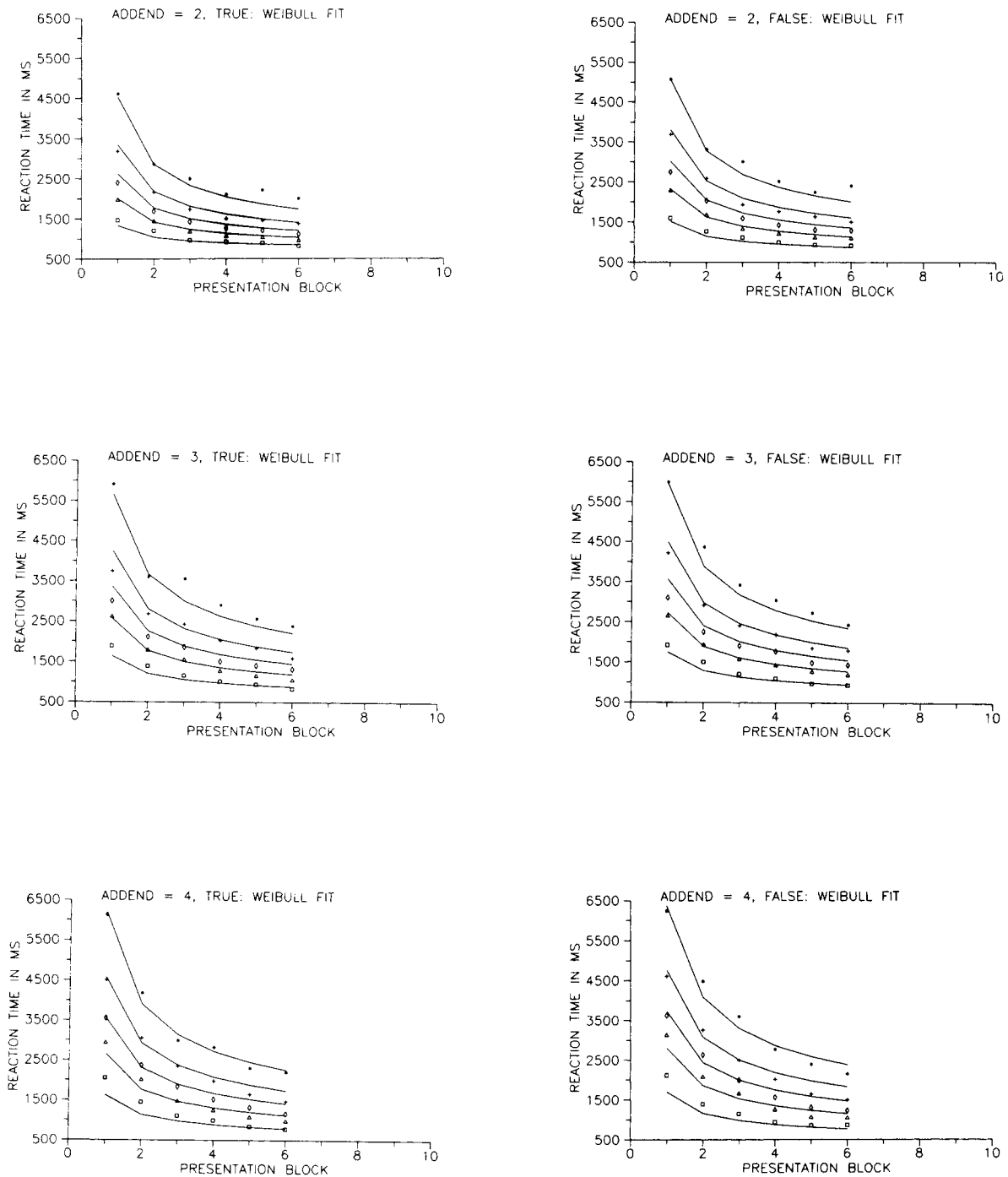


Figure 9. Weibull distributions from Equation 6 (lines) fitted to quintiles of the distributions (dots) of alphabet-arithmetic reaction times as a function of practice for each combination of digit addend and true versus false equation. (The five functions in each panel were fitted with the same scaling parameters and the same exponent.)

function exponents increased over decade, and the reciprocals of the power function exponents overestimated the Weibull exponents, as they did in the alphabet-arithmetic data, in the following cases: (1) memory Weibull with $a = b = 1,200$ and $c = 1.5$; intercept Weibull with $a = b = 600$ and $c = 1$; (2)

memory Weibull with $a = b = 1,200$ and $c = 2.5$; intercept Weibull with $a = b = 600$ and $c = 1$; and (3) memory Weibull with $a = b = 1,200$ and $c = 3$; and intercept Weibull with $a = b = 600$ and $c = 2$. The important factor seems to be that the intercept distribution is more sharply skewed (i.e., c is

smaller) than the memory distribution. Early in practice, the memory distribution dominates the intercept distribution in determining the shape of the reaction-time distribution. As practice progresses, the memory distribution shrinks (following a power function), and eventually, the intercept distribution dominates the memory distribution. The reaction-time distribution initially resembles the memory distribution and finally resembles the intercept distribution, which accounts for the reduction in the Weibull exponent. The power function exponent may increase with decile because the largest changes occur in the higher deciles. These conclusions are highly speculative. More analysis will be necessary to confirm them. However, the simulations do demonstrate that variability in intercept processes may account for the failure of the instance theory predictions.

Counting Dots

Counting Task

The second data set came from a dot-counting task reported in Lassaline and Logan (1991). Subjects saw 6 to 11 dots presented in random positions in a 7 × 7 grid and reported the numerosity by pressing keys. Four subjects were tested for 13 sessions each consisting of 480 trials. In all, there were 30 dot patterns, 5 at each numerosity level, and subjects saw each pattern 16 times per session for 12 sessions for a total of 192 exposures. In Session 13, subjects were transferred to new patterns to determine whether they could generalize what they had learned in Sessions 1 to 12. Lassaline and Logan focused their analyses on slopes of linear functions relating reaction time to numerosity. It is well established that reaction times increase linearly with numerosity, for levels of numerosity above the subitizing range, with slopes of 300 ms to 400 ms per item (e.g., Chi & Klahr, 1975; Jensen, Reese, & Reese, 1950; Mandler & Shebo, 1982).

The slopes from the 13 sessions of Lassaline and Logan's (1991) experiment are plotted in Figure 10. The slope for

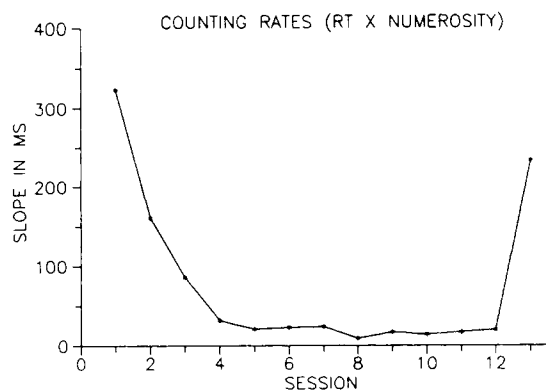


Figure 10. Slopes of linear functions relating reaction time (RT) to the number of dots presented in the dot-counting task, plotted as a function of training session. (The data are from Lassaline & Logan, 1991.)

Session 1 was 324 ms per dot, which is typical of counting studies. The slope dropped rapidly over sessions, asymptoting at 17 ms per dot over Sessions 7–12. On Session 13, when new patterns were introduced, the slope increased to 233 ms per item, suggesting the item-specific learning that was predicted by the instance theory.⁴

Reaction-time distributions were constructed for each numerosity level at each session, thus producing six data sets each reflecting changes in distributions over 12 practice sessions. Each subject contributed 80 trials to each distribution at each practice level. The distributions were averaged over subjects by calculating deciles (i.e., the value of the 5th, 15th, 25th, 35th, 45th, 55th, 65th, 75th, 85th, and 95th percentiles) for each subject and averaging the deciles over subjects (see Ratcliff, 1979; Thomas & Ross, 1980).

Power function fits. To test the first instance-theory prediction, the distributions were fitted by power functions (Equation 1). I allowed the power functions for each decile to have separate scaling parameters (*a* and *b*) and separate exponents (*c*). The fits are displayed in Figure 11. Measures of goodness of fit are presented in Table 4. As with alphabet arithmetic, the counting data appear to be well fit by power functions. The 120 data points in each panel of Figure 11 are fit by 30 parameters, 2 scaling parameters, and 1 exponent for each function.

The instance theory predicts that the power functions fitted separately to each decile should have the same exponent. The exponents averaged across numerosity were -1.082, -1.074, -.926, -.849, -.780, -.721, -.660, -.614, -.591, and -.533 for Deciles 1 to 10, respectively. The exponents decreased in magnitude as the decile increased, suggesting that the lower deciles changed at a faster rate than the higher deciles. The significance of the differences between deciles was tested by a one-way ANOVA with decile as the independent variable and the six numerosity levels as "subjects." The main effect of

Table 3
Measures of Goodness of Fit and Parameter Values for Fits of Equation 6 to Quintiles of Reaction-Time Distributions from Alphabet-Arithmetic Data for Weibull Distributions Constrained to Have the Same Exponent and the Same Scaling Parameters

Measure	True			False			M
	Digit addend			Digit addend			
	2	3	4	2	3	4	
rmsd	119	202	174	135	175	214	170
r ²	.979	.965	.979	.979	.975	.971	.975
a	4,909	6,337	7,164	5,534	6,631	7,426	6,334
b	656	506	394	575	548	276	493
c	1.766	2.040	1.993	1.984	2.060	2.123	1.994
1/c	.566	.490	.502	.504	.485	.471	.503

Note. The value of *a* in the table is *N*^{1/c} times the value for the distribution in each practice block; rmsd = root-mean-square deviation between predicted and observed values; *c* = Weibull exponent; 1/*c* = power function exponent.

⁴ Alphabet arithmetic also shows this item specificity. Transfer to new items is poor, even after 12 sessions of practice (see Logan & Klapp, 1991, Experiments 1 and 3).

