

# The Power Law repealed: The case for an Exponential Law of Practice

Andrew Heathcote<sup>1</sup>

Scott Brown<sup>1</sup>

&

D. J. K. Mewhort<sup>2</sup>

<sup>1</sup>Department of Psychology  
The University of Newcastle, Australia

<sup>2</sup>Department of Psychology  
The Queen's University, Canada

Address for Correspondence

Dr. Andrew Heathcote  
Department of Psychology,  
The University of Newcastle,  
Callaghan, 2308, Australia

Email: [heathcote@psychology.newcastle.edu.au](mailto:heathcote@psychology.newcastle.edu.au)

Phone (1-61) 249215952

Fax (1-61) 249216980

**Abstract**

The power function is treated as the law relating response time to practice trials. However, the evidence for a Power Law is flawed because it is based on averaged data. We report a survey that assessed the form of the practice function for individual learners and learning conditions in paradigms that have shaped theories of skill acquisition. We fit power and exponential functions to 40 sets of data representing 7910 learning series from 475 subjects in 24 experiments. The exponential function fit better than the power in all the unaveraged data sets. Averaging produced a bias in favour of the power function. A new practice function based on the exponential, the APEX function, fit better than a power function with an extra, pre-experimental practice, parameter. Clearly, the best candidate for the Law of Practice is the exponential or APEX function, not the generally accepted power function. Theoretical implications are discussed.

**The Power Law repealed: The case for an Exponential Law of Practice.**

"Curve fitting without benefit of a model is notoriously a black art."  
(Newell & Rosenbloom, 1981, p.23)

The benefits from practice follow a nonlinear function: improvement is rapid at first but decreases as the practitioner becomes more skilled (e.g., Thorndike, 1913). The idea that a simple nonlinear function might describe practice effects in a broad range of tasks was championed by Newell and Rosenbloom's (1981) influential chapter entitled Mechanisms of Skill Acquisition and the Law of Practice. The "Law of Practice" in the title concerns the relationship between response time (RT) and number of practice trials. Newell and Rosenbloom examined data from a wide range of tasks. When they compared power and exponential functions as possible forms for the Law of Practice, power functions provided better fits than exponential functions in every case.

The power function is now treated as the law of practice. In Anderson's (1982) words, "one aspect of skill acquisition ... distinguished ... by its ubiquity ... is the log-linear or power law for practice" (p. 397). A decade later, Logan (1992) echoed the same conviction: "The power law is ubiquitous. It occurs in virtually every speeded task" (p. 883). In accord with its status as a law, most research subsequent to the publication of Newell and Rosenbloom's (1981) findings has assumed a power function rather than testing to determine if it provides a better description than other functions (e.g., Cohen, Dunbar, & McClelland, 1990; Kramer, Strayer, & Buckley, 1990; Logan, 1988, 1992), or assumed that a power function holds for each component of performance (e.g., Delaney, Reder, Staszewski, & Ritter, 1998; Rickard, 1997).

The power function's status as a law has also made it a gold standard by which to judge the success of models of skilled performance, including: ACT and related models (Anderson, 1982; Anderson & Schooler, 1991), the Component Power Laws model (Rickard, 1997), network models (MacKay 1982; Cohen, et al., 1990), instance theories (Logan, 1988, 1992; Nosofsky &

Palmeri, 1997), and Newell and Rosenbloom's (1981) chunking model (see also Rosenbloom & Newell 1987a, 1987b). Logan (1988) leaves no doubt about the importance of the form of the practice function for theories of skill acquisition: "the power-function speedup [is] a benchmark prediction that theories of skill acquisition must make to be serious contenders." (p. 495; see also Cohen et al., 1990, and Palmeri, 1997, for similar views).

However, we contend that the evidence supporting a Power Law of Practice is flawed. Although theories of skill acquisition model learning in individuals, the bulk of the evidence favouring the Power Law is based on fits to averaged data. There is little empirical evidence from individual learners for individual learning conditions that a power function describes skill acquisition better than an exponential function. Data from all but one of the tasks examined by Newell and Rosenbloom (1981), for example, were averaged over subjects, conditions, or practice blocks. In the few published comparisons that report analysis of data from individual subjects, the exponential function fit better than the power function (Josephs, Silvera, & Giesler, 1996; Rosenbloom & Newell, 1987a).

The mismatch between theory and evidence is more than a minor technicality: It has been known to the psychological literature for almost 50 years that average curves need not take the same form as the individual curves comprising the average (e.g., Estes, 1956; Kling, 1971; Sidman, 1952). Hence, the form of the average practice function does not unambiguously indicate the form of the components of the average. Moreover, recent work shows that linear averaging yields a composite that is systematically biased towards the power function when compared with the exponential function (e.g., Anderson & Tweney, 1997; Myung, Kim, & Pitt, 1998). Hence, evidence once thought to favour the Power Law may be artefactual.

The form of the practice law may also seem to be an unsolvable technical issue rather than an important psychological question. Estes (1997, personal communication) has indicated that the form of practice law does not constrain theory enough. In his words, a "generation of budding learning theorists (Bower, Bush, Estes, Greeno, Hunt, Restle) produced mountains of

analyses showing how easily the forms of particular performance curves can be mimicked by many alternative models”.

Nevertheless, the form of the practice law does carry an important implication about the nature of learning. As we will note more fully later, an exponential function implies a constant learning rate relative to the amount left to be learned. By contrast, the power function implies “a learning process in which some mechanism is slowing down the rate of learning” (Newell & Rosenbloom, 1981, p. 18). The question at issue, then, is whether the slowing implied by the power function is part of skill acquisition. Repealing the Power Law of Practice in favor of an exponential (or other) law has serious implications for all theories of skill acquisition – particularly those developed in order to account for the Power Law.

In light of the ambiguity concerning empirical support for the Power Law, we report the results of a survey that systematically assessed the form of the practice function for individual learners and learning conditions in paradigms that have shaped theories of skill acquisition (see Table 1 for a summary of the paradigms and data sets). In the next section we review the properties of candidate practice functions and propose a new practice function, the APEX function, that expedites our analysis. Following sections describe our methods and results. Results can be easily apprehended from Table 2 (which tabulates the average fit of the candidate practice functions in each data set) and Figures 1, 3, and 4 (which present the percentage of cases in each data set that are best fit by a candidate practice function). Finally, the discussion section examines the implications of our results for the measurement of practice functions and for theories of skill acquisition.

### Practice Functions

Equations 1 and 2 are the power and exponential functions used by Newell and Rosenbloom (1981) to fit practice data.  $E(\mathbf{RT}_N)$  is the expected value of  $\mathbf{RT}$  on practice trial  $\underline{N}$ . The bold notation indicates that  $\mathbf{RT}$  is a random variable. When referring to observed response

times (i.e., to samples from **RT**), we will use the notation  $RT$ .  $\underline{A}_P$  and  $\underline{A}_E$  are the expected values of **RT** after learning has been completed for the power and exponential functions, respectively. An asymptote parameter is necessary when modeling response time, even in a highly skilled subject, because performance is limited by physical constraints, such as neural integration time and motor response time.  $\underline{B}_P$  and  $\underline{B}_E$  are the change in the expected value of **RT** from the beginning of learning ( $\underline{N} + \underline{E} = 1$  for the power function, or  $\underline{N} = 0$  for the exponential function) to the end of learning (the asymptote). Hence,  $\underline{B}_P$  and  $\underline{B}_E$  indicate the range over which practice speeds responding.

$$E(\mathbf{RT}_N)_P = A_P + B_P (N + E)^{-b} \quad (1)$$

$$E(\mathbf{RT}_N)_E = A_E + B_E e^{-aN} \quad (2)$$

The amount of nonlinearity displayed by the practice function is controlled by its rate parameter:  $\alpha$  for the exponential and  $\beta$  for the power. The power function has one extra nonlinear parameter:  $\underline{E}$ . It represents the subject's prior learning due to practice before experimental measurement (the rationale for the  $\underline{E}$  parameter was first suggested by Seibel, 1963). Most subsequent researchers, however, have fit the simpler three-parameter version of the power function with  $\underline{E}$  fixed at zero (e.g., Logan, 1988, 1992).

An extension of the exponential function to include a parameter corresponding to  $\underline{E}$  is redundant because the exponential function is translation invariant, i.e.,

$$B'_E e^{-a(N+E)} = B'_E e^{-aE} e^{-aN} = B_E e^{-aN}$$

where  $B_E = B'_E e^{-aE}$ . Hence, the effect of prior practice is incorporated into the estimate of  $\underline{B}_E$  for the three-parameter exponential function.

Newell and Rosenbloom's (1981) evidence for a Power Law of Practice was based on a comparison of the fit of the three-parameter exponential function with the fit of the four-

parameter power function, which they call the general power function. We will adopt the terms “general power function” and “power function” when referring to the four- and three-parameter ( $\underline{E}$  fixed at zero) versions of Equation 1, respectively.

Apart from the difficulties introduced by averaging, Newell and Rosenbloom’s (1981) analysis is open to criticism on two technical grounds:

1. The first technical criticism concerns the number of parameters in the equations considered by Newell and Rosenbloom (1981). They compared the three-parameter exponential function against the four-parameter general power function. One might think that the latter function would be more flexible as it has an extra parameter. In particular, the extra flexibility may allow a general power function to mimic an exponential function, making the general power function almost impossible to falsify by a comparison with the simpler exponential.

The relationship between the general power and exponential functions is illuminated by expressing them as differential equations.

$$\frac{\mathcal{I}(E(\mathbf{RT}_N)_P)}{\mathcal{I}N} = \frac{-b}{N+E} (E(\mathbf{RT}_N) - A_P) \quad (3)$$

$$\frac{\mathcal{I}(E(\mathbf{RT}_N)_E)}{\mathcal{I}N} = -a (E(\mathbf{RT}_N) - A_E) \quad (4)$$

Equations 3 and 4 can be compared using their relative learning rate (RLR), defined as minus the rate of change of expected  $\mathbf{RT}$  divided by the amount left to be learned (i.e., relative learning rate equals the multipliers of the bracketed terms on the right side of Equations 3 and 4). The defining characteristic of an exponential function is a constant relative learning rate ( $\alpha$ ) at all levels of practice. For the general power function, by contrast, relative learning rate is a hyperbolically decreasing function of practice trials ( $\beta/(\underline{N}+\underline{E})$ ).

Strictly, the general power function can exactly mimic an exponential function only when  $\alpha = 0$  (i.e., a flat function). However, for practice series of finite length, large values of  $\underline{E}$  can make the relative learning rate almost constant, and any value of  $\alpha$  can be approximated,

particularly smaller values. In experimental measurement, practice trials vary over a limited range. If the estimate of  $\underline{E}$  is much larger than the range, the effective relative learning rate for the general power function is a constant, approximately equal to  $\beta/\underline{E}$ . Consequently, the general power function can mimic exponential data using large estimates of  $\underline{E}$ , although such fits are invariably poorly behaved because the parameter estimates are highly correlated, i.e., large estimates of  $\underline{E}$  are associated with large estimates of  $\beta$  and, particularly, of  $\underline{B}_P$ .

Newell and Rosenbloom (1981) noted the association between large estimates of  $\underline{E}$  and large estimates of  $\underline{B}_P$  when they fit the general power function to simulated exponential data. We reported the same behavior in fits to several data sets from a visual-search paradigm (Heathcote, 1990; Heathcote & Mewhort, 1995). When  $\underline{E}$  is large, large values of  $\beta$  will occur to allow the general power function to approximate a constant relative learning rate greater than zero. Very large values of  $\underline{B}_P$  allow the general power function to approximate the decrease in the expected value of  $\mathbf{RT}$  from the beginning to the end of measurement. This counteracts the tendency of a general power function to become flat when it mimics exponential data.

2. The second technical criticism concerns the way in which Newell and Rosenbloom (1981) fit practice functions. To save computation, they fit by minimizing squared deviations in  $\log(\mathbf{RT} - \hat{A}_P)$  and  $\log(\mathbf{RT} - \hat{A}_E)$ . Given that  $\underline{A}_P$  and  $\underline{A}_E$  estimate expected values it would not be unusual to observe samples from  $\mathbf{RT}$  which are less than these expected values; in such cases, the measures  $\log(\mathbf{RT} - \hat{A}_P)$  and  $\log(\mathbf{RT} - \hat{A}_E)$  are undefined. It is difficult to know how to deal with undefined values without biasing or distorting fits (Newell and Rosenbloom did not describe their approach). Undefined values will not occur, of course, if near-asymptotic performance is not measured or if the variability of  $\mathbf{RT}$  shrinks to zero with practice. The latter condition is unlikely as, even in very fast and simple tasks, response time remains variable (see Luce, 1986, for numerous examples). Such problems are likely to be more pronounced in individual subject data as it is noisier than averaged data.



To check whether the fitting method biased their results, we obtained a subset of the data in Newell and Rosenbloom's (1981) survey. We refit their data by minimising squared deviations in RT (i.e., the generally accepted method of ordinary least squares) instead of  $\log(\text{RT} - \hat{A}_p)$  and  $\log(\text{RT} - \hat{A}_E)$ . For the averaged data, the power function still fit better than the exponential function. However, for the two data series that were not averaged (times to win and to loose the Stair card game), the exponential function provided a better fit than the power<sup>1</sup>. In other words, the only evidence favouring the power function for unaveraged data in Newell and Rosenbloom's survey turns on the adequacy of their fitting method.

The two technical criticisms not only call Newell and Rosenbloom's (1981) findings into question<sup>2</sup> but also raise a dilemma. The general power function is clearly a plausible extension of

---

<sup>1</sup> For games won  $R^2$  for the Exponential was 0.339 and only 0.253 for the power. For games lost  $R^2$  was 0.183 and 0.173 for the exponential and power functions respectively. The fit of the general power function was somewhat better ( $R^2$  of 0.334 and 0.185) but it was still less than the exponential function for games won and less than the APEX function (with  $R^2$  of 0.339 and 0.186 respectively) in both cases. We would especially like to thank Paul Rosenbloom for sending us these data sets.

