Journal of Experimental Psychology:
Learning, Memory, and Cognition
1999, Vol. 25, No. 2, 532–542

0278-7393/99/$3.00

A CMPL Alternative Account of Practice Effects in Numerosity Judgment Tasks

Timothy C. Rickard
University of California, San Diego

A reanalysis of the numerosity judgment data described in T. J. Palmeri (1997) showed that the mean latency exhibits clear deviations from the power function as predicted by the component power laws (CMPL) theory of strategy shifting (T. C. Rickard, 1997). The variance of the latency systematically increases and then decreases with practice for large numerosities, a result that is also consistent with the CMPL theory. Neither of these results are predicted by existing versions of either the exemplar-based random walk or the instance theories. These findings suggest that numerosity judgment, like other skills, reflects one at a time rather than concurrent execution of algorithmic and memory retrieval strategies.

Palmeri (1997) reported results of three numerosity judgment experiments designed to test the exemplar-based random walk (EBRW) model (see Nosofsky & Palmeri, 1997) of categorization and automaticity, which synthesizes critical components of Nosofsky's (1986) generalized context model and Logan's (1988) instance theory of automatization. The experiments involved as many as 208 repeated presentations, over multiple sessions, of sets of randomly constructed patterns of 6 to 11 dots. The task was to determine the numerosity as quickly as possible. Overall, the EBWR model provided an impressive account of the data. Consistent with this theory, (a) there was clear evidence of a transition with practice from using a sequential dot counting strategy to retrieval of the answer directly from memory, (b) the mean reaction time (RT) depended in part on the similarity of each stimulus to other stimuli, and (c) speedup with practice appeared to follow a three parameter power function,

$$\mathrm{RT} = a + b \cdot N^{-c}, \tag{1}$$

where N is number of practice trials, a is asymptotic RT, and b and c are scaling parameters (see Newell & Rosenbloom, 1981).

Both the EBRW and instance theories assume that memory retrieval and the algorithm race independently on each performance trial. Rickard (1997) proposed an alternative component power law (CMPL) model that embodies the opposite assumption that memory retrieval and the algorithm cannot be executed in parallel. Rather, one each trial, either memory retrieval or the algorithm, but not both, are executed.[1] Rickard (1997) showed through mathematical analysis and simulation that this assumption, combined with auxiliary assumptions, yields a model that predicts unique power function speedup in mean RT within each strategy, but systematic deviations from the power function in the overall data (i.e., when collapsed across strategy).

The CMPL theory was originally developed to account for arithmetic and related tasks. It is an open question whether the predictions of that theory hold for other tasks, like numerosity judgment, that differ from arithmetic in several nontrivial ways. As one example, the counting algorithm used in numerosity judgment is probably not particularly challenging, nor is it very demanding on working memory. Under such conditions, there may be sufficient resources to execute algorithm and retrieval strategies simultaneously. In contrast, the pound arithmetic task used by Rickard (1997) involves execution of a multistep arithmetic algorithm that may be subjectively more demanding and clearly imposes at least some load on working memory. Under these conditions, participants may not be able to, or may choose not to, execute both strategies in parallel. This fact, considered along with other task differences, makes it far from obvious that the CMPL theory will generalize to numerosity judgment.

CMPL Predictions Expressed as Equations

The strategy execution assumptions of the CMPL theory imply a mixture of two distinct strategies on each

I am indebted to Thomas Palmeri for graciously making the data sets available.

Correspondence concerning this article should be addressed to Timothy C. Rickard, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0109. Electronic mail may be sent to trickard@ucsd.edu.

[1] Siegler (1988) and Lemaire and Siegler (1995) made similar assumptions in their theory of children's strategy choices in arithmetic and related domains.

block of practice. Speedup in the overall mean RT is thus governed by a mixture equation (Townsend & Ashby, 1983),

$$RT = RT_{(algorithm)} \cdot (1 - p) + RT_{(retrieval)} \cdot (p), \quad (2)$$

where p is the proportion of trials on which the retrieval strategy is used on a given practice block, $1 - p$ is the proportion of trials on which the algorithm strategy is used, and $RT_{(algorithm)}$ and $RT_{(retrieval)}$ describe the mean RTs for algorithm and retrieval strategies, respectively, as a function of practice. Speedup reflects both faster execution of each strategy with practice and a steady increase in the proportion of trials on which the faster retrieval strategy is used.

Simulation results of Rickard (1997) showed that, given a reasonable set of auxiliary assumptions, the mean RTs for each strategy follow power functions,

$$RT_{(algorithm)} = b1 \cdot (N + pre)^{-c1}, \text{ and} \quad (3)$$

$$RT_{(retrieval)} = b2 \cdot N^{-c2}, \quad (4)$$

where pre is previous learning, which must strictly be included in the model for the algorithm. For simplicity, asymptotes can be assumed to be zero with negligible consequences in most cases (see Rickard, 1997; Newell & Rosenbloom, 1981).

The CMPL simulation model discussed by Rickard (1997) was designed for arithmetic and related tasks. New simulation results described in the Appendix demonstrate that a version of the model adapted to the numerosity judgment task makes the same basic quantitative predictions. The results in the Appendix also demonstrate that the simulation prediction for p, the proportion of trials on which the algorithm strategy is used as a function of practice, can be closely approximated by a simple one parameter negative exponential function,

$$p = 1 - e^{-r(N-1)}, \quad (5)$$

where r is the rate constant. When r is large, this function approaches asymptote quickly and the transition to retrieval occurs in only a few trials. When r is small, the function approaches asymptote slowly and the transition to retrieval occurs only after many trials.

Substitution of Equations 3, 4, and 5 into Equation 2 yields a close approximation of the CMPL simulation predictions for the overall mean RT collapsed across items and participants.[2] Figure 1 shows graphically this CMPL RT function in a hypothetical (and noise free) data set. Also shown in the figure are the component power functions for each strategy in the hypothetical data. For comparison, the best fitting three-parameter power function to the CMPL function is also shown. The EBRW and instance theories can predict speedup overall which

