Memory & Cognition
1991, 19 (2), 151-158

The transition from algorithm to retrieval in memory-based theories of automaticity

BRIAN J. COMPTON and GORDON D. LOGAN
University of Illinois, Urbana, Illinois

Two memory-based theories of automaticity were compared. The mixture model and the race model both describe automatization as a transition from algorithmic processing to memory retrieval. The mixture model predicts that, with training, the variability of reaction time will initially increase, and later decrease in a concave downward manner, whereas the race model predicts the variability will decrease only in a concave upward manner. The mixture model predicts that using both algorithm and retrieval on a single trial will be slower than using the algorithm alone, whereas the race model predicts the reverse. The experiments used an alphabet arithmetic task, in which subjects verified equations of the form H + 3 = K and made subjective reports of their strategies on individual trials. Both the variability of reaction times and the pattern of reaction times associated with the strategy reports supported the race model.

Memory-based models of automaticity describe novice performance as governed by relatively slow algorithmic computations and that expert performance is governed by much faster memory retrieval processes. In these models, automatization is a transition from algorithm-based performance to memory-based performance. At issue in this paper is the process by which the transition occurs.

Two different mechanisms for the transition from algorithm to memory have been proposed. In the *mixture model*, only one process, either algorithmic or memorial, is selected for the initial attempt at obtaining the solution on a given trial. The transition then reflects a shift in the probability of initially selecting retrieval as opposed to selecting the algorithm. In the *race model*, both the algorithm and memory retrieval operate simultaneously, and the selection of the process is determined by the outcome of a race, in real time, between the two processes. The transition reflects a shift in the probability that retrieval will win the race.

In one version of a mixture model, the determination of whether to use an algorithm or to retrieve the answer is made before the process itself (be it algorithmic or memorial) is executed. At the onset of training, all trials are solved by the algorithm. As training progresses, the probability that a trial will be solved by retrieval increases. Although performance on a single trial is determined by only one process, performance over a number of trials reflects a mixture of the two processes. In this simple mixture model, only one process is to be used on a particular trial. Such a model is overly simplistic, since it does not consider the possibility that subjects may sometimes use the counting algorithm as a backup strategy if memory retrieval fails to produce an acceptable answer.

The mixture model that will be considered in this paper is an elaboration of the simple mixture model. The elaborated mixture model is similar to a learning mechanism proposed by Siegler (1987, 1988; Siegler & Shrager, 1984) as part of a theory of children's acquisition of arithmetic skills. In Siegler's model, children first try to retrieve the answer to a given problem; if that fails, they use an algorithm to compute the answer. If the trial is presented very early in training, retrieval will not be attempted before the algorithm is executed. Training serves to strengthen the memory representation for a given problem, which increases the probability that retrieval will be used successfully the next time the problem is presented.

The elaborated mixture model borrows the notion that a single process is initially selected and is followed by a backup process when it is unsuccessful. This mixture model differs from the simple one in that both algorithmic and memorial processes can operate in sequence on a single trial. As in the simple model, however, the two processes never operate simultaneously. This stands in contrast to the race model, in which the algorithmic and memorial processes operate in parallel.

Our intention was to test mixture models in general rather than Siegler's model in particular. There are many important aspects of Siegler's model besides the mixture hypothesis; it is not clear that the idea of mixture is essential to his model. It may be possible to reformulate it as a race model. Thus, results that reject a mixture model in favor of a race model need not be viewed as rejecting Siegler's model in general.

This research was supported by National Science Foundation Grant BNS 88-11026 to Gordon Logan and was conducted as a first-year project by Brian Compton under the direction of Gordon Logan. We would like to thank Art Kramer and Gary Dell for comments on an earlier version of the paper and Gail Heyman, "The Math Lady," for help with analyses. We would also like to thank Peggy Intons-Petersen, Anders Ericsson, and two anonymous reviewers for comments on the manuscript. Address correspondence concerning this article to Gordon D. Logan, Department of Psychology, University of Illinois, 603 East Daniel St., Champaign, IL 61820.

The race model is a component of Logan's (1988) instance theory of automatization. In instance theory, each exposure to a stimulus adds another memory trace whose retrieval will be initiated the next time the same stimulus is presented. As the number of relevant memory traces increases, the probability increases that one of them will win the race against the algorithm. With enough training, the set of instances will be so large that a memory trace will produce the fastest response on virtually every trial.

The present experiment distinguishes race models from mixture models by examining changes in the variability of reaction time and by examining the pattern of reaction times associated with strategy reports. Although many other models of practice have been proposed (Anderson, 1982; Crossman, 1959; MacKay, 1982; Newell & Rosenbloom, 1981; Schneider, 1985), the mixture model was chosen for comparison because it makes predictions that most closely parallel those of the race model.[1] Both models make specific, parallel, but opposite, predictions about variability of reaction time, which allows the two models to be clearly differentiated. The mixture and race models each assume that reaction times are selected from stable parent distributions of algorithm and memory finishing times that do not change over practice. Changes in performance are caused by changes in the probability of using one process or the other.