<sup>2</sup> Analysis of relative learning rates also shows that Mazur and Hastie's (1978) results, which Newell and Rosenbloom (1981) claim agree with their results in "rejecting exponentials" (p. 34), are not relevant for response time. Mazur and Hastie (1978) fit power and exponential functions to rate of response data. A nonlinear transform, the inverse, is required to convert rate to RT. The transform changes the relative learning rate of the exponential rate function to a relative learning rate on the RT scale ( $K/(e^{kN}-1)$ , where  $K$  and  $k$  are parameters greater than zero) that decreases more quickly than the power function's relative learning rate on the RT scale. Hence, Mazur and Hastie's (1978) comparison does not test a true exponential function on the RT scale.

power function, at least when estimates of  $\underline{E}$  are reasonable, but it cannot be fairly compared directly to the exponential function. To escape the dilemma, we propose a new<sup>3</sup> four-parameter practice function, the APEX function that nests (i.e., contains as special cases) both the power and exponential functions.

$$E(\mathbf{RT}_N) = A + B e^{-a/N} N^{-b'} \quad (5)$$

The APEX function has a relative learning rate that is the sum of the relative learning rates for the power and exponential functions:  $\alpha' + \beta'/N$ . Consequently, its relative learning rate decreases like a power function early in practice but approaches an asymptotic value potentially greater than zero ( $\alpha'$ ) later in practice.

We propose the APEX function for two reasons:

1. The APEX function nests both the power and exponential functions. Hence, fitting the APEX function to exponential data will likely be better behaved than fitting the general power function to exponential data. Because the APEX function can exactly fit both power and exponential data, it can be used to adjudicate between the two alternatives: If the APEX function provides a better fit than the power function, an exponential component is supported. If the

---

<sup>3</sup> A related function was proposed by Wickelgren (1975, p. 326) to model retention of memories. Note that a prior-practice parameter analogous to that in the general power function could also be added to the APEX function (i.e., substitute  $(N+E)$  for  $N$ ). This five-parameter function nests both the APEX and general power functions. It was not used because the general power function was over-parameterised for the practice data sets examined and often provided ill-conditioned objective functions that made minimisation difficult. Hence, the five-parameter version of the APEX function could only be more problematic.

APEX function provides a better fit than the exponential function, a power component is supported. If both conditions apply, the full APEX function is supported.

Tests comparing the fit of APEX, power and exponential functions can be applied to individual subjects and conditions. This allows for the possibility that the form of the practice function might differ between individuals or conditions. The significance of any improvement in fit can be assessed for each individual subject and condition by a straightforward nested-model change-of- $R^2$  test. Nested-model tests have an important advantage when dealing with nonlinear models; they need assume only a linear approximation to the intrinsic curvature of the function and are not effected by nonlinear effects of changes in parameters (see Bates & Watts, 1988).

2. Comparison of the fit of APEX and general power functions not only provides a fair test but also illuminates a theoretically crucial point. Many models of skill acquisition acknowledge that performance may be the result of a sum of component processes (e.g., Kirsner & Spelman, 1996) or the result of a mixture of processes, such as algorithmic and memory-based processing (e.g., Logan, 1988, 1992; Rickard, 1997).

In the case of a sum, processes with a relatively faster learning rate will approach their asymptote quickly and then will cease to affect the rate of change of RT; as a result, asymptotic learning will be dominated by the slower learning processes. The transition should be evident as a decrease in the relative learning rate early in practice, even if all component processes are exponential (i.e., even if all have constant relative learning rates). Asymptotic learning rates, however, will reveal the true relative learning rate function of (at least) the slow component processes. A superior fit for the APEX function supports an asymptotic relative learning rate greater than zero, a finding inconsistent with component power functions. A superior fit for the general power function supports an asymptotic relative learning rate that approaches zero, a finding inconsistent with component exponential functions.

A similar argument can be made for a mixture of memory and algorithmic processes. Even if each component learns exponentially, relative learning rate can decrease early in practice because of the transition from trials controlled by slow algorithmic processing to trials controlled by fast memory-based processing. After sufficient practice, however, learning should be dominated by the memory-based process. If memory-based processing follows a power function, the asymptotic relative learning rate will approach zero, and the general power function will provide a superior fit to the APEX function. If memory-based processing follows an exponential function, the APEX function will provide the superior fit with an estimate of  $\alpha$  greater than zero.

Several experiments in the survey required subjects to report their processing strategy as algorithmic or memory-based. Hence, we were also able to test the function describing each component explicitly in these experiments (assuming, of course, that the subjects' self report accurately describes their processing strategy).

## Method

### Analysis Techniques

We fit power, exponential, general power, and APEX functions using ordinary least-squares minimization on correct trial data. The fits were obtained using a Simplex search algorithm (Press, Flannery, Teukolsky, & Vetterling, 1986) because it is robust when fitting is ill conditioned. Fits were validated by comparison to the outputs of nonlinear regression programs provided in the SPSS and S+ statistical packages.

When data are exponential, fits of the general power function tend to be ill conditioned. In particular, fitting is problematic because estimates of  $\underline{B}_p$  diverge to very large values that cannot be represented with accuracy, even in double precision. To avoid the problem, estimates of  $\underline{B}_p$  were constrained to be less than  $10^{10}$  ms (around 115 days). Estimates of the asymptote parameters were constrained to be greater than zero for all fits in order to ensure that the estimates were plausible.

To ensure that the best fits of general power function were found, multiple fits were performed with starting estimates for  $\underline{E}$  equal to 1%, 10%, 50% and 100% of the length of the series. The best fit was then selected and compared to the fit of the power function for the same series. Where the general power function fit was worse than the power function fit, as sometimes happened due to correlated parameters, the power fit was substituted for the general power fit. Ill conditioned power, exponential, and APEX fits almost never occurred, and starting points for fitting were easy to obtain by heuristics. Where the fit of an APEX function was worse than either the fit of the power or exponential for the same series, refits were performed using starting values close to both the exponential and power solutions. The best fit among the refits and the power and exponential fits was selected.

Experimenters often censor their data to delete outliers. Whenever the experimenter censored their data, we followed the same procedure. If an experimenter did not report censoring, we removed obvious outliers (see Table 1 for the criteria used). Across all the data sets, however, very few observations were censored.

Because RT can vary as a function of accuracy, we also calculated the main effect of practice on errors for each block of trials. Most frequently, errors decreased with practice, but they also increased, or remain unaffected by practice in some data sets. Hence, the results for RT are not correlated in any simple pattern to changes in accuracy, at least across different data sets.

All fits used learning series broken down by subjects and by within-subject factors or learning examples. Where strategy reports were available, the learning series were also broken down by strategy for supplementary analyses. In some cases, data sets broken down by strategy produced series that were too short to obtain reliable fits. Such data sets were excluded. The data for production of key sequences were divided into trials on the first and subsequent days because new instructions that clearly influenced learning were given on the second day of practice (see Verwey, 1996, Figure 3, p. 548). Exponential, power, APEX and general power functions were then fit to each learning series.

Several approaches were used to compare the overall performance of the different functions. The proportion of learning series for which an exponential function provided a better fit than a power function (on the basis of  $R^2$ ) was tallied for each data set (see Figure 1). A similar comparison was made between the four-parameter functions (APEX and general power, see Figure 3). Binomial confidence intervals were used to determine if preference for either member of each pair of functions was significant in each data set. For all functions the average value of  $R^2$  for each data set was calculated and compared (see Table 2). Additionally, the improvement in fit of the APEX function relative to both exponential and power functions was used to provide nested-model tests. Table 2 reports the percentage of learning series for each data set where significant results for the nested model tests supported an exponential and/or power component.

#### Data Sets

We fit 40 sets of data; collectively, the data represent 7910 learning series from 475 subjects in 24 experiments taken from 13 published and 3 unpublished sources. The unpublished data (Brown & Heathcote, 1997; Ringland & Heathcote, 1998; Smith & Mewhort, 1994) were collected in our laboratories and were analyzed to clarify and expand on results from published data sets.

Table 1 summarizes the characteristics of the data sets used in the survey. Each data set was given a unique acronym used to index the summaries of results. For data sets that were broken down by strategy, the Length column indicates the criterion length used to exclude short series. Table 1 also reports the results of tests on the effect of practice on error rates. The following sections describe both the paradigms from which the data were drawn and the experimental factors used to produce separate series for each data set.

Memory Search. In the memory-search tasks, subjects studied a list of words and then were asked to indicate whether or not a probe word had appeared in the study list. The words used in the list were selected to represent particular semantic categories, and semantic category

was mapped consistently to the word's use as a target or distracter item. Fits used series broken down by subjects and within-subject factors.

Table 1. Summary of data sets fit in the survey.

Source	Name	N <sup>a</sup>	Length <sup>b</sup>	Ss. <sup>c</sup>	Errors <sup>d</sup>	Censor (ms) <sup>e</sup>
Strayer & Kramer (1994a, 1994b)	<b>MS1</b>	36	606-711	6	Increase*	150<RT<1500
	<b>MS2</b>	192	654-716	32	Increase	150<RT<1500
Strayer & Kramer (1994c)	<b>MS3</b>	132	625-860	22	Decrease	150<RT<1500
Palmeri (1997)	<b>Count1</b>	120	175-208	4	Decrease*	None
	<b>Count2</b>	288	125-160	4	Decrease*	None
	<b>Count3</b>	360	129-160	5	Decrease*	None
Rickard & Bourne (1996)	<b>Math1</b>	384	63-90	24	Decrease*	200<RT<5000
Rickard (1997)	<b>Math2</b>	228	62-90	24	Decrease*	None
	<b>Math2a</b>	125	> 9	-	-	None
	<b>Math2m</b>	228	> 9	-	-	None
Reder & Ritter (1992)	<b>Math3</b>	79	8-20	20	NA	None
	<b>Math3a</b>	44	> 7	-	-	None
	<b>Math3m</b>	38	> 7	-	-	None
	<b>Math4</b>	63	8-20	16	NA	200<RT<18000
	<b>Math4a</b>	50	> 7	-	-	200<RT<18000
	<b>Math4m</b>	14	> 7	-	-	200<RT<18000
Schunn et al. (1997)	<b>Math5</b>	65	8-28	22	NA	None
	<b>Math5a</b>	57	> 7	-	-	None
	<b>Math5m</b>	35	> 7	-	-	None
Rickard (1997)	<b>AA1</b>	504	25-84	21	Decrease	200<RT<10000
	<b>AA1a</b>	157	> 9	-	-	200<RT<10000
	<b>AA1m</b>	489	> 9	-	-	200<RT<10000
Smith & Mewhort (1994)	<b>AA2</b>	288	80	24	Decrease*	None
Heathcote & Mewhort (1993)	<b>VS1</b>	192	200	24	Decrease*	None
Carrasco et al. (1998)	<b>VS2</b>	120	63-88	10	Increase	None
Heathcote & Mewhort (1993)	<b>VS3</b>	128	160	8	Decrease	None
Verwey (1996)	<b>Key1t</b>	72	45-613	36	Increase*	None
	<b>Key1c</b>	180	45-613	-	-	None
	<b>Key1k</b>	648	45-613	-	-	None
	<b>Key2t</b>	72	67-1353	36	Flat	None
	<b>Key2c</b>	180	67-1353	-	-	None
	<b>Key2k</b>	648	67-1353	-	-	None
Brown & Heathcote (1997)	<b>Key3</b>	56	228-300	4	NA	200<RT<10000
Anderson et al. (1997)	<b>Rule1</b>	26	32	26	Decrease*	None
	<b>Rule2</b>	88	32	22	Decrease	None
	<b>Rule3</b>	180	32	45	Decrease*	None
Kail & Park (1990)	<b>MR1c</b>	96	35	8	NA	None
	<b>MR1a</b>	96	35	8	NA	None
Ringland & Heathcote (1998)	<b>MR2c</b>	576	18-32	12	Decrease	None
	<b>MR2a</b>	576	18-32	12	Decrease	None

<sup>a</sup>The number of practice series in the data set.

<sup>b</sup>The range of lengths for practice series in each data set.

<sup>c</sup>The number of subjects in each data set.

<sup>d</sup>The results of tests on the main effect of practice on accuracy, with a \* indicating significance at the 95% confidence level. Where results were not significant trends are indicated.

<sup>e</sup>Criteria used, if any, to censor outliers from the data set.

The data in set MS1 are the consistently mapped trials from mixed consistent/variable mapping training blocks from Experiment 2 of Strayer and Kramer (1994a). The data indexed by the label MS2 refer to the consistently mapped training blocks from Experiment 2 of Strayer and Kramer (1994a), Experiments 4, 6, and 7 from Strayer and Kramer (1994b), and to an unpublished two-alternative forced-choice version of the task. The experiments in the MS2 data set came from very similar paradigms, and individually produced the same pattern of results as the overall data set, so were grouped together. For both MS1 and MS2, two factors were manipulated within-subject: target/distracter probe and memory load (2, 4, or 6 items).

The data in set MS3 are consistently mapped trials from young subjects (ages 18-21) from Strayer and Kramer (1994c). Two factors were manipulated within subject: target/distracter probe and memory load (2, 4, or 6 items), and one factor was manipulated between subjects: speed versus accuracy instructions.

Counting. In the counting tasks, subjects were shown different patterns of six to 11 dots and a spelled-out number; they were asked verify whether the number of dots in the pattern matched the spelled-out number. All data were taken from Palmeri (1997). Each experiment used a number of unique patterns, and fits included series from each pattern. Fits used data broken down by subjects and dot pattern.

The data in set Count1 are the training series from Experiment 1. The number of dots was manipulated within-subjects. There were 30 patterns with 5 patterns of array size.

The data in set Count2 are the training series from Experiment 2. The number of dots and the similarity of dot patterns (no, low and moderate) was manipulated within-subjects. There were 72 patterns with 4 patterns for each level of similarity per array size.

The data in set Count3 are the training series from Experiment 3. The number of the dots and similarity (similar to a same or different number pattern) was manipulated within-subjects. There were 72 patterns with 6 patterns for each level of similarity at each array size.



Mental arithmetic. The mental arithmetic tasks included a diverse set of problem types. Fits used data broken down by subjects and problem examples.