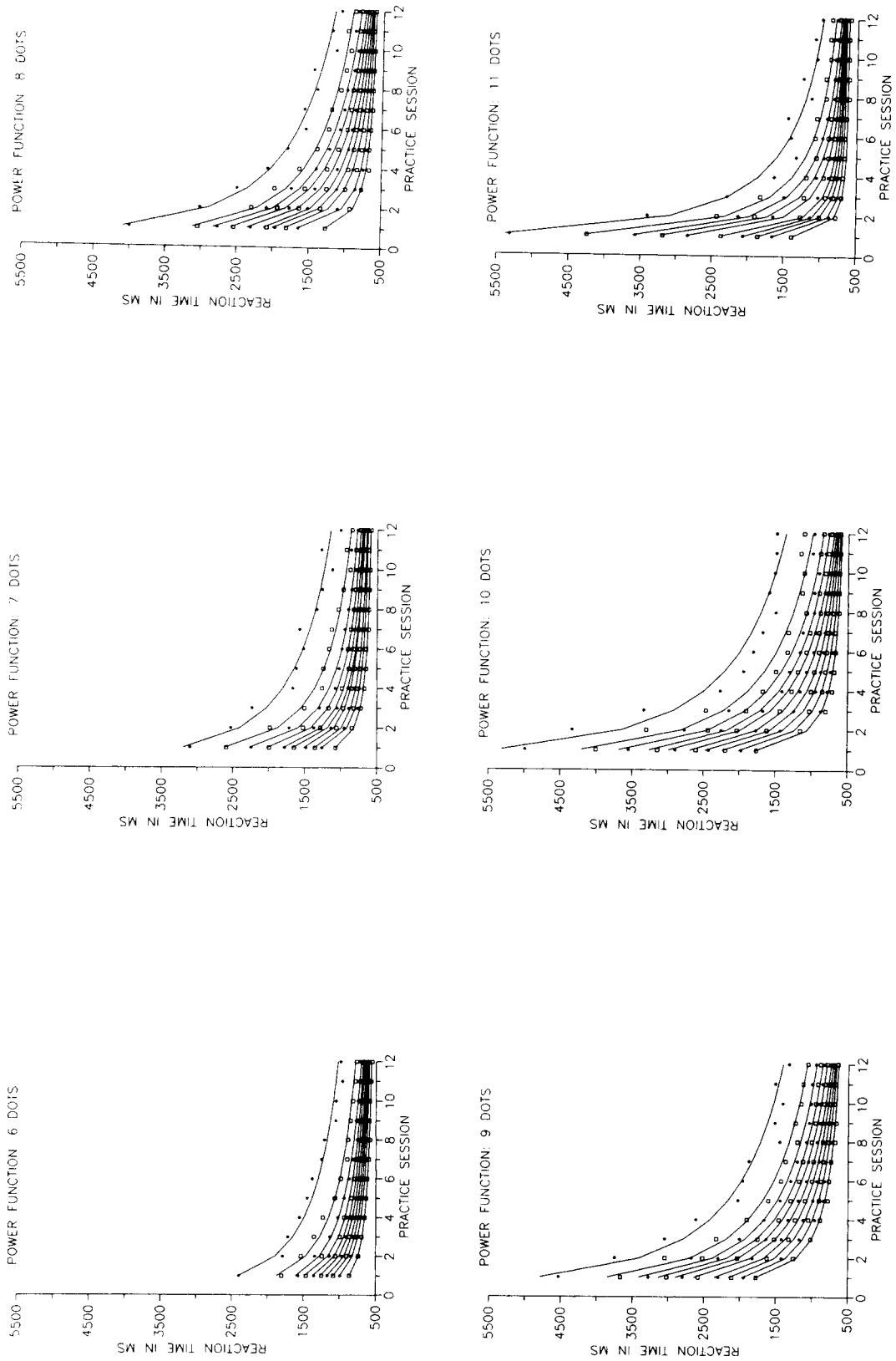


Figure 11. Power functions from Equation 1 (lines) fitted to deciles of the distributions (dots) of dot-counting reaction times as a function of practice for each numerosity level. (The 10 power functions in each panel were allowed to have separate exponents and separate scaling parameters.)

Table 4
Measures of Goodness of Fit and Exponent Parameter (c) for Fits of Equation 1 to Deciles of Reaction-Time Distributions from Dot-Counting Data of Separate and Constrained Power Functions

Measure	Numerosity						M
	6	7	8	9	10	11	
	Separate						
rmsd	31	49	58	99	143	69	75
r ²	.990	.988	.991	.982	.970	.992	.985
c	.475	.762	.705	.562	.746	1.452	.784
	Constrained						
rmsd	32	57	65	100	148	83	81
r ²	.990	.984	.988	.981	.968	.987	.983
c	.351	.561	.508	.514	.582	1.107	.604

Note. Separate functions have separate exponents and scaling parameters; constrained power functions have the same exponent but are allowed different scaling parameters; rmsd = root-mean-square deviation between predicted and observed values.

decile was not significant, $F(9, 50) = 1.68, p < .15, MS_c = .1372$, suggesting that the apparent differences did not appear consistently in the six data sets. Although the differences seem large and systematic enough to raise doubts about accepting the null hypothesis, the lack of statistical significance provides no grounds for rejecting it.

Another way to test the significance of variation in the power-function exponents is to fit constrained power functions to the data in which the different deciles are allowed to have separate scaling parameters but are required to have a common exponent. The predicted and observed values from these fits are displayed in Figure 12. Measures of goodness of fit are presented in Table 4.

The constrained power functions fit the data quite well. The squared product-moment correlation between predicted and observed values was only slightly lower than the one for unconstrained power functions and the root-mean-square deviation between predicted and observed values was only slightly higher. Only 21 parameters were required for the constrained fits in each panel of Figure 12: 2 scaling parameters for each decile and 1 common exponent. This is 9 fewer parameters than were required for the unconstrained fits in Figure 11. The reduction in the quality of the fits is small in comparison with the reduction in the number of parameters required. The goodness of fit of power functions constrained to have the same exponent suggests that variation in the exponent is not a very important feature of the data.

An important point to be taken from Figure 12 is that the 10 functions in each panel have the same shape, differing only in scale. All 10 functions share the same exponent, *c*; they differ only in the scaling parameters, *a* and *b*. The fits confirm the instance-theory prediction that the entire distribution of reaction times should decrease as a power function of practice.

Weibull distribution fits. A test of the second instance-theory prediction was done by fitting the Weibull distributions (Equation 3) to the data for each numerosity level in each

practice session. Predicted and observed values are presented in Figure 13. Measures of goodness of fit are presented in Table 5. The fits were quite good, both absolutely and in relation to the power function fits. The 120 points in each panel required 36 parameters, 2 scaling parameters, and 1 exponent for each practice session. This is a few more parameters than were required for the unconstrained power function fits, but not many more.

The instance theory predicts that the Weibull exponent should be the same for each practice session. Averaged across numerosity, the exponents were 1.374, 1.437, 1.110, 0.967, 0.999, 0.905, 0.837, 0.915, 0.800, 0.888, 0.869, and 0.986 for Sessions 1 to 12, respectively. As before, the significance of these differences was tested in a one-way ANOVA with sessions as the independent variable and the six numerosity levels serving as subjects. The main effect of sessions was significant, $F(11, 60) = 8.19, p < .01, MS_c = .0304$, suggesting that the differences in exponents were reliable across numerosity levels, contrary to the instance-theory prediction.

The instance theory also predicts that the scaling parameters of the fitted Weibull distributions should decrease as power functions of practice. Power functions were fitted to the mean values of *a* and *b*, averaged across numerosity level. The predicted and observed values are plotted in the top panel of Figure 14. The fits were quite good: For *a*, *r*² was .935, and rmsd was 94 ms; for *b*, *r*² was .993, and rmsd was 13 ms. These fits confirm the instance theory prediction that the scale of the Weibull distribution decreases as a power function of practice.

As a further test of the significance of the variation in Weibull exponents, the data were fitted by Weibull distributions (Equation 3) that were allowed to have separate scaling parameters but constrained to have a common exponent. A separate fit was carried out for each numerosity level. Predicted and observed values are plotted in Figure 15. Measures of goodness of fit are presented in Table 5. The fits were good. The product-moment correlation between predicted and observed values was almost as large as it was in the unconstrained fits, and the root-mean-square deviation between predicted and observed values increased by 18 ms. The 120 points in each panel of Figure 15 were fitted by 25 parameters, 2 scaling parameters for each practice session, and 1 common exponent for all practice sessions. That is 11 fewer parameters than were required for the unconstrained Weibull fits. The reduction in goodness of fit is small in relation to the reduction in the number of free parameters. This suggests that variation in the exponent of the Weibull distribution across practice is not very important.

Power functions (Equation 1) were fitted to the scaling parameters from the constrained Weibull fits, averaged across numerosity level, to test the instance-theory prediction that the distribution of reaction times over practice would remain Weibull with the scale reduced by a power function. Predicted and observed values are plotted in the bottom panel of Figure 14. Again, the fits were good: For *a*, *r*² was .960, and rmsd was 61 ms; for *b*, *r*² was .986, and rmsd was 24 ms. These fits confirm the instance-theory prediction.

Constrained power-function Weibull fits. The third instance-theory prediction, that the exponent of the power

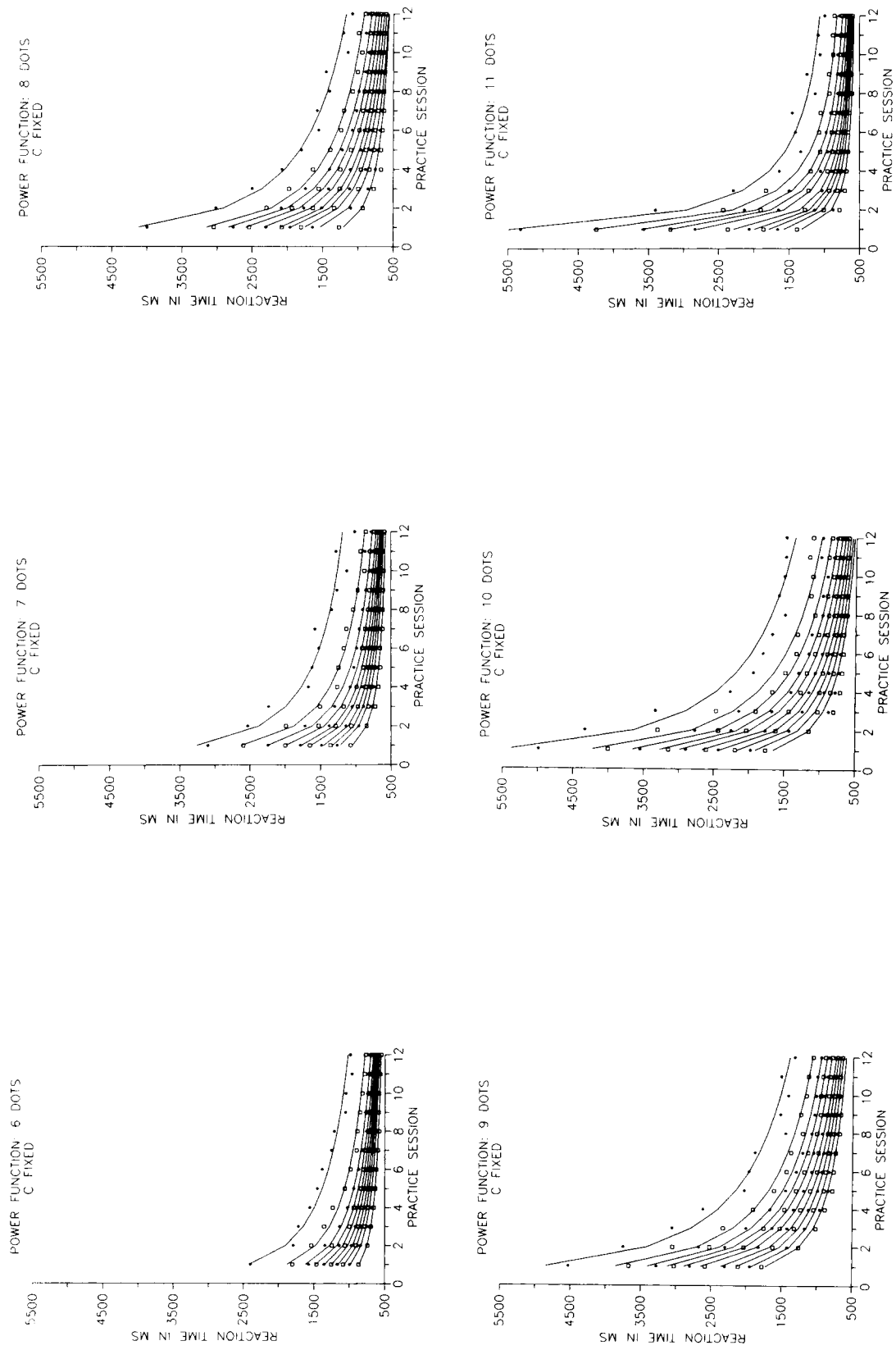


Figure 12. Power functions from Equation 1 (lines) fitted to deciles of the distributions (dots) of dot-counting reaction times as a function of practice for each numerosity level. (The 10 power functions in each panel were allowed to have separate scaling parameters but were constrained to have the same exponent.)