The mixture model predicts that, at the onset of training, subjects will use the algorithm on all trials and that the variability of reaction times for the first trials will be equal to that of the algorithm. After training, as the relevant memory representations become strengthened, most trials will be solved by memory retrieval, and the variability of reaction time will approach the variability for memory retrieval alone. However, during the transition from counting to remembering, the variability should increase, and then decrease, as the probability of remembering increases from zero to one. The formula for the variance of reaction times under the mixture model is:

$$\sigma^2_{\text{mix}} = p\sigma^2_{\text{alg}} + (1-p)\sigma^2_{\text{mem}} + p(1-p)(\mu_{\text{alg}} - \mu_{\text{mem}})^2. \quad (1)$$

The increase in variability can be seen only if subjects use the algorithm predominantly early in practice and memory retrieval predominantly later in practice. However, the mixture model makes a more general prediction that can be tested, even if one process or the other predominates throughout the practice period. It requires only that the probability of using one process, as opposed to the other, changes during practice. The quadratic component in Equation 1 implies that the function relating variance to practice will be concave downward.

The race model also predicts that, at the onset of training, the variability will be equal to that of the algorithm alone. However, the race model predicts that variability will decrease with training, and never increase; as more instances are added to memory, more traces are retrieved quickly, and the finishing time of the fastest trace becomes increasingly constrained by the finishing times of other fast traces and, as a result, variability is reduced. Logan (1988) has shown that, under the race model, the standard deviation of reaction time decreases as a power function of the number of instances, with an exponent equal to the exponent for the power function reduction in mean reaction time. The formula for the variance of reaction time under the race model is:

$$\sigma^2_{\text{race}} = N^{-2c}\sigma^2_{\text{mem}}, \quad (2)$$

where N is the number of instances and c is a constant reflecting the learning rate. The power function reduction predicted by the race model implies that the function relating variance to practice will be concave upward.

Thus, mixture models and race models can be distinguished in two ways: First, race models predict that variability will decrease monotonically with practice, whereas mixture models predict an increase, then a decrease, variability with practice, provided that the algorithm dominates early in practice and retrieval dominates later. Second, and more generally, race models predict that the function relating variability to practice will be concave upward, whereas mixture models predict it will be concave downward. This can be tested by examining the sign of a quadratic trend fitted to the practice data. Such a procedure involves finding the standard deviation of reaction time at several different points in training, multiplying these values by the coefficients for a quadratic trend, summing the results, and examining the sign of the sum.

Race models can also be distinguished from mixture models by examining reaction times associated with subjective reports of strategy. In the mixture model, the algorithm and memory retrieval operate sequentially whenever they are both used on a single trial. Consequently, the mixture model should predict that reaction times on trials for which the subject reports both counting and remembering will be slower than reaction times on trials for which the subject reports counting only, since the former condition represents two separate processes in sequence, whereas the latter represents only one. In contrast, the race model predicts that reaction times on trials for which the subject reports both counting and remembering will be faster than reaction times for the "pure" counting case, since, under the race formulation, counting and remembering operate in parallel, and memory can only serve to speed up trials on which it is used.

Alphabet Arithmetic

The present experiments used an alphabet arithmetic task (Logan & Klapp, in press). Subjects determined whether equations of the form H + 3 = K were true or false (this one is true since K is 3 letters down the alphabet from H). Alphabet arithmetic is an analogue of the acquisition of addition skills in children. It allows adults to perform as novices in the task domain, in a situation similar to that of a child learning to add for the first time.

In both alphabet arithmetic and regular addition, the learner already possesses knowledge about the relevant sequence of numbers or letters and has the ability to count through that sequence. This information allows the subject to use a counting algorithm, in which the sum of the addends is obtained by counting through one or both addends. A similar counting algorithm is used by novice subjects in the alphabet arithmetic task (see Logan & Klapp, in press).

An advantage of alphabet arithmetic is that different strategies for performing the task are empirically and phenomenally distinct. When the counting algorithm is used, the number of counting steps is determined by the size of the digit addend, and reaction time is a linear function of the digit addend, with a slope of 400-500 msec per counting step. The slow, serial counting process is readily detectable by subjects. When alphabet arithmetic is solved by remembering, the magnitude of the digit addend no longer affects performance (the slope approaches zero) and subjects experience the answer "popping into their heads" (Logan & Klapp, in press).