The data in set Math1 are from a single-digit multiplication task taken from Experiment 1 of Rickard and Bourne (1996). Either the subjects were shown 2 digits and asked to calculate the product, or they were shown a digit and a product and asked to divide the product to compute the dividend. RT was recorded as the time between the presentation of the problem and the keystroke of the first digit of the answer. There were 16 problem examples. Problem type (compute product or compute dividend) and range of digits was manipulated within-subjects.

The data in set Math2 are from a three-step arithmetic task (Experiment 1 of Rickard, 1997). Subjects were shown two numbers and asked to calculate their difference, to add 1 to the result, and, then, replacing one of the numbers with the result so far, to compute the sum of it with the remaining original number. RT was recorded as the time between the presentation of the problem and the keystroke of the first digit of the answer. Subjects reported using one of two strategies, recalling the answer from memory or computing the answer (algorithm) on every third trial. We split the data into sets defined by the subject's strategy using the logistic method described by Rickard (1997)<sup>4</sup> and fit series with more than 9 trials.

The data in set Math3 are from two-digit multiplication and addition tasks reported as Experiment 1 by Reder and Ritter (1992). Subjects were shown two two-digit numbers and asked first to indicate which of two strategies, recall (which we label "memory") or calculate (which we label "algorithm"), they intended to use. They then answered the problem. We fit only to data from the four problems (two addition and two multiplication) that were presented 20 times, and excluded one series because it had less than eight correct answers. Also, we fit data for each strategy separately, again excluding the short series.

---

<sup>4</sup> We also used a number of simpler schemes for partitioning trials with the similar results.

The data in set Math4 are from a multi-step multiplication task with an initial rapid strategy report as for the Math3 (Experiment 2 by Reder & Ritter, 1992). We fit only data from the four problems that were presented 20 times, and excluded one series that had less than eight correct answers. The data were also classified by strategy and fit separately, again excluding the short series.

The data in set Math5 are from a two-digit task combining multiplication and addition, preceded by a rapid strategy report, as for Math3 (Experiment 1 by Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997). We fit only the three problems presented 28 times and, again, excluded one series with length less than 8. The data were also classified by strategy and fit separately, again excluding the short series.

Alphabetic arithmetic. Subjects were required to verify equations of the form  $A + 2 = C$ , or  $A + 3 = C$ , true and false equations respectively. We broke down the data by subjects and by problem example.

The data in set AA1 are from Experiment 2 of Rickard (1997). Two factors were manipulated within subjects: addend (3, 5, and 7) and trial type (true/false), with four examples of each type. Data for each of the 24 problems were fit separately. Subjects reported strategy as for the Math2 data set. We also analyzed data for the two strategies separately, excluding series with less than 10 responses.

The data in set AA2 are from an unpublished experiment by Smith and Mewhort (1994). Three factors were manipulated within subjects: addend (2, 3, 4), arithmetic operator (+, -) and trial type (true/false), with one example of each type. We fit the data from each of the 12 examples separately.

Visual search. In the visual search tasks, subjects were required to indicate whether or not a target appeared in a visual display. In VS1 and VS3, the target was defined by the relative position of two features; in VS2, the target was defined by a conjunction of colours. Stimuli used for targets and distracters were consistently mapped over trials in VS1 and VS2. Targets and

distracters were variably mapped in VS3, and a target cue was given before each trial. Fits used data broken down by subjects and by within-subject factors.

Data in set VS1 are from Experiment 1 of Heathcote and Mewhort (1993). Two factors were manipulated between-subjects: feature type (brightness or colour) and display area (small or large). Two factors were manipulated within subjects: display size (2, 4, 6 or 8 objects) and trial type (target/distracter).

The data in set VS2 are from Experiment 3 of Carrasco, Ponte, Rechea, and Sampedro (1998). Two factors were manipulated within subject: display size (2, 6, or 10, 14, 18, 22 objects) and trial type (target/distracter).

The data in set VS3 are from Experiment 3 of Heathcote and Mewhort (1993). Three factors were manipulated within subject: display size (2, 4, 6 or 8 objects), target type, and trial type (target/distracter).

Motor learning. In the motor learning tasks, subjects were required to press combinations or sequences of keys in response to compatible stimulus displays. All fits were broken down by subjects. Fits for production of combinations of key presses used series broken down by combination. Fits to sequence production used data broken down by within-subject factors but not sequence, as each subject used the same sequence.

The data in sets Key1t, Key1c, Key1k, Key2t, Key2c and Key2k, were taken from Verwey (1996). Subjects executed the same 9-key sequence in response to a compatible display. Only the data where all 9 key responses were correct were analyzed. On most blocks of trials, the key sequences were divided into segments (calls chunks) by pauses between stimulus onset. The segment structure was manipulated between subjects: The 9 keystrokes were divided into 3 equal segments (3:3:3) or into 2 unequal segments (3:6). For the remaining blocks there was no pause between segments. Otherwise the next stimulus occurred immediately after the preceding key press. Subjects were instructed at the beginning of the second day of practice to use the pauses to group their responses temporally.

Trials from Day 1 and from later days, and trials from blocks with and without pauses, were fit separately. Fits were performed for the time to complete each key press (Key1k and Key2k), to complete each chunk (Key1c and Key2c), and to complete the total sequence (Key1t and Key2t).

The data in set Key3 were taken from an unpublished experiment by Brown and Heathcote (1997). Subjects pressed combinations of 1 to 3 keys from a set of four keys in response to a compatible visual display (bright rectangle presented on a screen above the response keys). All 14 possible combinations were practiced in random order.

Learning rules from examples. The data were taken from the three experiments reported by Anderson, Fincham, and Douglass (1997). Subjects studied examples of the form “Skydiving was practiced on Saturday at 5 p.m. and Monday at 4 p.m.”, with the underlying rule being that the second practice occurred 2 days later and one hour earlier. Subjects then indicated the missing parts of similar examples by clicking on one of a set of choices using a mouse.

The data in sets Rule1, Rule2, and Rule3 are from Experiments 1, 2, and 3 respectively. All experiments started with blocks using 8 different rules and required the same part (first or second) to be filled in for each rule. Over groups of blocks, examples were introduced that required the other part to be filled in. The manipulation defined the within-subject factor practiced/unpracticed rule direction. The data for the unpracticed rule direction came from the rules in each group of blocks that had not previously been seen in the unpracticed direction. Experiment 2 introduced a second within-subject factor: Rules could either have unique or repeated examples. Experiment 3 made the example repetition a between-subjects factor, with either 0, 1, or 2 repeated examples per block.

For fitting, the data were broken down by all within-subject factors. However, data were averaged over rules, and over examples of rules, for groups of four, eight and eight blocks for Experiments 1, 2, and 3 respectively.

Mental Rotation. Subjects were presented with one of four letters (F, G, P, R) either in their normal or mirror-image form and rotated by 0, 30, 60, 90, 120, or 150 degrees. The subjects' task was to indicate whether the letter was normal or a mirror-image form. Letter type, letter orientation and normal- versus mirror-image were manipulated within-subject. Participant age (child or adult) was a between-subjects factor.

Data in sets MR1c and MR1a, child and adult subjects, respectively, were taken from Kail and Park (1990). The data were averaged over letter and block (two examples per block); otherwise, we broke down the series by within-subject factors. Data in sets MR2c and MR2a were from child and adult subjects, respectively, and came from an unpublished replication of Kail and Park (Ringland & Heathcote, 1998). The latter data were not averaged: Letter type was used to break down the series, and we fit individual trial data rather than block averages.

## Results

Note that where we report a statistic averaged over data sets, the average was calculated weighted by the number of series in each data set (see Table 1 for the number of series in each data set). Table 2 and Figures 1, 3, and 4 give statistics for individual data sets.

### The Shape of the Practice Function

Comparing power and exponential functions. Figure 1 reports the percentage of series in each published<sup>5</sup> and unaveraged data set that were better fit by the exponential function than by the power. Power and exponential functions provided an equally good fit in 2.5% of series, and these series were excluded from the calculation of the percentages shown in Figure 1. Of the remaining series, the exponential function provided a better fit than the power in 82.2% of cases, ranging from a minimum of 64% for the MS2 data set to a maximum of 93% for the Count3 data

---

<sup>5</sup> The single exception is the unpublished two alternative forced choice experiment included in the MS2 data set. These data were included, as they are very similar to all others in MS2.

set. In every case, we could reject, at the 95% confidence level, the null hypothesis that power and exponential functions were equally likely to win.

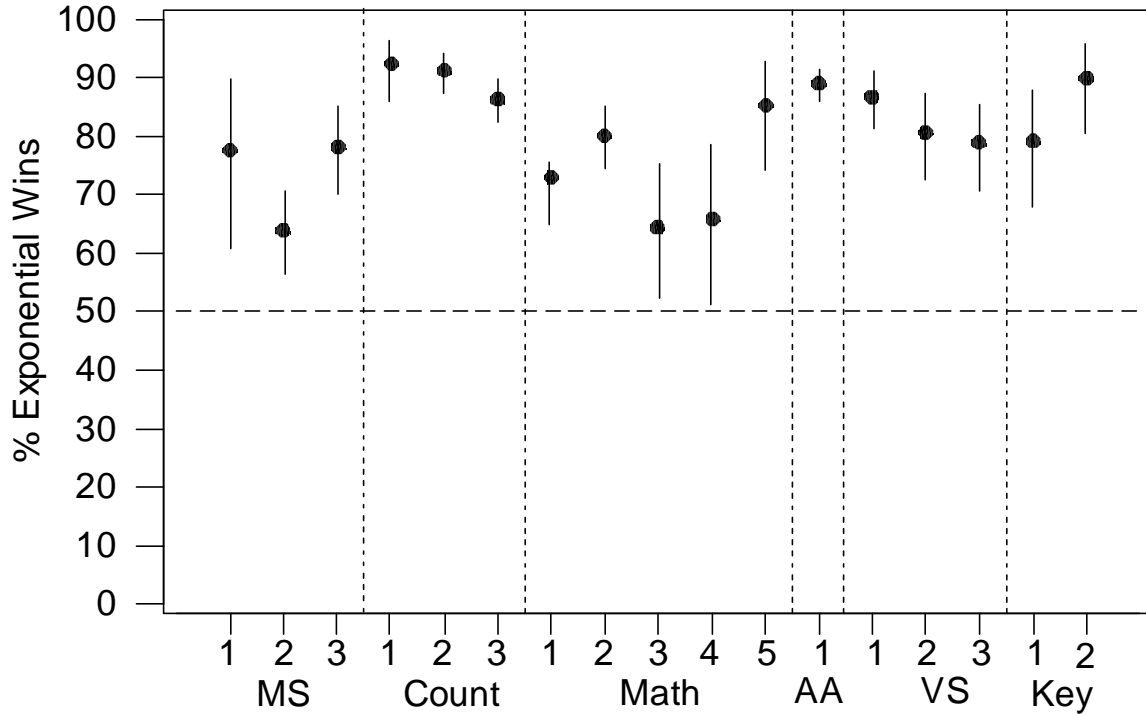


Figure 1. Percentages of cases where the exponential function provided a better fit than the power function (solid circles) and exact binomial 95% confidence intervals for the published and unaveraged data sets in the survey.

Table 2 gives the average across all learning series of the proportion of variance ( $R^2$ ) accounted for by the power and exponential functions. In every data set, the exponential function accounted for more variance than the power function. Overall, the average  $R^2$  was 0.498 for the exponential and 0.426 for the power. The exponential fits provided, on average, a 17% improvement relative to the power function, ranging from 3.7% for the Key1t data set to 28.6% for the Count2 data set. In many cases, the absolute increase in  $R^2$  was quite large; that is, the improvement in fit provided by the exponential was not trivial. Figure 2 reports one such case. Note that the power function’s decreasing relative learning rate forces it to approach asymptote very quickly at first and then very slowly, so that it briefly over estimates then under estimates data in early practice trials. For later practice trials, it over estimates the data. Beyond the range of

practice trials measured, it makes a clearly implausible under estimate, that response time reduces to zero.

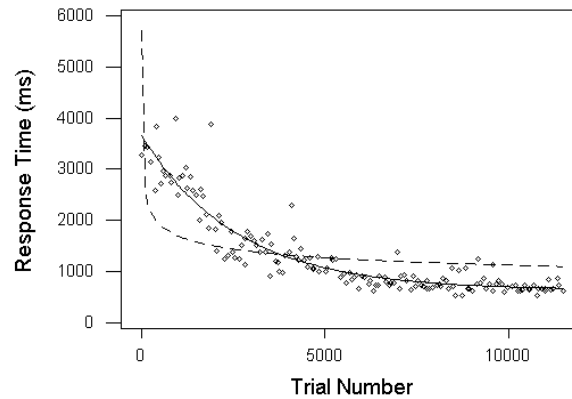


Figure 2. A learning series from the Count3 data set (subject number 2, stimulus number 39, “enemies” trials). Also shown are the best fitting exponential (solid line,  $\underline{A}_E = 840.56$ ,  $\underline{B}_E = 3800$ ,  $\alpha = 0.00142$ ,  $R^2 = 0.576$ ) and power (dashed line,  $\underline{A}_P = 0.00$ ,  $\underline{B}_P = 17037$ ,  $\beta = 0.33922$ ,  $R^2 = 0.472$ ) functions.

Sometimes, however, the advantage for the exponential over the power was small, and, in most of such cases, the  $R^2$  values themselves were also small. Interpretation of small  $R^2$  values for nonlinear regression is tricky, however. The expected value of the nonlinear  $R^2$  can depend on the length of the series. Specifically, the nonlinear  $R^2$  should decrease as series length increases because extra trials at asymptote add noise but do not add additional signal to the correlation (here, signal means a change in the expected value due to learning). In the memory-search experiments, for example, the practice series were very long, and the absolute values of  $R^2$  were relatively small. Nevertheless, learning, especially early in practice, was quite strong.

Over all data sets, the difference between  $R^2$  for the power and exponential functions was highly correlated with their average value ( $r = 0.752$ ), indicating that the magnitude of the difference was an increasing function of the variance accounted for by learning. Consequently, it appears that smaller advantages for the exponential function are associated with higher levels of noise, rather than being due to a systematic difference between data sets. Taken together with the

cautionary note, the correlation shows that the relative increase in  $R^2$  afforded by the exponential function was important in all data sets.