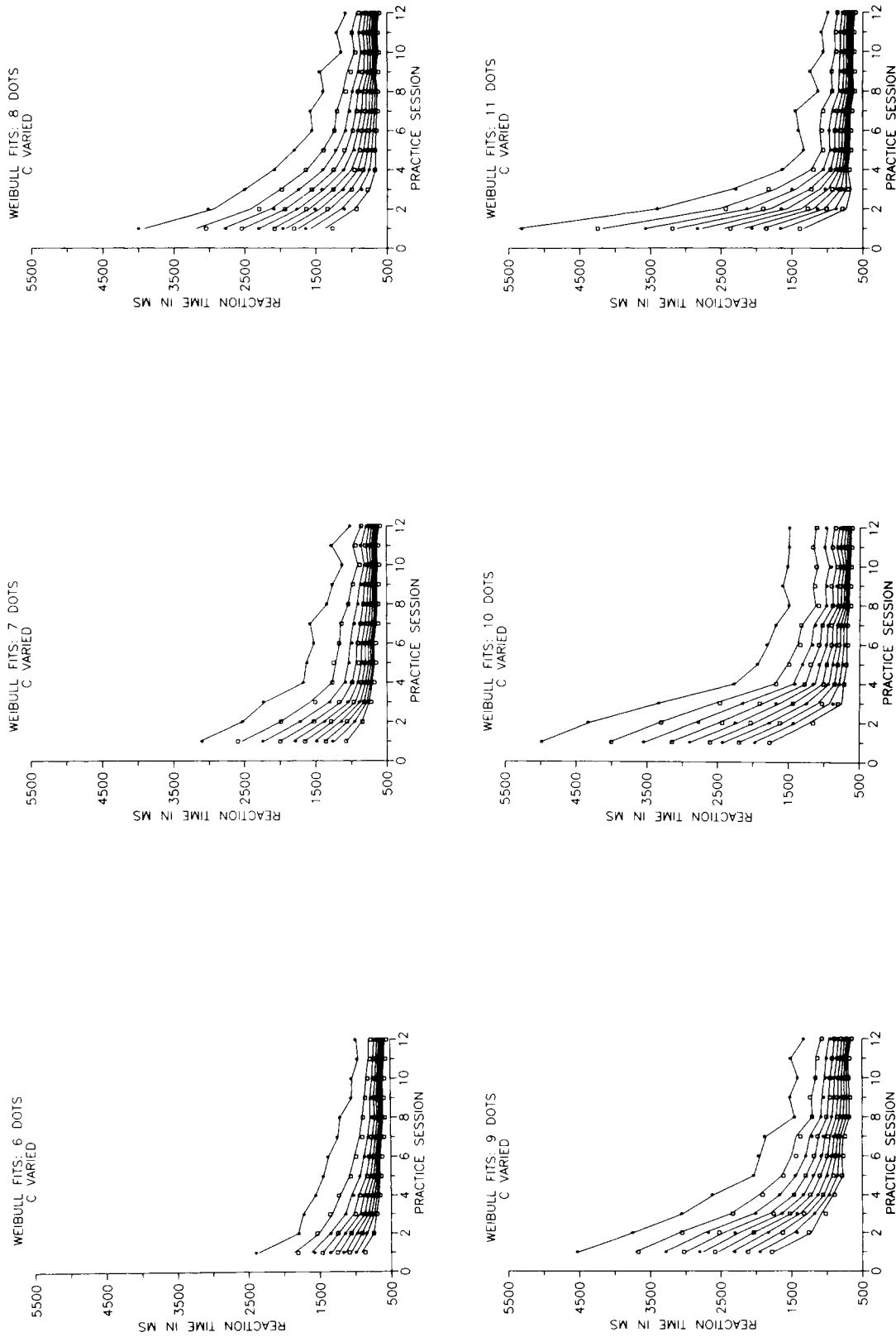


Figure 13. Weibull distributions from Equation 3 (lines) fitted to deciles of the distributions (dots) of dot-counting reaction times as a function of practice for each numerosity level. (The 10 functions in each panel were allowed to have separate scaling parameters and separate exponents.)

Table 5
Measures of Goodness of Fit and Exponent Parameter (*c*) for Fits of Equation 3 to Deciles of Reaction-Time Distributions from Dot-Counting Data of Separate and Constrained Weibull Distributions

Measure	Numerosity						<i>M</i>
	6	7	8	9	10	11	
Separate							
<i>rmsd</i>	18	18	30	24	23	33	24
<i>r</i> ²	.997	.998	.997	.999	.999	.998	.998
<i>c</i>	.874	.907	1.160	1.129	1.018	.956	1.007
Constrained							
<i>rmsd</i>	33	40	41	43	44	48	42
<i>r</i> ²	.989	.992	.995	.997	.997	.996	.994
<i>c</i>	1.114	1.148	1.340	1.299	1.312	1.321	1.256

Note. Separate Weibull distributions have different exponents and scaling parameters; constrained Weibull distributions have different scaling parameters but a common exponent; *rmsd* = root-mean-square deviation between predicted and observed values.

function should be the reciprocal of the exponent of the Weibull distribution, was tested in three ways. First, the reciprocals of the exponents from the unconstrained power-function fits ($M = 1.441$) were compared with the exponents from the unconstrained Weibull fits ($M = 1.007$). The significance of the difference was assessed with a *t* test for paired observations that compared the reciprocal of the average power function exponent at each numerosity level with the average value of the Weibull exponent for the corresponding numerosity level. The result was not significant, $t(5) = 1.96$, $p < .20$, $MS_e = .2208$, which is consistent with the instance theory.

Second, the reciprocals of the exponents from the constrained fits were compared (mean reciprocal of power function exponent = 1.860; mean Weibull exponent = 1.255). The reciprocal of the constrained power function exponent at each numerosity level was compared with the constrained Weibull exponent for the corresponding numerosity level. The result was not significant, $t(5) = 1.97$, $MS_e = .3071$, as the instance theory predicts.

Third, Weibull distributions (Equation 6) constrained to have the same scaling parameters and the same exponent for each decile and each practice session were fitted to the data from each numerosity level. Measures of goodness of fit are presented in Table 6. Observed and predicted values are plotted in Figure 16.

The fits were reasonable. Root-mean-square deviations were much larger than in the previous fits, and the product-moment correlations were lower, although they were still quite high. The reduction in goodness of fit is small considering that only three parameters were required for the fits in each panel of Figure 16, in comparison with the 21 to 36 parameters required for the previous fits. Three parameters for 120 points is very good.

The point to be taken from Figure 16 is that the shape of the functions is constrained within and between practice sessions by a single parameter, *c*. The shapes within each

session are the same except for a change of scale, and the shapes of each function across sessions are the same except for a change in scale. This confirms the instance-theory prediction that the shape of the reaction-time distribution determines the shape of the learning curve.

Inspection of Figure 16 reveals systematic deviations from predictions. In general, the predictions underestimated the observed values for the fastest and slowest reaction times (i.e., the 5th and 95th percentiles). Also, the predicted functions (across sessions) were often more sharply curved than the observed data. Once again, these discrepancies may be due to contamination from the counting algorithm used in the first few blocks.

A test of that hypothesis was done by fitting the data again, dropping out Session 1, then Sessions 1 and 2, and so on, until the first 6 sessions had been dropped. Table 7 presents measures of goodness of fit and parameter values averaged over numerosity for these fits. The measures indicate that goodness of fit improved substantially (root-mean-square deviation dropped to about one-third of its initial value) when the first 3 sessions were excluded and then did not improve

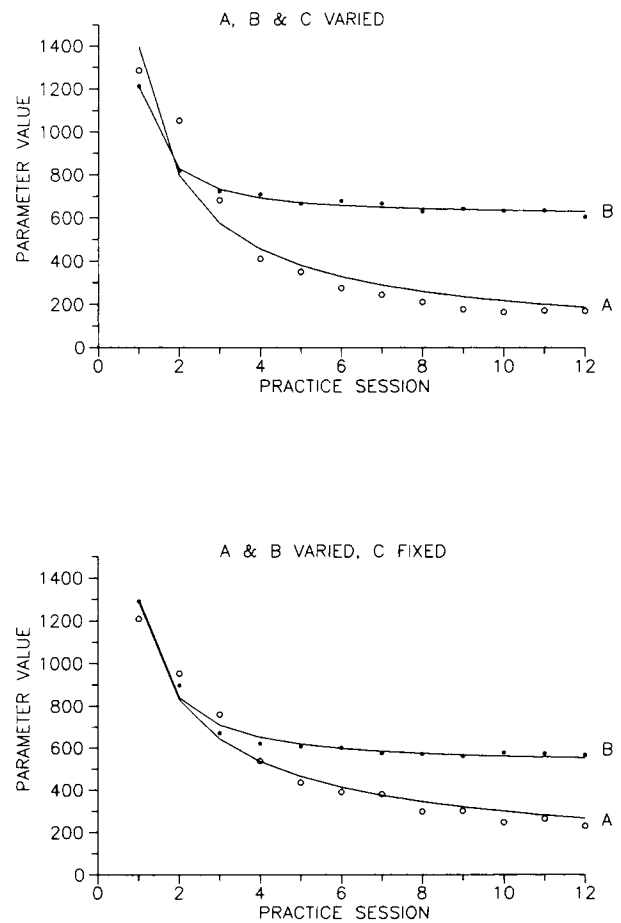


Figure 14. The *a* and *b* parameters from (top) unconstrained and (bottom) constrained Weibull distribution (Equation 3) fits to dot-counting data as a function of practice session. (Lines are predictions from fitted power functions; points are parameter values.)