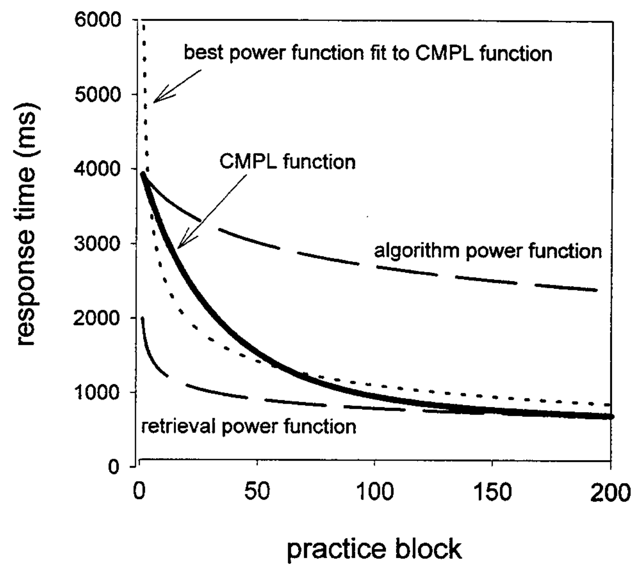

Figure 1. Predictions of the component power laws (CMPL) theory for the overall mean reaction time for a hypothetical data set.

is a close approximation to a three parameter power function, and this fact has been treated as providing support for the theories (Logan, 1988; Palmeri, 1997). It is important to note that these theories are only constrained to predict exact power function speedup for the retrieval strategy (due to the properties of the race among instances). When there is a transition from algorithm to retrieval, they can predict deviations from the power function during the practice interval over which the transition occurs. However, the specific deviations from the power function that are possible in these models have not been explicitly identified. In particular, neither of these theories have to date been shown to generate speedup effects that follow the CMPL function depicted in Figure 1. Thus, a demonstration that the numerosity judgment data do not follow the power function, but rather follow the CMPL function, would constitute a significant challenge.

Power Function Versus CMPL Fits to the Palmeri (1997) Data

Three parameter power functions, including asymptote, were fit to the overall data for each numerosity for each of

[2] Equations 2 through 5 are a close approximation to the CMPL simulation predictions. However, when data are collapsed over items and participants in the simulation, RTs are predicted to deviate from the power function for the retrieval trials during roughly the first half of the strategy transition (see Rickard, 1997). I assume for current purposes that these deviations from the power functions are negligible. If they are not negligible, they would only be expected to decrease the quality of the fits.

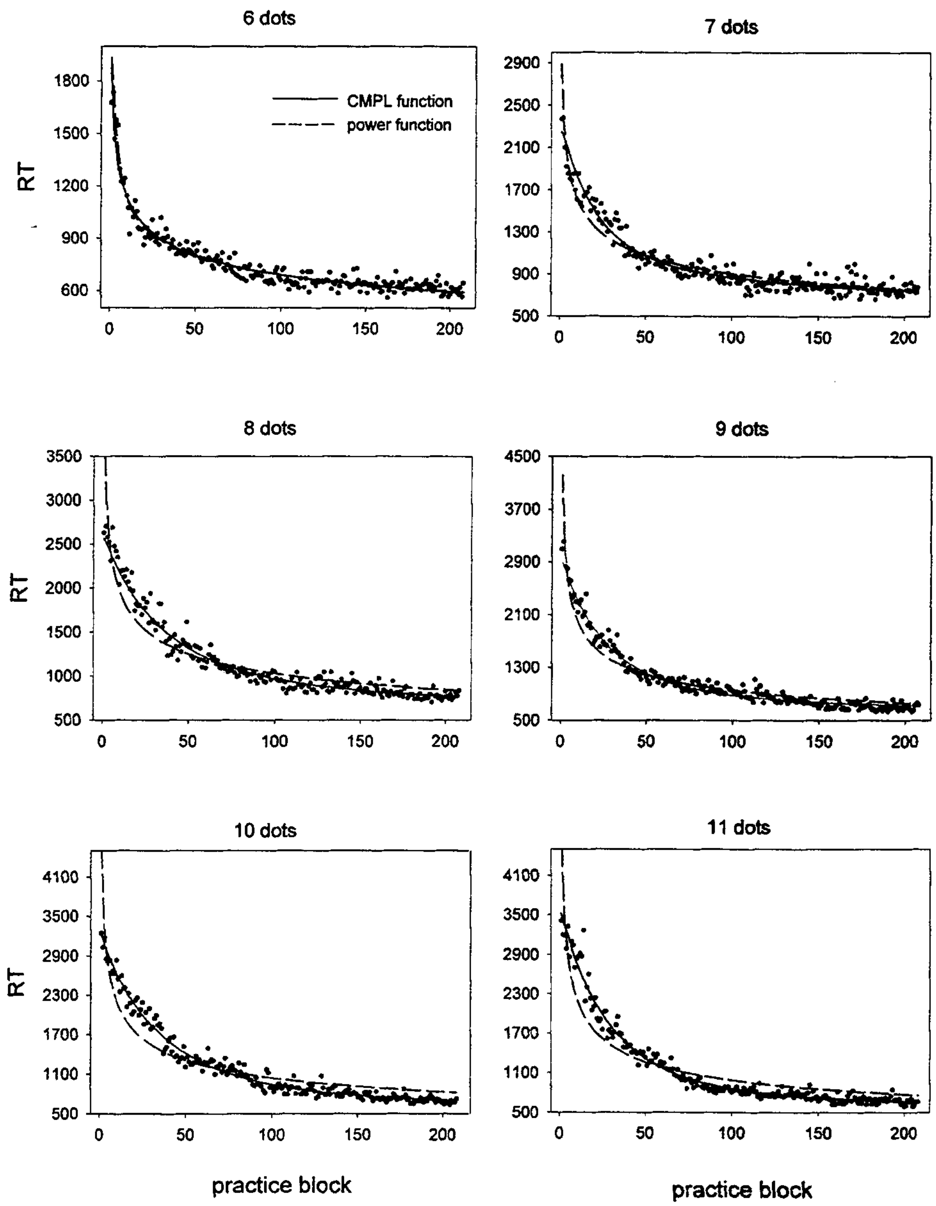

Figure 2. Power function and component power laws (CMPL) fits to the reaction time (RT) data of Experiment 2.

the three Palmeri (1997) experiments as shown by the hatched line in the plots in Figures 2, 3, and 4 (for Experiments 1, 2, and 3, respectively).[3] For Experiments 2 and 3, data are collapsed across item similarity. For each experiment, a total of 18 free parameters are required for these fits. The power function fits are good for the six dot patterns but become progressively poorer for larger numerosities. For Numerosities 8 through 11, the power function clearly is not providing a good characterization of the data. Indeed, for these numerosity levels in Experiments 1 and 2, the power function systematically underestimates actual RT by as much as 500 ms or more between about the 10th and 20th block of practice. These

[3] Data for individual items (correct trials only) were logged, then averaged across items, and then averaged across participants. The inverse log of these averages was then computed, and regression analyses were performed on these data. This procedure is equivalent to performing the analysis on the geometric mean of the data. This approach is less sensitive to distortions caused by possible differences in parameter values of the power functions for individual items. If the power function is true for the expected value at the item level, it will be preserved in the