Table 2. Results for published and unaveraged data sets. E, P, APEX, and GP refer to the exponential, power, APEX and general power functions respectively. Significance is assessed at the 95% confidence level.

Data Set	$R^2$				% $\hat{A} < 150\text{ms}$				% Significant			
	E	P	APEX	GP	E	P	APEX	GP	Overall	$\hat{a}'$	$\hat{b}'$	APEX
<b>MS1</b>	0.051	0.047	0.054	0.053	13.9	80.1	47.2	44.4	97.2	41.7	8.3	2.8
<b>MS2</b>	0.124	0.119	0.130	0.128	11.5	61.5	24.0	31.3	93.2	48.4	27.1	5.7
<b>MS3</b>	0.055	0.048	0.056	0.055	19.0	72.0	39.4	37.1	81.1	50.0	9.1	0.0
<b>Count1</b>	0.647	0.507	0.651	0.637	18.3	91.7	23.3	30.0	100	95.0	20.8	16.7
<b>Count2</b>	0.607	0.472	0.613	0.601	11.1	88.9	18.1	30.2	100	91.3	13.5	8.0
<b>Count3</b>	0.592	0.489	0.599	0.590	5.0	88.3	14.7	25.8	100	86.4	18.6	11.7
<b>Math1</b>	0.166	0.154	0.173	0.166	19.0	39.1	21.6	21.1	47.9	16.7	9.1	0.3
<b>Math2</b>	0.658	0.577	0.679	0.669	19.2	45.7	48.7	54.4	99.6	82.5	32.0	21.5
<b>Math3</b>	0.394	0.377	0.419	0.411	36.6	51.2	41.7	38.0	40.5	6.3	2.5	0.0
<b>Math4</b>	0.236	0.219	0.253	0.251	23.4	45.3	28.6	38.1	25.4	6.3	3.2	0.0
<b>Math5</b>	0.540	0.485	0.546	0.530	33.7	51.1	49.2	49.2	69.2	30.8	0.0	0.0
<b>AA1</b>	0.619	0.497	0.628	0.615	18.1	64.0	30.0	44.8	99.4	84.9	11.1	5.6
<b>VS1</b>	0.277	0.221	0.282	0.280	19.8	93.2	31.3	48.4	98.4	81.3	10.4	5.2
<b>VS2</b>	0.261	0.218	0.265	0.262	7.8	51.6	24.2	29.2	100.0	96.4	17.9	14.3
<b>VS3</b>	0.305	0.265	0.313	0.312	7.0	86.7	14.1	28.9	97.7	68.8	13.3	3.9
<b>Key1t</b>	0.624	0.602	0.633	0.624	1.4	52.8	8.3	13.9	95.8	52.8	36.1	36.1
<b>Key2t</b>	0.185	0.176	0.187	0.179	45.8	63.9	56.9	52.8	88.9	31.9	6.9	1.4

To assess the strength of learning, we tested each series using the null hypothesis that the series had a constant mean across practice. Table 2 shows the percentage of series in which learning was significant. For most data sets, the majority of series show significant learning. Importantly, the two data sets that showed the least significant preference for the exponential (Math 3 and Math4) also showed the weakest learning. Across data sets, the correlation between the percent of exponential winners and percentage of significant practice effects was  $r = 0.646$ . The strong positive correlation again indicates that the weaker advantage for the exponential shown by some data sets in Figure 1 reflects noise rather than the form of the underlying learning function<sup>6</sup>.

<sup>6</sup> In the limit of no learning, power and exponential functions will both win about 50% of the series each. Hence, a 50% result means either that the true shape of the practice function falls between the shapes of the power and exponential functions, or that learning is very weak.



Comparing APEX and general power functions. Although our analyses, so far, strongly support the exponential function over the power, the comparison may have been confounded by the effect of pre-experimental practice. The power function assumes pre-experimental practice has not occurred, whereas the exponential does not. When pre-experimental practice is important, the power function is forced to estimate a too-large decrease in relative learning rate early in practice, and, as a result, the exponential might dominate. The general power function takes pre-experimental practice into account. Hence, to consider a role for pre-experimental practice, we compared the fit of the general power function against the fit of the APEX function.

The APEX function can estimate an asymptotic relative learning rate greater than zero, whereas the general power function requires relative learning rate to decrease to zero across practice. Consequently, better fits for the APEX than the general power function indicate that relative learning rate does not decrease to zero (i.e., the function is asymptotically exponential). Conversely, better fits for the general power function indicate that relative learning rate does decrease to zero (i.e., the function is asymptotically power).

Figure 3 shows the percentage of series that were better fit by the APEX function than by the general power. In the few cases where APEX and general power functions provided equally good fits, the general power function was classified as the winner because these cases corresponded to a power function solution (i.e., the estimate of  $\alpha' = 0$  for the APEX function).

Overall, the APEX function won in 84.1% of series, ranging from a minimum of 70% for Math3 up to a maximum of 94% for Key2t. Importantly, all of the cases in which the exponential function was weakest in Figure 1 were more strongly won by the APEX function in Figure 2. Hence, many of the cases won by the power function in Figure 1 (reflecting a decrease in relative learning rate early in practice) do not support a further decrease to an asymptotic value of zero later in practice. These results suggest that the fundamental assumption of the power function – asymptotically negligible relative learning rates – is not supported by any of the data sets in this survey and that accounting for prior practice is not sufficient to rescue the power law of practice.

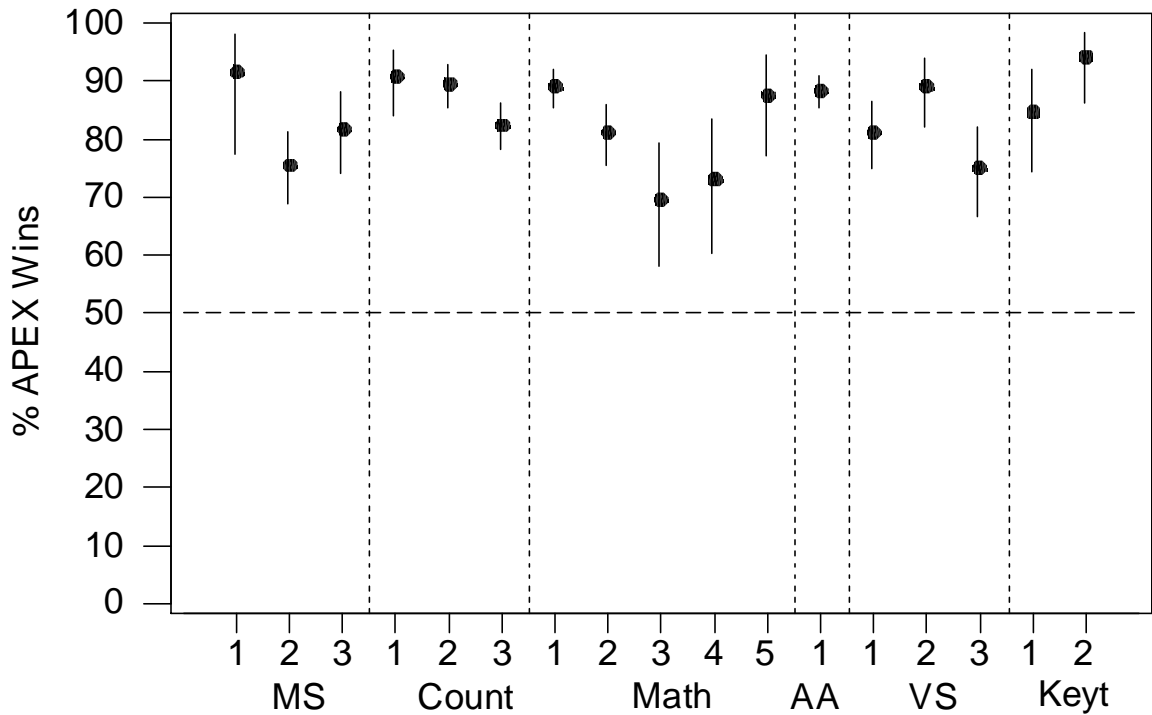


Figure 3. Percentages of cases where the APEX function provided a better fit than the general power function (solid circles) and exact binomial 95% confidence intervals for the published and unaveraged data sets in the survey.

Table 2 gives the average  $R^2$  values for APEX and general power functions for each data set. In every data set, the APEX function accounted for more variance than the general power function. Overall, the average  $R^2$  was 0.507 for the APEX and 0.498 for the general power. The APEX fits provided, on average, only a 2% improvement relative to the general power function. The small difference is to be expected, as the APEX function and the general power functions are quite flexible and hence able to imitate each other. However the advantage for the APEX function is very reliable, being evident in both the mean  $R^2$  values and number of individual learning series winners for every data set. Further, on average, the general power function provided no improvement in fit over the exponential, despite its extra parameter and ability to mimic exponential data. These results suggest that the better fit of the general power function over the

power function is due to its ability to mimic the exponential function rather than due to the effect of pre-experimental practice.

Across data sets, the correlation between the percent of APEX winners and percentage of significant practice effects was  $r = 0.405$ . The value was reduced somewhat by ceiling effects reflecting a very strong preference for the APEX function in some data sets. The difference between  $R^2$  for the general power and APEX functions was also highly correlated with their average value ( $r = 0.81$ ). Taken together, the results support the argument that the weaker advantage for the APEX function shown by some data sets in Figure 3 and Table 2 reflects noise, not the form of the underlying learning function.

Assessing asymptotic performance. As documented in Table 2, the power function tended to predict implausibly fast asymptotic performance. By implausibly fast, we mean an estimate of asymptotic expected RT ( $\hat{A}_p$  or  $\hat{A}_E$ ) less than 150 ms. Luce (1986), notes that "... minimal [simple] visual reaction times are of the order of 180 msec" (p. 63) and that "... choice reaction times ... are slower than the comparable simple ones by 100 to 150 msec" (p. 208). Hence, a criterion of 150 ms is a conservative estimate of plausible expected RT after extensive practice. Nevertheless, 87.7% of the estimates of asymptotic performance derived from the power function were less than 150 ms, compared to only 16.1% from the exponential.

Under prediction of asymptotic performance can be anticipated in a few cases, either because of noise associated with individual subject and condition series or because the series were too short to measure asymptotic performance adequately. The power function's systematic tendency towards under prediction, however, is a symptom of serious misfit. The reason is straightforward: The power function requires a large decrease in relative learning rate from the beginning to the end of practice. Apparently the power function cannot both match the relative learning rate occurring early in practice and maintain a large enough learning rate late in practice to predict plausible asymptotic performance. Figure 2 illustrates such a case. Note that the power

function's slow approach to its (implausibly fast) asymptote means that the under prediction is not evident in the range of practice trials measured.

Allowing for pre-experimental practice greatly improves the frequency of plausible asymptotic performance estimates. Overall, the general power function predicted asymptotic performance less than 150 ms in only 35.3% of series. Unlike the power function, the general power function does not have to predict a large decrease in relative learning rate from the beginning to the end of practice, although it does still predict that relative learning rate eventually decreases to zero. The improvement, however, required large estimates of the effect of pre-experimental practice. Overall, the estimates of pre-experimental practice were 90.4% of the length of practice series, indicating that almost half of the learning relevant to the experimental tasks occurred prior to the experiment.

Allowing for a constant relative learning rate greater than zero later in practice, but a decreasing relative learning rate early in practice, increased the frequency of implausible asymptotic performance estimates. Overall, the APEX function predicted asymptotic performance less than 150 ms in 27.4% of series. The result raises the possibility that an exponential function is biased to over predict asymptotic performance, at least when the relative learning rate decreases early in practice. Because the exponential function must predict that the relative learning rate is constant throughout practice, a decrease in relative learning rate early in practice causes over prediction of the relative learning rate later in practice and, hence, larger asymptote estimates. As noted earlier, under prediction of the asymptote may occur because of noise and because practice was not carried on sufficiently to obtain an accurate estimate of asymptotic performance. The results for the APEX function suggest that this may have been the case in about a quarter of the series contained in the published data sets.

Three parameters or four? Parsimony suggests that we should prefer a simpler model to a more complex model if the complex model does not provide a better explanation of the data. As noted previously, the general power function provides, on average, no improvement in  $R^2$

compared to the simpler exponential function. Hence, parsimony suggests that we should prefer the exponential function to the general power function as the Law of Practice.

The simple three-parameter exponential function also fits almost as well as the more flexible four-parameter APEX function. Indeed, in 49% of the data series, the fits of APEX and exponential functions were equal, because the estimate of  $\beta'$  for the APEX function was zero. However, the APEX function did provide, on average, a 2.4% improvement in  $R^2$ , relative to the  $R^2$  for the exponential function, with values ranging from 0.6% for the Count1 data set to 7.2% for the Math4 data set.

The latter results indicate that relative learning rate is either constant throughout practice or decreases slightly early in practice but then remains constant. Over data sets, the difference between  $R^2$  for the exponential and APEX functions was positively correlated with their average value ( $r = 0.334$ ). The correlation suggests that weaker improvements of the APEX function are due to noise rather than the underlying shape of the practice function. However, the association is much less than for comparison of the exponential function with the power function, and comparison of the APEX function with the general power function. Hence, other factors may be at work. The data sets that show the largest improvements in  $R^2$  for the APEX function (Math2, Math3, and Math4) all come from experiments designed to examine a mixture of algorithmic and memory-based processes. Consequently, a decrease in relative learning rate early in practice—hence, the need for an APEX Law of Practice rather than an Exponential Law of Practice—may often reflect a mixture of strategies.

Individual significance tests. For each series, we compared the fit of the APEX function against that of the power and exponential functions using change-of- $R^2$  tests (see Table 2).

Overall, goodness-of-fit decreased significantly when the exponential parameter ( $\alpha'$ ) was fixed at zero in 65.7% of series, ranging from 6.3% for the Math3 and Math4 data sets to 96.4% for the VS2 data set. By contrast, goodness-of-fit decreased significantly for only 15% of series when

the power ( $\beta'$ ) parameter was fixed at zero, ranging from no significant decreases for the Math5 data set to 36.1% for the Key1t data set. The difference in rates supports the previous evidence favouring an exponential function over a power function.

