Journal of Experimental Psychology:
Learning, Memory, and Cognition
1989, Vol. 15, No. 3, 517-526

0278-7393/89/$00.75

Practice and Working Memory Effects in Building Procedural Skill

Richard A. Carlson and Marc A. Sullivan
The Pennsylvania State University

Walter Schneider
University of Pittsburgh

Several theories assume that practice (a) results in restructuring of component processes and (b) reduces demand on working memory. Eight subjects practiced judgments about digital logic gates for over 8,000 trials. At two practice levels, subjects made judgments while retaining short-term memory loads irrelevant to the judgments, relevant but not accessed, or accessed to make the judgments. Four phenomena together provide constraints for theory: First, performance declined in moving from blocked practice to randomized practice. Second, gate and judgment type strongly affected latency. Third, these effects declined but did not disappear with practice. Fourth, the cost of accessing information in working memory remained substantial. These results are interpreted as reflecting a serial process with constant structure, while component processes become faster. The results challenge theories assuming that all learning results from restructuring or that restructuring is an automatic consequence of practice, and they support a distributed view of working memory.

Several current theories of skill acquisition (e.g., Anderson, 1982, 1983, 1987; Rosenbloom & Newell, 1987) share the assumptions that (a) practice results in the reorganization or *restructuring* of component processes, and (b) practice reduces the load on a working memory that serves as a single workspace for carrying out cognitive processes. In the study reported here, we observed the acquisition of procedural skill for judgments about causal rules over the course of approximately 20 hr (over 8,000 trials) of practice. The results challenge the above assumptions and suggest an alternative view in which the organization of component processes and use of working memory remain constant while the speed of component processes increases (and attentional load decreases) with extended practice (see Schneider, 1985).

Restructuring and the Acquisition of Cognitive Skill

The production system framework currently provides a popular theoretical approach to understanding learning (e.g., Klahr, Langley, & Neches, 1987). Although the idea of restructuring as a learning mechanism is not necessarily tied to the production system framework, production system theories provide the clearest descriptions of restructuring mechanisms. In Anderson's ACT* theory (1982, 1983, 1987), for example, a *composition* mechanism combines serially executed productions into single productions. Similarly, the chunking model described by Rosenbloom and Newell (1987) attributes performance improvements to the combination of serially processed chunks into higher order chunks that can be processed more rapidly than their constituents. In both cases, restructuring results in greater fluency by reducing the number of component processes that must be executed.

Learning curves for most skills follow a power-law function, with the marginal improvement decreasing with practice, and this characteristic function is often taken as the most basic phenomenon to be explained by learning theories (Newell & Rosenbloom, 1981). Rosenbloom and Newell (1987) predict this function solely on the basis of a restructuring mechanism (chunking), while Anderson's (1982, 1987) ACT* theory predicts a power-law speedup through a combination of restructuring (composition) and strengthening of productions. In some tasks, changes corresponding to both theoretical mechanisms—reduction in the number of steps (attributed to composition) and speedup of individual steps (attributed to strengthening)—can be observed (Neves & Anderson, 1981). The ACT* theory also includes a learning mechanism known as *proceduralization*, which builds domain-specific declarative knowledge into productions (Anderson, 1987), thus reducing the load on working memory.

Theories that include restructuring mechanisms thus are not committed to the position that all performance improvements are due to restructuring. However, these theories do seem to be committed to the prediction that, given certain constraints such as the goal-relatedness of sequential productions and complexity of resulting productions, restructuring will occur whenever component processes are repeatedly executed in series. As Anderson (1983) puts it, "Compositions will occur whenever there is an opportunity" (p. 239).

Working Memory and Procedural Skill

The concept of working memory is central to production system models of skill acquisition. Information active in working memory is matched to the conditions of productions to produce behavior, and it provides for the assembly of procedures requiring multiple productions (Anderson, 1983).

Portions of these data were first presented at the Annual Meeting of the Psychonomic Society in Seattle, Washington, November 1987.

The research was supported by the Army Research Institute, under Contract MDA903-86-C-0149 to Walter Schneider.

We would like to thank Mark Detweiler, Rondall Khoo, Paul McConnell, Merrill Noble, Stuart Klapp, John Anderson, and two anonymous reviewers for helpful comments on earlier drafts. Rondall Khoo also provided valuable assistance in analyzing the data.

Correspondence concerning this article should be addressed to Richard Carlson, Department of Psychology, The Pennsylvania State University, 613 Moore Building, University Park, Pennsylvania 16802.

In these models, working memory serves as a mental workspace and imposes seriality and other limitations on behavior (Anderson, 1982). No detailed links have generally been established between working memory assumptions in models of skill acquisition and empirical research on short-term memory. However, the general characteristics of working memory in these models, such as limited capacity and rapid forgetting, correspond to commonly accepted generalizations about short-term memory (for reviews, see Baddeley, 1986, or Schneider & Detweiler, 1988). In what Anderson (1983, pp. 13–18) calls "neoclassical" production systems, working memory assumptions are heavily influenced by simple slot models of short-term memory (e.g., Atkinson & Shiffrin, 1968). Anderson's (1983) ACT* theory defines working memory in a more flexible way, in terms of a limited amount of activation spread over a variable number of elements but retains the assumption of a single workspace. This view may be contrasted with recent models of working memory in which storage and processing are distributed over multiple systems or regions (Baddeley, 1986; Monsell, 1984; Schneider & Detweiler, 1988).

In the case of procedural skills, working memory must be used to hold current task information and to integrate that information with long-term knowledge. To carry out a procedural skill, appropriate information must be loaded into working memory in a sequence that fits the logical structure of the task. Hitch (1978), for example, has demonstrated that varying the order of operations in mental arithmetic may lead to errors due to rapid loss of information in working memory. It is also likely that some loss of information is due to content-based interference (e.g., Baddeley, 1986). Restructuring is believed to increase the effective capacity of working memory by chunking (Miller, 1956; Rosenbloom & Newell, 1987) or by eliminating the need to hold declarative knowledge for step-by-step application (Anderson, 1987). Improvements in performance with practice may therefore be important in overcoming the constraints imposed by working memory. Specifically, practice should increase the efficiency with which procedural skills operate on information in working memory.