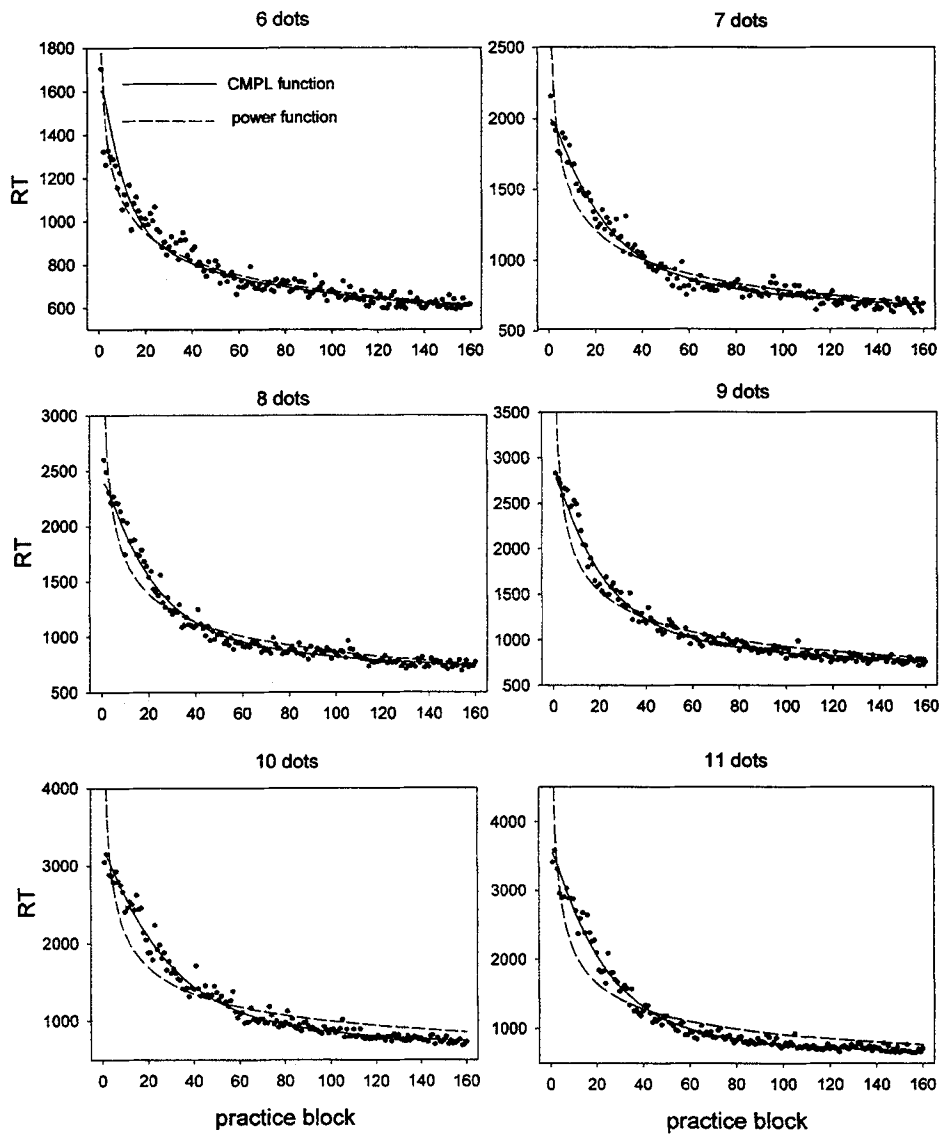

Figure 3. Predictions of the component power laws (CMPL) theory for the overall *SD* for a hypothetical data set. RT = reaction time.

are substantial deviations for RT curve fitting, and especially so in light of the fact that actual RTs were only about 2,000 ms during this interval of practice. Further, after about practice Block 50, the power function systematically overestimates the data. Indeed, for Numerosities 10 and 11 in Experiments 1 and 2, there is almost no overlap between the data and the power function prediction after about Block 50.

aggregate data by taking the geometric mean, provided the asymptote value is zero or is first subtracted, and that *pre* is roughly constant across items and participants. Thus, the geometric mean is susceptible to fewer sources of distortion than is the arithmetic mean.

If fit separately to each of the six numerosities of a given experiment, the CMPL RT function would require a total of 36 (6 × 6) free parameters. However, the theory suggests several reasonable constraints that allow the number of free parameters to be vastly reduced if data for a given experiment are fit simultaneously over all six numerosity levels. First, it is reasonable to assume that previous learning, *pre*, is the same value for all numerosities. Second, because the RT for the counting algorithm is assumed to be a linear function of numerosity, the *b1* parameter for the algorithm can be expressed as *b1* for Numerosity 6, plus a constant increment parameter, *inc*, which is multiplied for each numerosity by the difference between that numerosity and 6. Third, the