A very strong positive correlation was found between the percentage of significant practice effects and the percentage of significant  $\alpha'$  parameters,  $r = 0.85$ . The correlation suggests that cases where  $\alpha'$  was not significant were due to higher noise levels. The correlation between the percentage of significant practice effects and the percentage of significant  $\beta'$  parameters,  $r = 0.566$ , was weaker. The weaker correlation suggests some cases where  $\beta'$  was not significant were due to noise, but that in many other cases  $\beta'$  was not significant because the underlying practice function did not have a power component.

Only 7.8% of series had both  $\alpha'$  and  $\beta'$  estimates significantly greater than zero (i.e., required the full APEX function). The data set that provided the strongest evidence for a power component (Key1t) also provided the strongest evidence for the full APEX function. The percent of series with significant  $\beta'$  estimates was highly correlated ( $r = 0.879$ ) with the percent of series in which both  $\alpha'$  and  $\beta'$  estimates were significant; hence, even when relative learning rate decreased early in practice, asymptotic relative learning rate did not decrease to zero. The one exception was the Math2 data set; here, 21.4% of the series with a significant  $\beta'$  estimate did not have a significant  $\alpha'$  estimate. Even in the Math2 data set, however, significant  $\alpha'$  estimates predominated over significant  $\beta'$  estimates.

The individual significance tests support results reported earlier in suggesting that the exponential function provides a parsimonious model of the Law of Practice in most cases. Strong evidence for the full APEX function was obtained in only a minority of series and paradigms. Where evidence existed for a power component, it was usually associated with simultaneous evidence for an exponential component. Hence, a power function alone does not provide a good

model of the Law of Practice because it wrongly predicts that the relative learning rate decreases to zero with practice.

### The Effect of Aggregation

The analyses presented so far used data sets from individual learners and learning conditions to avoid confounding the shape of the function with the effects of aggregation over series with different learning rates. However, some level of aggregation could still have occurred, due either to a mixture of strategies or to summation of times for a series of responses. In the following sections, we examine the effects of such aggregation in detail. We then examine the effects of aggregation in data sets that were only available as an average across within-subject conditions.

Mixed strategies? Rickard (1997) has recently suggested that the Power Law of Practice fails because subjects use a mixture of algorithmic and memory-based processes, particularly early in practice. Although he maintains that both algorithmic and memory-based processes individually follow a power function, he argues that the mixture does not follow a power function. He has also speculated that an advantage for an exponential function over a power function might reflect the mixing of algorithmic and memory based processing, both of which separately follow a power function (Rickard, 1996, personal communication).

We tested his suggestions using data sets from Rickard (1997), Reder and Ritter (1992), and Schunn et al. (1997). In all three studies, subjects reported their processing strategy. For our analysis, the data sets from the latter two papers (Math3, Math4, and Math5) were combined, as each data set by itself had few series.

The results of fits, subdivided by processing strategy, are presented in Figure 4 (in the sections labelled Math and AA). Both algorithm (75.1%) and memory (79.2%) series were better fit by the exponential than power function (77.8% overall). An advantage for the APEX function over the general power was even stronger: 86.1% overall, with 86.4% of the algorithm series and 85.9% of the memory series won by the APEX function. As indicated by the 95% confidence

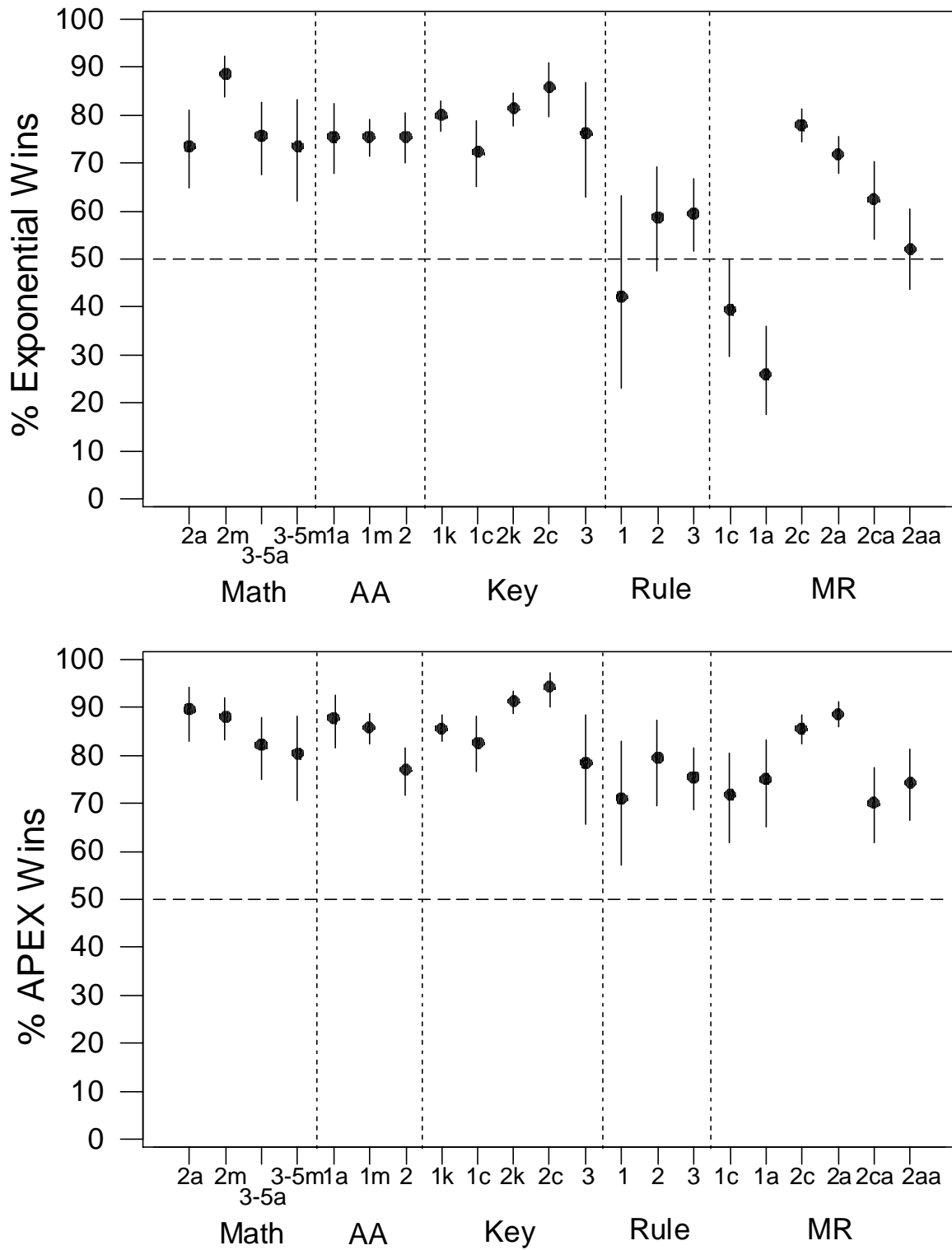


Figure 4. Percentages of cases (solid circles) where (a) the exponential function provided a better fit than the power function, and (b) the APEX function provided a better fit than the general power function, and exact binomial 95% confidence intervals, for data sets in the survey not presented in Figures 1 and 2.



intervals in Figure 4, the preference for exponential and APEX functions was highly significant. We conclude that the better fit for the exponential function is not due to a mixture of component power functions.

The results for the series separated by processing strategy, are almost identical to the data for the same data sets in Figures 1 and 3, where no distinction was made between processing strategy (83.4% exponential and 84.0% APEX overall). We take comfort from the similarity because it argues that mixtures of processing that might have taken place elsewhere in the survey did not distort the tests that we have reported.

Figure 4 also shows the results for a data set from an unpublished alphabet-arithmetic task (AA2) similar to Rickard's (1997) task, except that the subjects were not asked to report their strategy. The results for AA2 also clearly favour the exponential and APEX functions, confirming their dominance in the alphabet-arithmetic paradigm when strategy reports are not required.

Sequential responses. Figure 4 summarises the effect of a second type of aggregation: summing times from a sequence of responses. In the analyses of Verwey's (1996) data shown earlier, we examined the total time to perform nine sequential key presses. Because summing across individual key presses may have distorted the form of the practice function, we looked at the separate responses.

As Figure 4 shows, the time to produce individual chunks and the time to produce individual key presses were both fit better by the exponential and APEX functions than by the power and general power functions respectively. Combining over data sets from Day 1 and later days, there was a slight reduction in the preference for the exponential for individual key press (80.7%) and chunk (79.1%) compared to the total series (84.7%). There was also a slight reduction in the preference for APEX for individual key press (88.6%) and chunks (88.6%) compared to the total series (89.6%). The small reductions likely reflect greater noise in the key-

press and chunk compared to the total series. In any case, the results indicate that summing times to perform individual responses did little to distort the form of the practice function in these data.

Averaging across conditions. Seibel (1963) studied subjects who practiced production of all 1023 possible combinations of 10 key presses in response to a compatible visual display.

Newell and Rosenbloom (1981) re-analysed the data from one of Seibel's subjects, JK, from the first 75 blocks of practice. The data were averaged over the 1023 different combinations of key presses and blocks. When we re-analysed the data using ordinary least squares fitting, the power function ( $R^2 = 0.9858$ ) provided a better fit than the exponential ( $R^2 = 0.9584$ ) but the APEX function ( $R^2 = 0.9902$ ) provided a better fit than the general power function ( $R^2 = 0.9895$ ).

Some combinations of key strokes are harder than others and, as a result, may have been learned at different rates. If so, the shape of the practice function may have been distorted by averaging over conditions with different learning rates. The effect of averaging over conditions with different learning rates is the same as the effect of summing over components of performance with different learning rates; conditions with fast learning rates will dominate the relative learning rate early in practice but will soon approach their asymptote. Hence, late in practice, conditions with slow learning rates will control the relative learning rate. The transition will decrease the relative learning rate of the average function and could yield a better fit for the power than the exponential. However, if learning is exponential for the combinations with slower learning rates, the relative learning rate later in practice will be a constant greater than zero, and, as we found, the APEX function will fit better than the general power function. Further evidence against the general power function in this paradigm is reported by Rosenbloom and Newell (1987a). They found that that the exponential function fit better than both the power and general power functions for a single subject performing Seibel's (1963) task.

Because evidence on the form of the practice function for learning key combinations is both scant and used relatively short series for each combination, Brown and Heathcote (1997) examined learning key combinations in a larger sample of subjects with a larger number of trials

per combination. Simplifying Seibel's (1963) paradigm, they used only four response keys and the 14 possible keystroke combinations that involve 1 to 3 key presses. By reducing the number of combinations (relative to Seibel's 1023), they were able to increase the length of the series for each combination.

As shown in Figure 4, 76.4% of series were better fit by the exponential than power function. The result suggests that the power function fit better than the exponential function for Seibel's (1963) subject JK because of averaging. In Brown and Heathcote's (1997) data, 78.6% of the series were better fit by the APEX function than the general power function. The pattern suggests that the asymptotic relative learning rate did not decrease to zero and is consistent with the results for Seibel's subject JK and Newell and Rosenbloom's (1987a) subject. Hence, results for key combinations are consistent with results for other tasks in suggesting that relative learning rate is constant within experimental conditions.

The final analyses presented in Figure 4 concern data sets that were available only as averages across conditions. The first three data sets were from Anderson, Fincham, and Douglass (1997); subjects practiced the application of a number of different rules learned from examples. The data were averaged over the different rules and over groups of practice blocks. As shown in Figure 3a, only one data set produced significantly more exponential than power fits, and one data set displayed a (non-significant) preference for the power function.

As shown in Figure 4b, however, all data sets were fit significantly better by the APEX function than by the general power. Overall, the APEX function won in 76.4% of the series from these data sets. The advantage for the APEX indicates that power components to learning were largely restricted to the early part of practice and that asymptotic learning was exponential. We suspect that the rules differed in difficulty and, hence, in learning rate. Certainly, the pattern reported in Figures 4a and 4b is consistent with the effects of averaging across conditions with different learning rates.

The final averaged data set is from a developmental study of mental rotation of letters performed by Kail and Park (1990). The data were averaged over the four types of letter stimuli and the two presentations of the letters in each practice block. Figure 4 presents the results of analyses separately for children and adults. They show a strong preference for the power function over the exponential. However, like the other data that included averaging, the APEX function provided a better fit than the general power function, with preference for both children and adults being significant and on average 73.4%. As before, the advantage for the APEX indicates that power components to learning were largely restricted to the early part of practice and that asymptotic learning was exponential.

Because the mental rotation data provide the strongest evidence for the Power Law of Practice of any data set in the survey, Ringland and Heathcote (1998) replicated Kail and Park's (1990) study. As shown in Figure 4, when analyzed without averaging (MR2c and MR2a), mental rotation too significantly favored the exponential function over the power function for both ages. Figure 4 also shows Ringland and Heathcote's data averaged in the same way that Kail and Park's data had been averaged (MR2ca and MR2aa). Although a strong preference for power was not obtained, averaging reduced the number of exponential wins. Moreover, as before, the APEX function provided a better fit significantly more often than the general power function, for both averaged and unaveraged series. In other words, averaging pushes the results towards the power function. We suspect that the remaining differences between the Kail and Park's data and Ringland and Heathcote's data reflect differences in learning rates induced by differences in the font in which the stimuli were displayed.

### Individual Differences

Over almost all of the data sets analysed, both exponential and APEX functions dominate power and general power functions, but the dominance is not complete. Between 15-20% of learning series were better described by power and general power functions. One explanation is noise. We have provided evidence for this possibility by showing that preference for the

exponential and APEX functions increases as noise decreases. Nevertheless, the possibility remains that the power or general power functions might provide a better description of practice effects in at least some data sets for a minority of conditions or subjects.