The present experiments used a probe procedure to investigate how subjects' strategies changed with practice. On a subset of the trials, a probe was presented that asked subjects how they solved the preceding problem and that provided different response options. The probe procedure allowed the data from the main experiment to be grouped by strategy so that competing predictions of the race model and the mixture model could be evaluated.

In the main experiment, the response options were: counting, remembering, or counting and remembering at the same time. We have had extensive experience with the task and are in agreement that these three strategy possibilities can be differentiated phenomenally. The legitimacy of these strategy report options is also supported by an open-ended strategy report procedure conducted by Logan and Klapp (in press).

To address the possibility that the response options provided in the main experiment did not allow subjects to accurately report the strategies they had used, a follow-up experiment was conducted that was identical to the main experiment in all respects, except that eight, rather than three, response options were provided. This follow-up experiment replaced the "counted and remembered at the same time" report option of the main experiment with four more detailed options, as well as providing two new catchall report options.

The present experiments used a single-session experiment to investigate the acquisition of automaticity. Logan and Klapp (in press) have shown that it is the number of exposures to specific items, rather than the total amount of training, that is important in determining when automatization has been attained. In this experiment, a small problem set (six true and six false problems) allowed each problem to be presented 36 times in a single session, which should be sufficient for automaticity to develop, given the results of Logan and Klapp (in press). Consequently, results from a single-session experiment should generalize to multiple-session experiments.

METHOD

Subjects

The subjects in the main experiment were 36 introductory psychology students from the University of Illinois, who received course credit for their participation. The subjects in the follow-up experiment were 18 introductory psychology students, who received course credit, and members of the university community, who were paid for their participation.

Apparatus and Stimuli

There were nine stimulus sets, each consisting of 12 equations, 6 true and 6 false. The stimulus sets differed in the set of letter addends they contained and in the way letter addends were combined with digit addends. Three stimulus sets used the letters A through F, three used G through L, and three used M through R. The same three digit addends were used in all sets: 2, 3, and 4. Different stimulus sets that shared the same set of letter addends had different mappings of letter addend to digit addend.

Within each stimulus set, the set of six true equations included six different letter addends, which were combined with the three digit addends to create six unique equations. The answer for each true equation was the correct sum of the letter and digit addends (e.g., H + 3 = K). The six false equations were identical to the true equations, except that the answer was the sum of the letter and digit addend plus one letter (e.g., H + 3 = L).

The stimuli were presented on Amdek 722 color monitors with a display size of 24 rows × 80 columns, controlled by IBM AT or XT computers. The stimuli were presented horizontally on the screen beginning at row 12, column 35. Each stimulus was an equation made of a capital letter (the letter addend), a plus symbol (+), a single numeral (the digit addend), an equals sign (=), and another capital letter (the answer). The components of the equation were separated by single spaces; thus, the entire equation was nine characters long. Each equation measured 2.5 cm wide × .5 cm high when presented on the screen.

Prior to the display of each stimulus, a fixation pattern was displayed for 500 msec. This pattern consisted of two rows of five dashes (-) each, separated by spaces and displayed starting at row 11, column 35, and at row 13, column 35. The entire fixation pattern measured 2.6 cm wide × 1.5 cm high. Viewing distance was unconstrained but varied between 40 and 60 cm.

Following the display of the fixation pattern, the screen was cleared and the stimulus equation was displayed until the subject made a response. After the subject responded to the stimulus, feedback regarding the correct answer to the problem was given. The word "True" or "False" was displayed at row 14, column 37, for 1,000 msec. At this point, the screen was cleared for 1,000 msec prior to the beginning of the next trial.

On one-sixth of the trials, a probe was presented immediately following the subject's response. The probe was a special screen asking the subject to report which strategy was used to solve the preceding trial. In the main experiment, the following probe screen was displayed, beginning at row 12, column 10, in this format:

What did you do to answer the last problem?
If you counted through the alphabet, type "1"
If you remembered the answer without counting, type "2"
If you counted and remembered at the same time, type "3"

In the follow-up experiment, the following probe screen was displayed at row 7, column 0, in this format:

What did you do to answer the last problem?

Type:	If you:
1	counted through the alphabet only
2	remembered the answer without counting
3	first counted and then got the answer by remembering

4 tried to remember and then got the answer by counting
5 tried to count and remember simultaneously, and got the answer by counting
6 tried to count and remember simultaneously, and got the answer by remembering
7 used a strategy that is not listed above
8 made a mistake or do not know how you solved the problem

The subjects were told to type their response on the number keys on the top row of the keyboard. Following the subject's response to this query, the screen was cleared and the next trial began after a 2,000-msec delay.

At the end of each 72-trial block, the message "TIME FOR A BREAK; PRESS ANY KEY TO START" was displayed at row 12, column 10. The experiment paused during this period; the next trial began 2,000 msec after the subject ended the break by pressing a key.