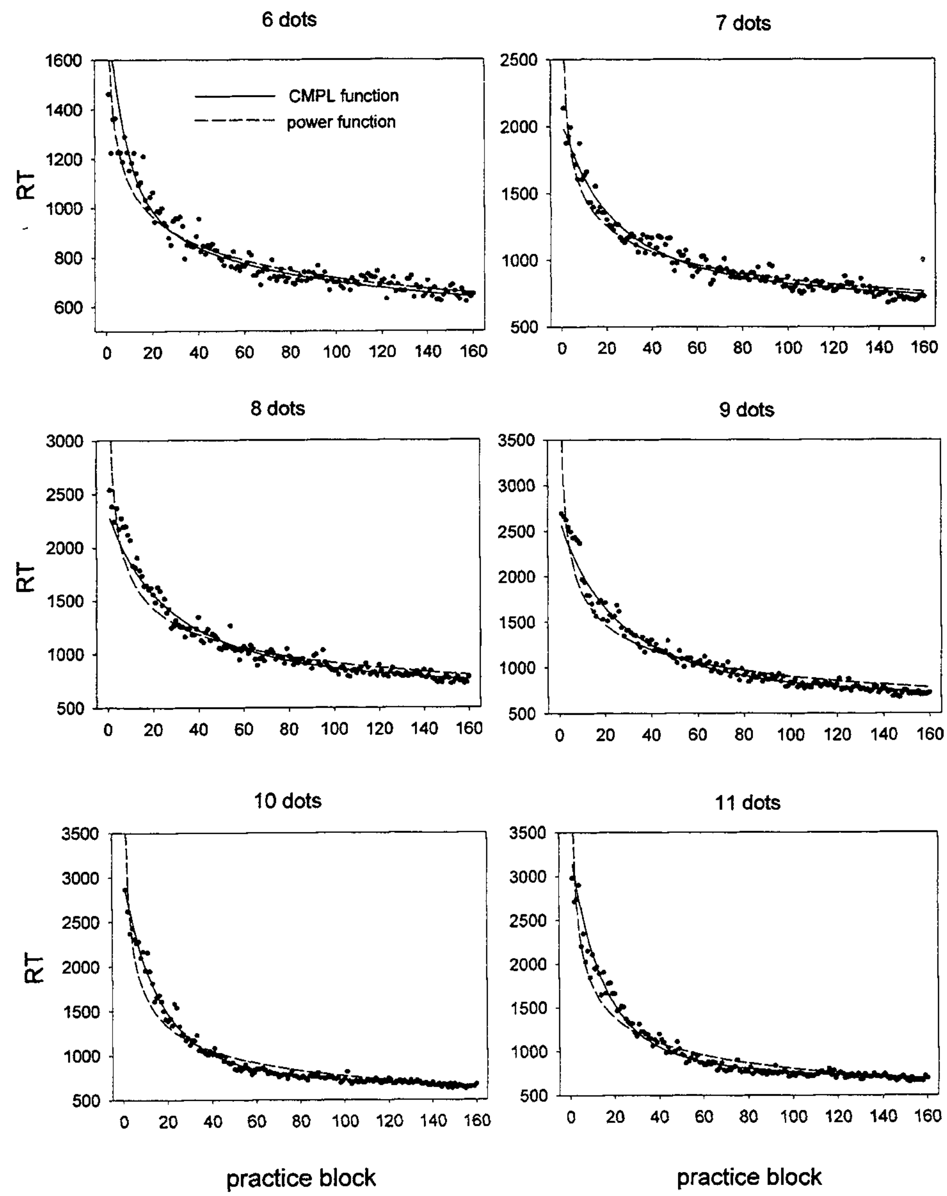

Figure 4. Power function fits to the standard deviation (*SD*) data of Experiment 2. CMPL = component power laws; RT = reaction time.

learning rate (c) can be assumed to be the same for all numerosities within the algorithm and retrieval strategies, respectively. Although the CMPL model as developed by Rickard (1997) does not require the rates to be the same for algorithm and retrieval strategies, this additional constraint appears reasonable and is included in the current fits to further reduce complexity. Finally, Palmeri (1997) produced evidence that the retrieval RTs and the number of trials necessary to make the transition to retrieval were different for different numerosities. Specifically, there appear to be anchor effects such that at least Numerosities 6 and 11 have faster retrieval RTs than do other numerosities. To accommodate the possibility of differing retrieval RTs, the values r and $b2$ were allowed to take different values for each numerosity. In summary, the above constraints allow the CMPL model to be fit simultaneously to all 6 numerosities with a total of 16 free parameters. These constraints put the number of parameters in the same range as that for the three parameter power function, and they reduce the complexity of the nonlinear regression equations to a more manageable level at which stable fits can be obtained. The corresponding equations for the predicted overall mean RT for each numerosity are as follows:

$$RT_6 = b1 \cdot (N + pre)^{-c} \cdot [1 - e^{-r6 \cdot (N-1)}] + b2_6 \cdot (N)^{-c} \cdot [e^{-r6 \cdot (N-1)}],$$

$$RT_7 = (b1 + inc) \cdot (N + pre)^{-c} \cdot [1 - e^{-r7 \cdot (N-1)}] + b2_7 \cdot (N)^{-c} \cdot [e^{-r7 \cdot (N-1)}],$$

$$RT_8 = (b1 + 2 \cdot inc) \cdot (N + pre)^{-c} \cdot [1 - e^{-r8 \cdot (N-1)}] + b2_8 \cdot (N)^{-c} \cdot [e^{-r8 \cdot (N-1)}],$$

$$RT_9 = (b1 + 3 \cdot inc) \cdot (N + pre)^{-c} \cdot [1 - e^{-r9 \cdot (N-1)}] + b2_9 \cdot (N)^{-c} \cdot [e^{-r9 \cdot (N-1)}],$$

$$RT_{10} = (b1 + 4 \cdot inc) \cdot (N + pre)^{-c} \cdot [1 - e^{-r10 \cdot (N-1)}] + b2_{10} \cdot (N)^{-c} \cdot [e^{-r10 \cdot (N-1)}],$$

$$RT_{11} = (b1 + 5 \cdot inc) \cdot (N + pre)^{-c} \cdot [1 - e^{-r11 \cdot (N-1)}] + b2_{11} \cdot (N)^{-c} \cdot [e^{-r11 \cdot (N-1)}].$$

These equations were fit to the RT data simultaneously for all numerosities for each experiment using the Proc NLIN program (SAS Institute, 1994). The fits are shown by the solid lines in Figures 2, 3, and 4. The CMPL function clearly characterizes the data better overall than does the three parameter power function, in terms of both the visual fit and overall r^2 (.972, .982, and .972 for the CMPL fits to Experiments 1, 2, and 3, respectively; .870, .871, and .920, for the three parameter power function fits to Experiments 1, 2, and 3, respectively). Indeed, across 18 data sets from the three experiments, there were no clear, systematic, and replicable deviations of the data from the CMPL predictions. Also as predicted, the improvement in fit provided by the CMPL function clearly increases with increasing algorithm RT (see also Rickard, 1997, Experiment 2).

Note that the r^2 for the power function fits are significantly lower than those reported by Palmeri (1997). The primary reason for this is that Palmeri reported r^2 based on data collapsed over practice blocks within each session, whereas the results reported here were not collapsed over session. Collapsing over session reduces noise, but at the potential expense of also greatly reducing sensitivity to any systematic deviations from the power function that may be present in the data.

The parameter estimates for the CMPL fits for each experiment are shown in Table 1. Several patterns are worth noting. First, all parameters take on reasonable values, indicating that the fits are not due to parameter flexibility outside of a psychologically reasonable range. Second, the rate estimate, r, which estimates how quickly the transition to retrieval occurs, indicates that the 50% strategy transition point (i.e., the point at which p in Equation 4 reaches a value of .5) occurred between about Block 10 and 25 for Numerosities 7 through 11. However, for Numerosity 6, the fit indicates a very fast transition to retrieval, within the first few blocks of practice. It is not clear why a super fast transition to retrieval might occur for six dot patterns alone. One reasonable possibility is that six dot patterns are much easier to map onto canonical forms, thus facilitating a fast transition to retrieval. Note that a fast transition to retrieval

Table 1
Parameter Estimates for the CMPL Fits for Each Experiment

Parameter	Exp. 1	Exp. 2	Exp. 3
pre	16.61	298	9.178
c	.2077	.2042	.1931
$b1_6$	3507	5133	2647
inc	576.7	1250	452.8
r_6	.3966	.1478	.1380
r_7	.0458	.0572	.0378
r_8	.0233	.0570	.0275
r_9	.0333	.0544	.0301
r_{10}	.0229	.0389	.0535
r_{11}	.0286	.0464	.0592
$b2_6$	1795	1707	1703
$b2_7$	2246	1923	1984
$b2_8$	2302	2091	2010
$b2_9$	2173	2175	1930
$b2_{10}$	2088	2080	1742
$b2_{11}$	1976	1888	1850

Note. CMPL = component power laws; Exp. = experiment.