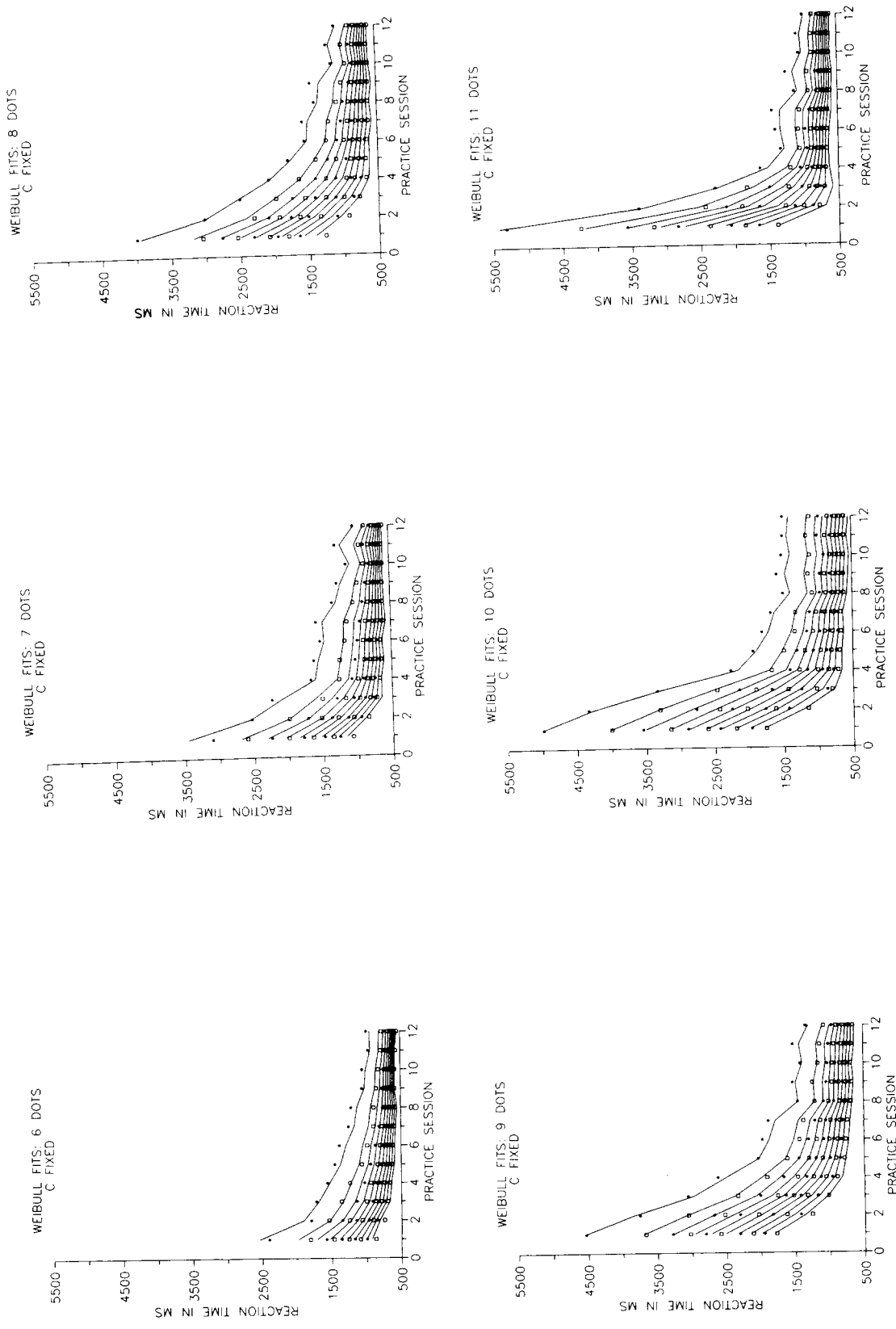


Figure 15. Weibull distributions from Equation 3 (lines) fitted to deciles of the distributions (dots) of dot-counting reaction times as a function of practice for each numerosity level. (The 10 functions in each panel were allowed to have separate scaling parameters but were constrained to have the same exponent.)

Table 6
Measures of Goodness of Fit and Parameter Values for Fits of Equation 6 to Deciles of Reaction-Time Distributions from Dot-Counting Data for Weibull Distributions Constrained to Have the Same Exponent and the Same Scaling Parameters

Measure	Numerosity						<i>M</i>
	6	7	8	9	10	11	
<i>rmsd</i>	85	104	129	198	229	143	148
<i>r</i> ²	.926	.947	.953	.928	.923	.963	.940
<i>a</i>	3,028	4,704	6,145	7,188	8,498	10,893	6,742
<i>b</i>	443	444	378	354	288	509	403
<i>c</i>	2.264	2.043	2.090	2.223	2.044	1.405	2.012
1/ <i>c</i>	.442	.489	.478	.450	.489	.712	.510

Note. The *rmsd* = root-mean-square deviation between predicted and observed values; *a* = $N^{1/c}$ times the value for the distribution in each practice session; *c* = Weibull exponent; 1/*c* = power function exponent.

much more when additional sessions were deleted. Dropping the first 3 or 4 sessions may have been enough to remove the influence of the counting algorithm. Inspection of Figure 10 indicates that the slope of the function relating reaction time to numerosity approached asymptote after 3 or 4 sessions. Subjects may not have counted to assess numerosity after that point.

Figure 17 displays the fits obtained when the first four sessions were deleted from analysis. The fits appear better than the fits to the entire data set plotted in Figure 16. Specifically, there is less underestimation of the extreme reaction times, and the curvature of the predicted functions seems to match better the curvature apparent in the data.

Discussion

The counting data supported each of the three instance-theory predictions. First, each decile of the reaction-time distribution was well fit by a power function (Equation 1), as the theory predicts. The exponents of power functions fitted to deciles separately were similar, and the fit was good when all quintiles were constrained to have the same exponent. Second, the distribution of reaction times in each session was well fit by a Weibull distribution (Equation 3), as the theory predicts. The exponents of Weibull distributions fitted to each practice block separately were not identical, but the fit was good when all blocks were constrained to have the same exponent. The scaling parameters decreased as power functions of practice, as the theory predicts. Third, the reciprocals of the exponents of the power functions did not differ from the exponents of the Weibull distributions in either the constrained or the unconstrained fits, as the instance theory predicts. Reasonable fits were obtained when all deciles and all practice blocks were fitted simultaneously with the constraint that the exponent of the power function equal the reciprocal of the exponent of the Weibull (Equation 6). The shape of the reaction-time distribution appears to determine the shape of the learning curve.

Note that the instance-theory predictions received more support in this data set than in the previous one. It may be that the instance theory applies more readily to dot counting than to alphabet arithmetic, but I doubt that it does (see Logan, 1988; Logan & Klapp, 1991). The difference is more likely due to the broader range of practice and the finer resolution of the distributions in the counting task. More data points allow more precise estimates of true parameter values.

The instance theory did not fit the counting task perfectly, however. There were substantial discrepancies between prediction and observation when the full model, represented by Equation 6, was fitted to the data (see Figures 16 and 17). Several factors may contribute to these discrepancies. First, the fits assumed no change over practice in the distribution of memory retrieval times or in the distribution of algorithm completion times. This may strike many readers as implausible. It may be possible to improve the fit by allowing both memory and the algorithm to improve with practice. However, there are no guarantees; allowing memory and the algorithm to change may make the fit worse.

Second, only three parameters were used to fit Equation 6: two scaling parameters (*a* and *b*) and one shape parameter (*c*). On the positive side, these three parameters in principle could account for an infinite amount of data. The distributions in each practice session are continuous and so could be described by an infinite number of data points. Practice could continue indefinitely. Regardless of the amount of data, however, only three parameters would be required.

On the negative side, using three parameters to fit the whole data set assumes that memory retrieval accounts for all of the variability in reaction time. This seems unlikely. Almost certainly, there are some sensory and motor "intercept" processes that add to the mean and the variance of the distribution. The parameter *b* can be interpreted as representing the mean of the intercept processes, but there is no parameter that represents the variability.

To assess the importance of variability in the intercept processes, I performed simulations in which an intercept distribution was added to a memory-retrieval distribution to produce distributions of reaction times at various stages of practice. As before, power functions (Equation 1) and Weibull distributions (Equation 3) were fitted to the simulated distributions. Ten deciles per session were simulated for 12 sessions, and each simulation was replicated 1,000 times. This time, however, I assumed that the intercept processes sped up with practice. That was reasonable for the counting task because there were 12 sessions of practice, in comparison with 1 session with the alphabet-arithmetic task. To simulate the speed-up, the experimenter multiplied the value for the intercept distribution by a power function (N^{-k} , which decreased from 1 to 0 over practice) before it was added to the value for memory retrieval that trial.

The main goal of the simulations was to find a combination of parameters that produced the qualitative pattern of the failed instance-theory predictions in the data from the counting task. In the counting task, Weibull exponents decreased over practice (although not significantly), power function exponents decreased (significantly) as decile increased, and reciprocals of power-function exponents were (nonsignifi-

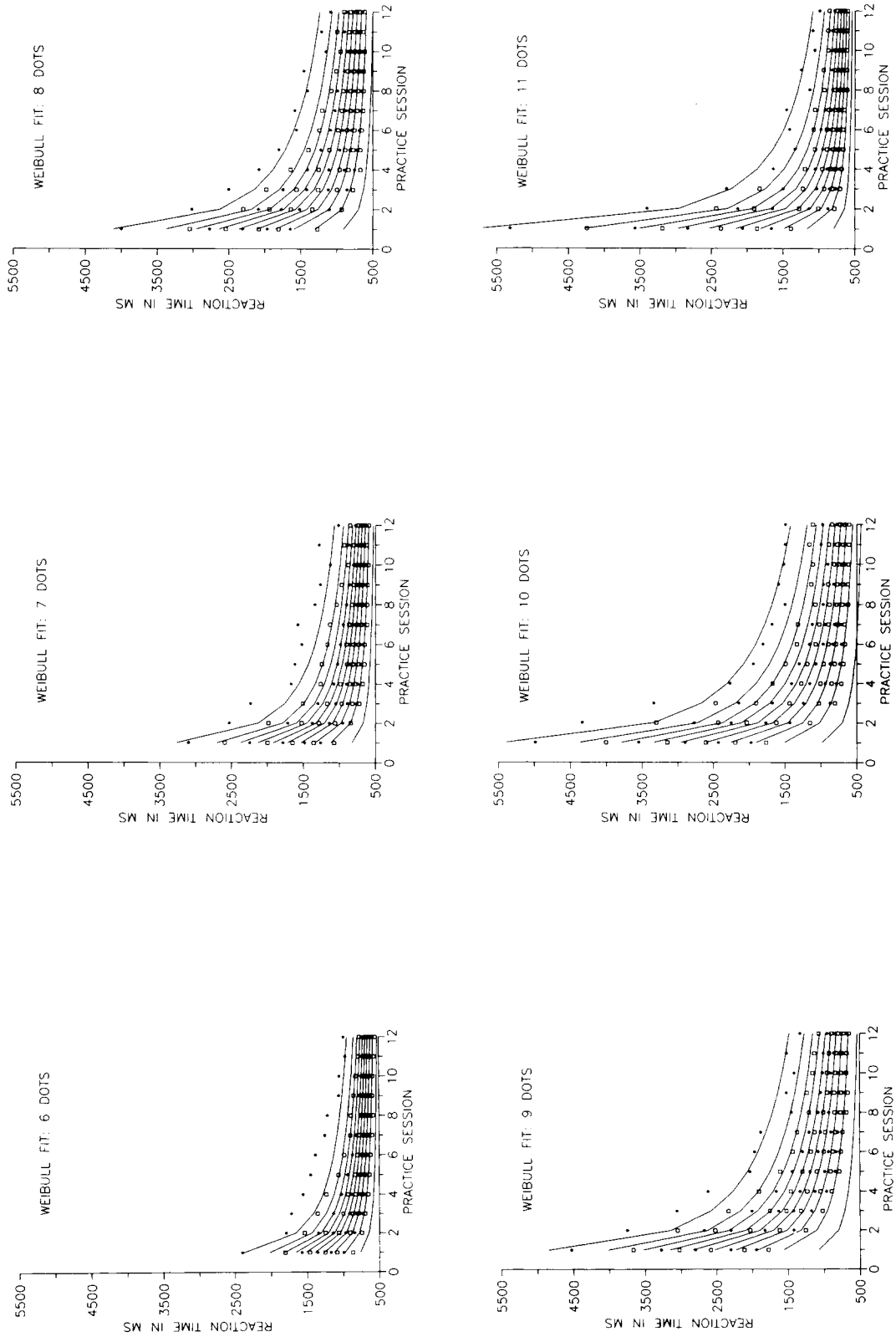


Figure 16. Weibull distributions from Equation 6 (lines) fitted to deciles of the distributions (dots) of dot-counting reaction times as a function of practice for each numerosity level. (The 10 functions in each panel were fitted with the same scaling parameters and the same exponent.)