The subjects entered their alphabet arithmetic responses on the bottom row of the computer keyboard by pressing the "z" and "/" keys, which are the leftmost and rightmost keys on an AT keyboard (nearly so on an XT). One-half of the subjects pressed the "z" key to indicate a judgment of true and the "/" key for a judgment of false; one-half did the reverse.

Procedure

The experimental session consisted of 432 trials, divided into six blocks of 72 trials each. Each block of 72 trials was further subdivided into six sub-blocks of 12 trials each. Each sub-block consisted of the 12 problems of the stimulus set in random order. One true problem and 1 false problem was probed in each sub-block. Each of the 12 problems of the stimulus set was probed once in each 72-trial block; therefore, at the end of each block, all problems had been probed the same number of times. Each problem was probed once per block, six times throughout the experiment. In an entire session, 72 trials were probed (one-sixth of the total).

In the main experiment, 4 subjects were assigned to each of the nine stimulus sets; in the follow-up experiment, 2 subjects were assigned to each of the stimulus sets. For each stimulus set, one-half of the subjects responded "true" with the "z" key and "false" with the "/" key, and one-half of the subjects did the reverse. The subjects were required to maintain 85% accuracy. In the main experiment, the 9 subjects who failed this criterion were replaced in their respective conditions with the subjects who passed the error criterion; in the follow-up experiment, 8 subjects were replaced in this manner.

Prior to the experimental session, each subject was given instructions on how to perform the alphabet arithmetic task and how to make the strategy report on the probe trials. The subjects were given an example of both a true and a false equation and shown how the truth value of these equations can be determined by counting through the alphabet. The subjects were told to answer the problems as quickly and as accurately as possible and to take only short breaks between blocks.

Data Analysis

Only correct responses were included in the data analysis. Trials in which reaction time exceeded 10,000 msec were judged to be errors, regardless of the actual response made. An average of 1.1% of trials per subject in the main experiment and 1.2% of trials per subject in the follow-up experiment were excluded from analysis for this reason. Reaction times for probed and nonprobed trials were analyzed. Error rates were calculated for both probed and nonprobed trials in each combination of addend and true/false conditions. Distinctions between the mixture model and the race model were made by analyzing the variability of reaction time and the pattern of reaction times associated with strategy reports as they changed over blocks. The validity of the subjective reports in the first experiment was determined by way of a comparison of the digit addend slope of the nonprobed trials and an estimate of that slope that used of data from the probed trials.

RESULTS AND DISCUSSION

Nonprobed Trials

Table 1 shows the percentages correct for each block, addend, and true-or-false condition in the main experiment. Table 2 shows the same data for the follow-up experiment. There was no evidence of a speed-accuracy trade-off. Mean reaction times for the nonprobed trials are shown for each block in Figure 1, with a solid line indicating results from the main experiment and a dashed line indicating results from the follow-up experiment. The decrease in reaction time over blocks was significant in the main experiment [$F(5,150) = 134.08, p < .01, MS_e = 917,765$] and also in the follow-up experiment [$F(5,60) = 48.64, p < .01, MS_e = 1,303,707$]. This result is consistent with a primary measure of automatization, the power function speedup in reaction time (Newell & Rosenbloom, 1981).

The effect of digit addend was significant in the main experiment [$F(2,60) = 15.26$, $p < .01$, MS_e =

Table 1
Percentages Correct for Nonprobed Trials by Digit Addend and True/False Condition for Main Experiment

Training Block	True/False Condition	Digit Addend 2	3	4
1	T	89	88	82
	F	90	85	82
2	T	95	92	92
	F	95	87	91
3	T	97	94	95
	F	94	90	94
4	T	95	95	93
	F	95	93	96
5	T	96	96	98
	F	95	97	95
6	T	96	95	95
	F	94	95	96

Table 2
Percentages Correct for Nonprobed Trials by Digit Addend and True/False Condition for Follow-up Experiment

Training Block	True/False Condition	Digit Addend 2	3	4
1	T	89	86	87
	F	91	86	84
2	T	92	97	92
	F	96	92	87
3	T	96	93	89
	F	96	92	87
4	T	98	87	97
	F	96	95	92
5	T	94	93	97
	F	95	92	94
6	T	91	89	94
	F	91	89	94

1,404,970] and in the follow-up experiment [$F(2,24) = 18.31, p < .01, MS_e = 667,801$]. Also, the effect of digit addend changed with training. Figure 2 shows reaction time by digit addend for each block, with the solid line indicating the main experiment and the dashed line indicating the follow-up experiment. This pattern suggests that subjects made a transition from an algorithmic counting process to a retrieval process in a single session (also see Logan & Klapp, in press). In the first block of the main experiment, when subjects reported counting on many trials, the slope was 487 msec per count. In Block 6, the addend slope fell to 18 msec per count, indicating that digit addend no longer had a substantial effect on reaction time. The interaction between block and digit addend was significant in the main experiment [$F(10,300) = 13.52, p < .01, MS_e = 194,144$] and in the follow-up experiment [$F(10,120) = 8.64, p < .01, MS_e = 667,801$]. The change in the slope of digit addend was not sudden but occurred gradually across the six training blocks.