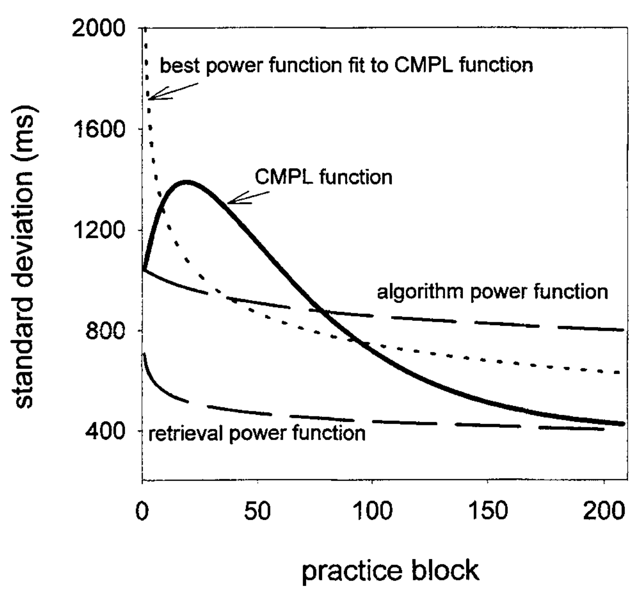

Figure 5. Power function and component power laws (CMPL) fits to the reaction time data of Experiment 1.

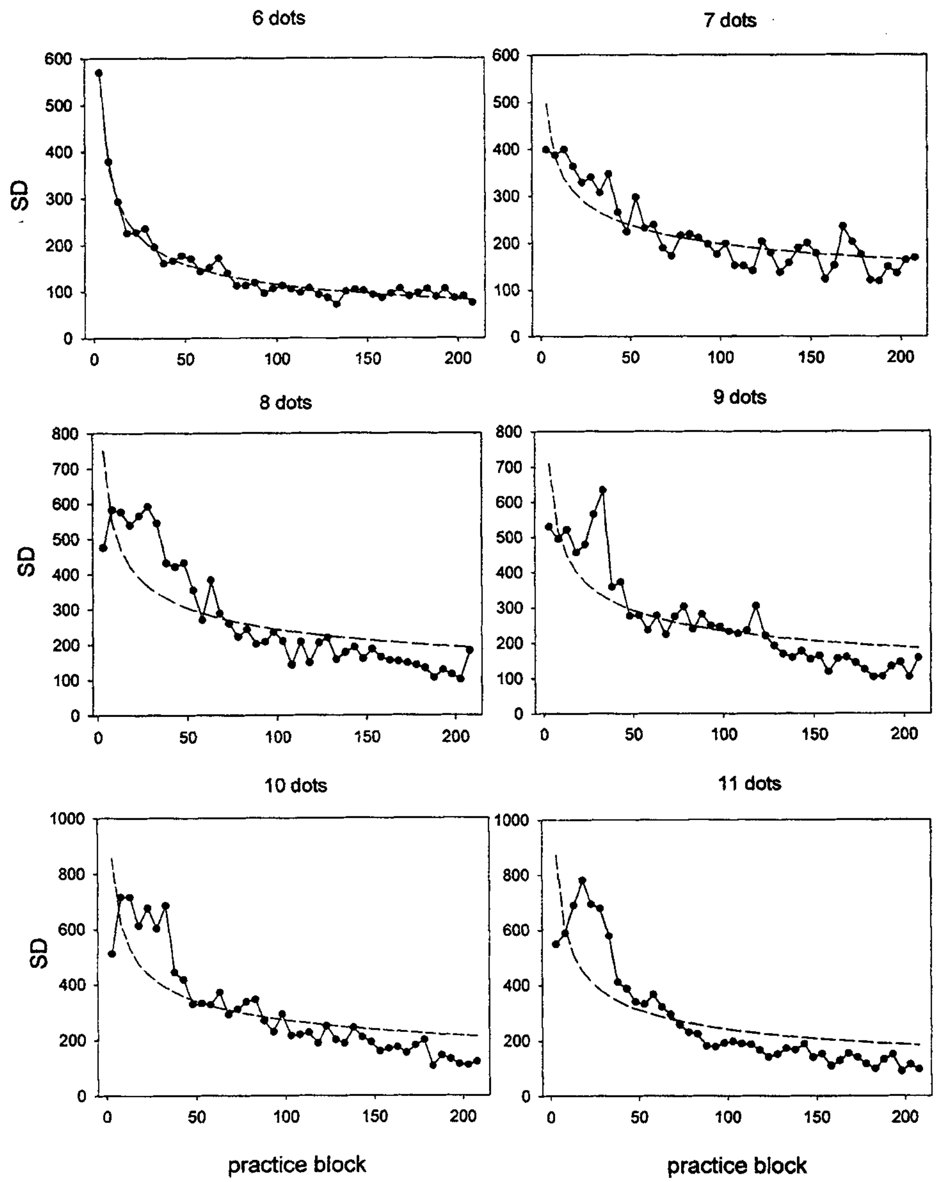

Figure 6. Power function and component power laws fits to the reaction time data of Experiment 3. *SD* = standard deviation.

for Numerosity 6 provides a candidate account for why no substantial deviation from the power function was observed for that numerosity level: After the first couple of blocks of practice, performance for that numerosity was (according to the CMPL fits) dominated by the retrieval power function. Finally, note that estimates of *pre* (previous learning) vary considerably across the three experiments. However, provided that this parameter takes some nonnegligible positive value, its precise value is not critical to the quality of fit. For example, *pre* could be set to a value of 15 (or, alternatively, to a high value like 300) for all experiments without noticeable loss in quality of fit. Thus, although the algorithm power function is strictly required under the CMPL model, it does not appear to be doing much work in the fits for these data.