Table 7
Effects of Sessions Dropped on Measures of Goodness of Fit and Parameter Values Averaged Over Numerosity (1 to 6) for Fits of Equation 6

Sessions dropped	Goodness of fit		Parameter			
	r^2	$rmsd$	a	b	c	$1/c$
0	.940	148.10	6,742	403	2.012	.510
1	.953	95.51	12,987	565	1.370	.738
2	.956	69.28	13,857	576	1.295	.776
3	.956	53.36	12,452	574	1.317	.764
4	.951	50.88	11,573	563	1.359	.741
5	.941	52.57	10,815	553	1.408	.715
6	.931	54.41	9,601	538	1.472	.687

Note. The $rmsd$ = root-mean-square deviation between predicted and observed values; $a = N^{1/c}$ times the value for the distribution in each practice session; c = Weibull exponent; $1/c$ = power function exponent.

cantly) larger than the corresponding Weibull exponents. Note that power-function exponents decreased with decile in this data set, whereas they increased with quintile in the alphabet-arithmetic data set. I interpret this as an effect of the greater degree of practice in the counting task.

My exploration of the parameter space was not systematic, but I was able to find several combinations of parameters that produced Weibull exponents that decreased over practice and power function exponents that decreased as decile increased. One combination that produced these two effects and also produced reciprocals of power-function exponents that were larger than the Weibull exponents was the following: memory Weibull with $a = b = 1,200$ and $c = 1.5$; intercept Weibull with $a = b = 600$, $c = 1$; and power-function exponent $k = .5$. These simulations demonstrate that it may be possible to account for the failures of the instance-theory predictions by adding a variable intercept process to the model and assuming that the intercept process improves with practice. However, the simulations do not constitute a theory of the intercept process. Such a theory is well beyond the scope of this article.

Where does this leave us? I would argue that the fits should encourage readers to take the instance theory seriously. They may be incomplete, but they provide a good approximation to a large number of data points with only three free parameters. The power function fits (Equation 1) were better than the fits to Equation 6 in this data set and the previous one, and they serve to demonstrate that the whole distribution of reaction times decreases as a power function of practice. That remains a fact to be accounted for or predicted by other theories.

General Discussion

The three fundamental assumptions of the instance theory (obligatory encoding, obligatory retrieval, and instance representation) imply a learning mechanism, whereby the strength of the response from memory increases with consistent practice. This, together with three supplementary assumptions (retrieval time is a random variable; the first instance to be retrieved governs performance; and the retrieval-time dis-

tribution is the same for all instances), implies that (1) the entire distribution of reaction times, not just the mean, will decrease as a power function of practice; (2) the distribution of reaction times in a well-practiced task will conform to a Weibull distribution; and (3) the exponent of the Weibull distribution will be the reciprocal of the exponent of the power function that describes the speed-up in reaction time (i.e., the shape of the distribution determines the shape of the learning curve). The fits reported in the last two sections of this article confirmed all three predictions. In this section, I ask whether other theories could account for these results.

I would argue that no other theory predicts these results, in that they can be deduced from fundamental assumptions. However, other theories may be able to account for some aspects of the results, in that they can be implemented in such a way as to produce the results. This difference is important. A theory that predicts results must stand or fall on the success of the prediction. A theory that accounts for results may stand when prediction fails; it is the implementation, not the theory itself, that stands or falls with the predictions. Moreover, a theory that predicts a phenomenon provides a more parsimonious account than one that requires arbitrary assumptions that go beyond its fundamental premises.

Power Law for Distributions

Most theories of skill acquisition and automaticity can account for the power-function speed-up in mean reaction time (e.g., J. R. Anderson, 1982; J. R. Anderson & Milson, 1989; Cohen et al., 1990; Crossman, 1959; Logan, 1988; MacKay, 1982; Newell & Rosenbloom, 1981; Schneider, 1985). Cohen et al. (1990) and J. R. Anderson (1992) have tried to account for the power-function speed-up means and standard deviations. But so far, only Logan (1988) provides an account of the power-function speed-up in the distribution of reaction times. In this section, I consider how theories might be modified to account for changes in distributions as well as means.

Mixture models. In probability mixture models, reaction time is determined by two parent distributions, one representing unpracticed performance and one representing well-practiced performance. Reaction times are sampled from one distribution with probability p and from the other distribution with probability $1 - p$. If p changes with practice appropriately, mixture models could predict a power-function speed-up for mean reaction time. Siegler (1987) proposed a model of children's acquisition of skill at addition that can be implemented as a mixture model and so could make this prediction.

Probability mixture models cannot account for the power law for distributions. In mixture models, the distribution does not contract monotonically with practice as the power law requires. It first expands and then contracts as practice progresses (as p goes from 0 to 1). This can be seen in the expression for the variance of a mixture distribution (see Townsend & Ashby, 1983, p. 264):

$$\sigma_{\text{mixture}}^2 = (1 - p)\sigma_{\text{slower}}^2 + (p)\sigma_{\text{faster}}^2 + p(1 - p)(\mu_{\text{slower}} - \mu_{\text{faster}})^2.$$

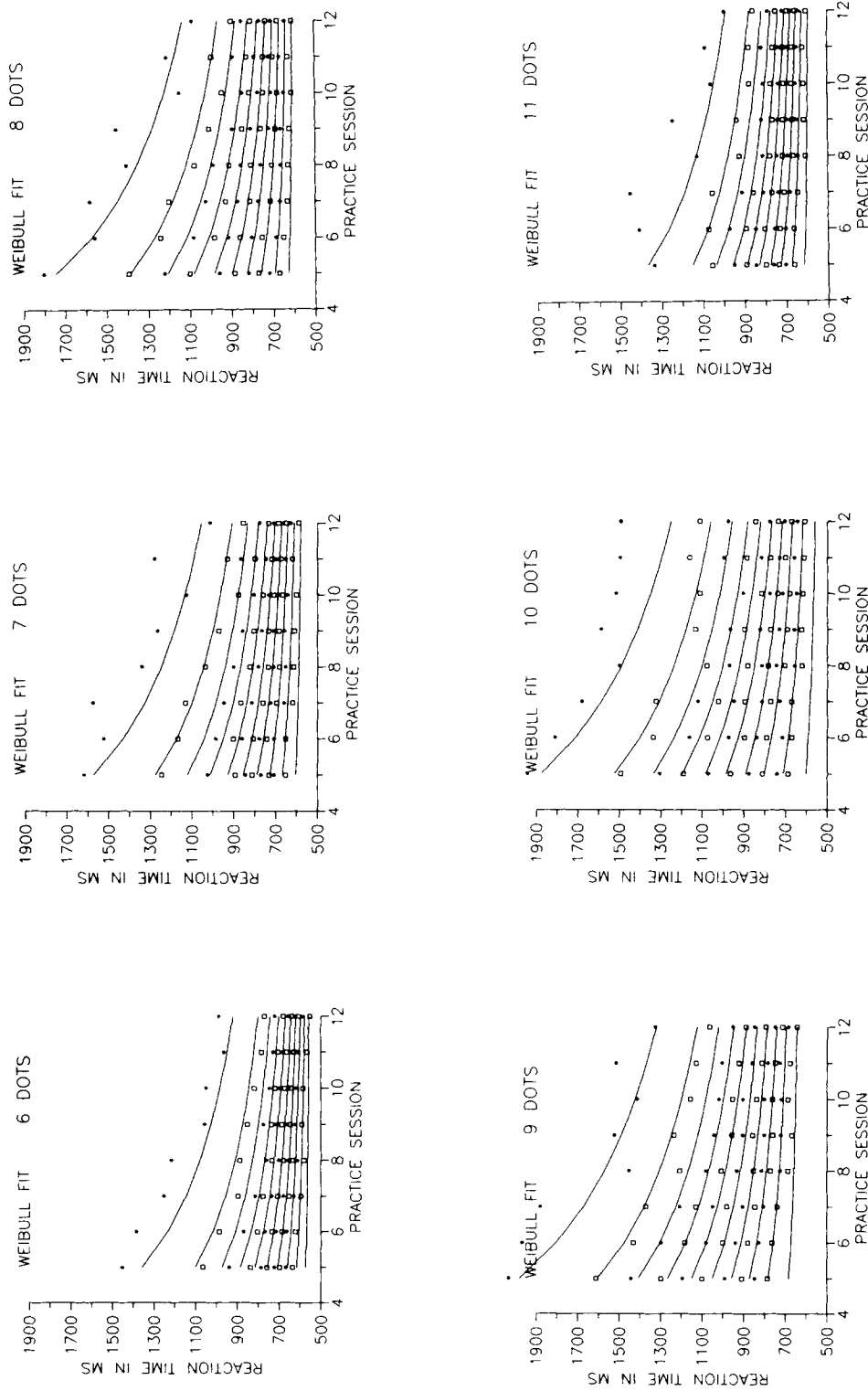


Figure 17. Weibull distributions from Equation 6 (lines) fitted to deciles of the distributions (dots) of dot-counting reaction times as a function of practice for each numerosity level, excluding the first four sessions of practice. (The 10 functions in each panel were fitted with the same scaling parameters and the same exponent.)

The third term on the right side of this equation represents the squared difference between the means of the parent distributions, multiplied by the product of the probabilities of sampling them. This term will grow from 0 as p increases from 0, reaching a maximum when $p = .5$, and diminish to 0 as p approaches 1. Thus, the distribution will expand and then contract as p goes from 0 to 1. Compton and Logan (1991) tested the variance prediction of mixture models and found evidence against it.

Mixture models that assume more than two parent distributions lead to the same predictions as long as subjects begin with one initial distribution and progress through a series of distributions to one final distribution. Analysis of variance provides an appropriate analogy: As long as there are differences between the means of the distributions, the variance within a single distribution must be smaller than the variance of the mixture of distributions, just as the mean square within groups must be smaller than the mean square between groups. A possible exception would be a model like Crossman's (1959), in which several methods are equally likely to begin with, and practice converges on one of them. In that case, variance would be maximal initially (when all methods were equally likely) and would diminish as a single method dominated. Whether the reduction would follow the power law remains to be seen.