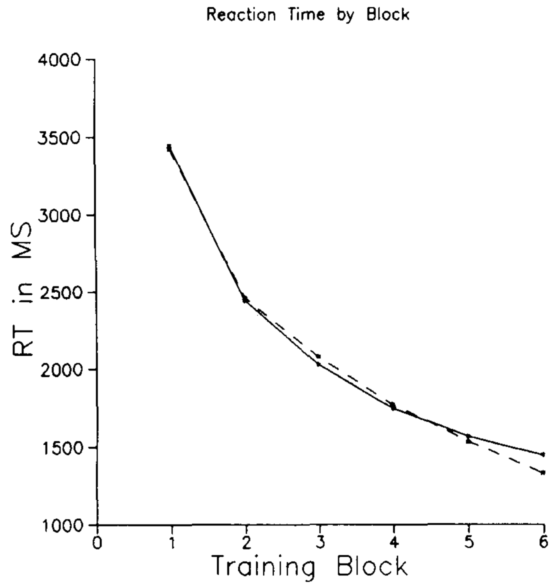

Figure 1. Reaction time by block for nonprobed trials. Solid line shows data from the main experiment; dotted line shows data from the follow-up experiment.

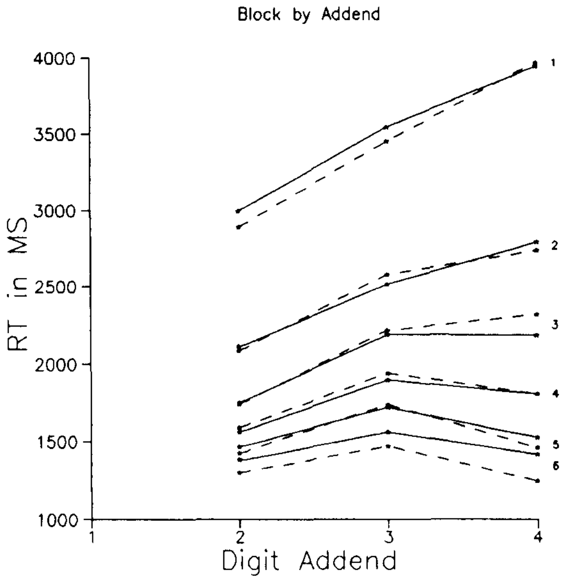

Figure 2. Reaction time by digit addend for each block, nonprobed trials only. Solid lines show data from the main experiment; dotted lines show data from the follow-up experiment.

Table 3
Percentages Correct for Probed Trials by Digit Addend and True/False Condition for Main Experiment

Reported Strategy	True/False Condition	Digit Addend 2	3	4
Counting	T	97	89	86
	F	86	89	79
Memory	T	96	94	92
	F	95	91	92
Counting and	T	96	88	88
Memory	F	98	94	97

Note that the difference between Addends 3 and 4 disappeared more rapidly than did the difference between Addends 2 and 3, producing a kind of "dog-leg" nonlinearity in Blocks 3 through 6. This effect follows from the race model. Memory traces race against the algorithm and are more likely to win when the algorithm is slow (i.e., addends of 4) than when the algorithm is fast (i.e., addends of 2 and 3). Subjective reports of strategy were consistent with this interpretation: Subjects reported responding on the basis of memory on 55%, 51%, and 63% of trials for Addends 2, 3, and 4, respectively.

A main effect for true/false problems was seen in the main experiment [$F(1,30) = 39.99, p < .01, MS_e = 195,068$] and in the follow-up experiment [$F(1,12) = 35.22, p < .01, MS_e = 251,170$], with true problems faster. A block × addend × problem set interaction was seen in the first experiment [$F(20,300) = 2.29, p < .01, MS_e = 194,114$] but not in the follow-up experiment. This interaction seemed to be due to a flatter addend slope in the first block for the subjects assigned to the first three problem sets, which used the first six letters of the alphabet as letter addends. A block × true/false interaction was seen in the follow-up experiment only [$F(5,60) = 2.95, p < .05, MS_e = 130,039$]. No other interactions were significant in either experiment.

Probed Trials: Subjective Reports

Table 3 shows the percentages correct for each report, addend, and true-or-false condition in the main experiment. There was no evidence of a speed-accuracy tradeoff. The subjective reports of strategy also indicated a transition from counting to remembering. Overall, subjects in the main experiment reported counting on 19.6% of the problems, remembering on 56.4% of the problems, and using both counting and memory on 24.0% of the problems. All 36 subjects reported using memory, 34 subjects reported counting, and 35 subjects reported both counting and remembering at the same time.

