The Strategy-Specific Nature of Improvement: The Power Law Applies by Strategy Within Task

By: Peter F. Delaney, Lynne M. Reder, James J. Staszewski, and Frank E. Ritter

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.

***** Note: Figures may be missing from this format of the document**

Abstract:

If strategy shifts speed up performance, learning curves should show discontinuities where such shifts occur. Relatively smooth curves appear consistently in the literature, however. To explore this incongruity, we examined learning when multiple strategies were used. We plotted power law learning curves for aggregated data from four mental arithmetic experiments and then plotted similar curves separately for each participant and strategy. We then evaluated the fits achieved by each group of curves. In all four experiments, plotting separately by strategy produced significantly better fits to individual participants' data than did plotting a single power function. We conclude that improvement of solution time is better explained by practice on a strategy than by practice on a task, and that careful assessment of trial-by-trial changes in strategy can improve understanding of the effects of practice on learning.

Article:

The generality and precision simultaneously achieved by expressing empirical regularities as mathematical functions facilitates theoretical development, testing, and the application of scientific knowledge. Although mathematical laws are more prevalent in the physical sciences than in the social sciences, psychology's search for quantitative laws that describe human behavior is long-standing, dating back to the 1850s. A few notable successes have been achieved, including Fitts's law (1954) and the Hick-Hyman law (Hick, 1952; Hyman, 1953).

Newell and Rosenbloom (1981) proposed another candidate for the status of quantitative psychological law. They argued that the *power law of practice*[1] offers a sufficiently accurate, general, and useful characterization of human skill acquisition. This article examines that proposal in the light of empirical evidence that strategy changes sometimes play an important role in cognitive skill acquisition. Such evidence raises questions about the adequacy of the regular power law as a complete descriptor of the temporal course of complex human learning. Our goal is to describe the tension arising between the general formulation of the regular power law and the data on strategy shifts and then to suggest a way to reconcile the two bodies of evidence.

PRACTICE AND SPEED OF PERFORMANCE IN SKILL ACQUISITION

It is well established that practice on a task almost always improves performance, both by reducing the number of errors and by reducing the time required to perform the task. Many longitudinal studies using performance time (e.g., solution time for problems, reaction time to stimuli) to measure skill acquisition have shown a remarkable regularity across a wide variety of tasks in numerous domains (e.g., motor, perception, and cognitive): The relation between practice and performance time is characterized by monotonically diminishing returns. That is, performance speeds up with practice, but the amount of trial-to-trial improvement decreases as practice continues (Anderson, 1983; Chase, 1986; Crossman, 1959; DeJong, 1957; Fitts & Posner, 1967; Nerb, Krems, & Ritter, 1993; Newell & Rosenbloom, 1981; Snoddy, 1926). Figure 1 a illustrates this relation using data cited by Newell and Rosenbloom (198 1) and originally collected by Seibel (1963).

Newell and Rosenbloom, in their efforts to formalize this relation with a variety of data sets, repeatedly found that the relation between practice and performance time could be modeled using a power function, whose general form is shown in Equation 1.

$$T = BN^{-\alpha} \qquad (1)$$

In Equation 1, the parameter T represents the performance time for a given trial, B is the time taken to perform the first trial, and N is the trial number. The rate at which performance time changes is represented by α, with the negative value indicating that performance time decreases at a rate specified by α. This interpretation of α is more readily understood when log transforms are applied to Equation 1, yielding a linear function with α as its slope, as seen in Equation 2.

$$\log[T] = \log[B] - \alpha \log[N] \qquad (2)$$

Figure 1b shows the effect of this transformation graphically when it is applied to the data plotted in Figure 1a. Newell and Rosenbloom (1981) achieved impressive quantitative fits to the data from a variety of skill acquisition studies, justifying their proposal that this functional relation be elevated to the status of psychological law. They were, however, careful to conclude that general power functions provided good approximations to the empirical data. This qualification was prompted in part by residuals that deviated systematically from the fitted functions.

IMPROVEMENT DUE TO STRATEGY SHIFTS

A growing body of evidence suggests that strategy changes can improve skilled performance, and that experts employ strategies that are more effective than those used by novices on the same tasks (e.g., Chi, Feltovich, & Glaser, 1981; Ericsson & Smith, 1991; Larkin, McDermott, Simon, & Simon, 1980; Staszewski, 1990). Studies of mental calculation provide some of the clearest evidence.

Improvements in problem solution times for a variety of arithmetic tasks have been linked to the adoption and adaptive use of different strategies (e.g., Compton & Logan, 1991; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996; Lemaire & Siegler, 1995; Reder & Ritter, 1992; Siegler & Jenkins, 1989; Staszewski, 1988). New strategies frequently reduce the number of intermediate steps in solutions, often by substituting a single memory retrieval step for several computational steps. Replacing multiple time-consuming steps with retrieval decreases solution time accordingly (e.g., Klapp, Boches, Trabert, & Logan, 1991; Staszewski, 1988).

Staszewski (1988), for example, found that extensive practice at solving multidigit multiplication problems without external memory aids led participants to adopt a variety of new strategies. These included replacement of a series of computational steps with direct retrieval of products, adoption of more efficient calculation algorithms, and efficient memory management strategies. Although the participants involved were ordinary undergraduates, their performance at the conclusion of their laboratory training reached the level of established mental multiplication experts.