Structure of Causal Judgments

Causal judgments have a structure consisting of at least two components: an input state and a causal rule. To make a *prediction* judgment, an individual combines knowledge of the input state and causal rule to generate a predicted *output state.* To *verify* an output state, an individual combines knowledge of the input state, output state, and causal rule to determine whether or not the rule is violated (or the output state correct). Initially, subjects are likely to consider the input state, causal rule, and (in the case of verification) output state sequentially; with practice, however, the sequence might be restructured so that an entire pattern of input state, causal rule, and output state is considered in a single cognitive step. For example, an electronics technician making a judgment about the operation of a particular component combines knowledge of the electrical state at the components' input and output, and of the rule governing operation of the component, to determine whether or not the component is operating correctly. In the case of the rules subjects learned in the present study (described below), an extra step of negation may be added. After extensive practice, these components might be recognized by the condition of a single production, if restructuring occurs.

Overview

For the present study, we taught subjects the rules governing the operation of a set of electronic components known as digital *logic gates.* A logic gate is a component that performs a logical operation on one or more binary inputs, producing a single binary output. For example, the output of an AND gate is 1 if and only if all inputs are 1 (i.e., Input A is 1 AND Input B is 1 AND Input C is 1). We abstracted from the electrical realization of these functions and described them in symbolic terms (see Figure 1).

The rules governing logic gate operation can be described in terms of two variables—the presence or absence of an operation for integrating multiple inputs and the presence or absence of negation. Thus, these rules range in complexity from an positive rule operating on a single input to negated rules integrating multiple inputs. In addition, the subject may be asked to *predict* the output of the gate or to *verify* the correctness of a given output.

This task is useful for studying the acquisition of procedural skill for several reasons. First, the causal (input–rule–output) nature of the task leads subjects to adopt a common, serial strategy for performing the task. Second, the binary nature of the output is convenient for collecting latency data. Third, and perhaps most important, the variables describing gate type and judgment type have consistent effects on latency that can be used to track changes in the structure of cognitive processes, even when performance becomes too rapid for introspection.

Each subject spent about 20 hr practicing speeded judgments about individual logic gates. Subjects received instruction about the rules governing the operation of logic gates and practiced predicting or verifying the outputs of these gates. Initially, practice was blocked with blocks each consisting of a single gate type. After this initial blocked practice, gate types were randomized within blocks. Judgment type (prediction or verification) was blocked. On the basis of previous research (Carlson & Schneider, 1988), we expected longer judgment latencies for rules combining multiple inputs, for negated rules, and for verification judgments. According to subjects' reports early in practice, these effects result from a sequential reasoning process. Restructuring theories of skill acquisition predict that these effects will disappear with extended practice; this prediction, however, was not supported by the data.

At two points in practice, subjects made judgments about logic gates while maintaining in working memory information relevant (but either accessed or not) or irrelevant to the judgments. Single-workspace models of working memory predict interference from memory loads that are not accessed. A single-workspace model of working memory that postulates a limited number of slots for information (e.g., Atkinson & Shiffrin, 1968) predicts substantial interference from near-span memory loads that leave few open slots for processing.

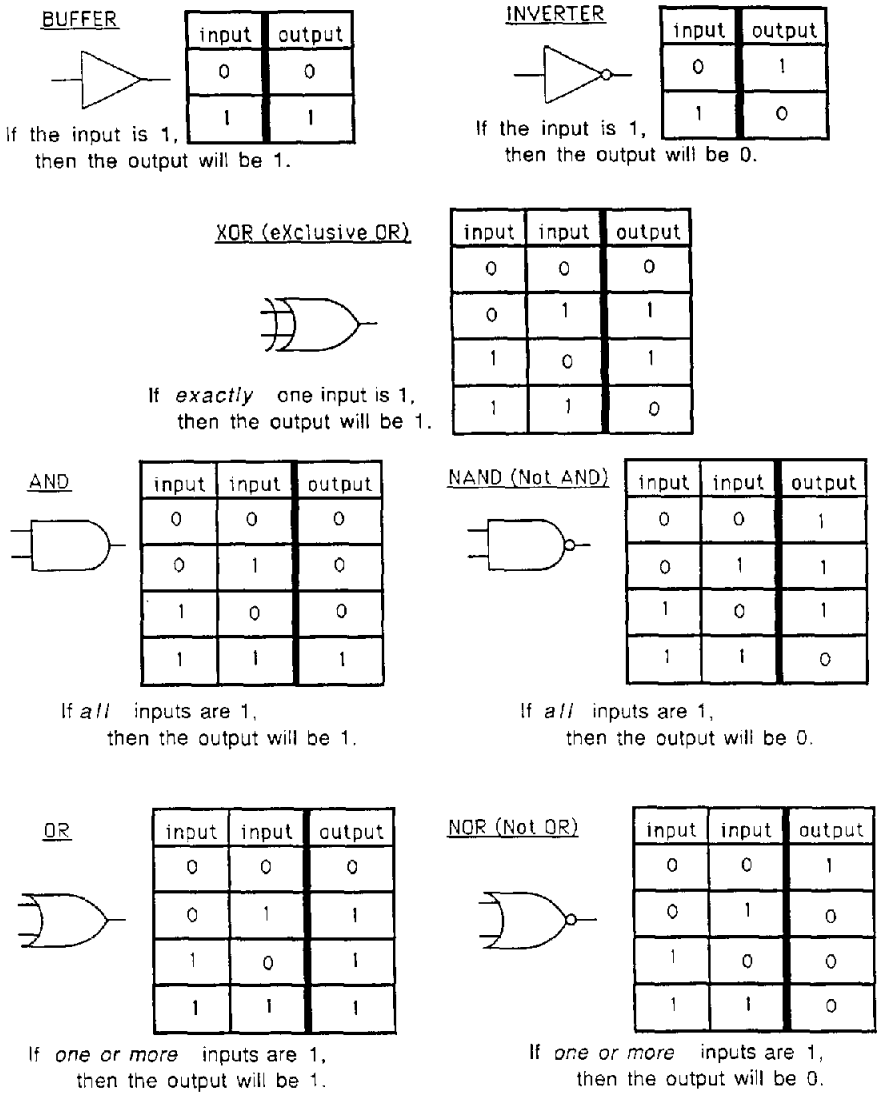

Figure 1. Logic gate symbols and functions.