Predictions of the CMPL Model for the Standard Deviation

The CMPL theory also makes the prediction that the variance of the latency will decrease as a power function

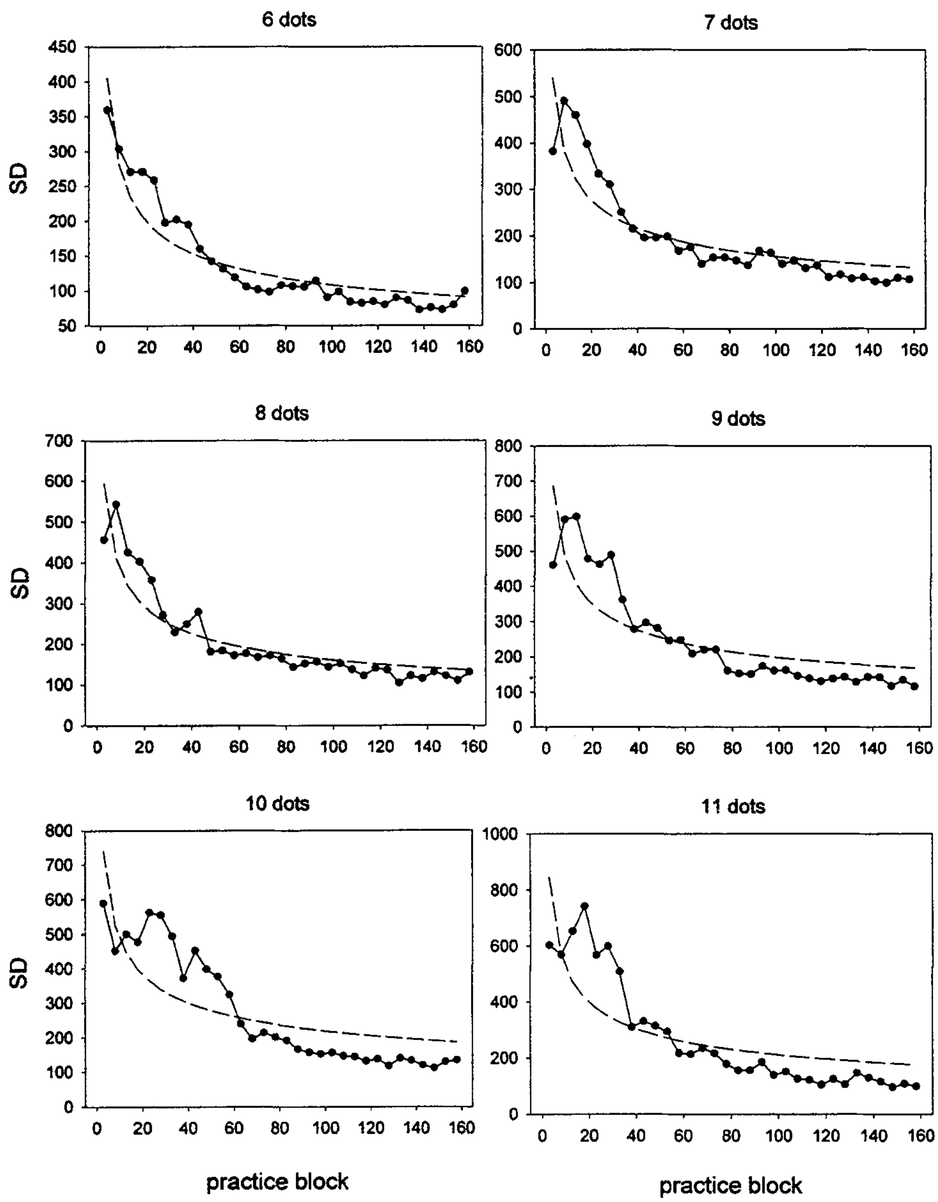

Figure 7. Power function fits to the standard deviation (*SD*) data of Experiment 1.

only within each strategy. The overall predicted variance is given by the more complex mixture equation,

$$\sigma^2 = \sigma^2_{\text{algorithm}} \cdot (1 - p) + \sigma^2_{\text{retrieval}} \cdot (p) + p \cdot (1 - p) \cdot (\text{RT}_{\text{algorithm}} - \text{RT}_{\text{retrieval}})^2, \quad (6)$$

where $\sigma^2_{\text{algorithm}}$ and $\sigma^2_{\text{retrieval}}$ are the power functions governing reduction in algorithm and retrieval variance, respectively. The standard deviation (*SD*) is then the square root of the variance computed in Equation 6. An example CMPL *SD* function is depicted in Figure 5. Note that there is a "bubble effect" in the overall *SD* caused by the $p \cdot (1 - p) \cdot (\text{RT}_{\text{algorithm}} - \text{RT}_{\text{retrieval}})^2$ term in Equation 6. The CMPL theory predicts that the overall *SD* can actually increase in absolute terms (that is, can become larger during the strategy transition than it is at the beginning of practice) for a brief interval, as is the case in the example function depicted in Figure 5. Note, however, that this absolute increase in *SD* is not required by the theory for any given data set. For example, if there is some decrease in algorithm RT and *SD* with practice and if the difference between the algorithm and retrieval RT functions is relatively modest (or if the transition to retrieval occurs very quickly) then the