Compton and Logan (1991) used an "alphabet arithmetic" verification task in which the participants' goal was to determine whether equations such as $A + 3 = D$ were true; in this case, the answer would be "true" because D is three steps away from A in the alphabet. On one sixth of the trials, after completing the verification, participants were asked to press a button corresponding to the strategy they employed. With experience, participants learned to use direct memory retrieval instead of a series of computational steps, thereby reducing solution times.

Reder and Ritter (1992) used an arithmetic task in which participants were to rapidly choose one of two procedures to solve two-digit x two-digit arithmetic problems (e.g., 44 x 18). Participants were told to imagine that they were in a "game show" and had 850 ms to choose to either compute the answer (calculation strategy) or retrieve the answer from memory (retrieval strategy). If they selected the retrieval strategy, they then had about 1 s to state the answer; if they selected the calculation strategy, they had effectively unlimited time to compute the answer. Participants were given an incentive for choosing to retrieve, provided that they could come up with the answer quickly enough. Over the course of the experiment, specific arithmetic problems were repeated numerous times, and participants' tendency to select "retrieve" rather than "calculate" increased accordingly.

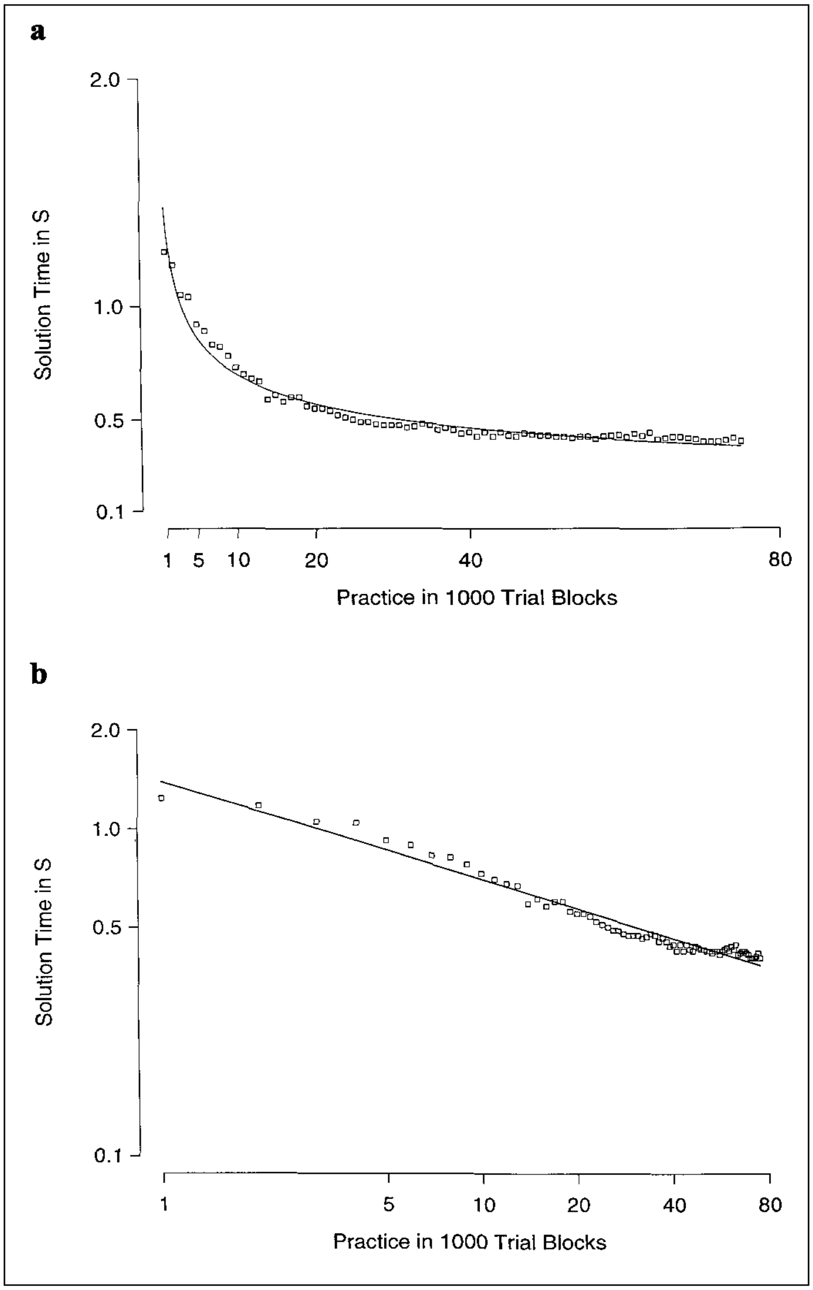

Fig. 1. A regular power function fit to an experiment, with untransformed axes (a) and with both axes log-transformed to linearize the relation (b). Data from Seibel (1963).