We calculated  $\Phi$  coefficients to quantify the relationship between a preference for exponential and APEX functions and (a) all of the between- and within-subject factors used in the various tasks in the survey, and (b) individual subjects. The  $\Phi$  coefficient is a non-parametric correlation that ranges between zero and one. It is a linear transformation of the  $\chi^2$  statistic used to test the contingency between frequencies in a two-way classification. Larger values of  $\Phi$  indicate a stronger contingency or association. If the power or general power functions provide a better description of learning for a subset of subjects or a subset of conditions, preference should be systematically related to those subjects and conditions, the corresponding  $\Phi$  coefficients should be large, and their corresponding p-values should be small. By contrast, if preference for the power or general power functions reflects noise, it should not be systematically related to any experimental factor or to particular subjects, and all  $\Phi$  coefficients should be small. Small values of  $\Phi$  and correspondingly large p-values indicate that the null model (i.e., no systematic relationship between factors and preference for a function) provides a good description of the data.

Table 3 presents both estimated  $\Phi$  coefficients and the significance levels of the associated  $\chi^2$  tests<sup>7</sup>. The values of the  $\Phi$  coefficients were very small in most cases, with a mean value of 0.11 and a median of only 0.07. Table 3 reports just the cases, out of more than 200, with  $\Phi$  estimates greater than 0.2. All but three of these estimates were associated with subjects

---

<sup>7</sup> The significance of each  $\Phi$  coefficient was tested with the corresponding  $\chi^2$  test. The significance levels for these tests must be treated with some caution, especially for the subjects factor, as they assume independence.

(i.e., individual differences), despite the larger number of other types of factors. Two of the three experimental factors with  $\Phi$  coefficients greater than 0.2 are between-subjects, and so may be due to individual differences. The only within-subjects factor in Table 3 has a  $\Phi$  coefficient only slightly greater than 0.2. Only two of over 200 tests indicated significant systematic contingency between function preference and an experimental factor, and in both cases, the experimental factor was a between subjects factor.

Table 3.  $\Phi$  coefficients greater than 0.2 for all data sets in the survey. Probability (p) refers to a test of the  $\chi^2$  value corresponding to each  $\Phi$ .

Exponential vs. Power				APEX vs. General Power			
Data Set	Factor	F	p	Data Set	Factor	F	p
Rule1	Subjects	0.664	0.638	Rule1	Subjects	0.470	0.994
MR1	Subjects	0.489	0.000	Key1t	3:3:3/3:6	0.281	1.000
Rule3	Subjects	0.440	0.861	MR1	Subjects	0.280	0.449
Math4	Subjects	0.434	0.674	Rule3	Subjects	0.274	1.000
Math1	Subjects	0.329	0.011	Math3	Subjects	0.263	0.999
Key1t	3:3:3/3:6*	0.307	1.000	Math4	Subjects	0.255	0.997
MS3	Subjects	0.279	0.975	MS2	Subjects	0.254	0.999
VS2	Subjects	0.261	0.275	VS2	Subjects	0.252	0.802
Key1c	Subjects	0.244	1.000	MS3	Subjects	0.235	0.997
AA2	Subjects	0.231	0.881	Key1c	Subjects	0.230	1.000
Math5	Subjects	0.224	1.000				
Key2t	Subjects	0.220	1.000				
Math4	Problem†	0.220	0.263				

\*Between subjects factor for chunk size.

†Within subjects factor for problems.

The largest estimates of  $\Phi$  occurred in the averaged data sets, suggesting that they may reflect individual differences in exponential learning rates across averaged conditions, rather than true power practice effects. The only significant associations were obtained for exponential compared to power fits. In the MR1 data set, a single child had exponential wins in all 12 conditions, and two adults had 8/12 exponential wins, whereas the majority of series were power wins. For the Math1 data set, subjects ranged from all exponential to all power across the 16 within-subject conditions.

The results suggest that deviation from the general finding of exponential practice is more likely to be associated with individual subjects than with particular within-subject

conditions. The finding reinforces our earlier caution that practice functions from different subjects should not be averaged, not only because variation of learning rates can distort the shape of the average function, but also because individual subject's learning may occasionally follow a function other than the usual exponential form.

### Discussion

Our results can be summarized as follows: The three-parameter exponential function provided a better description of learning than the three-parameter power function in more than 80% of cases. The four-parameter APEX function provided a better description than the four-parameter general power function in about 85% of cases. Hence, a mixture or sum of power-function processes early in practice cannot explain the power function's loss, as learning was exponential later in practice, when a single process should predominate. In experiments that identified algorithmic and memory based components, learning in both components was better described by the exponential and APEX functions than by the power and general power functions, respectively. The four-parameter general power function provided no improvement over the simpler three-parameter exponential function. Hence, the power function did not lose to the exponential function because of the effects of pre-experimental practice.

In about half of the cases considered, the four-parameter APEX function provided no measurable improvement over the three-parameter exponential function, indicating that the extra flexibility implied by a fourth parameter was not needed. Where the fit was improved, the improvement largely reflected a decrease in relative learning rate early in practice. Later in practice, learning was exponential, as indicated by the dominance of the APEX function over the general power function. Hence, the most parsimonious assumption appears to be that practice produces a simple exponential improvement and a constant relative learning rate, with the caveat that some change in the relative learning rate may occur early in practice.

Despite the ability of nonlinear practice functions to imitate each other, determining the form of the practice function, at least to a first approximation, has not proved to be an insoluble technical issue. For the theoretically important comparison of exponential and power functions, preference for the exponential function was clear and significant in all paradigms. The improvement in fit provided by the exponential function relative to the power was in many cases quite large, with an average value of 17%. Consequently, the description of practice effects by a power function is often substantially in error (e.g., Figure 2).

The estimated parameters of the power function are also misleading because of the problem of imitation. We have shown that at least one parameter of fitted power functions, the asymptote, is usually an under estimate. Under estimation of the asymptote likely reflects imitation of a constant relative learning rate, particularly later in practice. Because parameter estimates tend to be correlated, underestimation of the asymptote is likely to distort estimates of the other parameters of the power function. In short, parameter estimates from the power function are likely to be unrelated to the psychological processes underlying learning or, at best, related in a complex way determined by the best imitation that a fitting algorithm can find.

The minority of cases in which the exponential function was not the best description appear largely to be due to random variation. Preference for the exponential function was larger in paradigms with higher signal to noise ratios, and the frequency of exceptions to the exponential rule was not systematically related to within subject conditions. However, a small minority of individual subjects did show systematic deviation from the exponential rule.

Overturning an empirical law requires a high standard of evidence. Our results are clear that the most parsimonious Law of Practice is the exponential function. Nevertheless, it would be unwise to accept the exponential function without question. Instead, it should form a baseline for comparison with other possible forms, and comparison should be carried out using individual subjects' data via nested-model tests. Testing should not be carried out on averaged data, as averaging distorts the form of the practice function (Brown & Heathcote, in preparation;



Anderson & Tweney, 1997; Myung et al. 1998). The following sections examine these techniques in more detail. We then consider the implications of an Exponential Law of Practice for theories of skill acquisition.

### Measuring Practice Functions

Averaging and practice functions. Averaging usually distorts the form of the practice function. Our analyses for individual subjects and conditions stand in marked contrast to Newell and Rosenbloom's (1981) results for averages over subjects and conditions. Empirically, we found that averaging over within-subject conditions can produce a bias in favour of the power function. We also found that practice curves for a small minority of subjects may differ from the usual exponential form. Consequently, as Newell remarked, averaging "... conceals, rather than reveals. You get garbage or, even worse, spurious regularity." (1973, p. 295).

Our findings reinforce long standing analytic results that the learning curve for an arithmetic average need not have the same form as the functions contributing to the average (e.g., Estes, 1956; Sidman, 1952). They also confirm that these results are not a mathematical nicety: Arithmetic averages can be biased in favour of a power function and against an exponential function in real data. Myung, Kim, and Pitt (1998) have explored why arithmetic averaging of nonlinear functions distorts the average curve and have shown that other averaging techniques are required when dealing with nonlinear models; the appropriate average depends on the nature of the functions to be averaged. Our empirical results and the analytic results converge on the same conclusion: averaging cannot succeed without first taking into account the form of the functions to be averaged. Research workers can no longer afford to ignore, or worse to dismiss, the effect of averaging as irrelevant to real data from paradigms used in the study of learning.

Some variability among component learning rates is necessary for distortion of power and exponential function averages. When component learning rates are exactly equal, the average has the same functional form, and its parameters equal the average of the component's parameters, at least for purely deterministic functions. When the component learning rates vary,

neither condition need apply (see Myung et al., 1998). The degree of distortion is proportional to the degree of learning rate variability. In particular, for averages over exponential functions, a power-like decrease in the relative learning rate will occur if the sample contains fast and slow learners.

Rickard (1997) suggested that averages do not distort practice functions because the parameters of learning functions “do not have extremely large variance (a condition that probably holds in most real data sets)” (p. 295). Our experience, both with real data sets and with numerical simulations, differs. Most data sets in the survey contained both fast and slow learners and sometimes fast and slow learning conditions, with learning rates often varying over several orders of magnitude. Subjects or conditions that showed a substantial improvement with practice, but at a slow learning rate, were particularly likely to bias the fit toward the power function in the average. Not only were large variations in learning rates regularly observed in the survey data, but also simulations (Brown & Heathcote, in preparation) show that the amount of variation in learning rates required to produce significant bias is often as little as one order of magnitude. Hence, we think it is dangerous to assume that the parameters of learning functions are sufficiently homogenous to avoid substantial averaging distortion.

Geometric averaging (averaging on a logarithmic scale) has been widely suggested as a solution to the problem of distortion (e.g., Anderson & Tweney, 1997; Rickard, 1997). However, geometric averaging preserves the form of averages of power functions and averages of exponential functions only if the asymptote is negligible. It is unreasonable to assume a negligible (near zero) asymptote for practice functions because performance is limited by physical constraints, such as neural integration time and motor response time. Geometric averaging, therefore, will still distort the form of the average practice function. While the distortion may be less than for arithmetic averages (although Brown & Heathcote, in preparation, found that the benefits were negligible for the data typically collected in practice paradigms), its

effect may differ across experimental conditions or groups of subjects, potentially confounding comparisons.

In general the scale on which averaging should take place is data-dependent. With geometric averaging the scale depends on the (usually unknown) asymptote parameter, which must be subtracted from the data before logarithmic transformation. Consequently, practice functions must be fit to individual data to estimate the asymptote before averaging can be safely performed. Even then, averaging is problematic because the asymptote is an expected value, and hence will be greater than some data points, at least when a sufficiently long practice series is measured to produce a reliable estimate of the asymptote (cf. our critique of Newell & Rosenbloom's, 1981, fitting methods). Subtracting the asymptote estimate from such data points produces a negative number for which the logarithmic transformation is undefined. Hence, even when the appropriate transformation is known or can be estimated, averaging raw data, which is always contaminated by noise, will not be useful when fitting power and exponential functions.

The only case in which averaging is safe is when learning rates vary little between the components of the average. Given that individual analysis is needed to determine when averaging is safe (i.e., to determine the individual learning rates) analysis of the averaged data adds little of value. The survey, however, is encouraging for the study of individual learning. Individual learning effects were often very large, so that averaging was not needed to minimise the effects of trial-to-trial variability. In paradigms with large learning effects, nested-model tests on individual curves were usually decisive. Typically, they indicated the need for an exponential component only, but in a small number of cases, they implicated a power component or both components simultaneously. Such nested-model tests not only provide an inferential basis for conclusions about the form of individual learning curves but also provide a method of identifying unusual cases.

Despite these problems, most researchers would agree that “average data are useful because they often reveal general trends” (Anderson, 1999, personal communication). Given its

utility, averaging is not likely to surrender its place in data analysis – and nor should it.

Averaging itself is not intrinsically problematic; biases are only introduced when the scale on which data are averaged does not match either the data or the analyses. Because of noise, averaging of raw practice data is unlikely ever to be useful. However, analysis of average parameter estimates, such as performing ANOVA on the parameters of the best fitting practice functions as determined by individual analysis, remains viable (e.g., Heathcote & Mewhort, 1995). Such analyses maintain the beneficial effects of averaging, such as reducing noise and revealing general trends, without introducing the systematic distortions produced by averaging raw data. Representations of average performance can be obtained by plotting the practice function with parameters equal to the average of parameters for individual practice functions.

Nested models and relative learning rates. Detailed consideration of nested-model testing is beyond the scope of the present work (we recommend Bates & Watts, 1987, for further reading). However, a few points of clarification are in order. The benefits of nested-model tests are not limited to practice functions, and useful higher-order nesting functions are not limited to the APEX function. Commenting on the form of the forgetting function, for example, Wickens (1998) suggested the use of higher-order functions that isolate theoretically important characteristics in separate parameters. The idea is that theories can then be tested by measuring the effect of experimental manipulations on parameter estimates. We chose the APEX function because tests of the  $\alpha'$  and  $\beta'$  parameters determined the contributions of theoretically important power and exponential components. Other functions such as the Weibull, favored by Wickens<sup>8</sup>, or perhaps a sum or mixture of exponential functions, may prove useful in other applications.

---

<sup>8</sup> The Weibull function is a power transformation of the exponential function. With an exponent of one it equals an exponential function and so has a constant relative learning rate. Exponents greater than or less than one produce increasing and decreasing relative learning rates respectively. Wickens (1998) also used the Parto II function. The Parto II function is a special case of the general power function. Our evidence suggests that the general power function is inferior to the APEX function and plagued by ill-conditioned fitting for practice curves.

Again, nested-model tests provide an inferential basis for determining a parsimonious form for the function and for testing theories.

An important feature of our work is the use of relative learning to compare practice functions with different mathematical forms. Wickens (1998) used the hazard rate for forgetting functions in much the same way that we use relative learning rate. Relative learning rates and hazard rates are defined on different measures, expected **RT** and probability of forgetting, respectively, but are otherwise identical.