This interference should decline with practice, however, as restructuring results in a judgment procedure that requires fewer slots. A more flexible single-workspace view in which working memory is defined by activation (e.g., Anderson, 1983) also predicts interference from storage of irrelevant information, because activation must be distributed over more elements. Again, interference should decline with practice, because fewer elements need to be held in working memory in order to carry out a skilled procedure. In Anderson's (1983) theory, strengthening might also reduce the need for active rehearsal to maintain information in working memory. Distributed models of working memory (Baddeley, 1986; Monsell, 1984; Schneider & Detweiler, 1988) need not predict interference from memory loads not accessed during judgment, because separate subsystems are available for storage and for processing.

In one condition in the present study, information held in working memory (inputs to logic gates) was accessed in order to make judgments about logic gates. In single-workspace models, the cost of such access should be minimal because storage and processing take place in the same cognitive workspace. Only the coordinating of representations is required. Furthermore, the cost of coordinating representations should decline with practice as less space is required to represent procedures. Distributed models, on the other hand, predict that coordinating representations requires establishing communication between separate subsystems. Thus, these models predict substantial access costs and little decline in these costs with practice.

Method

Subjects

Eight undergraduate psychology majors at The Pennsylvania State University (5 males, 3 females) participated in return for course

credit. In addition to their participation as subjects, these students attended a weekly research meeting and assisted with other research projects under the direction of the first author. All subjects reported no previous experience with digital electronics.

Each subject participated in 24 experimental sessions of 90–120 min each, two sessions per week for 13 weeks with a 1 week break after 6 weeks. In addition to the tasks described here, these subjects participated in several sessions of problem solving at various points during the project. The problems involved reasoning about networks of logic gates, either to calculate output states of several interconnected gates or to locate faulty gates within networks. A total of 40 such problems was presented. Subjects also learned some basics of digital electronics (such as binary arithmetic) and the input–output functions of several digital circuits. After data collection for the present study was completed, the subjects participated in an additional problem-solving study. The purpose of the problem-solving tasks was to examine several issues concerning component process fluency in complex tasks, and the logic gate practice described here served as an independent variable for these problem-solving tasks. The problem-solving tasks provided subjects with a small amount of additional practice with logic gate judgments but did not provide any additional information about the tasks reported here. Further details concerning the problem-solving tasks may be obtained by writing to the first author.

Computer Training

An IBM PC/XT microcomputer with an IBM monochrome display was used to present stimuli, record answers, and provide feedback. Subjects practiced predicting and verifying the outputs of logic gates in a two-choice reaction time procedure. *Prediction* judgments required subjects to indicate the appropriate output (1 or 0), given a logic gate symbol and a set of inputs. *Verification* judgments required subjects to indicate whether a given output was correct or incorrect, given the logic gate symbol and inputs.

Each practice trial was initiated by the subject. A logic gate symbol appeared in the center of the screen, with each input labeled as 1 or 0, and the output labeled with a [?] (for prediction judgments) or with a 1 or 0 (for verification judgments). Accuracy and reaction time feedback followed each judgment. Figure 2 shows the displays for each type of judgment.

Assistance in remembering the function associated with each rule was available during the first blocks of computerized practice. By pressing a key labeled *H* (Help), the subject could display the truth table for the displayed symbol. This help facility was available throughout the trial, though subjects rarely used it after receiving feedback, even on incorrect responses.

Introduction to Rules

Rules were introduced one at a time, with 144 trials of blocked practice for each rule prior to introducing the next rule. Introduction of rules was in the constant order of *BUFfer*, *INVerter*, *AND*, *OR*, *NAND*, and *NOR*. The experimenter introduced each rule by starting the program to display the logic gate symbol and by pressing the help key to display the truth table. Subjects made only prediction judgments in these introductory practice blocks, and accuracy but not latency feedback was provided after each judgment.

The *XOR* (e*X*clusive *OR*) gate was introduced separately, after a total of 464 practice trials with each of the six gates mentioned above. This provided an opportunity to observe the effects of adding a new rule to an already well-learned set of rules.

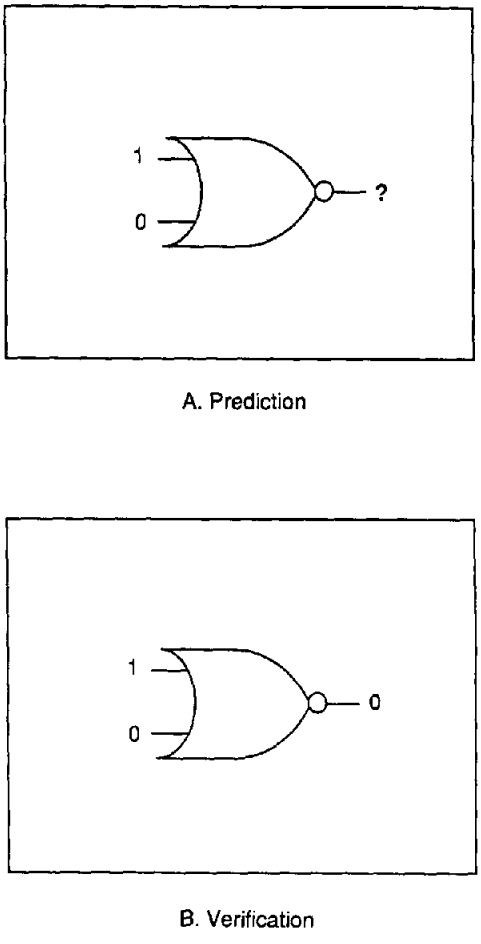

Figure 2. Panel A: Sample displays for prediction. Panel B: Verification judgments.