Several authors have shown similar effects of strategy changes on young children's arithmetic. For example, Siegler and Jenkins (1989) used concurrent verbal protocols and videos of young children who knew how to add using a simple counting-from-one rule. After 11 weeks of practice, almost all of the children had learned a more efficient rule that involved counting up from the larger addend rather than counting up from one. Children using this more sophisticated counting rule were faster at solving the problems because they had many fewer operations to perform to produce the answers.

In summary, several studies in the domain of arithmetic problem solving indicate that strategy shifts occur along with improvements in solution times. These findings suggest that at least within this task domain, strategy shifts could produce the systematic deviations seen when power functions have been fit to practice-related reductions in solution times. If so, when fitting power functions to practice data, investigators could obtain better accounts of practice-related changes in performance by supplementing practice measures with measures of strategy than by using power functions alone. In particular, our hypothesis is that strategies may change with practice, and that performance improves according to the power law for each individual strategy, even if performance on the task as a whole, with strategies not analyzed separately, does not.

It should be noted that our claim is similar in many ways to one made by Rickard (1997), who has independently pursued a related avenue of research. He fit the power law to data from an arithmetic task that used a synthetic operator similar to one used by Reder and Ritter (1992). The fits obtained were then compared with those produced by fitting strategy-sensitive functions to the same solution times. The latter were generated by first partitioning data points by strategy employed on a given trial.[2] A power function was then fit separately to only those trials associated with each individual strategy. The individual-strategy functions were then compared with the regular power function to determine which better explained the variability in participants' performance. Rickard's finding was that the individual-strategy functions were no better in terms of R^2. Visual inspection of plots of the data, however, suggested that the individual-strategy functions provided a more accurate fit to the data. On this basis, Rickard (1997) argued that individual functions should be fit to each strategy separately.

Our methods are similar in many respects to those of Rickard (1997), but also differ in important ways. We reanalyzed data from four mental arithmetic studies that assessed strategy on every trial. In addition to performing analyses similar to Rickard's, we analyzed the performance of individual participants by comparing, for every participant, the predictions of the regular power function and the appropriate individual-strategy functions. We were then able to compare the regular power law with an alternative form based on improvement of specific strategies.

THE FEELING-OF-KNOWING EXPERIMENTS

A series of three experiments performed by Reder and Ritter (1992) and Schunn, Reder, Nhouyvanisvong, Richards, and Stroffolino (1997) examined feeling of knowing in mental arithmetic problems. Although the present article is not concerned with feeling-of-knowing judgments per se, the data from these experiments provide a valid measure of participants' strategy use that is independent of the solution times for each trial. That information along with trial-by-trial solution times allowed us to trace the effects of strategy shifts on solution time.

Participants were presented with relatively novel two-digit x two-digit arithmetic problems (e.g., 45 x 27) in a random order. In each experiment, half of the problems used the multiply operator, and the other half used a different one. In one of these experiments, the other operator was plus (+); in the other two, the second operator, called *sharp* (#), was invented for the experiment.[3] Problems recurred a variable number of times, so that by the end of the experiment, participants had received different amounts of practice with different problems. Prior to attempting to answer each problem, participants were instructed to make a snap judgment (within 850 ms) as to whether they could retrieve the correct answer directly from memory or whether they would have to calculate the answer. Calculation is much slower than retrieval. This fact was used to verify that the quick decision, or feeling-of-knowing judgment, was an excellent predictor of strategy choice, or at least that the error rates and solution times corresponded closely to what one would expect given that strategy choice (see Reder & Ritter, 1992, or Schunn et al., 1997, for details).

In the analyses reported here, we excluded trials in which participants responded incorrectly because of the difficulty of interpreting solution times on those trials. Excluded trials were treated as missing values (i.e., they still contribute to the amount of practice participants received). However, we did include trials in which participants exceeded the solution time limit (approximately 1 s for retrieval or 18 s for calculation).

Analysis of Group Data

For each experiment and operator, the data were aggregated over participants. For each participant, we took the mean solution time of each problem after a given number of presentations of that problem; we then collapsed over participants. The regular power function was then fit to these aggregate data.

Solution times were also partitioned according to which strategy was used, calculation or memory retrieval. This classification was done on the basis of each participant's explicit strategy selection (i.e., the feeling-of-knowing button press). It should be noted that the strategy transition did not always occur in a stepwise fashion. Participants frequently chose to calculate an answer after they had retrieved it on a prior trial, averaging 2.39 and 2.12 strategy shifts per problem in Reder and Ritter's (1992) experiments and 3.23 switches per problem in the Schunn et al. (1997) experiment. This effect was not due solely to brief fluctuations in strategy choice while adopting a new strategy; for those problems for which more than one switch occurred, the region between the first and last strategy shift amounted to, on average, 52%, 48%, and 42% of the total trials for the three experiments, respectively. Hence, there was rarely a single point where a subject switched to a new strategy and then never switched back.[4] Power functions were then fit to these strategy-specific data subsets, after aggregating over participants.

The fits of the regular power function and individual-strategy power functions are summarized in Table 1. Although the specific strategies are well fit by power functions, there is little reason to prefer them to the non-strategy-specific regular power function, because all of the fits capture at least 95% of the variance in mean solution time. An example fit has been plotted as Figure 2. The figure plots the relationship of solution time to practice on a given problem for each strategy separately, as well as showing the more traditional plot of solution time against practice with a problem, regardless of strategy. As expected, the retrieval strategy is faster than the calculation strategy, and the regular power law predicts solution times somewhere in between.