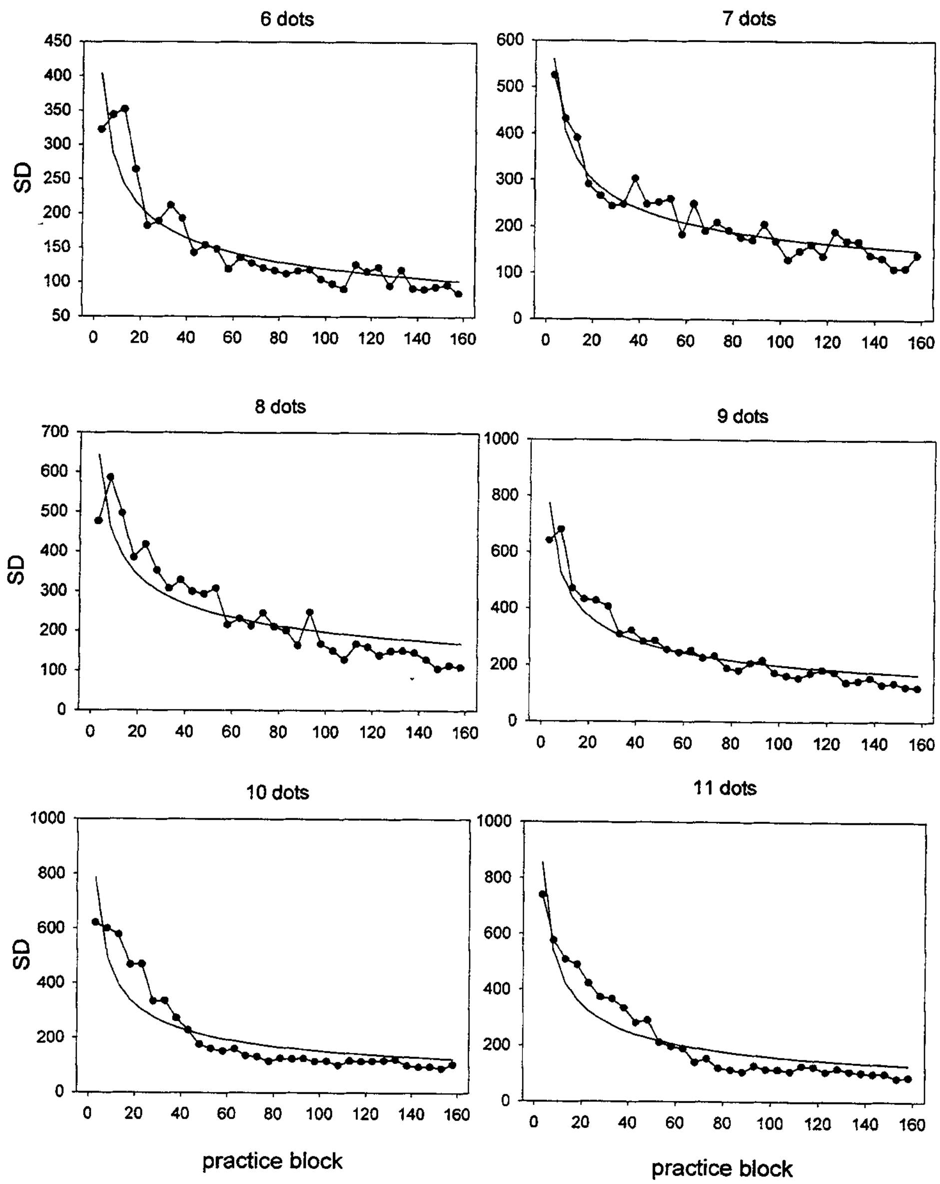

Figure 8. Power function fits to the standard deviation (*SD*) data of Experiment 3.

bubble effect will have less influence, and the *SD* could decrease monotonically with practice (although in most cases it would still deviate significantly from the power function). In contrast, race models such as the instance theory are mathematically required to predict a monotonic reduction in the overall *SD*, given a set of ancillary assumptions as described in Logan (1988; see also Compton & Logan, 1990).

*SD*s are much more variable than the mean RTs in any data set. To reduce noise in this analysis, *SD* data were computed for each block and then averaged over five consecutive block sequences. For Experiments 2 and 3, *SD*s for each pattern similarity type were then averaged to further control for noise.[4] Three parameter power functions were fit to each data set as shown in Figures 6, 7, and 8. The *SD*s are clearly not well fit by a power function, with the exception of some of the six and seven dot data sets. Further, the deviations have a shape generally consistent with the predictions of the CMPL model (see Figure 5) and for large numerosities there is typically a bubble in the

[4] *SD*s were first computed for each practice block and for each pattern similarity type, and then averaged across participants, practice blocks, and similarity types. Thus, differences in means across these variables are not a part of the variance estimates.

data.[5] For Numerosities 7 through 11, the bubble term is predicted (based on the RT fits described earlier) to take a maximum value between about Block 15 and Block 25. For Numerosity 6, it is predicted to take a maximum value within the first several blocks. These predictions are roughly consistent with the bubble effects observed in the data. Finally, for the larger numerosities, there is often a clear absolute increase in *SD* with practice. This result appears to constitute the first demonstrated case of absolute increase in *SD* with practice in any skill acquisition task.

Discussion

The preceding results make two fundamental points about human skill acquisition. First, the evidence is now clear that the power law of practice (Newell & Rosenbloom, 1981) does not hold for the overall RT or *SD* data for any arbitrary task. Rather, it appears to hold only within each strategy used to perform a task (see also Delaney, Reder, Staszewski, & Ritter, 1998). Second, as predicted by the CMPL theory, strategy execution appears to be an either–or phenomenon across a wide variety of skill domains. Either the algorithm or memory retrieval, but not both, are executed on each trial. There may be attentional limitations that preclude execution of multiple strategies concurrently in many if not all task domains.

Alternatively, it is conceivable that there are ways in which the instance or EBRW theories can account for the RT and *SD* results while preserving the assumption of parallel strategy execution. For example, the probability of learning an instance may not be the same for all items, as is assumed to date in the instance and EBRW theories. Consider the case in which an instance is encoded for Item 1 on the first block of trials but instances are not encoded for Item 2 until the third block. In this case, data for Block 2 are going to reflect a mixture of (a) a race between traces (instances) of Item 1 and the algorithm and (b) the algorithm alone applied to Item 2. This will produce a mixture of (a) versus (b) type trials that may generate the observed patterns in the data. However, at present it is at best unclear whether such modifications could account for the combined pattern of results for both RT and *SD* data. Explicit simulations testing candidate modifications to the instance or EBRW theories are needed.

One additional aspect of the fits merits additional discussion. The data from Experiment 3 deviated less from the power function at each numerosity level than did data from Experiments 1 and 2. An additional analysis separating data from Experiment 3 into friends and enemies revealed that for enemies, the RT and *SD* fits looked much like those of Experiments 1 and 2. For friends, however, the deviations from the power function, while still observable, were even less pronounced than those shown in Figures 4 and 8. Inspection of Figures 5, 9, and 12 of Palmeri (1997) gives some hint of a possible explanation for this effect. It appears that the transition to retrieval, as indexed in those figures by the slope relating RT to numerosity, occurred unusually fast for the friends condition in Experiment 3 compared with all other conditions of all experiments. A fast transition to retrieval will attenuate observed deviations from the power function in the overall RT and *SD* data, simply because if the transition occurs very quickly, the majority of the data will conform closely to the retrieval power function.

Finally, in drawing theoretical conclusions based on these results, it is important to first note that the EBRW theory provides an elegant account of dot patterns similarity effects at a given point during practice. The current version of the CMPL theory simply cannot account for such effects, and it is an open question whether it can be extended to do so. However, it is now also an open question whether the EBRW and instance theories can be modified to provide a comprehensive account of the skill acquisition effects in numerosity judgment and other tasks (e.g., Rickard, 1997).