Practice Sessions

After the initial blocked practice with rules taken one at a time, subjects practiced with the seven rules mixed together (six rules until the XOR gate was introduced). Each practice block consisted of 24 trials for each gate, with the entire set of 144 or 168 trials (6 or 7 rules × 24 trials) presented in a different random sequence for each block. Judgment type (*prediction* or *verification*) alternated from block to block.

Declarative Knowledge Test

Subjects' declarative knowledge of the logic gate rules was tested after the first block of mixed-gate practice. For each logic gate symbol (see Figure 1), subjects were asked to recall the gate name (e.g., *AND*) and to fill in a truth table for that rule. In addition, subjects made verification judgments in an unspeeded, paper-and-pencil test for each combination of symbol, pattern of two inputs (one in the case of *BUF* and *INV* gates), and output. Initially, we planned to repeat this test at several points in the practice sequence, but the nearly perfect recall on the initial test made this repetition moot.

Memory Load Task

At two points in learning (after 336 and 1,232 trials of practice per rule), subjects made logic gate judgments in the retention interval of

a working memory load. Working memory was loaded at the beginning of each trial by assigning values to either 3 or 6 variables (e.g., $A = 1$). Each variable appeared in the center of the computer display for 2 s, with a 1-s blank interval between variables. Not more than two-thirds (two of three, or four of six) variables were assigned the same value on any trial.

After the memory set was presented, a logic gate symbol appeared, and the subject made either a prediction or verification judgment as described above. After this judgment, a memory set probe (e.g., $A = 0$) was presented, and the subject indicated whether the probe was correct or incorrect with respect to the presented memory set. Figure 3 illustrates the sequence of events for these trials.

Three types of trials were defined by the relation between the memory load and the gate judgment. On *irrelevant* memory load trials, variables were assigned values of 7 or 8. Because these values cannot serve as inputs to logic gates, subjects knew that the memory load was irrelevant to the gate judgment task. On memory load *access* trials, variables were assigned values of 1 or 0. The logic gate symbol then appeared with inputs labeled with variables rather than values; thus, the subject was required to access the memory load in order to make the gate judgment. For example, with a memory set ($A = 1$, $B = 0$, $C = 1$), an *AND* gate might appear with inputs (A, C), and the subject should predict an output of 1. *Expect* trials were the same as access trials, but the gate inputs were labeled with values. Thus, subjects should expect that the memory load would have to be accessed, but in fact could make the gate judgments without accessing the memory set.

Results

Introduction to Logic Gates

Subjects requested help on only a small proportion (.003, or 22 of 6,864 trials) during the introductory practice blocks. Responses on these single-gate practice blocks, immediately after the introduction to each gate, were very accurate: .986 correct, and fast, mean latency = 673 ms. Table 1 shows the mean proportion correct and latency of correct responses for each gate in these initial practice blocks. No statistical analyses were conducted on these data because the gates were introduced in a constant order that confounded gate type with practice in the judgment task. We note, however, that judgments of positive and negated gates had very similar latencies (663 ms and 683 ms, respectively).

Transition to Randomized Practice

Mixing the gate types together for randomized presentation initially resulted in a sharp increase in response time, from a mean of 673 ms in blocked practice to 1,562 ms in the first block of randomized practice of prediction judgments. The level of performance in blocked practice is not reached in randomized practice until subjects have practiced for over 8,000 trials. This effect is shown in Figure 4, which displays the latency for correct responses in both blocked and randomized practice. Accuracy also declined, from .986 correct in blocked practice to .921 in the first block of randomized practice. Subjects requested help on about 13% of trials in this first block of randomized practice.

Randomized Practice

After the first block of randomized practice, declarative knowledge was essentially perfect. All subjects correctly completed truth tables for each rule and made 100% correct verification judgments in an unspeeded paper-and-pencil test, while reporting high confidence. All but 1 subject correctly

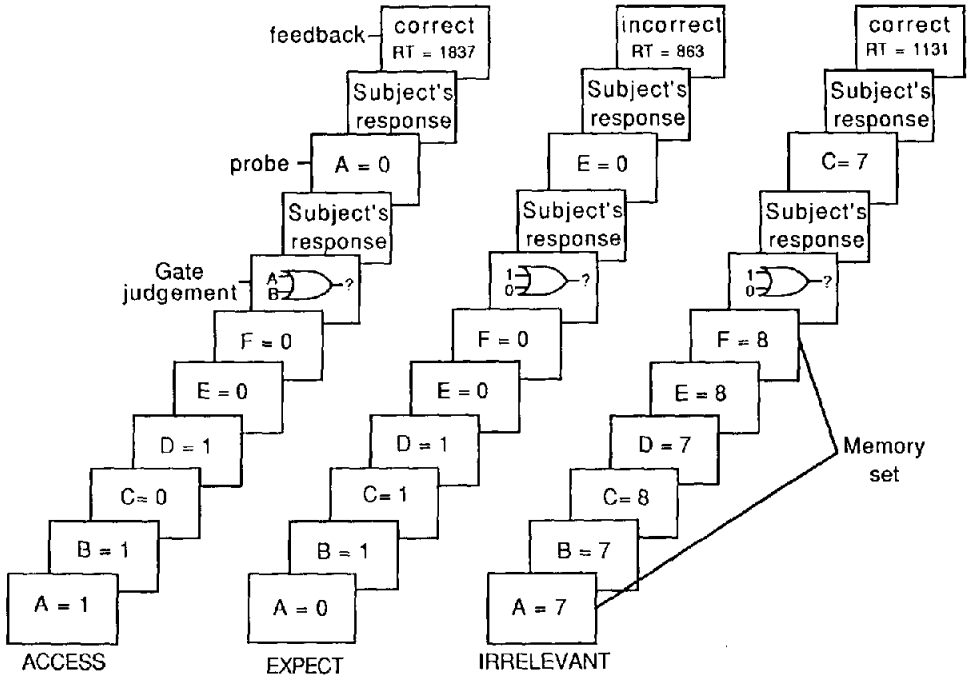

Figure 3. Sequence of events on each trial in memory load task.