Strength models. Most models account for the power law for means by assuming that the strength of connections between stimulus and response (or between internal representations) increases by a constant proportion of the strength remaining to be gained. That is,

$$\text{Strength}_n = \text{Strength}_{n-1} + c(\text{Strength}_{\max} - \text{Strength}_{n-1}), \quad (7)$$

where Strength_n is the connection strength on trial n , Strength_{\max} is the maximum possible strength, and c is the learning rate, usually varying between 0 and 1. This idea underlies the strength mechanism in J. R. Anderson's (1982) model, the Hebb learning rule used in Schneider's (1985) model, and the back propagation algorithm used in Cohen et al. (1990).⁵ In these models, strength changes quickly early in practice and slowly later on. Reaction time is assumed to be inversely related to strength, and this results in a speed-up that follows the power law.

Most strength models are deterministic rather than stochastic and, therefore, make no predictions about variability. They cannot account for the power law for distributions without adding some assumptions, without injecting noise into the system somewhere. I consider three possibilities: First, noise can be added to the output of the system. Reaction times can be computed deterministically and a random variable can be added to them. If the random variable were selected appropriately, this could produce a reaction time distribution with the right shape. However, adding noise at output is unlikely to work in general, because there is no mechanism by which the noise can change with practice. The distribution would stay the same with practice; only the mean would change.

Second, noise can be added to the input to the system or at various points along the way. This is the solution adopted by

Cohen et al. (1990) in their model of the Stroop effect. Their model consists of three layers of nodes, an input layer, a hidden layer, and an output layer. They add noise to the second and third layers. In this approach, noise influences performance by causing momentary fluctuations in activation level. Noise effects will diminish with practice because there is a limit to the maximum activation level. Early in practice, when connections are weak, activation takes a long time to approach the maximum level, and there is plenty of room for fluctuations to affect performance. Later in practice when connection strengths are near asymptote, activation reaches near-maximum levels very quickly, and there is little room for random fluctuations to affect performance. Consequently, the distribution of reaction times will contract as practice progresses.

There is no guarantee that adding noise in this fashion will produce the power law for distributions, however. Cohen et al. (1990) were able to produce power-function learning only if they fixed connection weights between the first two layers and allowed weights to change between the last two. Even then, the match between the exponent for means and standard deviations (i.e., the fit of the power law for distributions) depended on the amount of noise they added. With too much or too little noise, means and standard deviations followed different power functions. Cohen et al. did not report boundary conditions on successful fits; further research will be needed to establish them.

J. R. Anderson (1992) added noise to his ACT* model by treating strength as a rate parameter for an exponential distribution of reaction times. The mean and the standard deviation of the exponential distribution equal the reciprocal of the rate parameter, so Anderson can explain changes in means, standard deviations, and entire distributions of reaction time with practice. J. R. Anderson (1982) showed that strength increases as a power function of practice, so the mean, standard deviation, and the entire distribution of reaction times should all decrease as a power function of practice. However, as I pointed out before, the exponential distribution leads to predictions that are clearly falsified by empirical data (i.e., that the mean should equal the standard deviation and that the reduction in the mean with practice should equal the reduction in the standard deviation). Whether these data falsify an implementation of Anderson's model or the model itself remains to be seen. The answer depends on how central the assumption of exponential distributions is to ACT*. Perhaps some other distribution could be

⁵ Technically, in the Hebb learning rule and the delta rule used in back propagation, Strength_{\max} is defined as the strength required to produce a desired or optimal output. The delta rule describes the change in strength between units i and j as follows:

$$\delta \text{Strength}_{ij} = c(t_j - o_j)i_i,$$

where c is the learning rate, t_j is the desired or target output for unit j , o_j is the actual output for unit j , and i_i is the input from unit i (see Rumelhart, Hinton, & Williams, 1986). This expression will reduce to Equation 7 if $i_i = 1$ and t_j is interpreted as Strength_{\max} .

used instead. Even then, there is no guarantee that other distributions would lead to appropriate predictions with plausible interpretations (see Footnote 2).

Third, noise may be intrinsic in the input to the system. Stimuli may vary for a number of reasons, and this variation may propagate through the system, resulting in variation in reaction time (see J. A. Anderson, 1991). Variation among stimuli would be compressed as strength approaches asymptote and so would the resulting variation in reaction time. With appropriate parameters, the reduction in variation may follow the power law. Seidenberg and McClelland (1989) provided an example of the effects of practice on intrinsic noise in their model of word naming: Words vary in orthographic regularity and the number of words (neighbors) with which they share letters. Both of these factors affect the time it takes to name words, but the effects are much stronger with low-frequency words than with high-frequency words. Seidenberg and McClelland modeled these interactions by varying strength; differences apparent at low levels of practice diminished as strength increased. Whether effects like these are considered noise depends on the investigator's perspective; one person's noise may be another person's independent variable. All researchers collapse data across stimuli in some way or other, so intrinsic noise must be commonplace. In the fits reported above, for example, I collapsed across stimuli and small amounts of practice (6 or 16 presentations) to construct distributions. In principle, stimuli or practice effects or both could have produced the variation I observed.

It remains to be seen whether the reduction of intrinsic noise by strengthening will account for the power law for distributions, but it seems to be the most promising approach. It is reasonable and realistic to assume variation among stimuli (or among matches between stimuli and decision criteria, etc.), in contrast with arbitrary injection of noise into the system or with the instance theory's assumption that all stimuli have the same retrieval-time distribution.

Weibull Distribution

Will other theories of skill acquisition and automaticity predict that the distribution of well-practiced reaction times will conform to the Weibull? It seems unlikely that they will predict it, but they may well account for it. With appropriate exponents (i.e., between 1 and 3), the Weibull resembles typical reaction-time distributions. So to the extent that the theories can produce reaction-time distributions that resemble typical ones, they will fit the Weibull.

Probability mixture models may produce Weibull distributions in the initial and final stages of practice (i.e., when the parent distributions dominate). However, they are unlikely to produce Weibull distributions at intermediate stages of processing when the distribution is a mixture of the two parents. Mixtures of Weibulls are unlikely to be Weibulls.

Strength models may produce Weibull distributions, depending on the nature of the noise that is added to the system and on the transformation that relates strength to reaction time. One strategy, used by Cohen et al. (1990), is to use output strength to drive a random walk, such as Ratcliff's (1978) diffusion model. Ratcliff's diffusion model produces

reaction-time distributions that are well fit by the convolution of normal and exponential distributions, which in turn is well fit by the Weibull (see Appendix B).

Models that assume intrinsic noise may not produce Weibull distributions very easily. In principle, intrinsic noise can be attributed to systematic variation within the stimulus domain. This is a conceptual advantage over noise injected arbitrarily into the system. However, constraints on the stimulus domain must determine the distribution of intrinsic noise, and those constraints may not produce Weibull distributions of reaction times. There are fewer theoretical degrees of freedom with intrinsic noise than with arbitrary noise, and that may work against the theory.

Shapes of Distributions and Learning Curves

Will other theories of skill acquisition and automaticity predict that the shape of the distribution determines the shape of the learning curve? It is not likely. This prediction seems to be unique to the instance theory. Other theories may be able to account for the power law for distributions and the Weibull shape of reaction-time distributions. But I suspect other theories would have a hard time producing power law fits with exponents that equal the reciprocal of the exponent of the Weibull distribution of reaction times. The instance theory predicts that constraint because learning derives from assumptions about reaction-time distributions (i.e., learning reflects the outcome of a race between n samples drawn from the same distribution). In other theories, the assumptions that produce learning are separate from the assumptions about noise that produce reaction-time distributions. There is no necessary relation between them. It may be possible to select parameters such that the shape of the distribution is correlated with the shape of the learning curve, but it seems unlikely that the relation would be deterministic as it is in the instance theory. Other theories may account for the constraint, but it seems unlikely that they would predict it.

The constraint between the shape of the retrieval-time distribution and the shape of the learning curve is easy to grasp intuitively in the framework of the instance theory. Distributions shaped like the exponential, in which the minimum value is the modal value, should produce sharply inflected learning curves because few samples would be required to converge on the minimum value. By contrast, distributions shaped like the normal, in which the modal value is far from the minimal value, should produce more gradual learning curves because many samples would be required to converge on the minimum. It seems unlikely that other theories could account for the constraint in a manner that is as intuitively compelling as the instance theory's account.

Conclusion

The data presented in this article demonstrate that (1) the distribution of reaction times decreases as a power function of practice, (2) the distribution of reaction times is well described by the Weibull, and (3) the shape of the reaction-

time distribution determines the shape of the learning curve. All of these results were predicted by the instance theory. The ability to predict these results is a significant strength of the instance theory. The predictions derive from the fundamental assumptions, and the theory must stand or fall on the success of the predictions. The results suggest that the theory stands. Other theories may be able to account for some or all of these results, but so far, no theory predicts them by deduction from fundamental assumptions.

The results are important apart from their relevance to the instance theory. They generalize the power law to distributions, whereas previously it applied only to mean reaction times. They demonstrate a constraint between reaction-time distributions and learning curves that may prove to be a fundamental phenomenon in automaticity and skill acquisition, heretofore undiscovered. It is relatively easy to generate models that predict power function reductions in mean reaction time. It is harder to predict the power law for distributions and harder yet to predict the constraint between distributions and learning curves.

The discovery of these phenomena also constitute a significant strength of the instance theory. The worth of a theory is measured, in part, by its ability to generate new predictions and reveal new phenomena. The instance theory led to the discovery of the power law for distributions and the constraint between shapes of distributions and shapes of learning curves. If these phenomena are as fundamental as I suspect they may be, the instance theory will have been worthwhile, even if it is ultimately falsified on other grounds.

References

- Anderson, J. A. (1991). Why, having so many neurons, do we have so few thoughts? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays in honor of Bennet B. Murdock* (pp. 477–507). Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–406.
- Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, *105*, 165–180.
- Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*, 703–719.
- Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 517–526.
- Chandler, P. J. (1965). *Subroutine STEPIT: An algorithm that finds the values of the parameters which minimize a given continuous function* [Computer program]. Bloomington: Indiana University, Quantum Chemistry Program Exchange.
- Chi, M. T. H., & Klahr, D. (1975). Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology*, *19*, 434–439.
- 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 memory retrieval in memory based theories of automaticity. *Memory & Cognition*, *19*, 151–158.
- Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, *2*, 153–166.
- Gentner, D. R. (1983). Keystroke timing in transcription typing. In W. Cooper (Ed.), *Cognitive aspects of skilled typewriting* (pp. 95–120). New York: Springer-Verlag.
- Gnedenko, B. (1943). Sur la distribution limitée du terme maximum d'une série aléatoire [On the limiting distribution of maximum terms of a random series]. *Annals of Mathematics*, *44*, 423–453.
- Gumbel, E. J. (1958). *Statistics of extremes*. New York: Columbia University Press.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace 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.
- 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* (pp. 1–47). San Diego, CA: Academic Press.
- Jensen, E. M., Reese, E. P., & Reese, T. W. (1950). The subitizing and counting of visually presented fields of dots. *Journal of Psychology*, *30*, 363–392.
- Johnson, N. L., & Kotz, S. (1970). *Continuous univariate distributions: Vol. 1*. New York: Wiley.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*, 136–153.
- Kail, R. (1986). The impact of extended practice on mental rotation. *Journal of Experimental Child Psychology*, *42*, 378–391.
- Kramer, A. F., Strayer, D. L., & Buckley, J. (1990). Development and transfer of automatic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 505–522.
- Lassaline, M. L., & Logan, G. D. (1991). *Memory-based automaticity in the discrimination of visual number*. Unpublished manuscript, University of Illinois at Urbana-Champaign.
- Leadbetter, M. R., Lindgren, G., & Rootzen, H. (1983). *Extremes and related properties of random sequences and processes*. New York: Springer-Verlag.
- 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., & 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*. New York: Oxford University Press.
- MacKay, D. G. (1982). The problem of flexibility, fluency, and speed-accuracy tradeoff in skilled behavior. *Psychological Review*, *89*, 483–506.
- MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 126–135.
- 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.
- 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.
- Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 136–153.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
- Ratcliff, R. (1979). Group reaction time distributions. *Psychological Bulletin*, *86*, 446–461.
- Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recog-

niton memory. *Psychological Review*, 83, 190–214.
 Ross, B. H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology*, 16, 371–416.
 Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–364). Cambridge, MA: MIT Press.
 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.
 Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
 Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, 116, 250–264.