Table 1. R^2 *values for individual-strategy function fits and regular power function fits for the feeling-of-knowing data aggregated over participants*

Experiment	Calculation R^2	Retrieval R^2	Overall R^2
Reder and Ritter (1992)			
Experiment 1, multiply	.999	.993	.999
Experiment 2, multiply	.999	.979	.998
Experiment 2, sharp	.999	.974	.998
Schunn, Reder, Nhouyvanisvong, Richards, and Stroffolino (1997), Experiment 1			
Multiply	.999	.991	.999
Sharp	.999	.983	.999

Another measure of fit is the autocorrelation of the residuals. Systematic patterning of the residuals following a curve fit may indicate that a predictive variable has not been considered or that an incorrect curve has been fit. Table 2 shows the results of autocorrelating the residuals for both the regular power function and the two individual-strategy functions in each of the experiments. The analysis shows that in two of the five cases, the regular power function estimates deviated systematically from the observed data (and in a third case, the regular power function estimate showed a marginally significant deviation from the observed data). In contrast, only one of the individual-strategy functions showed systematic deviations from the observed data (in a case in which the regular power function fared well).

In summary, analyses of aggregated data suggest that strategy variation affects solution times as hypothesized, but not unequivocally so. The mixed findings could, however, be the result of losing information by aggregation. In the next section, we address this issue through analyses of individual participants' performance.

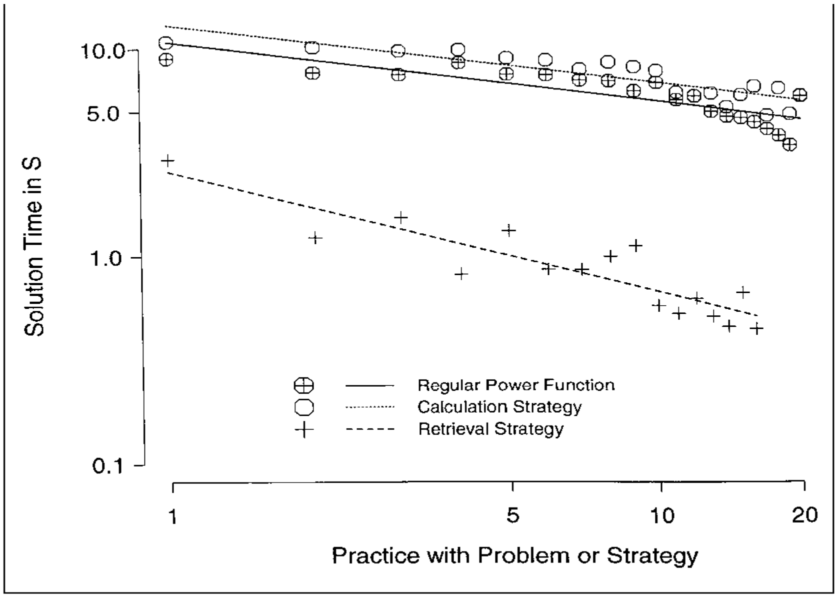

Fig. 2. Example plot of power law fits to the feeling-of-knowing data, with both axes log-transformed to linearize the relation. The retrieval strategy (bottom line) and calculation strategy (top line) curves are plotted as a function of strategy for a given problem, not practice with the problem per se, and have different intercepts from the regular power curve (center line). Data from Reder and Ritter (1992), Experiment 1, multiplication problems (aggregated over participants and problems).

Table 2. *Autocorrelation of residuals (lag = 1) by condition for the feeling-of-knowing data aggregated over participants*

Experiment	Calculation (r)	Retrieval (r)	Overall (r)
Reder and Ritter (1992)			
Experiment 1, multiply	.23	.13	.49**
Experiment 2, multiply	.05	–.12	.29
Experiment 2, sharp	.11	–.27	.45*
Schunn, Reder, Nhouyvanisvong, Richards, and Stroffolino (1997), Experiment 1			
Multiply	.04	.38	.63**
Sharp	.15	.58**	.21

*$p < .10$. **$p < .05$.

Analysis of Individual Participants' Data

Analyses of individual participants' data were conducted according to the general method outlined earlier. Each data point used in the analysis represented a single experimental trial; there was no aggregation over participants or trials in these analyses. Data were grouped for analysis by operator (i.e., all multiplication problems together, all sharp problems together).

First, each participant's solution times were log-transformed and regressed on the log-transformed number of trials of practice with a particular problem (i.e., we applied the power law to get an estimate of solution time). In order to generate a strategy-specific estimate of solution time, we created two single-strategy functions. These were power functions fit to log-transformed solution times and log-transformed number of trials on which the strategy had been used for this problem by the participant (as indicated by the feeling-of-knowing button press). Both of these estimates were then entered as predictors of solution time in a stepwise fashion to determine which provided a better account of improvement on the task.