Table 1
Proportion (p) Correct and Latency (in Milliseconds) for Correct Judgments as a Function of Gate Type During Initial Blocked Practice

Gate	*p* (correct)	Latency
BUF	.988	602
AND	.996	707
OR	.982	681
INV	.972	723
NAND	.988	675
NOR	.989	652

recalled the names of all seven gates. Further evidence of the rapid acquisition of declarative knowledge is provided by the rapid decline in the use of help, to less than 2% by the fourth block of practice. Subjects maintained a high level of accuracy—approximately 92% correct—throughout practice. The effects of the independent variables and their interactions on accuracy were very small, with a total range of .89–.95 correct over cells in the design. In all comparisons, latency and proportion correct were negatively correlated, indicating that subjects did not trade speed for accuracy as a function of experimental condition. We therefore focus on the latency of correct responses in the remainder of this section.

Gate and judgment type effects. Early in practice, subjects unanimously reported a serial judgment process of classifying the input pattern, applying the rule, negating the result (if necessary), and comparing the result with the displayed output (for verification judgments). As would be expected from these reports and from previous research (Carlson & Schneider, 1988), more time was required for multiple-input gates, for negated gates, and for verification judgments. Judgment latency for single-input gates was 762 ms, compared with 1,045 ms for multiple-input gates, $F(1, 7) = 62.6$, $MS_e = 102{,}265$, $p < .01$. Latency for positive gates averaged 839 ms, while latency for negated gates was 967 ms, $F(1, 7) = 180.7$, $MS_e = 7{,}335$, $p < .01$. The time to predict an output was 861 ms, and verification latency was 945 ms, $F(1, 7) = 34.8$, $MS_e = 16{,}331$, $p < .01$. This pattern of results is consistent with the serial process reported by subjects. In addition, these variables showed overadditive interactions with one another; for example, the extra time required for negated gates was greater for verification judgments. Because the magnitude of these interactions changed substantially with practice, they are described in the next section.

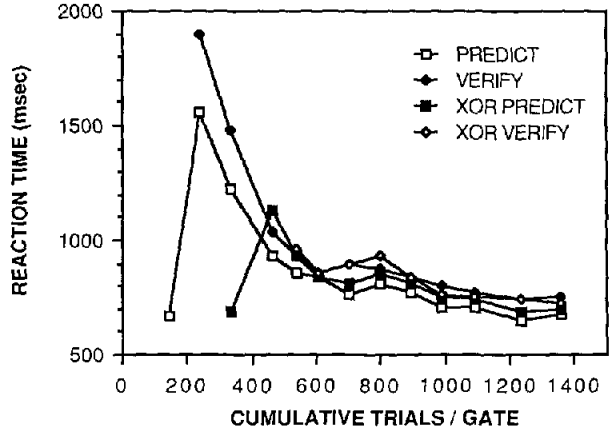

Figure 4. Latency for correct logic gate judgments as a function of judgment type and practice.

Practice. As shown in Figure 4 (above), judgment latency in randomized practice declined, following approximately a power-law function (the correlation of log reaction time (RT) and log trial is .96). The effects of gate type and judgment type declined but did not disappear with practice, as shown in Table 2. As the table shows, the magnitude of effects is compressed with practice, but the pattern remains the same.

To confirm these effects, we divided the randomized practice trials into five groups. Each of the first three groups contained about 200 trials per gate, while groups four and five contained about 300 and 250 trials, respectively. The divisions between groups were chosen to correspond with natural "breaks" in the semester-long project (e.g., spring break). All of the three-way interactions (*df*s = 4, 28) in this table are significant at the .01 level, with a mean square error of 693.

Adding a rule. After 464 trials per rule with six logic gates (*BUF*, *AND*, *OR*, *INV*, *NAND*, and *NOR*), the *XOR* rule was introduced in a separate block of 144 trials of prediction judgments. As in the initial blocked practice with other rules, subjects made these judgments very rapidly: mean latency = 689 ms, and accurately, mean proportion correct = .957.

When the *XOR* rule was mixed with other rules on the next block of practice, there was a large decrement in performance on the *XOR* rule. Mean RT increased to 1,233 ms, with mean proportion correct = .875. As shown in Figure 4, however, performance on the remaining six rules was not disrupted. With additional practice, performance on the *XOR* rule improves at approximately the same rate as the other rules.

Memory Load Task

This task required subjects to make logic gate judgments in the retention interval of a short-term memory load that was either irrelevant to the judgment, relevant but not used (*expect* condition), or relevant and used to make the judgment (*access* condition). This procedure was repeated at two levels of practice: 336 and 1,232 trials per gate. Gate type, judgment type, and size of the memory set also served as independent variables, as described above.

Overall, subjects made more than 91% of gate judgments correctly in this task. Only trial type had a substantial effect on accuracy, with proportion correct of .94 for irrelevant and expect trials, and .86 for access trials, $F(2, 14) = 23.9$, $MS_e = .035$, $p < .01$. Several marginal effects of other variables on gate judgment accuracy are consistent with the latency data (lower accuracy in cells with longer latencies); we therefore focus on latency for correct judgments in the remainder of this section.

Trial type and memory load. The main effect of trial type was substantial. Mean latency was 929 ms for irrelevant memory load trials, 970 ms for expect trials, and 1,737 ms for access trials, $F(2, 14) = 54.7$, $MS_e = 1{,}456{,}755$, $p < .01$.

Table 2
Latency (in Milliseconds) for Correct Gate Judgments as a Function of Gate Type and Judgment Type

Gate type	Prediction		Verification	
	Positive	Negated	Positive	Negated
First block of randomized practice				
Single input	995	1,134	1,281	1,539
Multiple input	1,527	1,800	1,948	2,332
After 1,200 trials of practice				
Single input	554	609	601	678
Multiple input	637	736	713	830

These data show little difference (41 ms) between the effects of holding an irrelevant memory load and holding a memory load that subjects expect to use. However, there is a large cost of accessing the memory set in order to make a judgment. This cost remains very substantial even after extended practice with the logic gates, as shown in Figure 5.