The importance of the relative learning and hazard rates suggests the desirability of a direct estimate that does not assume a parametric function. However, as Wickens (1998) notes “Although the use of empirical estimates of the hazard function to select among candidate functions or explanations is attractive, adequate precision is hard to obtain ...” (p.382). Non-parametric relative learning rates can be obtained by dividing estimates of the derivative, such as the difference between RT for adjacent practice trials, by an estimate of the expected value of **RT**. Like other researchers (e.g., Luce, 1986, p. 60-63), we have found such methods to be inefficient for response time measures. The problem is that such estimates of the derivative tend to be unacceptably noisy, especially with data that are not averaged. Because of these problems, the approach we have taken is fitting simple parametric models with different relative learning rate functions.

Recent advances in non-parametric regression may provide an alternative approach. These techniques reduce the effects of noise by local smoothing with kernels or splines. By estimating the regression function using local polynomial regression, for example, direct estimates of the derivative are obtainable at every point. These methods have been shown to often provide reliable estimates of differential metrics (Wand & Jones, 1995), such as the relative learning rate.

---

However, simple parametric functions, particularly the exponential function, provide an important advantage for empirical investigations: They allow the estimation of a single learning rate parameter that applies for all levels of practice. Because the practice function's slope changes with practice, slope cannot provide a single parameter summary of learning rate. The exponential function's  $\alpha$  parameter does provide such a summary by assuming that slope is proportional to the amount left to be learned. The  $\beta$  parameter of the power function also provides a single parameter summary by assuming that slope is hyperbolically related to the amount left to be learned. However, the results of our survey suggest that the power function does not provide an accurate model of practice effects, and so inferences based on estimates of  $\beta$  may be misleading.

Fitting methods and RT distribution. Least-squares fitting, as used in the survey, assumes normally and independently distributed residuals around the expected value function (i.e. the practice function). This assumption is violated by response time data. Response time distributions are usually positively skewed and their means and variances are often positively correlated across conditions (Luce, 1986). Changes in the residual distribution with practice have been analysed previously (e.g., Logan, 1988; Rickard, 1997) and large decreases in standard deviation with practice observed.

These violations of the assumptions of least-squares fitting may have biased our results. To check for such bias, we developed a more sophisticated regression technique (APEXL) and applied it to the survey data. Space restrictions do not allow a full explication of the APEXL technique, or the results of its application. Briefly, APEXL fitting uses a special case of Box and Cox's (1964) two-parameter transformation family and implements Carroll and Ruppert's (1988) "transform both sides" approach to regression. It simultaneously estimates both the expected value function's parameters and a data-dependent transformation parameter. The transformation is a shifted logarithm,  $\ln(\text{RT}-\lambda)$ , where the transformation parameter,  $\lambda$ , is an estimate of the

lower bound of **RT** distribution. APEXL fitting iteratively re-weights residuals during fitting to maximise normal distribution and homogeneity of variance on the transformed scale.

The model underlying APEXL fitting is that **RT** follows a Lognormal distribution. The Lognormal distribution is positively skewed and has been found to provide a good account of response time data (Ratcliff & Murdock, 1976). The APEXL model assumes that Lognormal distribution is combined multiplicatively with the expected value function, so that response time mean and variance are positively correlated. A detailed analysis by Heathcote and Mewhort (1995) found that the model provided a good description of response time distribution for Heathcote and Mewhort's (1993) visual search practice data. When we applied APEXL fitting to the unaveraged data sets<sup>9</sup> in the survey, 79.0% of practice series were better fit by exponential than power functions. Hence, the APEXL analysis supports the conclusion from the least-squares analyses that the exponential function is the best simple candidate for the Law of Practice. We are confident, therefore, that our results were not biased by violations of the assumptions underlying least squares fitting.

While the results of APEXL and least squares fitting are consistent in their selection of the best practice function, APEXL fitting has a number of advantages over the ordinary least squares approach. APEXL fits simultaneously estimates the median function along with the expected value function (the two are identical on the transformed scale as residuals are normally distributed). It provides not only estimates of the expected value function's parameters but also characteristics of variability around the expected value function, such as the variance and lower bound of the Lognormal distribution. When its assumptions hold, APEXL fitting is more efficient than ordinary least squares fitting, a critical advantage when fitting noisy individual data. The assumptions of nested model tests are better fulfilled on the transformed scale and so Type 1 error

---

<sup>9</sup> The averaged data sets were not analysed using the APEXL technique because the central limit theorem implies that averages tend to be normally distributed.

probability is better estimated. Finally, the form of the expected value function is the same on the transformed and natural scale, a property that is not true of the more commonly used power transformations (Miller, 1984). Heathcote, Brown and Mewhort (in preparation) examine the performance of the APEXL technique in detail, both for data from the survey and for simulated data.

Approaches such as APEXL fitting help to illuminate interesting properties of practice data sets beyond the expected value function, such as the change in response time variance as a function of practice. Such properties can provide useful constraint for theories of skill acquisition (Logan, 1992; Rickard, 1987). Proper measurement of variance, however, relies on prior estimation of the expected value function. Previous attempts to measure variance as a function of practice, for example, used variance estimates calculated from successive blocks of raw data (Kramer, Strayer & Buckley, 1990; Logan, 1988). Block variance has two components, variance around the expectation function and variance caused by a decrease in expected values across the block. The contribution to block variance from the change in expected value function is unequal in different practice blocks. Early in practice, the expected value function changes rapidly, and, as a result, block variance is greatly inflated by it. After extensive practice, however, the function is relatively flat, so block variance is an almost pure measure of variance around the expected value function.

Predictions from skill acquisition theories have been derived for pure variance around the expected value function (e.g. Logan, 1988, 1992). Consequently, these theoretical predictions cannot be tested by measurement of block variance, which does not purely measure variance around the expected value function. The problem can be remedied, by analysing residuals obtained by subtracting the expected value function from the raw data, but the correct expected value function must be determined first. A similar procedure is required for measurement of other

---



interesting properties of practice data sets, such as autocorrelation between responses. Hence, the results of our survey, which bear on the form of the expectation function, provide the first step in the proper assessment of these other interesting measures.

Before closing this section we wish to caution against a widely used technique for fitting practice functions: log~log and log~linear plots. In detail, the goodness of fit of power and exponential functions are determined by comparing the linearity of data in  $\log(\text{RT})\sim\log(N)$  and  $\log(\text{RT})\sim N$  plots respectively. Such fits implicitly assumes that the asymptote of the practice function is zero. The assumption is not only wrong, but also produces a bias in favor of the power function. The power function approaches its asymptotic value more slowly than the exponential function and, so, its fit is less effected by an underestimated asymptote. In the survey data, small estimates of the asymptote were more commonly associated with the best fits for the power function than the best fits for the exponential function. Hence, our results show that assuming a zero asymptote will likely hurt the fit of the exponential function more than the fit of the power function. Misspecification of the asymptote is also likely to make estimates of other practice function parameters, such as learning rates, at best difficult to interpret and at worst misleading, because parameter estimates are usually correlated.

### Theoretical Implications

Our results repeal the Power Law in favour of an Exponential Law of Practice. Many of the data sets included in the survey were collected in order to test specific theories of skill acquisition. We will now consider the implications of an Exponential Law of Practice for these theories. Some of the theories are tied to the power function more tightly than other theories. Although it is possible for some theories to retain their fundamental assumptions, all require at least some modification in the way these assumptions are applied. The mathematical details of the required modifications are beyond the scope of the present work. In the following we provide only heuristic details in order to illuminate the theoretical implications of an Exponential Law of Practice.

Chunking theory. The first response to Newell and Rosenbloom's (1981) Power Law of Practice was their chunking theory, which was later elaborated by Rosenbloom and Newell (1987a, 1987b). The theory does not exactly predict a Power Law of Practice, but it does predict that relative learning rate decreases to zero with practice. To obtain a decreasing relative learning rate, Newell and Rosenbloom assume (a) chunks are learnt hierarchically, (b) that larger chunks necessarily practice their smaller components every time the larger chunk is practiced, and (c) that no larger span chunk is acquired until all chunks of smaller span are acquired, at least in combinatorial learning environments. A combinatorial learning environment is one in which larger chunks are encountered less often than smaller chunks. The prototypic example is Siebel's (1963) key-press combination task.

Newell and Rosenbloom's (1981) assumptions can be replaced by a single simple assumption to achieve an exponential practice function: Chunks are executed as a single unit and, so, practice only themselves, not their constituents. Newell and Rosenbloom implicitly make a similar assumption by claiming that the execution time for a chunk is independent of its size. Neves and Anderson (1981) also note that such a chunking mechanism, which they call composition, produces an exponential practice function.

Even if Newell and Rosenbloom's (1981) theory is not modified, it predicts a decreasing relative learning rate only with practice in a combinatorial learning environment. Arguably, in most of the paradigms examined in the survey, the learning environment was not combinatorial. Hence, chunking theory predicts the observed exponential practice functions. The only data from clearly combinatorial environments come from Siebel's (1963) subject JK, Rosenbloom and Newell's (1987a) single subject, and Brown and Heathcote's (1998) subjects. Brown and Heathcote's data strongly favoured an exponential function, but their paradigm used a smaller set of combinations than the other experiments; hence, the effect of the combinatorial environment was likely attenuated. Verwey's (1996) paradigm is also likely to be only weakly combinatorial as subjects practiced exactly the same sequence on all trials and chunk structure was consistently

defined by cues. For simple environments our results are unambiguous that the relative learning rate does not change much with practice, a pattern predicted by a simple chunking mechanism.

Clearly, more work is required to determine if complex combinatorial environments yield a decreasing relative learning rate. At present, however, the weight of evidence favours a constant relative learning rate and, hence, the assumption that chunks are seamless units that do not practice their constituents. More work is also required on Newell and Rosenbloom's (1981) derivation of a decreasing relative learning rate (it is not exact). In contradiction of the derivation, several of the simulations of their model presented in Rosenbloom and Newell (1987a) were better fit by an exponential function than a power function.

Aggregated component theories. Neves and Anderson (1981) suggested a second mechanism based on chunking that produces a decreasing relative learning rate with practice: The summation of many components that learn exponentially. The mechanism is a generic one. For example, Rickard (1997) suggests that response strength for memory based processing is the sum of strengths of a collection of neural connections that individually learn exponentially, so that the sum is a power function. At a coarser scale, Kirsner and Spelman (1996) argue that many tasks rely on several components and so a sum of component learning functions provides a superior model of practice effects.

The critical factor for producing a decreasing relative learning rate in the sum of exponential components is that some components learn more quickly than others. A decreasing relative learning rate occurs for the same reason that averaging across subjects with different learning rates produces a decreasing relative learning rate. Fast learning components produce a large overall relative learning rate early in practice but soon reach asymptote; hence, learning later in practice is controlled by slow learning processes with smaller relative learning rates.

A modified assumption—that the variation amongst learning rates is small—is required if theories based on summed exponential components are to predict the exponential practice functions found in the survey. Where all components of the sum have exactly the same learning

rate, the sum is exactly exponential. Where learning rates vary, the sum will not be exactly exponential, but if the variation is small, it will approximate an exponential function. Such a model can also accommodate APEX practice functions if learning rates vary but the smallest learning rates are appreciably greater than zero. Later in practice, the smallest learning rate component or components control learning, so that relative learning rate is asymptotically greater than zero, as in the APEX function. This may account for the evidence for an APEX function that was found in the survey.

It is also possible that a sum of components that learn according to power functions can approximate an exponential or APEX practice function. It is well known that the sum of an infinite number of power functions (i.e., a Taylor series) can approximate nearly any function. However, an infinite number of components is not plausible for psychological processes. For finite sums, approximating an exponential function by the sum of power components also requires assumptions about the weight for each component in the sum. In particular, the weights must vary over at least as many orders of magnitude as there are components. Consequently, power components require justification for the change in weights across many orders of magnitude.

The assumption of power components itself also needs justification. Exponential components can be naturally derived from simple mechanisms. For continuous mechanisms, one need only assume that learning is proportional to the time taken to execute the component. That is, a component that takes longer to execute presents more opportunity for learning. As learning proceeds, the time to execute the component decreases; hence, the learning rate decreases, resulting in exponential learning. For discrete mechanisms, such as chunking, exponential learning occurs for similar reasons. As responses are produced by larger and larger chunks, fewer opportunities for further composition are available. Similar justifications are needed if power components are to be plausible.

Rickard's (1997) Component Power Laws theory also claims that the practice function is the result of an aggregate of component functions. Aggregation occurs through a mixture, rather than by summation, and only two component processes, algorithmic and memory based processing, are assumed. Our analysis of Rickard's data, combined with analyses of other data sets in which subjects were required to identify responses controlled by algorithmic and memory based processes, suggest that each component learns exponentially. As already discussed, the exponential nature of learning for memory based processing could result from homogenous learning rates amongst its component processes.

The mixture assumption in Rickard's (1997) theory can result in a complex change in the relative learning rate with practice, depending on the form of the mixture function. In particular, Rickard's assumption of a logistic mixture usually results in a non-monotonic change in the relative learning rate of the aggregate, first increasing and then decreasing. This occurs because the change in mixture proportions first accelerates then decelerates with practice. Rickard's Math2 data set did provide reasonably strong evidence for an APEX function and, hence, a decrease in relative learning rate early in practice. Explicit fits of the mixture model based on exponential components, and consideration of a range of possible mixture functions, are required to clarify this issue.

Sums of decaying traces and the forgetting function. Anderson's (1982) ACT model added a second mechanism based on strength of learning to the composition or chunking mechanism already discussed. This mechanism assumes that response time is a linear function of the reciprocal of learning strength. Learning strength equals the sum traces from each practice trial, and the strength of each trace is assumed to decay as a power function of time. The sum of the decaying traces increases approximately as a power function of practice trials, and hence the model predicts an approximately power decrease in response time with practice. Anderson, Fincham, and Douglass (in press) develop these ideas more fully and use them to explain slowed performance after a break in practice.

There has been much recent debate on the form of trace-decay functions, as measured by forgetting of memorized items at a range of study-test delays. Many of the quantitative issues in this debate reflect the issues we have discussed for practice functions. As we previously noted, Wickens (1998) promotes the use of hazard rates, which are similar to relative learning rates, to compare and interpret different forms for the forgetting function. His analyses of short-term memory data supported a relatively constant hazard rate and hence an exponential function. Most analyses of long-term memory data favour a power function (Wixted & Ebbesen, 1991) or a Weibull function (Rubin & Wenzel, 1996), both of which have decreasing hazard rates. Anderson and Tweney (1997) suggested that previous analyses of forgetting functions might have been confounded by averaging over subjects. However, Wixted and Ebbesen (1997) presented a re-analysis of their 1991 data that showed that the power function also applied for individuals.