We reasoned that we would be able to adequately distinguish between the two estimates only if participants used each strategy on at least 30% of the total trials. As we were interested only in those cases in which the two estimates would differ, we made an a priori decision to analyze only the data from participants who used each strategy on 30% or more of all trials for a particular operator. Using this criterion, we made 46 individual participant-by-operator comparisons, 28 for multiply and 18 for sharp (recall that the sharp operator was not used in one of the experiments).

The results of this analysis are presented in Table 3. In every case, the strategy-specific estimate was superior to the regular power function estimate. Furthermore, the regular power function did not account for a significant amount of variance once the strategy-specific estimate was entered, except in one case in which it explained an additional 2% of the variance. Table 3 also shows the incremental variance accounted for by the strategy-specific estimate after the regular power function estimate is forced in. In every case, the strategy-specific estimate accounts for additional variance.

Differences Between Strategies

As the slope and intercept of the power function can vary considerably from task to task (Newell & Rosenbloom, 1981), it seemed probable that the slopes and intercepts for different strategies would also vary. The slopes and intercepts are given in Table 4. However, the different strategies in these tasks did not have significantly different slopes. Two-tailed pair-wise comparisons revealed no significant differences between the slopes of the retrieval and calculation power law learning curves for multiplication, $t(28) = 1.63$, or for sharp, $t(17) = 1.21$. The slopes of the calculation curves for sharp and multiply were not significantly different across participants, $t(15) = 0.90$; neither were the slopes of the retrieval curves for these operators, $t(15) = 0.78$.

Table 3. *Summary of individual-strategy estimates' ΔR^2 values for stepwise regressions for the feeling-of-knowing data*

Experiment	Free-entry $R^{2\,a}$	Forced-entry $\Delta R^{2\,b}$	*N*
Reder and Ritter (1992)			
Experiment 1, multiply	.537 (.125)	.369 (.136)	6
Experiment 2, multiply	.602 (.083)	.510 (.071)	5
Experiment 2, sharp	.620 (.152)	.453 (.144)	4
Schunn, Reder, Nhouyvanisvong, Richards, and Stroffolino (1997), Experiment 1			
Multiply	.633 (.122)	.278 (.179)	17
Sharp	.665 (.156)	.265 (.155)	13

Note. Values enclosed in parentheses represent standard deviations. All values that contributed to these means were significant, $p < .01$ or better.
[a]Estimates for individual-strategy solution time entered the regression model first in the free-entry analyses for all conditions. Estimates based on the regular power function augmented the model for only 1 participant, who contributed to the mean for the sharp problems in Schunn et al. The incremental variance accounted for by this variable in this case was .020.
[b]ΔR^2 values in the forced-entry analyses represent the additional explanatory power achieved by entry of the individual-strategy estimates after variance accounted for by the regular power function estimates has been removed.

The intercepts did differ, with retrieval being faster than calculation for both multiplication, $t(27) = 18.67$, $p < .001$, and sharp, $t(17) = 9.13$, $p < .001$.

EXPERIMENT WITH AN EXPERT MENTAL CALCULATOR

The analyses just described show that the natural transition from calculation to direct memory retrieval that occurs with practice on a problem is better fit by two power functions than by one. To test the generality of this conclusion, we reanalyzed Staszewski's (1988) data. In this study, a single subject (G.G.) practiced mental multiplication for more than 600 sessions. After the 500th session, G.G. was taught a new strategy. This data set allowed us to look at long-term strategy changes following explicit instruction. It also differs from the feeling-of-knowing experiments in that strategies were not interleaved; rather, the introduction of a new strategy occurred at a single transition point.

Method

Problems from nine problem-size categories were presented either orally or visually. For visually presented problems, the problem operands remained available to the participant throughout his computation. The problem-size categories were differentiated by the size of problem operands: Multipliers were one or two digits, and multiplicands ranged from one to five digits. An example from the problem-size category called 2 x 4 would be 69 x 4,957. Problems were randomly generated for each problem size during the first phase of the experiment, so repetitions were rare (unlike in the feeling-of-knowing experiments). During the second phase, problems from the last 100 practice sessions of the first phase were repeated in order.

During the first phase, G.G. was instructed to use an unconventional general solution strategy commonly used by expert mental calculators (Smith, 1983). Unlike the usual strategy, which works from right to left, the experts' strategy works in the opposite direction. That is, they start computations with the digits of the highest magnitude in both operands and proceed to the lowest. G.G. was instructed to work as quickly and accurately as possible, although accuracy was emphasized if his aggregate error rate exceeded 10%. He took 20 sessions to achieve this criterion, which he never again exceeded, maintaining an average error rate of 5.2% over the remainder of his training. His consistently high accuracy minimizes any interpretation problems arising from speed-accuracy trade-offs.

Table 4. *Mean slope and intercept values of individual-strategy power functions for the feeling-of-knowing data*

Condition	Mean slope (α)	Mean intercept (log B)
Multiply, retrieval strategy	0.54 (0.53)	3.21 (0.25)
Multiply, calculation strategy	0.37 (0.18)	3.99 (0.13)
Sharp, retrieval strategy	0.11 (1.15)	2.97 (0.90)
Sharp, calculation strategy	0.25 (0.88)	3.77 (1.14)

Note. All means are based on the slope and intercept for the linear form of the power function (Equation 2). Values in parentheses represent standard deviations.