Memory set size had a small effect on gate judgment latency, except in the access condition, where the effect was substantial. This pattern is substantially the same at both levels of practice. These data are shown in Table 3. Analysis of variance confirms that the interaction of memory set size and trial type is significant, $F(2, 14) = 30.9$, $MS_e = 56{,}797$, $p < .01$.

Although the results of major interest here are the large effect of trial type and the presence of a substantial memory set size effect only for the access condition, the effect of nonaccessed memory loads is of some theoretical interest. We did not include a no-load control condition intermixed with the memory load trials, but a comparison can be made with the practice trials immediately preceding and following each memory load session. The mean latency for the practice blocks just before the first memory load test was 1,141 ms; reflecting the fact that this part of the learning curve is still quite steep, latency for the practice blocks immediately after the first memory load test was 899 ms. Performance was more stable by the time of the second memory load test, with mean latency of about 650 ms both before and after that test. Comparing these means with those in Table 3 suggests that unaccessed memory loads did slow gate judgments to some degree. This effect is very small, however, compared with the cost of accessing information in working memory in order to make judgments.

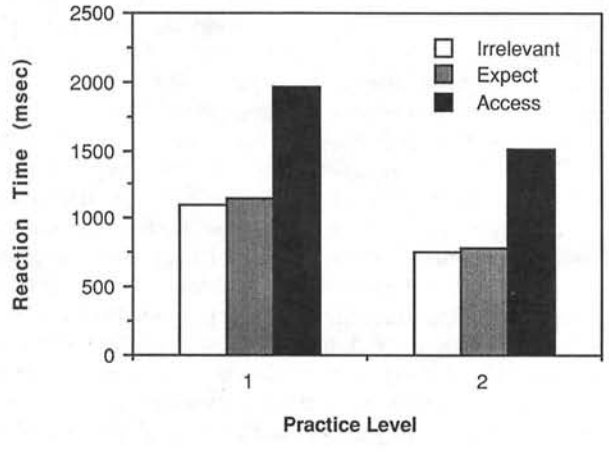

Figure 5. Latency for correct logic gate judgments on memory load trials as a function of trial type at two practice levels. Test 1 was presented after 336 trials of practice with each logic gate judgments, and Test 2 after 1,232 trials.

Judgment and gate type effects. As in the practice procedure, more time was required for verification than for prediction, for multiple input than for single input gates, and for negated than for positive gates. More important, the magnitude of these effects is greater in the access condition than in the other two conditions. Table 4 shows the effect of each of these variables as a function of trial type. Each of these variables participates in significant interactions with trial type ($p < .01$), though in some cases the two-way interactions are of marginal significance while three-way interactions are significant at the $p < .01$ level. For example, the two-way interaction of trial type with single versus multiple inputs is of marginal significance, but the three-way interaction of these variables with positive versus negated gates is significant ($p < .01$).

Practice. Subjects practiced for approximately 900 trials per gate between the two replications of the memory load procedure. Judgment latency declined from a mean of 1,407 ms on the first replication to 1,017 ms on the second, $F(1, 7) = 171.6$, $MS_e = 255{,}916$, $p < .01$. Figure 6 shows latency for correct gate and memory probe judgments as a function of practice both within and between the two replications of the memory load task. There is a clear discontinuity in the curve for gate judgment latency, but not for probe judgment latency. The effects of practice on gate judgment latency can thus be attributed largely to the intervening practice, not simply to practice with the memory load procedure.

The pattern of results was very similar at the two levels of practice. As described above (see Figure 5), the cost of accessing information in working memory to make gate judgments remained substantial, though this effect did decline slightly. Similarly, the effects of gate type (positive vs. negated, single vs. multiple inputs), judgment type, and memory set size declined slightly with practice. The interactions of each of these variables with practice were at least marginally significant, Fs = 4.2 to 16.1, all ps $< .10$, except for judgment type, $F(1, 7) = 2.25$, $p < .18$. The largest effect was a reduction in the difference between single and multiple-input gates from 340 ms to 145 ms.

Probe responses. Accuracy on the memory probe judgments was remarkably constant over experimental conditions, with means within cells ranging from .89 to .91 correct. There was a marginal effect of judgment type, $F(1, 7) = 5.5$, $p = .051$, $MS_e = .018$, with mean p(correct) of .889 for prediction and .905 for verification judgments. No other effects on accuracy approached significance (all ps $> .25$).

Probe latency did not vary as a function of trial type, gate type, judgment type, or their interactions ($p > .25$). This result suggests that strategies for short-term storage did not differ over experimental conditions. Probe response times did decline significantly from the first to the second replication of the memory load task: Means = 1,115 and 850 ms, $F(1,$

Table 3
Latency (in Milliseconds) for Correct Gate Judgments on Memory Load Trials as a Function of Practice Level (PL) Memory Set Size (MSS), and Trial Type

PL and MSS	Trial type		
	Irrelevant	Expect	Access
PL 1			
MSS 3	1,087	1,137	1,819
MSS 6	1,125	1,160	2,115
PL 2			
MSS 3	743	786	1,399
MSS 6	760	797	1,616

7) = 34.3, MS_e = 483,153, $p < .01$. This effect is probably due to practice with the probe task rather than the intervening practice in making gate judgments, because the learning curve for probe RT shows no discontinuity corresponding to the break between the two replications (Figure 6).

Probe response time also varied as a function of memory set size. Mean response time was 906 ms for memory set size 3, and 1,058 ms for memory set size 6, $F(1, 7) = 57.2$, MS_e = 94,327, $p < .01$. This difference—about 50 ms per item—is typical of short-term memory tasks.