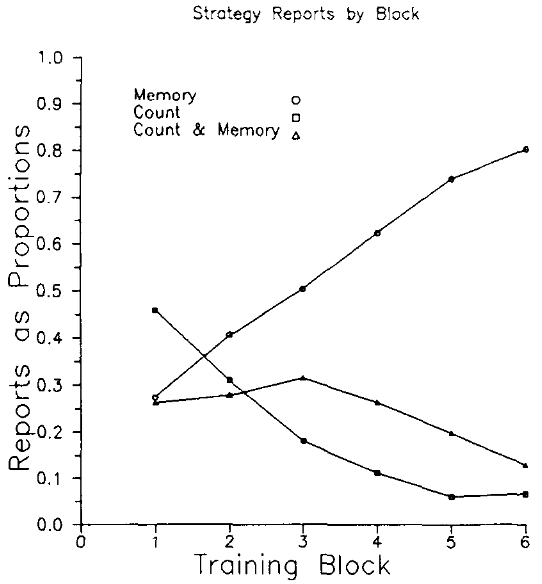

Figure 3. Strategy reports by block for the main experiment. Each report is shown as a proportion of all reports for each block. Nonprobed trials only.

Figure 3 shows the proportion of probed trials in the main experiment in which subjects reported each strategy, plotted by block. Since incorrect responses were not analyzed, the reports are based on slightly different numbers of trials for each block. Reports of remembering increased across blocks [$F(5,175) = 43.61, p < .01, MS_e = .034$]. Reports of counting decreased across blocks [$F(5,175) = 38.04, p < .01, MS_e = .024$]. Reports of both counting and remembering increased slightly over Blocks 1 to 3, then decreased over Blocks 4 to 6 [$F(5,175) = 5.44, p < .01, MS_e = .030$]. Subjects did not report counting exclusively in the first block, because a particular problem may have been probed after it had already been seen as many as five times, which could allow for memory retrieval in some cases.

Table 4 shows results for the follow-up experiment. For each strategy, mean reaction times for correct trials are shown, along with the percentages correct, averaged over subjects. The frequency of each report (for correct trials only, aggregated over subjects) is shown as a proportion of all correct probed trials. The number of subjects who made each report at least once on a correct trial is also shown.

The mean reaction times and proportion of occurrence for the memory only and counting only trials in the follow-up experiment closely resembled the results from the main experiment. However, the other report categories were used inconsistently, with some subjects making certain reports many times more frequently than other subjects. No strategy, other than counting or remembering, was reported more than 5% of the time. Only rarely did any subjects report either using a strategy not listed or not knowing what strategy they used. One possible explanation for this lack of consistency in the pattern of strategy reports is that individuals differ greatly in the strategies they use when performing a task such as alphabet arithmetic. However, given the authors' experience with the task, it seems more likely that subjects are not able to make fine distinctions about the strategies they are using when they both count and remember on the same trial.

Variance on Nonprobed Trials

The race and mixture models can be distinguished by the variability of reaction time during the transition from counting to memory. The variance of reaction time on nonprobed trials is plotted by block in Figure 4, with the solid line indicating the main experiment and the dashed line indicating the follow-up experiment. The decrease of variance with training was significant in the main experiment [$F(5,150) = 34.59, p < .01, MS_e = 3{,}126{,}269{,}569$] and in the follow-up experiment [$F(5,60) = 11.88, p < .01, MS_e = 2{,}584{,}409{,}964$].

Since the subjects reported counting on slightly fewer than half of the trials in the first block of the main experiment (45.9%), it was necessary to examine the shape of the decrease in variance to distinguish between the mixture and the race models. A test for quadratic trend was performed in the following manner: The variance of nonprobed trials was obtained in each block for each subject. For each subject, these six values were multiplied by the coefficients for a quadratic trend (5, −1, −4, −4, −1, 5) and then summed to obtain a positive or negative value. This value was positive for 81% of the subjects in the main experiment (29 out of 36, sign test, $p < .01$) and 78% of the subjects in the follow-up experiment (14 out of 18, sign test, $p < .05$), indicating that the variance decreased in a concave upward manner.