Recently, Rubin, Hinton and Wenzel (in press) reported recall and recognition experiments that were specifically designed to determine the form of the forgetting function using more trials and delays than in previous data sets. Their analysis supported an exponential function for both long- and short-term forgetting, with the rate of short-term forgetting more than an order of magnitude greater than the rate for long-term forgetting. It is possible, therefore, that decreasing hazard functions for forgetting data are due to a transition between a large hazard rate due to a short-term exponential component and a smaller hazard rate due to long-term exponential forgetting. Clearly, however, further work is needed to check this speculation.

Given the present uncertainties, the mechanism suggested by ACT to explain the form of the practice function, a sum of decaying memory traces, cannot be ruled out. Indeed, it is an attractive mechanism because it can potentially unify results for forgetting and practice functions. However, if ACT is to explain the exponential practice functions found in our survey, and particularly the constant relative learning rate found later in practice, it must modify its assumptions about the trace decay function.

Instance theories. The theory most closely tied to a power practice function is Logan's (1988, 1992) theory based on the minimum time for race among instance retrievals. As Logan (1995) has acknowledged, "A major goal in developing the theory was to account for the power function speedup." (p.751). Logan's theory uses a weak learning mechanism in comparison to other theories of skill acquisition. Each new learning trial speeds performance only because of random variation amongst retrieval time for traces. Hence, it is suited to predicting the power function's decreasing relative learning rate. Our survey included two of the tasks that Logan's theory was directly developed to account for, alphabet arithmetic and counting patterns of dots. These tasks provided some of the strongest evidence for an exponential function in the survey, with an average improvement in fit of almost 25% relative to the power function.

Logan (1988) originally claimed that a power practice function could be derived by asymptotic arguments for racing instances with any retrieval time distribution. However, Colonius (1995) showed that the asymptotic argument is flawed, because the asymptotic distribution of minimum times is degenerate except under linear rescaling that is not justified by Logan's theory. Although Colonius's point reduces the generality of the theory, Logan (1995) countered that a power function is still predicted using non-asymptotic arguments combined with the extra assumption that instance retrieval times have a Weibull distribution.

To accommodate our results, Logan's (1988, 1992) theory must either assume a different instance retrieval-time distribution or add auxiliary mechanisms. One such mechanism could be a race between algorithmic and retrieval processes. However, Logan's (1988) simulations of such a race mechanism did not deviate much from a power function. Secondly, a race between algorithm and retrieval cannot predict exponential learning later in practice, a clear finding of our survey, unless a substantial proportion of responses continue to be algorithmic throughout practice. Further, we found that that both algorithmic and retrieval processes were exponential in paradigms in which subjects indicated the type of processing that they used. It remains an open

question as to whether an alternative retrieval-time distribution or other auxiliary mechanisms can allow Logan's theory to predict an exponential practice function.

Logan's (1988, 1992) theory suggests that learning is quite specific, and so should benefit from pre-experimental practice only if that practice is on a very similar task. Pre-experimental practice is unlikely to be a strong influence on performance in most of the tasks examined in the survey, because the tasks were probably quite unique in the subject's experience. Logan implicitly assumes a negligible pre-experimental practice effect because he fits the power function rather than the general power function. Hence, his instance theory cannot take advantage of the improved fit of the general power function. In any case, the general power function achieved no better fit than the simpler exponential function and required average estimates of pre-experimental practice almost equal to the amount of experimental practice.

EBRW (Nosofsky & Palmeri, 1997) is also based on an instance race, but it includes extra mechanisms, specifically mechanisms reflecting similarity between instances and the accumulation of information via a random walk. The similarity mechanisms make larger estimates of pre-experimental practice, and hence the general power function, more plausible than for Logan's (1988, 1992) theory.

The extra mechanisms may also be able to accommodate an exponential practice function. For example, when similarity is negligible, EBRW's predictions follow Logan's (1988, 1992) theory, but the predictions may diverge as similarity increases. Detailed investigation of this issue is beyond the scope of the present work. However, we note that in Palmeri's (1997) Experiment 2, which used three levels of similarity between stimuli, overall preference for the exponential and APEX function was high (91.3% and 89.6%) and decreased slightly with increasing similarity (96%, 90%, and 88% for exponential preference and 93%, 92% and 84% for APEX preference). If similarity effects explain exponential practice functions, the power function should be preferred more often with low similarity stimuli, not with high similarity stimuli.



### Final Word

We do not claim that the practice function is exactly exponential, or that theories of skill acquisition must exactly predict an exponential function to be taken seriously. The flexible nature of nonlinear functions means that it is difficult to determine the exact form of the practice function. However, our results indicate that the ability of nonlinear functions to imitate each other does not make the form of the practice function an insoluble technical issue. Further, our results indicate that the form of the practice function can provide a useful constraint for theories of skill acquisition.

The most important characteristic of the exponential function is that it has a constant relative learning rate. Any theory that predicts an approximately constant relative learning rate is supported by the results of our survey. The exponential function defines a baseline against which more subtle theoretical predictions can be tested. It provides a single parameter for the rate of learning that can be used to test hypotheses about factors effecting the efficiency of learning. It also provides plausible estimates of asymptotic performance that can indicate the extent to which learning can improve performance.

If a more flexible practice function is required, our results support the APEX function. The APEX function has the added advantage that it contains the power and exponential functions as special cases, so examination of its parameter estimates and nested-model testing can indicate if either simpler function provides a more parsimonious model. The consistently superior fit provided by the APEX function also suggests that non-exponential theories will likely provide the best fit to data if they predict a relative learning rate that decreases only early in practice then remains constant at a value greater than zero later in practice.

The difficult nature of discriminating the correct form may have discouraged others from attempting work in this field: many researchers seem to agree with Anderson (1999, personal communication) that “the exact nature of the practice function will never be resolved.” However, the results of the present survey are encouraging – they allow a clear discrimination between

exponential and power functions as candidates for the Law of Practice. Clear results were made possible by the willingness of researchers to share their data and recent advances in the theory of nonlinear regression (Bates & Watts, 1988; Carroll & Ruppert, 1988). Future research using new techniques such as APEXL fitting (Heathcote & Mewhort, in preparation) and non-parametric regression (Wand & Jones, 1995) coupled with resampling analyses (e.g., Azzalini, Bowman & Hardle, 1989) promise to take these results further and may allow the form of practice functions to be identified with even greater precision. In order to facilitate this enterprise we will make practice data used in this survey available on the World Wide Web<sup>10</sup>. At the time of publication, most survey contributors have agreed to make their data available.

Despite its difficulties, we believe that determination of the mathematical form of empirical laws in psychology is a worthwhile enterprise. Mathematically specified empirical laws both expedite scientific inquiry and guide the development of theory. When we embarked on our survey of the practice function for individual subjects and conditions, we anticipated that the most likely outcome, if the least desirable, would be a variety of function forms for different paradigms and subjects. We were agreeably surprised, therefore, with the consistency of results across the experimental paradigms. The consistency supports Newell and Rosenbloom's (1981) contention that a simple nonlinear function can describe practice effects in a broad range of tasks. However, our survey clearly indicates that best candidate for a parsimonious Law of Practice is the exponential function rather than the power function.

---

<sup>10</sup> From <http://psychology.newcastle.edu.au/> follow the links to the first author's home page.

### References

- Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, *89*, 369-406.
- Anderson, J. R., Fincham, J. M., & Douglass, S. (in press). Practice and retention: A unifying analysis. Journal of Experimental Psychology: Learning, Memory & Cognition.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. Psychological Science, *2*, 396-408.
- Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. Journal of Experimental Psychology: Learning Memory & Cognition, *23*, 932-945.
- Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. Memory & Cognition, *25*, 724-730.
- Azzalini, A., Bowman, A.W., & Hardle, W. (1989). On the use of nonparametric regression for model checking. Biometrika, *76*, 1-11.
- Bates, D. M., & Watts, D. G. (1988). Nonlinear regression analysis and its applications. New York: Wiley.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society B, *26*, 211-246.
- Brown, S. & Heathcote, A. (1997). Unpublished experiment, University of Newcastle, Australia.
- Carrasco, M., Ponte, D., Rechea, C., & Sampedro, M. J. (1998). Transient Structures. Perception & Psychophysics, *60*, 1243-1258.
- Carroll, R. J., & Ruppert, D. (1988). Transformation and weighting in regression. New York: Chapman and Hall.

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

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

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. Ergonomics, *2*, 153-166.

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

Estes, K. W. (1956). The problem of inference from curves based on group data. Psychological Bulletin, *53*, 134-140.

Heathcote, A., (1990). Learned pop-out in search for relative position. Unpublished Ph.D. Thesis, Queen's University at Kingston.

Heathcote, A., & Mewhort, D. J. K. (1993). Selection and representation of relative position. Journal of Experimental Psychology: Human Perception & Performance, *19*, 448-515.

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

Heathcote, A., Brown, S., & Mewhort, D. J. K. (in preparation). Nonlinear regression for response time data: The APEXL technique.

Josephs, R. A., Silvera, D. H., & Giesler, R. B. (1996). The learning curve as a metacognitive tool. Journal of Experimental Psychology: Learning, Memory, & Cognition, *22*, 510-524.

Kail, R., & Park, Y-S. (1990). Impact of practice on speed of mental rotation. Journal of Experimental Child Psychology, *49*, 227-244.

Kirsner, K., & Spelman, C. (1996). Skill acquisition and repetition priming: One principle, many processes? Journal of Experimental Psychology: Learning, Memory, & Cognition, *22*, 563-575.

Kling, J. W. (1971). Learning: An introductory survey. In J. W. Kling & L. A. Riggs (Eds.), Woodworth and Schlosberg's Experimental Psychology (pp. 551-613). New York: Holt, Rinehart, & Winston.

Kramer, A. F., Strayer, D. L., & Buckley, J. (1990). Development and transfer of automatic processing. Journal of Experimental Psychology: Human Perception & Performance, *16*, 505-522.

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

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

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

Luce, R. D. (1986). Response times. New York: Oxford University Press.

Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. Psychological Bulletin, *85*, 1256-1274.

MacKay, D. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. Psychological Review, *89*, 481-506.

Miller, D. M. (1984). Reducing transformation bias in curve fitting. The American Statistician, *38*, 124-126.

Myung, I. J., Kim, C., & Pitt, M. A. (1998). The power artifact: Explained. Paper presented at the Annual Meeting of the Psychonomic Society, Dallas.

Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 1-55). Hillsdale, NJ: Erlbaum.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), Visual information processing (pp. 283-308). New York: Academic Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 1-55). Hillsdale, NJ: Erlbaum.

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

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

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). Numerical recipes The art of scientific computing. New York: Cambridge University Press.

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

Reder, L. M. & Ritter, F. E. (1992). What determines initial feeling of knowing? Journal of Experimental Psychology: Learning Memory & Cognition, 18, 435-451.

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

Rickard, T. C. & Bourne, L. E. (1996). Some tests of an identical elements model of basic arithmetic skills. Journal of Experimental Psychology: Learning Memory & Cognition, 22, 1281-1295.

Ringland, C. & Heathcote, A. (1998). Unpublished experiment. The University of Newcastle, Australia.

Rosenbloom, P., & Newell, A. (1987a). Learning by chunking: A production system model of practice. In D. Klahr, P. Langley, & R. Neches (Eds.), Production system models of learning and development (pp. 221-286). Cambridge, MIT Press.

Rosenbloom, P. S., & Newell, A. (1987b). An integrated computational model of stimulus-response compatibility and practice. In G. H. Bower (Ed.), The psychology of learning and motivation Vol. 21 (pp. 1-52). New York: Academic Press.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. Psychological Review, 103, 734-760.

Rubin, D. C., Hinton, S., & Wenzel, A. E. (in press). The precise time course of retention. Journal of Experimental Psychology: Learning, Memory & Cognition.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R. & Stroffolino, P. J. (1997). To calculate or not calculate. Journal of Experimental Psychology: Learning Memory & Cognition, 12, 1-27.

Seibel, R. (1963). Discrimination reaction time for a 1,023-alternative task. Journal of Experimental Psychology, 66, 215-226.

Sidman, M. (1952). A note on functional relations obtained from group data. Psychological Bulletin, 49, 263-269.

Smith, D. G., & Mewhort, D. J. K. (1994). Unpublished experiment, Queen's University, Canada.

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

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

Strayer, D. L., & Kramer, A. F. (1994c). Aging and skill acquisition. Psychology & Aging, 9, 589-605.

- Thorndike, E.L. (1913) Educational psychology: The psychology of learning (Vol. 2).  
New York: Teachers College Press
- Verwey, W.B. (1996). Buffer loading and chunking in sequential keypressing. Journal of Experimental Psychology: Human Perception & Performance, 22, 544-562.
- Wand, M.P. & Jones, M.C. (1995). Kernel smoothing. London: Chapman and Hall
- Wickelgren, W. A. (1975). Memory storage dynamics. In W. K. Estes (Ed.), Handbook of learning and cognitive processes Vol. 4 (pp. 321-361). Hillsdale, NJ: Erlbaum.
- Wickens, T. D. (1998). On the form of the retention function: Comment on Rubin and Wenzel (1996): A quantitative description of retention. Psychological Review, 105, 379-386.
- Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. Psychological Science, 2, 409-415.
- Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. Memory & Cognition, 25, 731-739.



Author Notes

Thanks to everyone who contributed data to the survey and to reviewers of previous version of this paper including, J. R. Anderson, W. Estes, G. Logan, R. D. Luce, T. Rickard, and J. Wixted.

We would like to acknowledge financial assistance from Australian Research Council Grant 42/280/022 to R. Heath and A. Heathcote, a University of Newcastle Research Management Committee Grant to A. Heathcote, and Grants from Natural Science & Engineering Research Council of Canada to D. J. K. Mewhort.