General Discussion

Over more than 8,000 trials of practice, latency for logic gate judgments continued to decline, following approximately a power-law function. This result is common to many skills (e.g., Anderson, 1987; Newell & Rosenbloom, 1981) and by itself is not very diagnostic with respect to current theories. However, more detailed aspects of the data together may place important constraints on theories of cognitive skill. Four phenomena stand out: First, there was a large decrement in performance in moving from initial blocked practice with each logic gate to randomized practice in which logic gates were randomized within blocks. Second, in randomized practice judgment latency was strongly affected by gate type (single vs. multiple inputs, positive vs. negated rules) and judgment type (prediction vs. verification). Third, these effects declined but did not disappear with practice, despite the large overall speedup. Figure 7 illustrates the continuing effects of negation and judgment type over practice. Fourth, the cost of accessing information in working memory to make logic gate judgments remained substantial even with very extended practice in making those judgments.

Table 4
Effects of Gate Type and Judgment Type on Latency (in Milliseconds) for Correct Gate Judgments as a Function of Trial Type for Memory Load Trials

Judgment/gate type	Trial type		
	Irrelevant	Expect	Access
Judgment type			
Prediction	881	934	1,663
Verification	952	988	1,774
Gate type			
Positive	877	896	1,609
Negated	981	1,044	1,865
Gate type			
Single input	830	857	1,586
Multiple input	1,027	1,083	1,889

Taken together, these results appear to challenge two common assumptions of current theories of skill acquisition: (a) that practice results in continuing restructuring of cognitive procedures and that (b) capacity demands on a single-workspace working memory are reduced with practice, beyond a few hundred trials. Instead, we suggest, subjects continued to execute sequential procedures that make use of distributed working memory capacities. On this view, the observed changes in performance result not from restructuring but from speedup of component processes.

Practice and Restructuring

Subjects' reports and the presence of strong effects of gate and judgment type on latency indicate that the logic gate task has a sequential structure, at least early in practice. More time was required for gates with multiple inputs, for negated gates, and for verification judgments. These effects declined but did not disappear with practice (see Table 2 and Figure 7), posing a challenge to restructuring views of learning.

The problem for restructuring theories is that this seems like a clear case in which practice should lead to restructuring. The steps are both temporally and logically contiguous, conditions for composition in Anderson's ACT* theory (Anderson, 1983). In addition, the task has a clear combinatorial structure, in which combinations of input patterns, symbols, and (for verification judgments) output values are consistently associated with responses, meeting the conditions for chunking in Rosenbloom and Newell's (1987) model. Although the general speedup might be handled by other mechanisms such as strengthening (Anderson, 1982, 1983), such alternatives do not explain why restructuring did not occur.

An alternative explanation for the persisting effects of gate and judgment type (pointed out to us by an anonymous reviewer) is that the longer latencies are associated with displays of greater complexity. For example, negated gates (see Figure 1) might require more time to match the condition of a single production, even if composition has occurred (Anderson, 1983). We examined this possibility in a simple control experiment. Each of four logic gate symbols (*AND*, *OR*, *NAND*, and *NOR*) was assigned to a different response (the index and middle fingers of left and right hands), and subjects were trained to respond as quickly as possible to each symbol. Response assignments were counterbalanced over subjects, and each subject was trained for 120 trials per symbol (approximately $^1/_{10}$ the amount of practice presented in the main study). Symbols were randomized within blocks of 120 trials (30 trials per symbol). Eight subjects participated in this procedure. The results were clear: Subjects responded correctly to 94% of the negated symbols, with a mean latency of 607 ms, compared with 95%, with a mean latency of 618 ms for positive symbols. Thus it is clear that the longer latency for negated gates cannot be attributed to pattern-matching complexity.

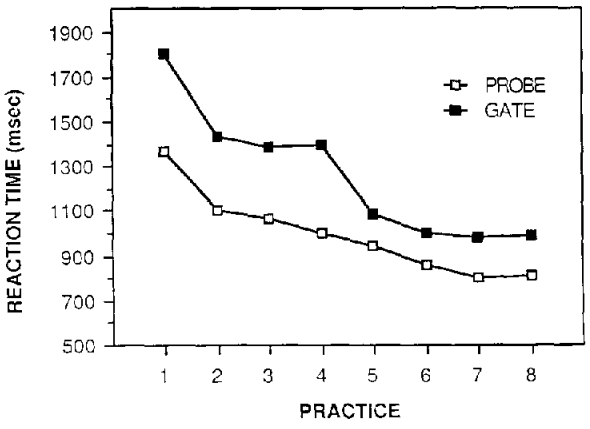

Figure 6. Latency for correct gate and probe judgments on memory load trials as a function of blocks within Test 1 (practice levels 1–4) and Test 2 (practice levels 5–8).

The large decrement in latency that occurred when subjects shifted from blocked to mixed-gate practice initially came as a surprise, though we have now replicated this result several times in unpublished studies. An obvious explanation is that in blocked practice, subjects need only to establish associations between possible input patterns and output values, ignoring the symbol itself (and hence the need to discriminate among symbols). Restructuring is not at issue here because of the single-step nature of the suggested process. In randomized practice, however, the entire process must be assembled in working memory on each trial. This effect is similar to the *contextual interference* effect found in verbal (e.g., Battig, 1979) and motor learning (e.g., Shea & Morgan, 1979), and the present explanation is similar in spirit to that offered by Lee and Magill (1983) for contextual interference in motor learning. Although this effect may provide only a weak constraint on theories of skill acquisition, it is a very large effect that has not received a great deal of theoretical consideration.

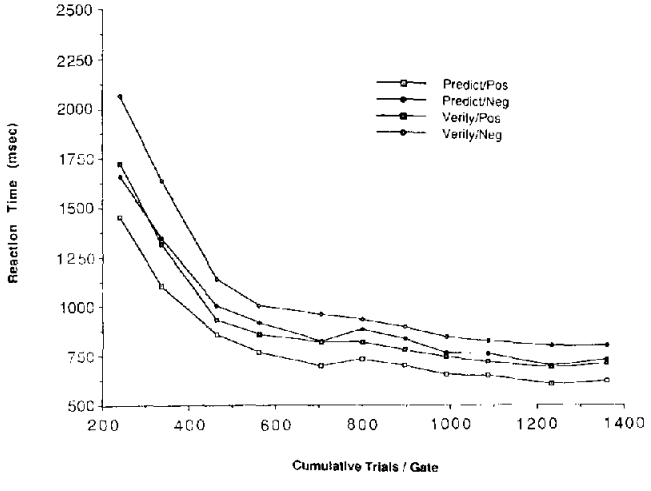

Figure 7. Latency for correct gate judgments as a function of practice, trial type, and gate type (positive vs. negated).