These results support the race model, which predicts that variability will decrease in a concave upward manner, with training, and never increase. They are incon-

Table 4
Results of Follow-up Experiment, Showing Strategy Report Made, Mean Reaction Time, Percentage Correct, Frequency of Report as a Proportion of All Reports Made, and Number of Subjects Making Report at Least Once

Strategy Report	Mean RT	Percentage Correct	Proportion	Subjects Reporting
Not probed	2089	92	—	—
Count	3170	96	0.25	18
Memory	1422	98	0.54	18
Memory, then count	3084	96	0.05	10
Count, then memory	3709	99	0.02	10
Simultaneous count and memory; answer by counting	3348	98	0.04	8
Simultaneous count and memory; answer by memory	2764	99	0.05	9
Other strategy	2023	98	0.02	8
Made error or don't know	2159	2	0.02	2

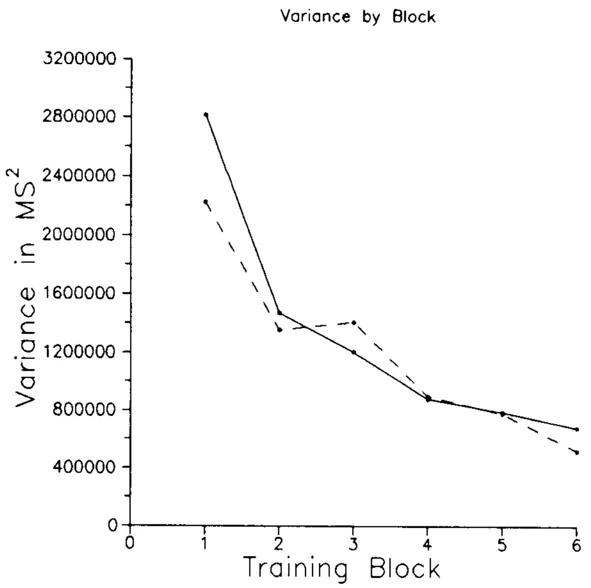

Figure 4. Variance by block, nonprobed trials only. Solid line shows data from the main experiment; dotted line shows data from the follow-up experiment.

sistent with the mixture model, which predicts an initial increase in variability, followed by a later decrease that is concave downward throughout the range of probabilities.

Subjective Reports and Reaction Time

Figure 5 shows the effect, in the main experiment, of digit addend on reaction time for probed trials, collapsed over block and broken down by strategy. Count trials produced the slowest reaction time (3,371 msec) and steepest slope (664 msec per count), and memory trials the fastest reaction time (1,531 msec) and flattest slope (−8 msec per count). Reaction time (2,513 msec) and slope (301 msec) were intermediate when subjects reported both counting and remembering. The difference in mean reaction time between counting and both counting and remembering was significant in the main experiment [$t(32) = 5.53, p < .01, MS_e = 54$]. In the follow-up experiment, no significant differences in reaction time were seen between the counting strategy and any of the four strategy options that involved a combination of counting and remembering.

The intermediate position of the count and memory trials in the main experiment is consistent with the race model but not with the mixture model, which predicts that counting and remembering will take longer than counting only. This pattern of slopes was seen as early as the first block, in which count trials showed a slow mean reaction time (3,747 msec) and a steep slope (697 msec), memory trials showed fast reaction times (1,531 msec) and a relatively flat slope (11 msec), and count and memory trials showed intermediate reaction times (3,190 msec) and a relatively flat slope (15 msec).

Validity of Subjective Reports

The validity of subjective reports in the main experiment was tested by estimating the addend slope for the nonprobed trials by combining the slopes of the different reported strategies, weighted by their frequency of occurrence in each block. The first step in creating this slope estimate was to find, separately for each strategy report, the slope of reaction time over digit addend, collapsed over block. The estimated slope was then calculated separately for each block by weighting the slope corresponding to each strategy by its frequency of occurrence in each block. For example, the estimate for Block 1 was made by multiplying the slope for each report type (slope = 664.24 for counting, −7.73 for memory, 301.45 for counting and memory) by its proportion of occurrence in Block 1 (proportion for counting = .459, memory = .275, counting and memory = .265, as shown in Figure 4) and then summing these weighted slopes to obtain the predicted slope for Block 1 (prediction = 382.64). This process was repeated for the other five blocks.

The slope estimate is plotted against the slope of the actual data in Figure 6. The estimate underpredicted the actual slope in the first blocks and overpredicted it in the last blocks. This may have been due, in part, to the assumption that the algorithm did not change over practice. In fact, the reaction times conditionalized on strategy reports showed that the slope on trials on which subjects reported counting did decline somewhat over blocks, suggesting the possibility of some process-based learning. The

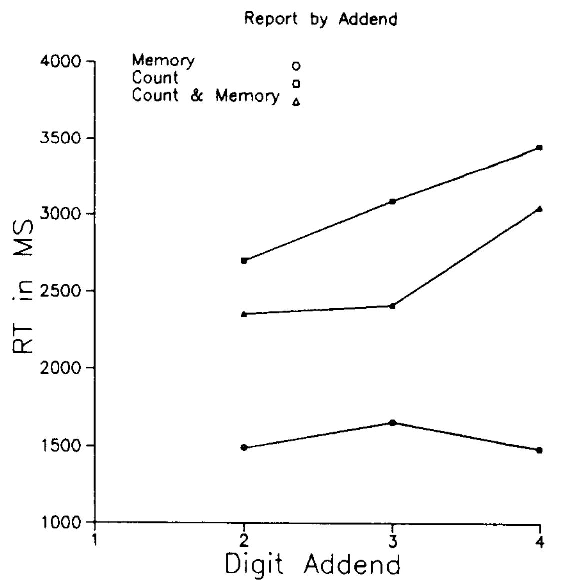

Figure 5. Reaction time by addend for each strategy report for the main experiment.