Starting with the 501st practice session, the second phase began, and G.G. was instructed to use a new strategy that applied only to the problems with two-digit multipliers (i.e., 2 x *n*s, where *n* is 2, 3, 4, or 5). This new strategy offers greater efficiency because it reduces memory load relative to the original two-place multiplication procedures (consult Staszewski, 1988, for further details). During his 1 day's practice with this new strategy, G.G. gave concurrent verbal reports. In later sessions, retrospective reports were used to monitor strategy use, and he consistently reported using the new strategy. Moreover, G.G. was also required to give concurrent verbal protocols of how he was solving the problem during approximately half of these later sessions, and these protocols also revealed consistent use of the new strategy.

Comparison of Performance by Problem Condition

The data were fit separately to a power function for each mode of presentation for each of the nine problem-size categories. Each data point in our analysis is the median correct solution time over five experimental sessions for that problem size (15 presented problems). The analysis procedure is identical to that used for the feeling-of-knowing data sets, with one exception: The power function used in fitting included an empirically determined asymptote for each problem size (based on Staszewski, 1988). In other words, we generated two estimates of solution time, one based on the regular power law (the regular power function estimate) and the other based on fitting two separate functions to the data set, splitting the trials into those from the first phase using the old strategy and those from the second phase using the new strategy (the multiple-strategy estimate). These two estimates were then employed as predictors of solution time in stepwise regressions.

We predicted that for 2 x 2 to 2 x 5 problems, to which the new strategy applied, we would see an improved fit by taking into account strategies, while in the 1 x 2 to 1 x 5 problems, for which the new strategy had no effect, we would not see an improved fit. The results of the analysis are shown in Table 5. In analyses in which predictors with the highest F value entered the regression models first (column 1), the multiple-strategy estimate entered first for each problem condition. The regular power function estimate accounted for incremental variance in only one condition (2 x 5 oral) in which strategy was manipulated (i.e., 2 x 2, 2 x 3, 2 x 4, 2 x 5), and the amount was small (0.5%). In analyses that forced the regular power function estimate into the model first, the multiple-strategy estimate always accounted for a significant amount of incremental variance for the conditions in which the new strategy applied (column 2).

For the majority of those cases in which the new strategy did not apply (i.e., the new strategy should not have influenced performance), the multiple-strategy estimate did not account for incremental variance once the regular power function estimate was forced in. The small improvement due to the multiple-strategy estimate in some of these cases may be due to other concurrent endogenous local strategy changes found in G.G.'s solutions to specific subproblems. The evidence indicates that subproblems can be identified in problems with multidigit operands, and that direct retrieval supplants multistep computation for an increasing proportion of 1 x 2 and 1 x 3 subproblems with practice (Staszewski, 1988). Additional small endogenous strategy shifts cannot be ruled out.

Figure 3 provides example plots for G.G.'s solution times in the 2 x 5 visual and 1 x 5 visual conditions, blocked for display in 15-session practice blocks. Figure 3a shows the fits of the regular power function estimates (including asymptotes computed by Staszewski, 1988, p. 115) for the 1 x 5 problems and for the 2 x 5 problems. Additionally, a third line has been included to show how improvement on the new strategy deviates from improvement on the old. There are two things to note about this additional line. One is that the first few points after the new strategy is introduced are above the regression line for the older strategy. This is because when people first learn a new strategy, it often takes longer to use it than a more practiced strategy. The other observation is that the slope of the new function implies that by Session 1,000, G.G. would be answering 2 x 5 problems as quickly as 1 x 5 problems, which seems unlikely. This is an artifact of starting the plot of a new strategy far along in the practice curve, however. When the same data are replotted using our individual-strategy formulation in Figure 3b, the slope of the new strategy function is more sensible. Here, the new-strategy data for the 2 x 5 problems are plotted starting from Block 1, because if it is assumed that practice occurs by strategy rather than by task, then there is no practice on the new strategy until it is introduced.