Smith, E. R., Branscome, N. J., & Bormann, C. (1988). Generality of the effects of practice on social judgment tasks. *Journal of Personality and Social Psychology*, 54, 385–395.
 Smith, E. R., & Lerner, M. (1986). Development of automatism of social judgments. *Journal of Personality and Social Psychology*, 50, 246–259.
 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.
 Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology*, 21, 136–152.
 Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, England: Cambridge University Press.
 Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General*, 117, 319–331.

Appendix A

Power Function Predictions for Nonidentical Distributions

The power function proofs assume that the retrieval times for the various traces in the race come from identical independent distributions. Identity is a strong assumption. It requires that the distributions have the same form and the same parameter values, which seems unlikely to be true in practice. In this section, I attempt to generalize the proof to independent distributions of the same form with different parameter values. Thus, the distributions need not be identical; they need only come from the same family. That may be a more plausible assumption than complete identity, especially considering the flexibility of families of distributions such as the Weibull (see Appendix B).

Weibull Distributions With Fixed Exponents

It is possible to prove mathematically that the distribution of minima sampled from independent Weibull distributions with identical exponents but different scaling parameters (different values of a) will decrease asymptotically as a power function of the number of samples.^{A1} The proof assumes that the additive constant, b , is zero. The multiplicative constant, a , can vary randomly.

I begin by restating the proof for independent, identically distributed Weibull distributions. Let $\mathbf{X}(\alpha, c)$ denote a Weibull distributed random variable:

$$F(x) = \text{Prob}\{\mathbf{X}(\alpha, c) \leq x\} = 1 - \exp[-\alpha x^c],$$

which is essentially Equation 4 with $a = \alpha^{-1/c}$ and $b = 0$. Let $\mathbf{X}_{\min}(N, \alpha, c)$ denote the minimum of N independent identically Weibull-distributed random variables, $\mathbf{X}_1(\alpha, c), \mathbf{X}_2(\alpha, c), \dots, \mathbf{X}_N(\alpha, c)$, with common parameters α and c . For the random variable $\mathbf{X}_{\min}(N, \alpha, c)$ one has

$$\begin{aligned} \text{Prob}\{\mathbf{X}_{\min}(N, \alpha, c) > x\} &= \prod_{i=1}^N \text{Prob}\{\mathbf{X}_i(\alpha, c) > x\} \\ &= \prod_{i=1}^N \exp[-\alpha x^c] = \exp[-N\alpha x^c], \end{aligned} \quad (\text{A1})$$

which means it is Weibull-distributed with parameters $(N\alpha, c)$:

$$\text{Prob}\{\mathbf{X}_{\min}(N, \alpha, c) \leq x\} = 1 - \exp[-N\alpha x^c]. \quad (\text{A2})$$

The p th quantile of this distribution is $X_p(N, \alpha, c)$. It is obtained by choosing a value of p ($0 < p < 1$) and solving Equation A2 for x :

$$\begin{aligned} p &= 1 - \exp[-N\alpha x^c] \\ 1 - p &= \exp[-N\alpha x^c] \\ \log(1 - p) &= -N\alpha x^c \\ \log(1 - p)/(-N\alpha) &= x^c \\ x &= \{\log(1 - p)/(-N\alpha)\}^{1/c} = X_p(N, \alpha, c). \end{aligned}$$

With a little rearrangement,

$$X_p(N, \alpha, c) = N^{-1/c} \{\log(1 - p)/-\alpha\}^{1/c}. \quad (\text{A3})$$

According to Equation A3, as N increases, $X_p(N, \alpha, c)$ decreases strictly proportionally to $N^{-1/c}$; each quantile of the distribution decreases as a power function of N with an exponent equal to the reciprocal of the exponent for the original Weibull distribution. This is essentially what I showed earlier in Equation 6. It follows that all k moments about zero of this distribution will also decrease as power functions of N with exponents equal to $-k/c$, where c is the exponent the original Weibull distribution.

Now I generalize this proof to nonidentical distributions: Suppose that the exponent c is fixed but the values $\alpha_1, \dots, \alpha_N$ for N Weibull-distributed random variables are selected at random and independently from a certain population of α values. This makes the selected α_i a random variable with $\text{Prob}(\alpha_i < \alpha^*)$ being a fixed value for any given α^* , independent of i and N . Let $E(\alpha)$ denote the mean value of α in the population.

Let $\mathbf{X}_{\min}(\alpha_1, \dots, \alpha_N, c)$ denote the minimum of N random variables selected in this fashion. For any particular choice of $\alpha_1, \dots, \alpha_N$ one has (by analogy with Equation A1)

$$\begin{aligned} \text{Prob}\{\mathbf{X}_{\min}(\alpha_1, \dots, \alpha_N, c) > x\} &= \prod_{i=1}^N \text{Prob}\{\mathbf{X}_i(\alpha_i, c) > x\} \\ &= \prod_{i=1}^N \exp[-\alpha_i x^c] = \exp[-\sum_{i=1}^N \alpha_i x^c], \end{aligned}$$

which is Weibull-distributed with parameters $(\sum_{i=1}^N \alpha_i, c)$:

$$\text{Prob}\{\mathbf{X}_{\min}(\alpha_1, \dots, \alpha_N, c) \leq x\} = 1 - \exp[-\sum_{i=1}^N \alpha_i x^c].$$

^{A1} I would like to thank Ehtibar Dzhafarov for providing this proof.

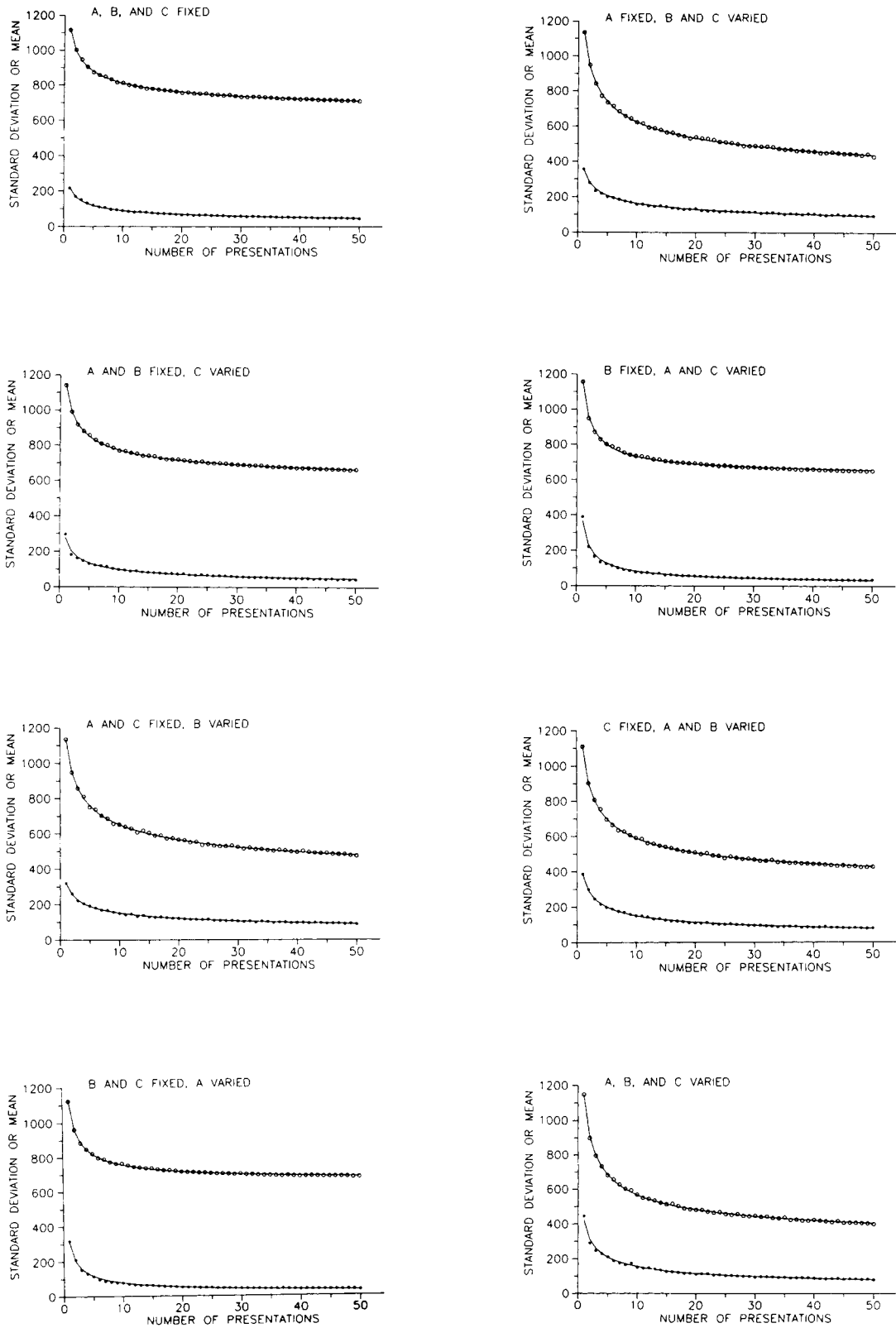


Figure A1. Power functions (lines) fitted to means (top line in each panel) and standard deviations (bottom line in each panel) of simulated data (dots) sampled from Weibull distributions in which parameters were fixed or free to vary. (The two functions in each panel were allowed to have different asymptote and multiplicative parameters but were constrained to have the same exponent.)