[5] In principle, Equation 6 could be fit to the overall *SD* data analogously to the way Equation 2 was fit to the RT data. However, as evident in Equation 6, overall *SD*s are dependent in part on the strategy specific RTs. Thus, in order to produce optimized fits to both of these variables, RTs and *SD*s would need to be fit simultaneously. This fact, combined with the intrinsically much greater noise in the *SD*s relative to the RTs, precluded arriving at stable nonlinear regression fits for Equation 6 for these data.

References

Compton, B. J., & Logan, G. D. (1990). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition, 19,* 151–158.

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.

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.

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

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

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

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

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

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

SAS Institute. (1994). SAS/STAT user's guide, Version 6 (4th ed., Vol. 2). Cary, NC: Author.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skills. *Journal of Experimental Psychology: General, 117,* 258–275.

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

(Appendix follows)

Appendix

Simulation Results

In the CMPL simulation model for arithmetic and related tasks described in Rickard (1997), the critical first step of processing involved a competition between a problem node mediating the first step of the algorithm and a problem node mediating direct retrieval of the answer. The problem node in both cases corresponded to a particular interpretation of the stimulus. For example, consider the pound arithmetic problem 4 # 17 = ? The first step of the algorithm strategy activated a subtraction node based on the two numbers contained in the stimulus (17 − 4). The direct retrieval strategy, on the other hand, activated a problem node corresponding to the pound arithmetic problem (4 # 17). The node with the most activation at a specified threshold point then continued to accrue activation, while the activation of the other node was set to zero.

For numerosity judgment, in contrast, no preexisting problem node corresponding to each specific novel pattern of dots would exist prior to the first practice block. Further, execution of the dot counting algorithm does not depend on accessing such a preexisting representation. Thus, a competition between a problem node representing the first step of the algorithm and a problem node representing direct retrieval would make no sense for this task domain. For this reason, the competition at the problem level in Rickard (1997) needs to be eliminated to simulate numerosity judgment.

A CMPL simulation model for the critical strategy decision stage in the numerosity judgment task, modified to eliminate the problem level competition, is shown in Figure A1. In this simulation, the first processing step is a competition between the "retrieve" subgoal and the "execute counting algorithm" subgoal, which is analogous to the subgoal competition described in Rickard (1997).[A1] If the algorithm subgoal wins this competition, then the counting algorithm is executed and memory retrieval is suppressed. If the retrieval subgoal wins, retrieval proceeds exactly as in the Rickard (1997) simulation and the algorithm is suppressed. No attempt is made to model the dot counting algorithm explicitly beyond the strategy decision point. Rather, it is assumed that any speedup in dot counting with practice follows a rough power function of the number of blocks of practice. There is no theoretical precedent for this claim in the model. Rather, it is simply the most reasonable default expectation. There are no published data from skill acquisition tasks to date that call into serious question the assumption that the power law holds to a very good approximation within a single strategy (i.e., within a single sequence of cognitive steps performed in the service of some goal).

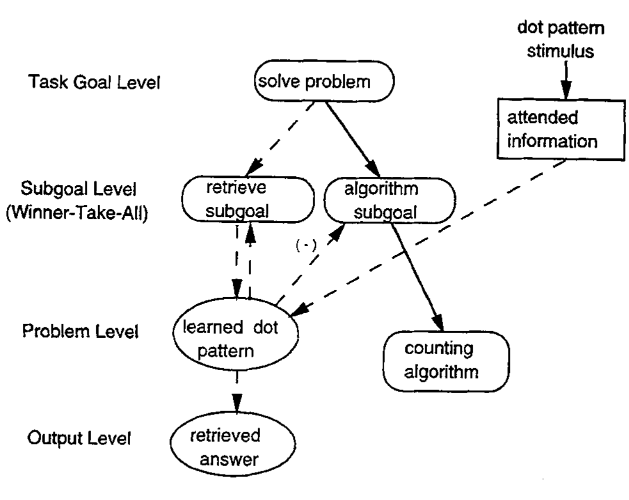

Figure A1. Component power laws simulation architecture for numerosity judgment.

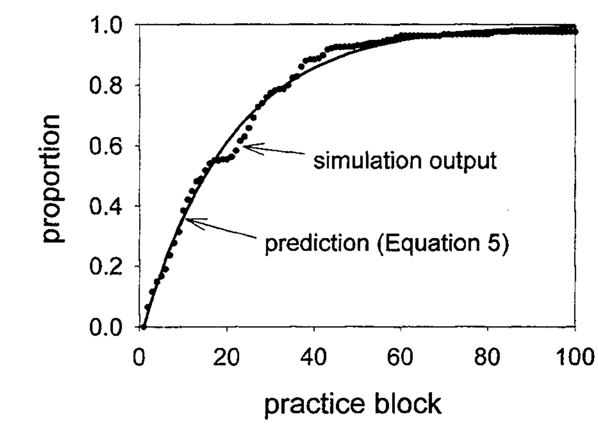

Figure A2. Simulation results for proportion of algorithm trials as a function of practice. Also shown is the best fitting negative exponential function (Equation 5). The simulation is for 18 participants with 12 problems each over 100 blocks of practice.

Parameter settings for this simulation are the same as those described in Rickard (1997), with two exceptions. First, the beta distribution that governs the values of $c2$ across items in the simulation had parameter values alpha = 4 and beta = 16. These values yielded an average rate parameter for the retrieval power function of 0.2, which matches closely with the estimates of the regression fits to the data. Second, the previous learning for the algorithm was set to 325.

The proportion of items on which the retrieval strategy was selected (p) in the simulation is shown for a set of 12 items for 18 participants in Figure A2 (for more details of the simulation, see Rickard, 1997). The fit of Equation 2 is quite good, yielding an r^2 of 0.992. Most of the major deviations between the data and the prediction in Figure A2 appear to reflect random fluctuations, as they do not replicate over repeated simulations. The chosen value of pre (325) yielded a strategy transition that roughly matches that which appears to have occurred in the data. Values of pre smaller than 325 result in a faster transition to retrieval that is well described by Equation 5, and values somewhat greater than 325 result in a slower transition that is also well described by Equation 5. However, as pre becomes substantially larger than 325, the form of this curve begins to take more of a sigmoidal shape. As pre becomes extremely large, Equation 5 no longer provides a close fit.

[A1] Note that the equation specifying the relation between $c1$ and $c2$ in Rickard (1997) implies that the random fluctuation in $a(t)$ is incorporated into the value of $c1$ on each trial. This reflects an oversight in the description of the simulation model in that article. The value of $a(t)$ used in calculating $c1$ was in fact set to the mode of the beta distribution for $a(t)$, which has been .8 in all simulations to date.

Received April 14, 1997
Revision received December 22, 1997
Accepted January 6, 1998 ■