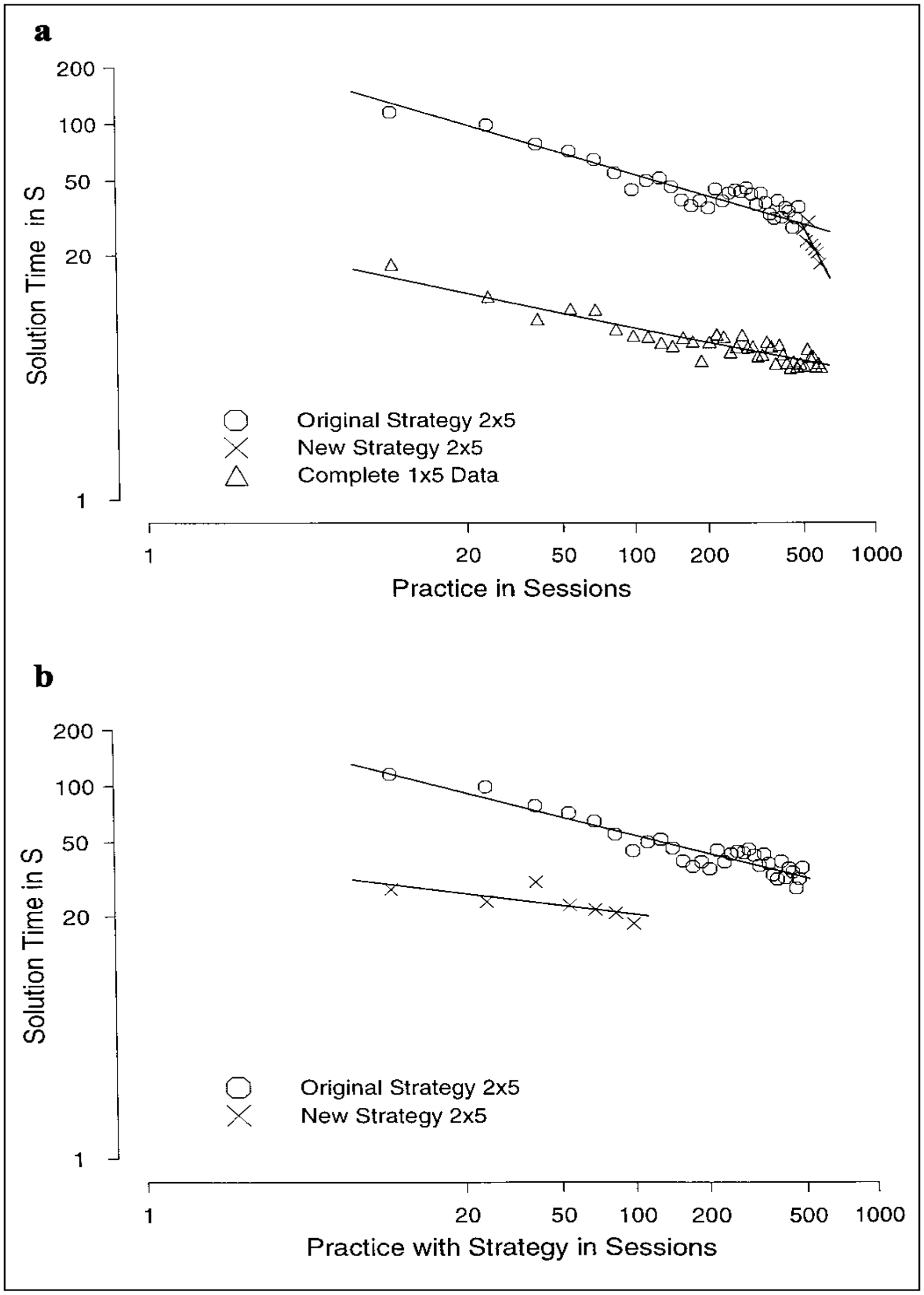

Fig. 3. Comparison of regular power functions and individual-strategy power functions. Shown in (a) is an example plot of the regular power function estimates of solution times produced by an expert mental calculator (G.G.) for 1 × 5 (one-digit × five-digit) and 2 × 5 (two-digit × five-digit) visually presented mental multiplication problems, as a function of practice. Each data point represents 15 sessions. Also shown is a power function fit to the subset created by introducing a new strategy at Session 501, plotted to show the nonlinearity of the log-log plot when the new strategy is introduced. Shown in (b) are example plots of the individual-strategy power function estimates of the same data, plotted as a function of practice with the strategy rather than with the task. Data from Staszewski (1988).

Table 5. *Summary of individual-strategy estimates' ΔR² values for stepwise regressions for the expert mental calculator data*

Condition	Free-entry $R^{2\,a}$	Forced-entry $\Delta R^{2\,b}$
1 × 1 oral	.274	.066
1 × 2 oral	.779	.008
1 × 3 oral	.757	—
1 × 4 oral	.713	.011
1 × 5 oral	.797	—
2 × 2 oral	.883	.012
2 × 3 oral	.931	.010
2 × 4 oral	.889	.018
2 × 5 oral	.888	.060
1 × 1 visual	.679	.013
1 × 2 visual	.801	—
1 × 3 visual	.734	—
1 × 4 visual	.781	—
1 × 5 visual	.772	—
2 × 2 visual	.848	.017
2 × 3 visual	.892	.004
2 × 4 visual	.822	.031
2 × 5 visual	.794	.049

Note. All numerical entries in this table were significant, $p < .05$.
[a] Estimates of individual-strategy solution time entered the regression model first in the free-entry analyses for all conditions. Estimates based on the regular power function augmented the model only for the 2 × 5 oral problems. The incremental variance accounted for by this variable in this case was .005.
[b] ΔR^2 values in the forced-entry analyses represent the additional solution time variance accounted for by the individual-strategy estimates after the variance accounted for by the regular power function estimate has been removed.

GENERAL DISCUSSION

Two general concerns motivated this work: theoretical unification and progress. Cognitive psychology's apparent acceptance of the power law of practice is inconsistent with growing evidence that strategy shifts play an important role in cognitive skill acquisition. We hoped to reconcile the two typically unrelated bodies of data that produce this tension. Our results indicate that such a reconciliation is possible, and we offer one form that it can take. The explanatory capability of the power law can be improved by assuming that power-law improvement occurs relative to strategies, not tasks. We found that composite functions interleaving point predictions generated by fitting strategy-specific power functions to partitioned sets of solution times yield better accounts of observed behavior than the regular power law. This improvement was illustrated both when the strategy shift was discrete and explicitly taught and when the shift from one strategy to another was gradual and motivated by task constraints. As noted in the introduction, this conclusion has been independently verified by Rickard (1997). The generality of these findings is also strengthened by the work of Lovett and Anderson (1996), whose analyses of simple problem-solving tasks yielded additional converging results.