By analogy with Equation A3, the p th quantile of this distribution is

$$X_p(\alpha_1, \dots, \alpha_N, c) = \{\log(1-p)/-\sum_{\alpha_i}^N\}^{1/c}. \tag{A4}$$

Now, $\sum^N \alpha_i$ can be rewritten as $N\bar{\alpha}$, where $\bar{\alpha}$ is the mean of $\alpha_1, \dots, \alpha_N$ (because by definition, $\bar{\alpha} = \sum^N \alpha_i/N$). This can be substituted into Equation A4 to yield

$$X_p(\alpha_1, \dots, \alpha_N, c) = \{\log(1-p)/-N\bar{\alpha}\}^{1/c} = N^{-1/c}\{\log(1-p)/-\bar{\alpha}\}^{1/c}. \tag{A5}$$

Unlike the quantiles described in Equation A3, the quantiles described in Equation A5 do not decrease strictly proportionally to $N^{-1/c}$ because the value of $\bar{\alpha}$ will generally change as N increases. It is well-known, however, that as N increases, $\bar{\alpha}$ converges to $E(\alpha)$ both “in probability” and “almost certainly.” The former means that for any positive ξ , however small,

$$\text{Prob}\{|\bar{\alpha} - E(\alpha)| < \xi\} \rightarrow 1.$$

The second meaning of convergence (“almost certainly”) is even stronger:

$$\text{Prob}\{\bar{\alpha} \rightarrow E(\alpha)\} = 1,$$

that is, with probability 1, for any positive ξ , however small, one can find a value of N such that beginning with that value,

$$|\bar{\alpha} - E(\alpha)| < \xi.$$

Since the p th quantile $X_p(\alpha_1, \dots, \alpha_N, c)$ is a continuous function of $\bar{\alpha}$, one can conclude that both in probability and almost certainly

$$X_p(\alpha_1, \dots, \alpha_N, c) \rightarrow N^{-1/c}\{E(\alpha)^{-1}\log(1-p)\}^{1/c}. \tag{A6}$$

The right side of Equation A6 decreases strictly proportionally to $N^{-1/c}$, so one can say that $X_p(\alpha_1, \dots, \alpha_N, c)$ decreases asymptotically proportionally to $N^{-1/c}$. This means that with N sufficiently large, the quantiles of a distribution of minima sampled from independent Weibull distributions with a common exponent but separate scaling parameters will decrease approximately as a power function of N with an exponent equal to the reciprocal of the exponent for the Weibull distribution. It follows that asymptotically, all k moments about zero of this distribution will also decrease as power functions of N with an exponent equal to $-k/c$, where c is the exponent for the original Weibull distributions. Thus, in this case, the assumption of identical distributions can be relaxed somewhat without changing the power function prediction.

Weibull Distribution With All Parameters Varying

Unfortunately, it is not easy to generalize the preceding proof for Weibull distributions with different exponents. As an alternative to mathematical proof, I performed a number of Monte Carlo simulations with Weibull distributions (i.e., Equation 3) that varied in all three parameters to test the importance of assuming identical distributions. In each simulation, N independent samples were drawn from Weibull distributions, and the minimum value was calculated. N varied from 1 to 50. Sampling was replicated 1,000 times, and the mean and the standard deviation of the minimum was calculated as a function of N . Power functions were fitted to the means and standard deviations using STEFIT, allowing them to have separate asymptotes and multiplicative parameters but constraining them to have the same exponent, as the instance theory predicts.

The Weibull parameters were either fixed or varied. When a and b were fixed, they were set at 600, which was the mean of the range

of values they took when they varied; when c was fixed, it was set at 2.5, which was the mean value of the range it took when it varied. When the parameters varied, each of the N samples was drawn from a distribution with a different parameter value, selected randomly from a rectangular distribution of parameter values. The parameters a and b varied between 200 and 1,000; c varied between 1 and 4. The range for a and b were representative of the range of fitted values in the data sets reported in the text. The range for c represents the range of exponents over which the shape of the distribution varies from exponential to normal. These ranges produced distributions with means that ranged from 382 to 2,018 and standard deviations that ranged from 52 to 1,056. These values spanned the range of means and standard deviations in the actual data I fitted.

Altogether, eight simulations were performed. In the first simulation, all three Weibull parameters were fixed. In the next three, two parameters were fixed, and one was allowed to vary (e.g., a and b were fixed, and c was allowed to vary). In the next three, one parameter was fixed, and two were allowed to vary (e.g., a was fixed, and b and c were allowed to vary). In the eighth simulation, all three parameters were allowed to vary.

The means and standard deviations from each simulation are plotted in Figure A1. The points represent the simulated data, and the lines represent the power function fits. Measures of goodness of fit and the power-function exponents appear in Table A1. In each simulation, the fit was excellent. The goodness of fit was not affected much by variation in the Weibull parameters. When all three parameters were fixed, as the power-function proofs assume, $r^2 = .99997$, and $rmsd = 2.2$. When all three parameters varied randomly, violating the assumption of identical distributions, the fit remained excellent: r^2 decreased in the fourth decimal place (to .9995), and $rmsd$ increased by 2.7 (to 4.9). There were no systematic effects of varying the Weibull parameters on the power-function exponents. On average, the exponent was .438 (reciprocal = 2.282) when the Weibull c was fixed at 2.5; on average, the exponent was .480 (reciprocal = 2.083) when the Weibull c varied. This difference was not significant, $t(6) = .44$, $MS_e = .095$. Thus, the assumption of identical distributions does not seem to be crucial to producing power-function reductions in means and standard deviations nor does it seem to be crucial in producing the predicted relation between the exponent of the power function and the exponent of the Weibull, provided that the form of the distribution remains the same.

Table A1
Measures of Goodness of Fit for Power Function (Equation 1)
Fits to Means and Standard Deviations of Simulated Data

Fixed	Varied	r^2	$rmsd$	Exponent
a, b, c	none	.99997	2.18	.388
b, c	a	.99992	3.11	.621
a, c	b	.99978	3.73	.336
a, b	c	.99987	3.88	.446
a	b, c	.99967	4.26	.341
b	a, c	.99982	4.46	.681
c	a, b	.99978	3.37	.408
none	a, b, c	.99950	4.89	.453

Note. Means and standard deviations were fitted simultaneously. They were allowed to have separate scaling parameters but constrained to have the same exponent. The $rmsd$ = root-mean-square deviation between predicted and simulated values.

Appendix B:

Relations Between the Weibull and the Ex-Gaussian

I performed several Monte Carlo simulations comparing the shape of the Weibull distribution with the shape of distributions formed by the convolution of the normal distribution with the exponential distribution. The convolution is called the *ex-Gaussian* distribution. It is important because Ratcliff and Murdock (1976) showed that it provides accurate quantitative descriptions of reaction-time distributions, and Ratcliff (1978) showed that it provides an accurate description of the finishing-time distribution for certain continuous random walk (diffusion) models. To show that the Weibull is shaped like the ex-Gaussian is important because it means that the Weibull can also provide a good quantitative description of observed reaction-time distributions.

Mathematically, the ex-Gaussian distribution represents a two-stage process, in which a sample from a normal distribution is added to a sample from an exponential distribution. Repeated sampling leads to a distribution that is characterized by three parameters: μ , σ , and τ . The parameters μ and σ represent the mean and the standard deviation, respectively, of the normal distribution, and τ represents the reciprocal of the rate constant for the exponential distribution, which is also the mean and standard deviation of the exponential. The two-stage character of the ex-Gaussian is meaningful mathematically but not empirically. In general, data are not consistent with the idea that the normal distribution represents one processing stage and the exponential distribution represents another (Ratcliff & Murdock, 1976).

The shape of the Weibull was compared with the ex-Gaussian by generating quantiles from ex-Gaussian distributions with various combinations of parameters and fitting the Weibull to the generated distributions. Each simulation involved generating a distribution of 200 trials in which one sample was drawn from a normal distribution and another from an exponential distribution and the two were added together. The 200 data points in each distribution were rank ordered and reduced to 20 quantiles, representing successive increments of 5 percentile points, beginning with 2.5 (i.e., 2.5, 7.5, 12.5, etc.).

One set of simulations used two values of μ (200 ms and 400 ms), two values of σ (50 ms and 100 ms), and two values of τ (100 ms and 200 ms). The eight ex-Gaussian distributions formed by combining these parameters factorially were fit by three-parameter Weibull distributions (Equation 3) using STEPIT. Across the eight distributions, the mean r^2 between simulated and fitted points was .993, and the mean *rmsd* was 11.6 ms. The Weibull parameters were all affected by variation in the parameters of the ex-Gaussian, in ways that were readily interpretable: Increasing μ , the mean of the normal, increased b in the Weibull. Both of these parameters represent a kind of

intercept that moves the leading edge of the distribution away from zero. Increasing σ , the standard deviation of the normal, increased a in the Weibull, which makes sense because both σ and a affect the spread of the distribution. Increasing σ decreased b because it broadened the distribution and moved the leading edge closer to zero. Increasing σ also increased the exponent, c , of the Weibull. This follows because the ex-Gaussian is shaped more like the normal as σ increases, and the Weibull is shaped more like the normal as c increases. Finally, increasing τ decreased c in the Weibull. Increasing τ makes the ex-Gaussian more exponential in shape, whereas decreasing c makes the Weibull more exponential in shape.

Two simulations fitted special cases of the Ex-Gaussian. In one, τ was set to zero so that the ex-Gaussian would be normal, and in the other, μ and σ were set to zero so that the ex-Gaussian would be exponential. Both fits were good (mean $r^2 = .995$, and mean *rmsd* = 8.5). In the fit to the exponential, $c = 1.047$, which was close to the expected value of 1.0; in the fit to the normal, $c = 3.480$, which was close to the expected value of 3.6. These fits suggest that the Weibull can describe a range of reaction-time distributions that is similar to the range that the ex-Gaussian describes.

Finally, five simulations were performed using parameters from Ratcliff and Murdock (1976, Table 2 and Figure 14) and four simulations using parameters from Ratcliff (1978, Figure 7). The parameters from Ratcliff and Murdock were taken from fits to data, and those from Ratcliff were taken from fits to theoretical distributions generated by a continuous random walk (diffusion) process. The μ ranged from 89 ms to 500 ms, σ ranged from 22 ms to 37 ms, and τ ranged from 96 ms to 300 ms. The Weibull fits were quite good. The average $r^2 = .996$, and the average *rmsd* = 9.4 ms.

These analyses suggest that the Weibull may describe reaction-time distributions nearly as well as the ex-Gaussian does. This is encouraging because the instance theory predicts that well-practiced reaction times should conform to the Weibull and because theoretical calculations are easier if the Weibull can be used at all stages of practice. I do not mean to suggest that the Weibull should replace the ex-Gaussian as a way to describe reaction-time distributions. The ex-Gaussian is easier to work with. Its mean ($\mu + \tau$) and variance ($\sigma^2 + \tau^2$) are easier to compute than the mean ($a\Gamma(1 + 1/c) + b$) and variance ($a^2[\Gamma(1 + 2/c) - \Gamma(1 + 1/c)]^2$) of the Weibull.

Received April 1, 1991

Revision received November 7, 1991

Accepted December 12, 1991 ■

1993 APA Convention "Call for Programs"

The "Call for Programs" for the 1993 APA annual convention appears in the October issue of the APA Monitor. The 1993 convention will be held in Toronto, Ontario, Canada, from August 20 through August 24. Deadline for submission of program and presentation proposals is December 10, 1992. Additional copies of the "Call" are available from the APA Convention Office, effective in October. As a reminder, agreement to participate in the APA convention is now presumed to convey permission for the presentation to be audiotaped if selected for taping. Any speaker or participant who does not wish his or her presentation to be audiotaped must notify the person submitting the program either at the time the invitation is extended or prior to the December 10 deadline for proposal submission.