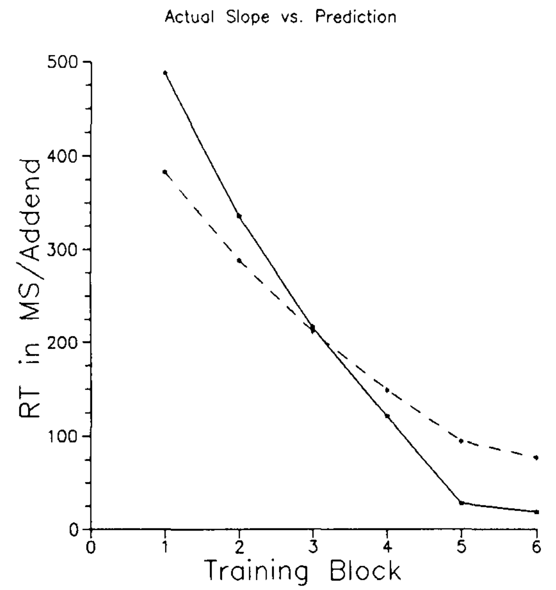

Figure 6. Estimate of addend slope by block plotted against observed addend slope by block, for the main experiment.

mixture model and the race model both assume that reaction times are sampled from a stable distribution of algorithm finishing times.

GENERAL DISCUSSION

Three pieces of evidence suggested that automatization, a transition from algorithmic- to memory-based processing, occurred with training. Reaction times decreased with training. The effect of digit addend was reduced, and subjective reports of strategy showed a transition from algorithm to memory.

Two mechanisms were considered for the transition from algorithm to memory. In the mixture model, only one process operates at a time, whereas in the race model, the algorithm and memory operate in parallel. Two types of evidence supported the race model. First, the variability of reaction times decreased consistently in a concave upward manner, with training, as the race model predicts, instead of increasing and then decreasing in a concave downward manner, as the mixture model predicts. Second, reaction times for trials in which subjects reported counting only were greater than those in which subjects reported both remembering and counting, suggesting the parallel nature of the process.

The validity of the subjective reports was assessed. An estimation procedure supported the hypothesis that the process underlying performance on the nonprobed trials was composed of the strategies identified by the subjects' reports mixed in proportion to the number of reports.

The use of strategy reports and the analysis of variability have only recently been applied to the study of automaticity and skill acquisition. Several studies have used subjective strategy reports, but these studies have generally involved asking questions after all trials have been completed (Logan & Klapp, in press). Siegler (1987) clarified the issue by asking subjects to report their strategies immediately after the completion of individual trials. This approach is likely to yield more accurate reports than those that query subjects about strategies they have used over multiple trials (Ericsson & Simon, 1980); this approach also allows performance to be analyzed by strategy.

The present experiment also showed that parameters other than mean reaction time can be useful in discriminating among models. Research on skill acquisition and automaticity has advanced to the state where different models make virtually the same predictions about mean reaction time. In many cases, however, they may make different predictions about reaction time distributions. Our analyses suggest that variability is an important parameter to predict. In the future, tests of models must be more detailed and analytic than they have been in the past.

REFERENCES

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, **89**, 369-406.

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

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, **87**, 215-251.

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

Logan, G. D., & Klapp, S. T. (in press). Automatizing alphabet arithmetic: Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory & Cognition*.

MacKay, D. G. (1982). The problem of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, **89**, 483-506.

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

Schneider, W. (1985). Toward a model of attention and the development of automatic processing. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 475-492). Hillsdale, NJ: Erlbaum.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, **116**, 250-264.

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

Siegler, R. S., & Shrager, J. (1984). A model of strategy choice. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.

Townsend, J. T., & Ashby, F. G. (1984). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.

NOTE

1. There are several ways to test for mixture models, besides the variance tests we propose (see, e.g., Townsend & Ashby, 1984). However, most of them do not compare mixture models with specific alternatives, such as the race model, so they are not as useful for our purpose as the tests we propose.

(Manuscript received December 29, 1989;
revision accepted for publication September 7, 1990.)