Our second concern involves the de facto status of the power law of practice as a benchmark for evaluating and comparing unified theories of cognition otherwise known as cognitive architectures. Explanations of the smooth, continuous changes in solution time consistent with the regular power law have become a standard for evaluating such large-scale theories (e.g., Anderson, 1983; Newell, 1990), despite reservations about the power law's accuracy as a descriptor of human learning (Chase, 1986). It now seems clear that a composite of strategy-sensitive power functions represents a more precise description than the regular power law (and may help indicate where such strategy shifts occur). To remain viable, candidate architectures that seek to explain the acquisition of complex skills must accommodate strategy change and strategy variation, and account for the independent speedup of each strategy. Clearly, this effort will require understanding the conditions that determine strategy changes as well as how and why new strategies emerge. This more accurate prediction comes at a price of more detailed analysis.

Notes:

1. This regularity was first noted by Lewis (1976) in an unpublished manuscript that Newell acknowledged reading.

2. Rickard had strategy information for only one third of all trials. He used logistic regression to interpolate for the remaining trials.

3. Sharp was defined for two-digit x two-digit arithmetic problems by the pseudoequation *AB* # *CD* = [(*A* + *C*) x (*B* + *D*) x 3] modulo 100. For example, 34 # 56 = [(3 + 5) x (4 + 6) x 3] modulo 100 = 40. Application of this rule creates an operation roughly equivalent in performance time to multiplication (Reder & Ritter, 1992). Note that this operation is not identical to Rickard's (1997) pound operator, which is defined by the equation $a \# b = [(b - a) + 1] + b = 2b - a + 1$.

4. Monotonic shifts do occur for some subjects, however, as reported by Ritter, Reder, and Newell (1989). This finding was one of the direct precursors to our work reported here.

REFERENCES

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Chase, W.G. (1986). Visual information processing: Acquisition of visual-spatial skills. In K.R. Boff, L. Kauffman, & J.P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 53–69). New York: John Wiley & Sons.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.

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

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

DeJong, R.J. (1957). The effects of increasing skill on cycle-time and its consequences for time-standards. *Ergonomics*, *1*, 51–60.

Ericsson, K.A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge, England: Cambridge University Press.

Fitts, P.M. (1954). The information capacity of the human motor system in controlling amplitude of movement. *Journal of Experimental Psychology*, *47*, 381–391. Fitts, P.M., & Posner, M.I. (1967). *Human performance*. Belmont, CA: Brooks/Cole.

Hick, W.E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*, 11–26.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*, 188–196.

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

Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335–1342.

LeFevre, J., Bisanz, J., Daley, K.E., Buffone, L., Greenham, S.L., & Sadesky, G.S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General*, *125*, 284–306.

LeFevre, J., Sadesky, G.S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 216–230.
Lemaire, P., & Siegler, R.S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*, 83–97.
Lewis, C. (1976). *Speed and practice*. Unpublished manuscript, University of Michigan, Ann Arbor.
Lovett, M.C., & Anderson, J.R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology*, *31*, 168–217.
Nerb, J., Krems, J.F., & Ritter, F.E. (1993). Rule learning and the power law: A computational model and empirical results. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 765–770). Hillsdale, NJ: Erlbaum.
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the power law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1– 55). Hillsdale, NJ: Erlbaum.
Reder, L.M., & Ritter, F.E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435–45 1.
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.
Ritter, F.E., Reder, L.M., & Newell, A. (1989). Modeling a feeling of knowing. In *Proceedings of the Soar VI Workshop* (pp. 1–22). Marina del Rey, CA: University of Southern California, Information Sciences Institute.
Schunn, C.D., Reder, L.M., Nhouyvanisvong, A., Richards, D.R., & Stroffolino, P.J. (1997). To calculate or not calculate: A source activation confusion (SAC) model of problem-familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1–27.
Seibel, R. (1963). Discrimination reaction time for a 1,023 alternative task. *Journal of Experimental Psychology*, *66*, 215–226.
Siegler, R.S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum. Smith, S.B. (1983). *The great mental calculators: The psychology, methods, and lives of calculating prodigies, past and present*. New York: Columbia University Press.
Snoddy, G.S. (1926). Learning and stability: A psychophysiological analysis of a case of motor learning with clinical applications. *Journal of Applied Psychology*, *10*, 1–36. Staszewski, J.J. (1988). Skilled memory and expert mental calculation. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), *The nature of expertise* (pp. 71–128). Hillsdale, NJ: Erlbaum.
Staszewski, J.J. (1990). Exceptional memory: The influence of practice and knowledge on development of elaborative encoding strategies. In W. Schneider & F. Weinert (Eds.), *Interactions among aptitudes, strategies, and knowledge in cognitive performance* (pp. 252–285). New York: Springer-Verlag.