Furthermore, the possibility that subjects use a single-step associative process in blocked practice and a serial judgment process in randomized practice suggests that the procedures subjects develop for using rules, as well as the rate of acquisition, may depend on acquisition context in theoretically important ways. In another study (Carlson & Schneider, 1988), we examine some aspects of acquisition context for logic gate rules, and we are currently examining the effects of randomized practice on the acquisition of procedural skill in more detail.

Procedural Skill and Working Memory

Short-term memory loads had little effect on gate judgment latency if those loads were not accessed to make the gate judgments (irrelevant and expect conditions) at either level of practice. The difference in gate judgment latency for memory set sizes of 3 and 6 (presumably near span) in these conditions was very small, and no major change was observed as a function of extended practice with logic gate judgments. These results appear to conflict with single-workspace models in which working memory is defined either by slots (e.g., Atkinson & Shiffrin, 1968) or by activation (e.g., Anderson, 1983). A comparison of gate judgment latencies in the memory load task with no-load practice latencies suggests some slowing in the unaccessed memory load conditions. However, the small memory set size effect and lack of change with practice suggests that this slowing might be attributed to changes in the goal structure of the task rather than to memory loading per se. Furthermore, short-term retention of memory set items was very good, suggesting that subjects did not trade off memory performance for rapid judgments. In any case, the effects of unaccessed memory loads are clearly small in comparison with the cost of accessing the memory set to make judgments.

These results initially appear to conflict with other results in the literature; for example, Logan (1980) showed interacting effects of memory set size and task complexity in a similar paradigm. Logan suggests, however, that his results are dependent on a task that requires "strategic adjustment of existing abilities" (p. 205) on each trial. This suggestion and the present data fit nicely with our claim that no restructuring is occurring—there is no memory load effect in these conditions, and no interaction with practice, because subjects are using the same processing sequence throughout training. Our emphasis on speedup of component processes corresponds roughly to perceptual learning in Logan's task.

There was a large cost (approximately 800 ms) of accessing the memory set in order to make logic gate judgments, and this cost remains very substantial even after extended practice. These results support a distributed view of working memory capacity. The extra time required in the access condition by itself might be accounted for by an added stage of scanning short-term memory for the appropriate information. However, the additional time required is far greater than estimates of scanning time from other studies or from the probe judgment latencies here. Furthermore, the effects of gate type variables such as negation are exaggerated in the access condition. These results seem to be incompatible with the single-

workspace working memory in most production system models, because even in other conditions all production matching occurs in working memory (which must hold the same number of elements in all conditions). The results might be handled with appropriate assumptions about the decay of working memory representations (Anderson, 1983). Although we have not modeled the use of working memory in this task, it seems unlikely that decay parameters that account for the trial type effect on logic gate judgments would also be consistent with the lack of differences in probe judgment latency.

In summary, these data suggest that the use of working memory, in terms of either processing or storage requirements, changes little with practice on the logic gate task, at least over the levels of practice examined here. They appear to disconfirm a single-workspace model of working memory, consistent with other results (e.g., Klapp, Marshburn, & Lester, 1983), and to complement those previous results by examining performance at widely separated levels of practice.

Theoretical Implications

We believe that these results have several implications for theories of cognitive skill. First, the present data pose a challenge to theories of skill acquisition that rely purely on restructuring, such as the chunking model of Rosenbloom and Newell (1987). On the basis of this model, we would expect subjects to develop single chunks for each combination of input pattern and logic gate symbol, resulting in null effects for gate type. Second, these results point to the need for more detailed specification of when restructuring will occur. The present task seems to meet the conditions—temporal and logical contiguity of a sequence of processes—for composition (Anderson, 1983), yet composition seems not to have occurred. An interesting possibility is that restructuring is strategic, depending on subjects' metacognitive access to feedback directly relevant to the organization of cognitive steps. Third, practice did not result in major changes in the ability to share processing and storage capacity or in the ability to coordinate representations in working memory. These results are incompatible with single-workspace views of working memory, suggesting instead some kind of distributed model. Specifically, the results suggest a distinction between storage and central processing capacity, with different (but perhaps typically correlated) mechanisms for changes with practice (Schneider & Detweiler, 1988). Extended practice may simply increase the speed of a cascade of sequential processes, with little restructuring or change in the use of working memory. The component speedup may, however, be critical in developing highly skilled performance.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic Press.

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

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

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94*, 192–210.

Baddeley, A. (1986). *Working memory.* Oxford: Clarendon Press.

Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 23–44). Hillsdale, NJ: Erlbaum.

Carlson, R. A., & Schneider, W. (1988). *Learning and using causal rules.* Manuscript submitted for publication.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmentic. *Cognitive Psychology, 10*, 302–323.

Klahr, D., Langley, P., & Neches, R. (Eds.). (1987). *Production system models of learning and development.* Cambridge, MA: The MIT Press.

Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General, 112*, 240–264.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 730–746.

Logan, G. D. (1980). Short-term memory demands of reaction-time tasks that differ in complexity. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 375–389.

Miller, C. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Monsell, S. (1984). Components of working memory underlying verbal skills: A "distributed capacities" approach. In H. Bouma & D. G. Bouwhis (Eds.), *Attention and performance X* (pp. 327–350). Hillsdale, NJ: Erlbaum.

Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Eds.), *Cognitive skills and their acquisition* (pp. 57–84). Hillsdale, NJ: Erlbaum.

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.

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

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

Schneider, W., & Detweiler, M. (1988). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21, pp. 54–119). New York: Academic Press.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 179–187.

Received November 30, 1987
Revision received August 18, 1988
Accepted July 27, 1988 ■